Yu Wang

Intelligent condition monitoring of bearings for early material damage detection using acoustic emission signals

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Engineering
Department of Mechanical and Industrial
Engineering

NTNU
Norwegian University of
Science and Technology

Yu Wang

Intelligent condition monitoring of bearings for early material damage detection using acoustic emission signals

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2023

Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU

Norwegian University of
Science and Technology

Preface

This PhD thesis is submitted in partial fulfillment of the requirements of the degree of Philosophiae doctor (Ph.D.) at the Norwegian University of Science and Technology (NTNU) in Trondheim. The presented research was carried out at the Department of Mechanical and Industrial Engineering between September 2020 and September 2023. The Ph.D. research was directed by Professor Alexey Vinogradov as the main supervisor and Associate Professor Andrei Lobov initially acted as the co-supervisor and later assumed the role of the main supervisor.

The Ph.D. work was supported from the Norwegian Research Council by the RCN Project No 296236. The aim of the Ph.D. research is to investigate a reliable and effective method for early material damage detection of bearings in online intelligent condition monitoring systems, which can improve the safety and efficiency of marine machinery, reduce downtime, and minimize maintenance costs.

Acknowledgement

I would like to express my deepest gratitude to all those who have supported me throughout my PhD journey and contributed to the completion of this thesis.

First and foremost, I wish to express my sincere gratitude to my esteemed supervisors, as they offered me the opportunity to join in this research team and pursue my Ph.D. degree, which is one of the most important turns in my life. I am immensely thankful to Prof. Alexey Vinogradov, a wise and handsome gentleman. He told me that doing research is like walking through a dark forest with no set method, which encouraged me not to be afraid of making mistakes and to keep an open mind when I was confused. He also gave me full freedom to conduct my own research. Despite his busy schedule, he was always responsive and provided prompt feedback on my questions and revisions to my research papers. I am truly blessed to have had the opportunity to work with such a knowledgeable and helpful supervisor. I am also grateful to my another supervisor, Assoc. Prof. Andrei Lobov, for his unwavering support and guidance during the final year of my Ph.D. journey. His passion, energy, and dedication have been truly inspiring, and his invaluable suggestions and assistance have significantly contributed to the completion of my thesis and will continue to shape my future work.

I would also like to extend my appreciation to all the members from AEMON project. Thanks to Rune Harald Hestmo, Hans Lange, and Szymon Bernat, for providing me all the valuable experimental data to support my research. Thanks to Vigdis Olden, an excellent and amiable leader of this team. I am not a conversational person, but she would often pay special attention to me at parties. And Einar Løvli Hidle, although you only worked with me for a short time as a Master student, your work on this project inspired me a lot. As well as other group members, Andre Böhme, Ove Sagen Adsen, Magnus Eriksson, Stian Arnesen, Mette Lokna Nedreberg, John Olav Skogås, Anette Brocks Hagen, Nikolai Marhaug. The group trip to Bergen is a valuable memory in my life.

Yu Wang

June 2023, Trondheim, Norway

Abstract

Rotating machinery is a vital component in maritime vessels and wind turbines, and bearing is one of the most important parts in rotating machinery providing support, reducing friction, and facilitating smooth operation. However, bearings are susceptible to wear, degradation, and various other faults as they are subjected to continuous operation, escalating into catastrophic failures over time. Hence, early detection of incipient damages is an effective way to avoid downtime and loss of revenue as well as to protect both assets and employees.

Today, manual inspections and condition monitoring (CM) are still frequently used, however, it is not easily feasible in autonomous and remote-controlled vessels and subsea installations. In such scenarios, remote CM becomes the preferred option, allowing for continuous monitoring over long periods of time. This requires sensor data processing and analysis techniques to identify patterns and anomalies that indicate the presence of damages. In recent year, significant advancements in Artificial Intelligence (AI) techniques have emerged, offering a promising solution to address this challenge. This has sparked considerable interest and discussion in the field.

Some research gaps are to be mentioned: (i) The majority of existing studies focus on vibration analysis; (ii) The unsupervised early damage detection is relatively understudied compared to the supervised paradigm; (iii) The real-time monitoring of bearing requires further investigation. It has been reported that the current vibration-based CM systems for rotating machinery have limited sensitivity and capability for detecting pre-failure damages. As a result, by the time damage is detected, catastrophic failure becomes imminent, leaving little to no time for adjusting operational parameters to prevent further damage. This often necessitates shutting down operations until repairs or component replacements can be carried out. As an alternative non-destructive monitoring technique, Acoustic Emission (AE) has been found superior to vibration monitoring, as it can pick up signals from early damage before it propagates to the

surface and become detectable by vibration sensors. Besides, AE detection also shows superiority in slowly rotating machinery where the hit energy is far too low to be detectable with vibration methods.

We aim to integrate the AE technology and power of AI algorithms, providing real-time insights into the condition monitoring of bearings in this Ph.D. work. The research process encompasses six research papers that contribute to the development of novel CM frameworks using AE signals and intelligent analytics. These frameworks are designed for early damage detection in machinery. By the introduction of sensitive and intelligent CM systems with real-time analytics capabilities, this research aims to advance the digitalization of the maritime sector. The goal is to reduce operational costs associated with maintenance.

The thesis is presented as a collection of publications that build towards the goal of intelligent early material damage detection of bearings with AE signals. The structure of the thesis is consisted of five components: The Chapter 1 provides a comprehensive overview of the background and research questions of this Ph.D. work. Chapter 2 investigates the existing literatures of related topics and outlines the research gaps and challenges of the present study. Chapter 3 summarizes the main contributions of each research paper. Chapter 4 elaborates the connections between the conducted research papers and the derived three research questions. A brief summary of the entire Ph.D. research and the prospect of future research are given in Chapter 5.

Abbreviations

AE	Acoustic Emission
AI	Artificial Intelligence
BC-GAN	Binary-Classification Generative Adversarial Network
CM	Condition monitoring
CNN	Convolutional Neural Network
CWRU	Case Western Reserve University
DL	Deep Learning
EFD	Early Fault Detection
EHI	Ensembled Health Indicator
EMT-GAN	Ensembled MT-GAN
FFT	Fast Fourier Transform
GAN	Generative Adversarial Network
HI	Health Indicator
HSE	History-State Ensemble
ICM	Intelligent Condition Monitoring
IEPF	Information Entropy Penalty Factor
iForest	Isolation Forest
IMS	Intelligent Maintenance System
KAt	Konstruktions-und Antriebstechnik
LSTM	Long Short-Term Memory
ML	Machine Learning
MT-GAN	Multi-task Generative Adversarial Network
MVW	Moving Variance Window
RF	Random Forest
RMS	Root Mean Square
RQs	Research Questions
RUL	Remaining Useful Life
SENC	Streaming data with Emerging new classes
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TL	Transfer learning
WPT	Wavelet Packet Transform
WT	Wavelet Transform

x

Figures

Contents

Chapter 1

Introduction

Condition Monitoring (CM) of bearings has been a hot topic for years in the field of prognostic and health management of rotating machinery. As an essential component in various types of machinery and equipment, such as maritime vessels, wind turbines, pumps, and motors, bearings are often subject to significant loads and stresses, which can cause wear, fatigue, cracks, and other damages over time. Studies have indicated that bearings are responsible for 30-70% of failures in various rotating components of the machine [1–3]. Taking the induction motor as an example, the investigation into bearing faults represents 51%, 41%, and 42% of the overall faults according to the research conducted by ASEA Brown Boveri, Institution of Electrical and Electronics Engineers, and Electric Power Research Institute, respectively, as depicted in Figure 1.1 [4]. It goes without saying that the CM of bearings is a crucial and significant task for maintaining the health and reliability of critical equipment, reducing costs and downtime, and improving overall operational efficiency and safety.
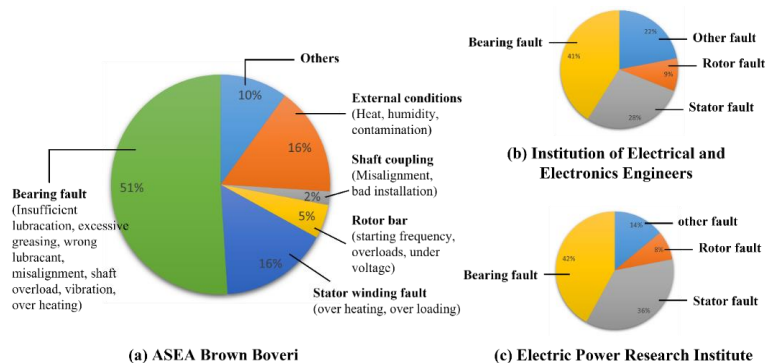
Figure 1. 1 A case study of failure rate of bearings in induction motors. The image is reproduced from [4].

In the past, manual CM methods such as frequent inspections were commonly involved to assess the health of bearings, and these techniques are still utilized today. Although these methods can yield valuable insights, they require substantial time and labor to implement and may lack precision. Furthermore, bearings are often located in hard-to-reach or inaccessible regions within machinery or equipment. For example, in autonomous and remote-controlled vessels as well as subsea installations, the systems are usually inaccessible and remote condition monitoring has become the only feasible option over long periods of time. As a result, there is a growing need for advanced tools to monitor bearings' health.

The Ph.D. work is targeted to explore differentiating technology for bearing CM system. The reminder of this chapter will provide a comprehensive overview of the background and research scope. Initially, we discuss the motivation behind our effort for Intelligent Condition Monitoring (ICM) systems as well as the utilization of the Acoustic Emission (AE) technique. We outline the main framework of the state-of-the-art intelligent data-driven analytics, which provides the reader a preliminary understanding of this topic. Subsequently, we introduce the research scope and highlight the primary research questions (RQs) that guide our work. Then, the main contributions and the corresponding research outcomes are briefly introduced. Finally, the thesis outline is provided.

1.1 Background

1.1.1 Intelligent condition monitoring

To begin with, let us have a look at the term 'intelligent' in ICM, which usually describes the capacity of a system or process to learn, adapt, and make decisions through the utilization of advanced artificial intelligence (AI) technologies, specifically Machine Learning (ML) and Deep Learning (DL) methods. The relationship between different AI disciplines and ICM is depicted in Figure 1.2. The ICM circle overlaps with both ML and DL. In practice, ICM involves collecting large amounts of data from various sensors that mounted on the targeted objects, such as vibration sensors, AE sensors, temperature sensors, and oil analysis data, etc. The acquired data is then processed by advanced ML or DL tools for analysis, for example classifying different

fault patterns or detecting anomalies to achieve the goal of monitoring and improving system performance and reliability. Therefore, ICM is an exemplary instance of a data-driven approach advocating that the sensor data itself provides the primary source of information for solving problems or making decisions, rather than relying on preconceived models or assumptions.
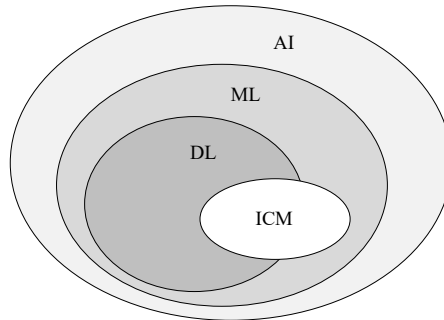
Figure 1. 2 The venn diagram of the relationship between different AI disciplines and ICM.

The general procedures of ICM comprise three steps, including *data acquisition*, *feature extraction*, and *decision making*, as illustrated in Figure 1.3. Initially, the data acquisition system in ICM works by collecting signals from various sensors that are mounted on the machine. A complete data acquisition system consists of three essential elements – Sensor, Signal Conditioning, and Analog-to-Digital Converter that converts the conditioned sensor data into digital values. The data are usually collected at regular intervals and stored in a database. Then, the acquired data can be processed by the utilization of ML and DL tools for feature extraction and decision-making.

In traditional ML paradigm, manually engineered features are often required to be extracted from time domain (such as kurtosis, skewness, shape factor, margin factor, etc.), frequency domain (such as mean frequency, frequency center, root mean square frequency, etc.), and time-frequency domain (involving signal processing techniques like Wavelet Transform (WT), Wavelet Package Transform (WPT) or Empirical Model Decomposition , etc. Then, feature selection or compression methods may be utilized to reduce the dimension and redundancy of the extracted features. In the decision-making step, the predictive model can be developed based on various ML algorithms, such as Support Vector Machine (SVM), Random Forest (RF), Extreme Learning

Machine, and artificial neural networks, depending on the specific application requirements. The output of the model can then be used to trigger alerts, schedule maintenance, or take other corrective actions to prevent catastrophic failures and optimize the system's performance.
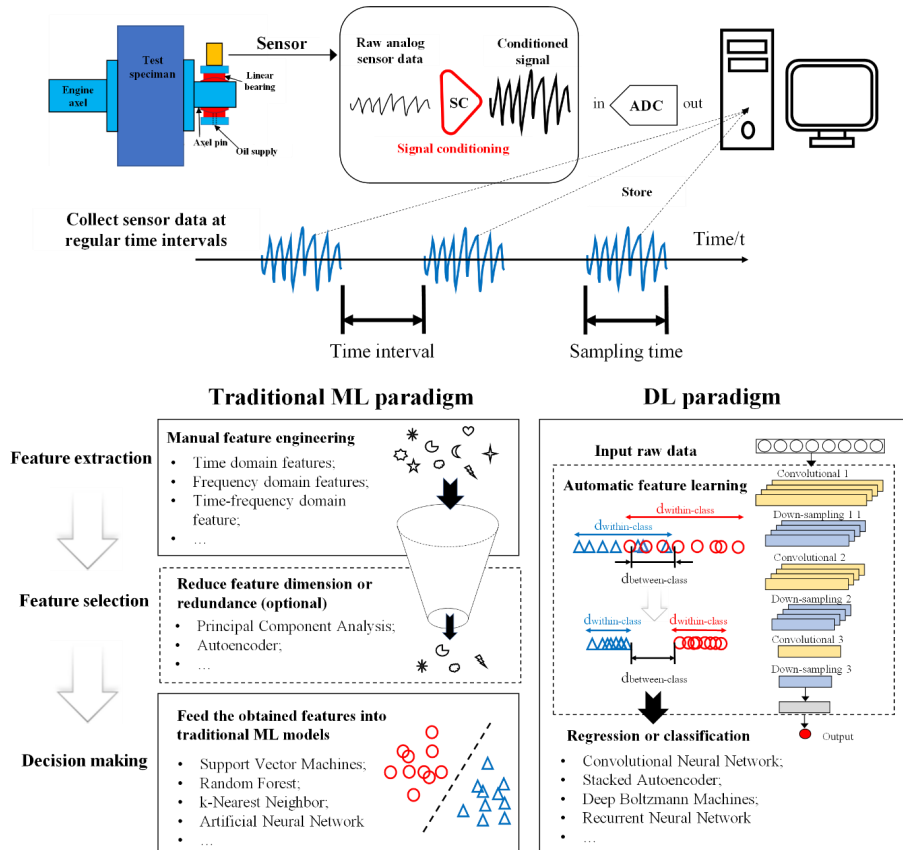
Figure 1. 3 General procedures of ICM.

As a sub-field of ML, the DL paradigm has become increasingly popular in the ICM. DL refers to a variety of artificial neural networks with numerous hierarchical layers, which enables the models to learn hierarchical and deep representations from sensor data. The highlight of DL paradigm is to eliminate the need for some data pre-processing steps that are typically required in traditional shallow ML methods, and enables *end-to-end modeling*, namely, the feature extraction and decision-making procedures are combined into a single process, as described in Figure 1.3. This

simplifies the ML pipeline and can lead to more accurate results by allowing the model to learn implicit features automatically from the raw input data.

According to a recent study investigating the applications of intelligent algorithms in the field of fault diagnosis [5], it has been found that DL paradigm has outperformed traditional ML methods. The conclusion is supported by the statistical study as presented in Figure 1.4, showing that the popularity of DL has surpassed that of traditional ML approaches since 2017, and this trend is projected to continue well into the future. This highlights the superiority of the DL approach and its potential in solving complex problems in various domains. In contrast to the traditional ML paradigm, DL paradigm offers the following advantages:

- *End-to-End modeling*: DL models are capable of conducting end-to-end learning, which enables them to take raw data as input and directly generate a decision without the need for prior knowledge and expert experience. This aligns with the pursuit of model intelligence in modern industry, where automation and efficiency are highly valued.
- *Versatility*: Without manual feature engineering, DL models can be easily applied to various types of data acquired from different sensors, including vibration signals, AE signals, current signals, etc., making them suitable for a wide range of applications in different domains.
- *Higher Accuracy*: DL models have the potential to achieve higher accuracy in fault detection and diagnosis compared to traditional ML methods due to the capacity of learning complex patterns and relationships in large datasets.

However, DL-based approaches are still in their infancy, there are still some remaining challenges to be addressed. For instance, DL models still face many training problems such as overfitting or parameter sensitivity etc. Therefore, improved pattern recognition approaches based on DL still need to be further developed.

1.1.2 Acoustic Emission Technology vs. Vibration Technology

Sensor data is the carrier of the bearing condition information. At present, vibration sensing is the most widely used technology for ICM, owing to its simplicity, robustness, and widespread availability. However, the vibration based ICM system can only

respond to the detectable vibrations caused by severe material failure such as large cracks or surface defects, and thus, have limited sensitivity and capability for detecting pre-failure damages. As a result, when damage is detected, full-scale failure is often imminent, leaving little or no time for adjusting the operating parameters to prevent further damage [6]. This is why we seek alternative technologies.
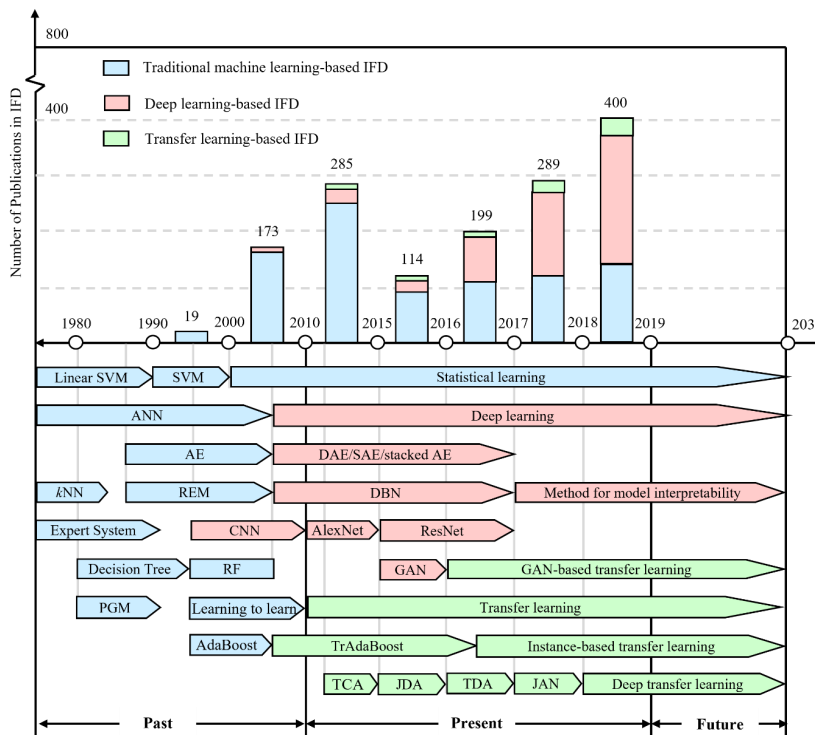
Figure 1. 4 Development and milestones of intelligent fault diagnosis using ML. The figure is cited from [5].

As an alternative non-destructive testing technique, AE signals exhibit a substantially broader frequency range (100 kHz to 1 MHz) [7], which does not significantly coincide with low-frequency mechanical vibration signals generated by machine component imbalance or misalignment. In addition, AE can be generated by microscopic flaws such as breaks in hard non-metallic inclusions or incipient cracks, providing insight into the dynamics of sources under load. Plastic deformation and fracture associated with the nucleation and growth of cracks represent the primary mechanisms of the sources releasing the elastic strain energy associated with AE transients [8]. The AE signal tends

to increase as the scale of the sources grows. A great deal of evidence has been accumulated, suggesting that AE parameters can reveal the faults in rotating equipment before they show up in the vibration acceleration range [9,10]. Hence, the potential of the AE technique for ICM is increasingly acknowledged in the industrial field.

1.1.3 Objectives

The objective of this project is to investigate the combined potential of AE technology and AI disciplines for ICM of bearings. While DL has exhibited remarkable success in end-to-end modeling with vibration signals in ICM, its application to AE signals is limited [11]. One of the challenges is that AE monitoring generates vast amounts of data that require extensive processing due to the high sampling frequency. This makes it difficult to apply DL directly to raw AE signals. As a result, single-parameter characterization of the AE waveforms, such as the Root Mean Square (RMS), AE counts, rise time etc. has been investigated by researchers. While they require little computational cost, they have been found unreliable for the characterization of the onset and propagation of damage. The limitations of these approaches have prompted researchers to move from time to frequency and then to time-frequency domain by using a variety of spectral transformations. Although the results from these later approaches appear more reliable, these manually engineered parameters require extensive prior knowledge and expert experience. Furthermore, a specifically designed feature for one machine is hardly applicable to other applications. These limitations call for further research and exploration of the potential of DL methods in AE waveform analysis.

To this end, this research aims to explore a more versatile method with the focus of early warning of damage based on AE technology and DL paradigm. We aim to develop an end-to-end ICM framework, which is capable to address the following practically relevant challenges:

- Extraction of implicit features from historical AE signals.
- Timely detection of the emerging pre-mature failure types.
- Continuous learning for monitoring the progression of failures.
- Pattern recognition of these detected failure types with high confidence.

With these capabilities, the model can be effectively applied to online ICM of bearings, enabling it to significantly reduce operational costs associated with maintenance activities.

1.2 Ph.D. project scope

The PhD project is a component of the 'AEMON - Novel Failure Monitoring System for Marine Applications by including Acoustic Emission ' project which aims to develop a cutting-edge failure monitoring system for marine applications using AE technology. The project involves research partners from NTNU (education) and SINTEF (industrial research), as well as industry partners Kongsberg Maritime, Rolls-Royce Marine, Equinor, and Island Offshore.
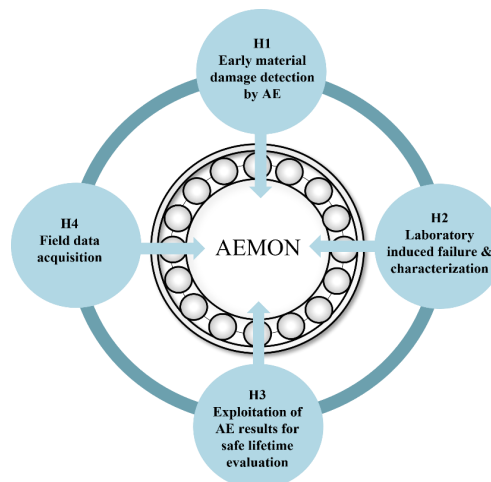
Figure 1. 5 Work packages of AEMON project.

In AEMON, we propose to investigate a collection of models that make few assumptions on the damage mechanism. By using such models, we aim to arrive at a collection of indicators, evolving over time, that cover a wide range of mechanisms under a variety of operating conditions. The indicators will flag the presence of damage, and the evolution of the indicators will provide an estimation of the time to damage escalation.

Figure 1.5 presents the main work packages under the entire project and the corresponding responsible partners. The Ph.D. project is part of the activity H1, which

is responsible for construction of intelligent models for *early material damage detection*. This requires the utilization of signal processing techniques, improved ML or DL tools, and other data analytics to identify patterns and anomalies that indicate the presence of damages. Additionally, the research will involve the design and testing of prototype software systems to acquire and process AE signals in real-time. The target is to build a reliable and effective method for early material damage detection of bearings in online ICM systems, which can improve the safety and efficiency of marine machinery, reduce downtime, and minimize maintenance costs. In addition, the results of this research have the potential of being extended beyond AE analysis and can be implemented in other types of sensor sources and industrial sectors involving rotating machinery.

The Ph.D. work is closely related with the activity H2, which is responsible for the production of subsurface cracks by rolling contact fatigue and propagate these to the material surface while monitoring the AE activity. Two lab-scale test rigs equipped with AE monitoring systems (a Roller Bearing test rig and a Ball Bearing test rig) have been built to generate sub-surface damages and acquire AE signals. The damage reproduced at lab scale will be compared to field data in H4 for confirming their relevance and similarities with real life damages. The H3 will analyze the damage produced in H2 and evaluate the effectiveness of the method developed in H1.

The thesis is guided by the following Research Questions (RQs):

- **RQ 1**: How to derive useful information from AE signals using intelligent analytics that can reveal the emergence and development of damages in bearings?
- **RQ 2**: How to perform end-to-end modeling that minimizes the model's dependence on expert assumptions and prior knowledge, while facilitating intelligent learning and decision-making processes?
- **RQ 3**: How to incorporate the model trained on a closed dataset into an open environment for real-time online monitoring?

1.3 Contributions and publications

The methods and discussions presented in this thesis are founded on the outcomes of five journal papers and a conference paper conducted in this PhD work during three years of study, from 2020 to 2023, as listed in the following.

- **[Paper A]** Yu Wang, R. H. Hestmo, A. Vinogradov, Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder, Measurement Science and Technology, vol. 34, pp. 064008, March, 2023 [12].

- **[Paper B]** Yu Wang, A. Vinogradov, Simple is good: Investigation of history-state ensemble deep neural networks and their validation on rotating machinery fault diagnosis, Neurocomputing, vol 548, pp. 126353, September, 2023 [13].

- **[Paper C]** Yu Wang, A. Vinogradov, Improving the Performance of Convolutional GAN Using History-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Signals, Applied sciences, vol. 13, no. 5, pp. 3136, February, 2023 [14].

- **[Paper D]** Yu Wang, S. Bernat, A. Vinogradov, BC-GAN: A Threshold-Free Framework for Unsupervised Early Fault Detection in Rotating Machinery, 2023. (Under review) [15].

- **[Paper E]** Yu Wang, Wang, Q.; Vinogradov, A. Ensembled multi-classification generative adversarial network for condition monitoring in streaming data with emerging new classes, 1st Olympiad in Engineering Science – OES 2023. Accepted. Presented. (Olympiad Medal/Best paper award)

- **[Paper F]**. Yu Wang, Wang, Q, S. Bernat, Vinogradov, A. Ensembled multi-task generative adversarial network (EMT-GAN): a deep architecture for classification in streaming data with emerging new classes and its application to condition monitoring of rotating machinery, 2023. (Under review) [16]

1.4 Thesis outline

The first chapter of this thesis introduced the research background and explained the motivation behind it. Furthermore, the subsequent chapters of the thesis are outlined as follows:

Chapter 2 presents an overview of the research focus of this thesis by outlining the main sub-tasks of the developed ICM framework and reviewing the existing literature. The chapter identifies the key challenges and research gaps in this field to provide a comprehensive understanding of the state-of-the-art research in this area.

Chapter 3 provides a comprehensive summary of the primary contributions made through the completion of the six research papers during the Ph.D. work.

Chapter 4 is dedicated to providing a comprehensive understanding of the connections between the developed methods and how they contribute to answering the proposed RQs.

Chapter 5 summarizes the main findings of the Ph.D. work and provides suggestions for future research activities in the field.

Appendix A contains the main contribution body.

Appendix B provides the link to the GitHub repositories of corresponding work.

Chapter 2

Literature review

In the previous chapter, we have briefly introduced the ICM paradigm and advantages of AE signal in early material damage detection over traditional vibration-based analysis. The significance and necessity of combining the two techniques has been revealed. This chapter will delve into the two techniques in terms of theoretical foundations, methodologies, and implementations in the existing literatures, highlighting their strengths, limitations, and challenges.

2.1 Acoustic emission-based monitoring of bearings

AE is commonly defined as a phenomenon whereby transient elastic waves are spontaneously emitted by the rapid stress relaxation within localized sources in material under load [17]. Defects occurring at different locations within a bearing, such as the inner race, outer race, or roller, exhibit distinct characteristic frequencies at which bursts of AE are generated [18]. Consequently, the AE signal from a damaged bearing comprises periodic bursts that correspond to these characteristic frequencies. These bursts serve as indicators of the presence of defects, including cracks, plastic deformation, or fractures, etc.

2.1.1 AE hit parameters

AE hit parameters refer to these widely used metrics for analyzing and interpreting AE waveforms. The most commonly used AE hit parameters for diagnostic include AE

counts, amplitude, duration, rise time, RMS, and some waveform shape descriptors, some of them are depicted in Figure 2.1.

Significant work has been dedicated to establishing the correlation between these parameters and AE events. Tan reported that AE events increase exponentially with defect size in bearings [19]. A study in [20] revealed that AE amplitude has a positive correlation with the defects in bearing outer race, however, this doesn't apply to inner race defects. Choudhury and Tandon conducted tests on bearings with different defect sizes. Their findings revealed that the ringdown counts of AE signals proved to be a highly effective parameter for detecting small-sized defects in both the inner race and roller of bearings [8]. In the work of Morhain and Mba, they performed the first known attempt to determine the appropriate threshold level for AE count diagnosis [21]. Besides, a clear relationship between the parameters RMS, amplitude and energy with speed, load and defect size has been reported in their research, particularly on outer race defects. He et al. further investigated the correlation between various AE parameters, including AE counts, amplitude, energy, and kurtosis, and the running condition of rolling element bearings [11]. They highlighted the impact of rotating speed to AE activities. Pachaud investigated two waveform shape descriptors, crest factor and kurtosis, and found that the kurtosis is more sensitive than crest factor in detecting the appearance of defects [22]. In addition, the successful application of AE parameters in slowly rotating rolling bearing can be found in [23,24]. Different time-domain features were extracted in [25] for fault diagnosis of low-speed bearing. Motahari-Nezhad and Jafari utilized sixty time-domain features, including kurtosis, energy, peak, to perform remaining useful life prediction of bearings [26]. Their results showed that kurtosis has the highest accuracy. Furthermore, it is worth mentioning that additional advanced AE hit parameters have been extensively studied to comprehensively characterize the AE waveforms from various perspectives. Detailed information regarding these parameters can be found in [27–32], here we do not to list all of them for the sake of brevity.
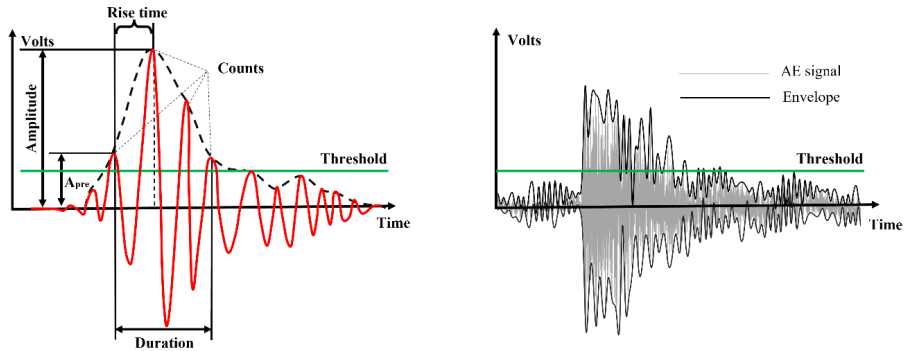
(a) Illustration of AE hit parameters on one event.

(b) Visualization of of a Genuine AE Event.

Figure 2. 1 Typical AE signal.

Although parameter-based AE descriptors are economical and facilitate high computation speed, they have been found unreliable in many practical applications. Raw AE waveforms encapsulate a blend of information derived from mechanical interactions, including rotating components, splashing oil, electrical interferences, and many other noise-like sources of unknown origin. Consequently, the true underlying AE events associated with damages are often masked by heavy ambient noise, making it challenging to identify them. Under such circumstances, utilizing AE hit parameters becomes impractical. In order to filter out noise and extract useful information, the raw AE signals are frequently required to be processed by advanced signal processing techniques.

2.1.2 AE-based signal processing

A wide range of research has been conducted to explore and develop signal processing methodologies that span across the frequency domain, time-frequency domain and other signal processing techniques, like cyclostationary analysis, for the extraction of valuable information from AE signals.

Fast Fourier Transform (FFT) is a widely used technique to interpret the AE signals from frequency domain. Einar et al. applied a high-pass filter based on FFT to denoise the raw AE signal, which amplifies the AE events in time domain [33]. The frequency domain characteristics, including peak frequency, frequency centroid, frequency band, and spectral kurtosis, provide valuable insights into the features of AE sources [17]. A

series of FFT features have been explored [34]. Gowid et al. developed a new FFT-based segmentation and features selection algorithm which yielded a high detection accuracy of 100% [35]. Wang et al. reported an experimental investigation into the relationship between the time–frequency patterns of AE signals and various cutting mechanisms under different speeds and fiber orientations [36].

Time-frequency domain analysis involving Short-Time Fourier Transform (STFT) and WT is another popular approach. Pham et al. adopted STFT to create spectrogram images as the representation method in the time-frequency domain, which is further analyzed by neural network [37]. Gao et al. utilized Empirical Wavelet Transform to decompose AE signals for extraction of the time-frequency domain features [38]. Liu et al. performed WPT to decompose bearing AE signal at different scales [39]. Bianchi et al. showed that WPT is a powerful technique to filter out noise from AE signals [40]. Law et al. proposed an advanced signal processing method combining the Wavelet Packet Decomposition and Hilbert–Huang transform for spindle bearings condition monitoring [41].

Besides, a number of methods have been investigated for signal processing of AE signals as alternative feature extraction approaches, including the utilization of Empirical Model Decomposition [42] and Variational Mode Decomposition [43], etc. Nikula et al. utilized autocorrelation to distill buried information from statistical features calculated from band-pass filtered signals based on short time windows for low-speed bearing fault diagnosis [25]. The analysis of cyclostationarity in AE signals had been demonstrated an effective method to interpret AE signal in bearing fault diagnosis. Kilundu et al. reported an experimental study of the Cyclic Spectral Correlation to characterize the cyclostationary aspect of AE signals recorded from a defective bearing [44]. Antoni provided a comprehensive introduction of cyclic spectral analysis methods, and validated the effectiveness of Spectral Coherence method in bearing damage detection [45]. Ni et al. proposed a novel fault energy based correntropy method to analyze the fault-irrelevant impulsive and cyclostationary interferences in bearings [46]. In [47], a novel Adapted Spectral Coherence method is investigated to identify faulty sources of wind turbine main bearing that are buried under multiple disturbances.

The AE-based signal processing technique is commonly reliant on the availability of prior knowledge regarding the specific machine and fault characteristics. This requirement poses a challenge as the approach heavily relies on the expertise of individuals, which can sometimes be vaguely defined. Especially in the case of online monitoring, unforeseen information or fault types may appear, so it is virtually impossible to design the best-suited features for unforeseen fault types. This case will be discussed in Section 3.

2.1.3 AE-based intelligent condition monitoring

We have outlined the traditional AE hit parameters and prevalent alternative feature extraction methods associated with signal processing techniques in the previous sections. The above-mentioned methods are classified as traditional methods in contrast to AI-based paradigms. The combination of AE technique and intelligent analytics can be found in existing literatures with applications ranging from EFD, fault diagnosis, and Remaining Useful life (RUL) prediction.

Applications to EFD: Extensive research has provided compelling evidence supporting the effectiveness of AE signals in detecting early damages that would otherwise remain undetectable by vibration signals [9,10]. Konig et al. introduced a ML-based anomaly detection method for wear monitoring in sliding bearing systems. In their approach, they employed an autoencoder as a feature extractor to identify anomalies [27].

Applications to fault diagnosis: Fault diagnosis has been the most studied sub-field. With traditional statistical features of AE signals as input, Omoregbee and Heyns leveraged SVM and Genetic Algorithms to classify different faults in low-speed bearings [48]. The k-Nearest Neighbor was adopted in [49] for fault diagnosis of rolling element bearing. Meserkhani et al. developed an artificial neural network based classifier to diagnose AE signals from four fault conditions in angular contact ball bearings [50]. The study aimed to compare the performance of different AE sensors in accurately detecting and classifying these fault conditions. It has been pointed out that traditional ICM paradigm requires handcrafted features that are independent of substantial human expertise and domain knowledge [51]. To this end, Hasan et al. applied Acoustic Spectral Imaging method to transform the one-dimensional AE signals into images. By leveraging the Convolutional Neural Network (CNN) for end-to-end

learning, the study enhanced the automated processing and interpretation of AE signals. Similarly, by representing the acquired AE signals as spectrogram, Pham et al. employed a CNN-based network to achieve accurate bearing fault diagnosis [37]. Inspired by sequential k-means, Pomponi and Vinogradov proposed a novel cluster method for real-time monitoring using AE signals [52].

Applications to RUL prediction: RUL refers to the estimation of the remaining operational lifespan or time until failure. Elforjani and Shanbr proposed a ML-based prognosis method based on AE signals [28]. A new fault indicator referred to as Signal Intensity Estimator was proposed in this study, and the authors employed three comparative methods, including artificial neural networks, SVM, and Gaussian Process Regression methods simultaneously, to highlight the feasibility of the artificial neural networks model for estimating the RUL of slow natural degrading bearings. Authors in [26] proposed an AE-based technique that extracts informative features from the time domain. Feature selection was firstly performed by an improved distance evaluation method, and then, the selected features are further possessed by k-Nearest Neighbor classifier to achieve RUL prediction by taking the problem as a binary classification between 'healthy' and 'faulty'.

The integration of AE technology with intelligent analytics has significantly contributed to informed decision-making in maintenance and repair strategies. Despite notable advancements, it is important to address certain limitations in current research. Firstly, the combination of AE technique with AI-based paradigms remains relatively understudied compared to the extensive research on vibration signals. While end-to-end learning models have been successfully established for vibration signals, their application to AE signals is hindered by the high dimensionality of AE data. Consequently, the reliance on manually engineered features and signal processing techniques is prevalent in AE analysis, necessitating a high level of expertise and experience from analysts. These limitations underscore the need for further research and development in the integration of AE technology with advanced AI methodologies to overcome these challenges and enhance the effectiveness of AE-based intelligent decision-making processes.

2.1.4 Summary

The first part of the literature review focuses on AE-based monitoring of bearings. We provide a concise introduction to the traditional parameter-based AE analytics, starting with conventional AE hit parameters and signal processing techniques. Subsequently, its applications in ICM of bearings are introduced, as well as the associated challenges with present AE-based monitoring. By addressing these aspects, this section aims to establish an overview understanding of traditional AE-based monitoring methods. In the following section, we will delve deeper into the existing ICM paradigms.

2.2 Intelligent condition monitoring of bearings

ICM is a continuous process of identifying changes that are indicative of the developing failures in the system using intelligent analytics. The signals acquired from the deployed sensors are fed into the monitoring system in real-time and analyzed to provide insights into the health and performance of the equipment. The core responsibilities of ICM encompass a range of sub-tasks, they include but not limited to: (i) Detecting the initial signs of early damage, which is also frequently referred to as 'early fault detection (EFD)' problem; (ii) identifying various fault patterns, commonly known as 'fault diagnosis'; (iii) monitoring the development of the damage rate; (iv) design and testing of prototype model to process the acquired signals in real-time in order to accomplish the aforementioned tasks. The main objective of this section is to outline the state-of-the-art methodologies utilized in ICM.

Table 2.1 ML-based paradigm vs. DL-based paradigm, reproduced from [53].

Method	ML-based paradigm	DL-based paradigm
End-to-end Optimization	×	√
Tailored Representation Learning	×	√
Intricate Relation Learning	Weak	Strong
Heterogeneity Handling	Weak	Strong

ICM can be roughly categorized into two paradigms: ML-based and DL-based, as presented in Figure 1.3. A summary of this discussion of the two paradigms is presented in Table 2.1. One can see that DL-based paradigm has the advantages over end-to-end optimization, tailored representation learning, intricate relation learning, and heterogeneity handling. In the following, we investigate the application of the two paradigms by focusing on three key aspects that are closely related to this Ph.D. project: EFD, intelligent fault diagnosis (IFD), and the challenges with online monitoring.

2.2.1 Early fault detection (EFD)

EFD is a process of detecting and diagnosing faults or anomalies in a machine at an early stage before they can cause significant damage or failure. When tackling such a problem, the key is to identify the unexpected events as early as possible.
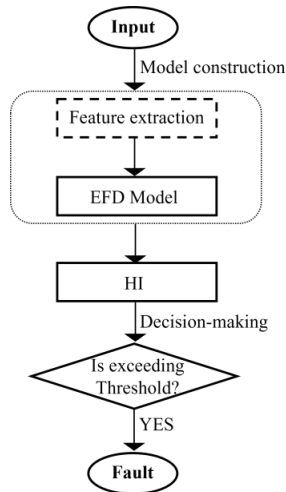
Figure 2. 2 Workflow of the general unsupervised EFD paradigm (**Paper D** [15]).

Unsupervised learning paradigm of EFD: Two major challenges are faced in the EFD of bearings: (1) The early faults tend to be too subtle for detection, making it more challenging to extract the underlying features; and (2) there are limited data that is accessible to train the model [54]. Hence, EFD is usually considered as a classical unsupervised problem based on the condition that only the data from the normal operation stage are available for training the surrogate models. The problem is therefore generally associated with novelty detection and anomaly detection. A conventional

approach involves extracting a HI that captures the degradation trend of the machine. By observing the changes in the associated HI value of AE signals, emerging fault can be detected. The decision-making process involves setting a threshold to define the boundary of normal operating conditions, as depicted in Figure 2.2. If the HI value of a signal exceeds this pre-set threshold, it is considered a fault alarm.

ML-based paradigm: Methods in this category focus on learning the boundaries of normal-state data using traditional shallow learning algorithms, which is typically followed after signal processing and feature extraction for downstream analysis. One-class SVM is a widely employed technique in the field of anomaly detection [55,56]. Saari et al. conducted a study where they employed One-class SVM for detecting bearing faults in windmills. In their approach, the model was trained using fault-specific features as input to automatically identify the faults [56]. Brito et al. utilized Principal Component Analysis to reduce the dimensionality of the extracted features in their study. Subsequently, they utilized the Isolation Forest (iForest) algorithm for the detection of early faults in bearings [57]. Mi et al. proposed a High-Level Extended iForest method in fault detection for rolling bearings [58].

DL-based paradigm: The category of methods leverages the power of DL algorithms to learn useful information from input data. Autoencoder is a common choice to tackle the problem. As mentioned in [27], authors utilized autoencoder to extract HI by learning a single hidden neuron in the bottle-neck layer. Another way to leverage the autoencoder is by making use of the reconstruction error between the input and output of the network. Related research can be found in [59], where the HI was constructed from the deviation of reconstruction errors. Similarly, Zhang et al. utilized the reconstruction error as HI based on multiscale CNN and gated recurrent unit Network with an Attention Mechanism [54]. Kong et al. leveraged a novel attention recurrent autoencoder to learn the most valuable features of input signal in an unsupervised way [60]. Shao et al. proposed complex wavelet packet energy moment entropy of vibration signals as indicator, they trained an Enhanced deep gated recurrent unit to map the relationship between the input data and the indicator for detection of early fault in bearings [61]. Song et al. converted the EFD as a sequence segmentation problem using CNN and a Simulated Annealing method to search for the optimal segmentation [62]. Their method is independent of manually designed features and pre-set threshold.

Limitations to be overcome: While AI analytics have found extensive use in the field of EFD, there are limitations in contemporary approaches: (1) As mentioned in the above sections, most existing methods are specifically designed for analyzing vibration signals and cannot be readily applied to other techniques, such as AE time series [11,63]. (2) Many existing methods rely on manually engineered features, and the performance of the trained surrogate models is highly dependent on the quality of these features. Consequently, the success of the approach heavily relies on the expertise of the analyst, and there is an increased risk of misinterpretation, particularly when inexperienced analysts are involved. (3) Most of the well-established models require pre-defined thresholds to distinguish between healthy and faulty states [62]. For one thing, determining an appropriate threshold value can be challenging, as it varies across different applications. For another, the threshold sets a hard boundary between normal and abnormal conditions, which can be easily affected by noises. These limitations highlight the need for further research and development to address these challenges and enhance the applicability of AI-based EFD methods.

2.2.2 Intelligent fault diagnosis (IFD)

IFD refers to the systematic process of identifying and determining the underlying causes or patterns of malfunction in a system or equipment using intelligent data-driven analytics to reduce human labor demand and cost [5]. It is the most extensively studied topic in ICM, allowing operators to not only detect mechanical irregularities buy also pinpoint the root cause of the issue and carry out repairs accordingly. In contrast to EFD, IFD is often approached as a supervised learning problem in many existing literatures [5,64–69]. These studies involve acquiring and labeling historical sensor data from various fault locations and conditions in order to create a comprehensive training dataset. Intelligent surrogate models are then trained to learn the correlation between the sensor data and corresponding labels, allowing them to accurately classify different fault conditions.

ML-based paradigm: Numerous ML-based fault diagnosis schemes have been proposed. k-Nearest Neighbor [49,70], SVM [71,72], RF [73–75], and Naive Bayes classifier [76] are to be mentioned amongst the most popular methods. A detailed review of the application of ML models targeted to practical machine fault diagnostics can be found

in [68]. In the past, the feature extraction and the fault recognition represented two separated procedures. Statistical features, which are derived from the raw signal acquired through various kinds of vibration sensors, include (but are not limited to) the mean voltage, RMS, shape factor, skewness, kurtosis, or entropy. These features may need manual tweaking before classification [49,71]. They are often linked to specific time, frequency or time-frequency domain-based signal processing techniques such as FFT, WPT, Empirical Model Decomposition and so on. The bearing fault diagnostics encounters two major challenges: (i) the noisy working environment, and (ii) variable operation conditions. The signal of interest, which is nonstationary and nonlinear, can be obscured by significant noise that is also nonstationary in nature. In such cases, the features selected and optimized for specific testing conditions may not necessarily be applicable to different situations. Therefore, the design and analysis of features, as well as their behavior, require careful consideration by an expert. This task is typically challenging and relies heavily on prior knowledge and experience.

DL-based paradigm: Since 2015, there has been a significant surge in the adoption and utilization of DL algorithms. and become the mainstream of IFD *de facto* [5]. In contrast to ML paradigms, the DL paradigms can extract features automatically from training data through multi-layered hidden neurons and achieve *end-to-end* modeling. We will delve into IFD of bearings based on DL paradigm from the following conceptual aspects: (i) deep feature learning, (ii) applications in variable working conditions, (iii) applications for imbalanced and insufficient data problem, and (iv) challenges.

Deep features learning: Deep features refer to the learned representations of data obtained from deep learning models, specifically deep neural networks. Autoencoder is one of the most popular network architectures widely employed in ICM. It employs an unsupervised approach to learn and capture underlying features by reconstructing input data. The use of a stacked network structure allows for the extraction of hierarchical deep features. Related works can refer to [77–80]. On the other hand, the Convolutional Neural Network (CNN) excels in extracting local features, leveraging sophisticated feature extractors called "kernels" or "filters". This has led to promising results in various applications of bearing fault diagnosis [62,81–86]. The Recurrent Neural Network [87] and its variants, such as Long Short-Term Memory (LSTM) [88–92] and Gate Recurrent Unit [54,93,94], are competitive in capturing the temporal correlations in time series data. These have also been found effective in bearing fault diagnosis

applications. Additionally, other network architectures like Sparse Filtering [95], Deep Belief Network [96], Deep Boltzmann Machine [97], Deep Fuzzy Neural Networks [98,99] are frequently utilized in this field to extract deep features.

DL-based applications in variable working conditions: With the successful applications of DL paradigms in IFD, increasing researchers are now shifting their attention towards tackling the problem under variable working conditions [81,100–104]. The dynamic nature of variable working conditions presents new challenges that need to be addressed in order to effectively apply DL techniques for fault diagnosis. Transfer learning (TL) has been identified as an effective solution for addressing this problem [105]. TL enables the transfer of knowledge acquired from a source machine or specific operational conditions to another machine or different operational conditions, which enhances the efficiency of fault diagnosis. Zhao et al. proposed a novel deep multi-scale convolutional transfer learning network to address the challenge [106]. Similar works can be found in [51,85,103,107–111].

DL-based applications in imbalanced and insufficient data problem: The issue of imbalanced data has garnered increasing attention in this field and has become a hot topic of discussion. It refers to the scenario where the distribution of samples across different fault classes is significantly unequal, which is a prominent concern in IFD. In real industrial settings, mechanical systems often operate normally for the majority of their service time, while the failure stage can be quick and catastrophic. It is, therefore, difficult to obtain faulty data. Zhang et al. conducted a review research over this problem based on intelligent analytics [112]. Generative Adversarial Network (GAN) has emerged as a potent solution for addressing this challenge by serving as a data augmentation method. GAN is utilized to generate authentic simulated data to supply these small-sized fault classes, effectively overcoming the insufficient data problem. Numerous studies have explored the application of GAN for this purpose, and relevant works can be found in [113–120]. In recent year, the emerging of Meta Learning and few-shot learning provide a new solution for this problem. Few-shot learning aims to learn from very limited number of labeled data to generalize and make accurate predictions on new, unseen classes or tasks. Zhang et al. leveraged Model-Agnostic Meta-Learning, a classical few-shot learning algorithms, to perform bearing fault diagnosis using only a few training samples for each fault class [121]. The readers are encouraged to explore more research findings in [122–125].

Challenges: Despite the reported remarkable success, the present paradigms of DL-based IFD still encounter many challenges. In this section, we would like to highlight the '*Generalization*' problem of the current DL methods, which can be crucial for practical application of DLs. As most of the proposed methods are tailored to specific datasets or application scenarios, which means the performance of a designed architecture can degrade on another task. An example can be found in [126], where a multi-scale cascade CNN witnessed a decline in classification accuracy from 99.7% to 96.9% when the authors tested the same network with varying kernel scales in the multi-scale learning layer. Furthermore, neural networks are susceptible to training parameters like the learning rate and the number of iterations. The question of enhancing the neural network's generalization and adaptability is still open and remains unresolved.

2.2.3 Challenges in online monitoring

Up to now, we have outlined some existing methodologies for EFD and IFD within ICM, let us delve into a specific problem with online monitoring that requires closer examination. Online monitoring refers to the continuous monitoring and assessment of the condition of equipment or systems while they are in operation, which is an especially difficult task. In traditional ICM paradigm, the surrogate models, as reviewed in the above, are trained on closed training dataset, namely, there are limited number of categories and samples in the training set. In practical scenarios, machines can exhibit various fault possibilities, each with distinct behaviors in sensor data. It is practically impossible to encompass all of them in the training dataset. Consequently, we can foresee that a model trained on a closed dataset will inevitably encounter new information during practical online monitoring that it has never encountered before. This issue highlights limitation of current methods in addressing online ICM, prompting us to seek new solutions to tackle this challenge. In the following, we will introduce a new concept that can be leveraged to address the problem.

2.3 Streaming data with emerging new classes (SENC)

The above-mentioned challenges faced with online ICM can be encompassed by the problem of streaming data with emerging new classes (SENC), namely, by addressing the SENC problem, we can effectively tackle these challenges within a unified framework. In the below, we will discuss the definition, applications, methodologies, and challenges of SENC problem.

2.3.1 What is SENC problem?

The SENC problem is generally described as: given a training dataset $T = \{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in R^d$ denotes the i-th training instance and the corresponding label $y_i \in \{1, 2, \dots c\}$, c stands for the number of initially known classes. The goal is to build a multi-classification model based on T so as to classify the streaming data $S = \{(\tilde{x}_t, \tilde{y}_t)\}_{t=1}^{\infty}$, where $\tilde{x}_t \in R^d$ is a streaming instance at time t, and the corresponding label $\tilde{y}_t \in \{1, 2, \dots c, c+1, c+2, \dots c+k\}$, k is the number of emerging new classes. As a result, a SENC model is required to fullfill the following tasks in an integrated framework: (1) *pattern recognition* of these already known classes with high accuracy; (2) timely *detection of emerging new classes*; (3) *model update* to adapt to new classes, and (4) *discriminating* between different new classes. Due to the growing need to minimize dependence on human labor and expert knowledge in modern industries, a significant challenge has emerged as (5) the model must be capable of extracting features and making *data-driven decisions* automatically on the basis of the original signals.

By addressing SENC problem, we can significantly enhance the intelligence and reliability of online monitoring systems as the model can dynamically evolve and adapt to changing tasks and conditions. Figure 2.3 summarizes the connections between SENC problem and ICM. In accordance with the sub-tasks of ICM, we divide the SENC problem into three scenarios with different settings of known classes c and emerging new classes k.

- When $c = 1$, and $k = 1$.

This particular scenario presents a unique case of the SENC problem with a single known class and a single emerging new class. In practical applications, this scenario is associated with EFD in the run-to-failure process. In this context, instances from the normal operating stage are utilized as the training data, and a model is trained to detect any deviations from the normal behavior that could indicate an impending fault. By addressing this problem, the CM system can improve its efficiency and reduce maintenance costs.

- When $c > 1,$ and $k = 0.$

This scenario is a special case of the SENC problem where there are no new classes that have emerged. It is directly related to the challenge of IFD in CM. The objective is to develop a model that can accurately classify the different fault types and differentiate them from each other based on the sensor data acquired from the system. Achieving accurate classification is crucial for timely and effective maintenance, repair, and replacement of the faulty components to avoid costly downtime and potential safety hazards.

- When $c \geq 1,$ and $k > 1.$

The most challenging case for real-time online monitoring is the standard SENC scenario where there are multiple known classes and emerging new classes in the streaming data. This scenario requires the model to continuously update itself to adapt to new emerging classes, which is also known as the incremental learning problem. To address this challenge, adaptive algorithms are needed to continuously update the model as new data arrives. This problem is associated with monitoring the emergence of new fault types or the development of damage rate, etc.
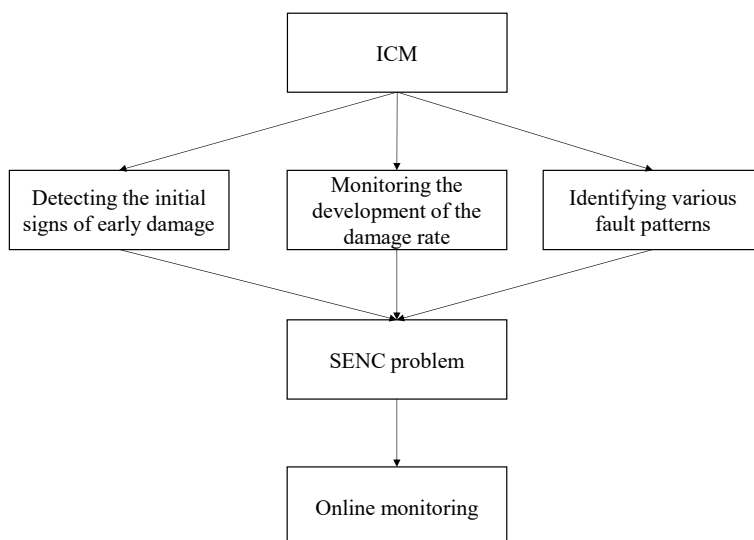
Figure 2. 3 The interconnections between SENC problem and ICM.

2.3.2 Existing methodologies

The SENC problem remains understudied especially within the field of ICM. Here we introduce some state-of-the-art frameworks to address this challenge. Janardan and Mehta investigated the concept drift problem in streaming data and explored various classification algorithms and platforms that are useful in their research [127]. They pointed out that many of these tools are very young, and further research is required to effectively benchmark and evaluate the wide range of available options. Din et al. conducted a comprehensive review research of this topic, and summarized a general workflow for SENC problem [128], as depicted in Figure 2.4. The basic idea is to train an initial model that can learn the boundary of known classes, and a pre-defined threshold is then utilized to identify emerging new classes. Upon detection of a new instance, it is stored in a fixed-sized 'buffer'. Once the buffer reaches its capacity, a model update is triggered to update the existing model. Based on the workflow, some of the state-of-the-art methods are to be mentioned.
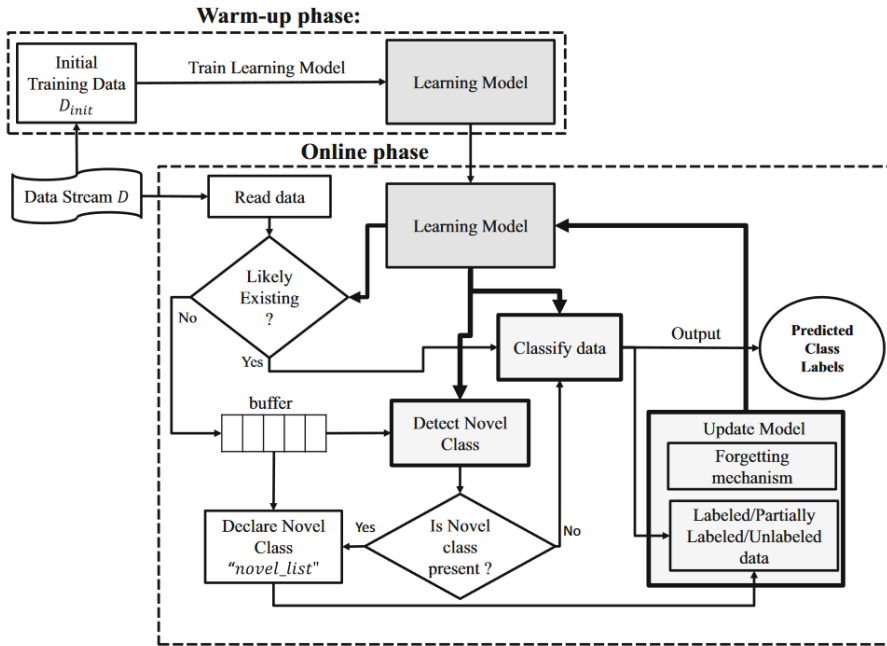
Figure 2. 4 The general workflow for SENC problem, cited from [128].

A group of tree-based methods have been investigated [129–131]. Gama et al. extended the Very Fast Decision Tree algorithm to deal with continuous data [129]. Mu et al. proposed a method known as SENCForest, which utilizes the assumption that anomalies from established classes are relatively more 'normal' compared to newly emerging classes. This approach involves constructing an 'outlying' anomaly subregion using a completely-random tree method, with the aim of detecting these emerging new classes [131]. To establish a unified framework for the SENC problem, the multi-classification process is realized by recording the class labels in each leaf of the pre-trained trees.

Cluster-based methods provide an alternative solution [132–135]. Gao et al. suggested that new instances should exhibit a distinct feature distribution locally in comparison to instances from the existing known classes [132]. They introduced a novel SACCO framework for SENC, which utilizes an ensemble of the fast graph-based clustering method. However, this assumption can be impractical in real-world applications, as new classes can often be in close proximity to the known classes. For instance, in the case of the EFD problem, distinguishing early fault information from normal operating

conditions can be very challenging. . Cai et al. pointed out this question and proposed the introduction of a Class Separation Indicator α to measure the difficulty of SENC tasks [133]. They emphasized that most existing methods perform optimally on high-α SENC tasks, where there exists a substantial geometric distance between the existing and new classes, and the performance of these methods degrade on low-α SENC tasks. To address this, a cluster-based ensemble method called SENNE is introduced in their study. Zhang et al. further improved their method by introducing a new KNNENS model [134].

Thus far, most existing methods rely on shallow learning approaches, which exhibit subpar performance when applied to original mechanical signals. Consequently, the incorporation of manually designed features becomes necessary. Zhou et al. introduced a neural network-based method that is capable of handling raw signal data, referring to as LC-INC [136]. However, this framework is only applicable when there are at least two known classes present in the training dataset. As a result, it is unable to address problems that involve only one known class at the initial stage, and EFD problem falls into this category.

It is worth noting that the aforementioned methods are mostly tailored for domains other than ICM of bearings or rotating machinery. The field of ICM in this context remains relatively new and understudied. A few relevant works are to be mentioned. Zhang et al. proposed an ENL-SAE framework that utilizes the sparse autoencoder for fault diagnosis, particularly with regards to emerging new classes [127]. In another study, a deep adversarial transfer learning network was introduced for emerging new fault detection [128]. This method was specifically designed to address the imbalance in transfer learning when the target domain encompasses more fault types than the source domain. However, it does not account for situations where the emerging new faults are completely absent from the training dataset.

2.3.3 Challenges

Some limitations of contemporary SENC frameworks are summarized:

(1) The lack of effective DL-based frameworks: Currently, most state-of-the-art SENC frameworks rely on shallow learning approaches and heavily depend on manually designed features. There is a need for more robust DL-based

frameworks that can effectively handle the challenges posed by SENC, eliminating the dependency on manually crafted features and improving overall performance.

(2) Limited applicability to ICM: Existing SENC frameworks primarily focuses on streaming image classification tasks, lacking specific methodologies tailored to the requirements of ICM. Consequently, there is a need for the development of SENC methodologies specifically designed for condition monitoring applications.

(3) Reliance on pre-defined thresholds: Existing SENC frameworks rely on pre-defined thresholds for novelty detection. However, these thresholds often need to be adjusted for different testing conditions, leading to challenges in rigorous threshold setting. Furthermore, the threshold-based mechanism shows limited performance in low-α SENC tasks and struggles to effectively differentiate between different emerging new classes.

Addressing these limitations is crucial to advancing the field of ICM, especially for online monitoring. Future research efforts should focus on developing DL-based frameworks, extending the applicability of SENC methodologies to online ICM of bearings.

2.4 Summary

This chapter encompasses the literature review of three topics: AE-based monitoring of bearings, ICM of bearings, and the introduction of SENC problem. These topics are aligned with the three RQs presented in Chapter 1. To provide a clear overview of the research gaps and challenges discussed, we present a summary in Figure 2.5, illustrating how these gaps and challenges are relevant to the RQs addressed in this study.
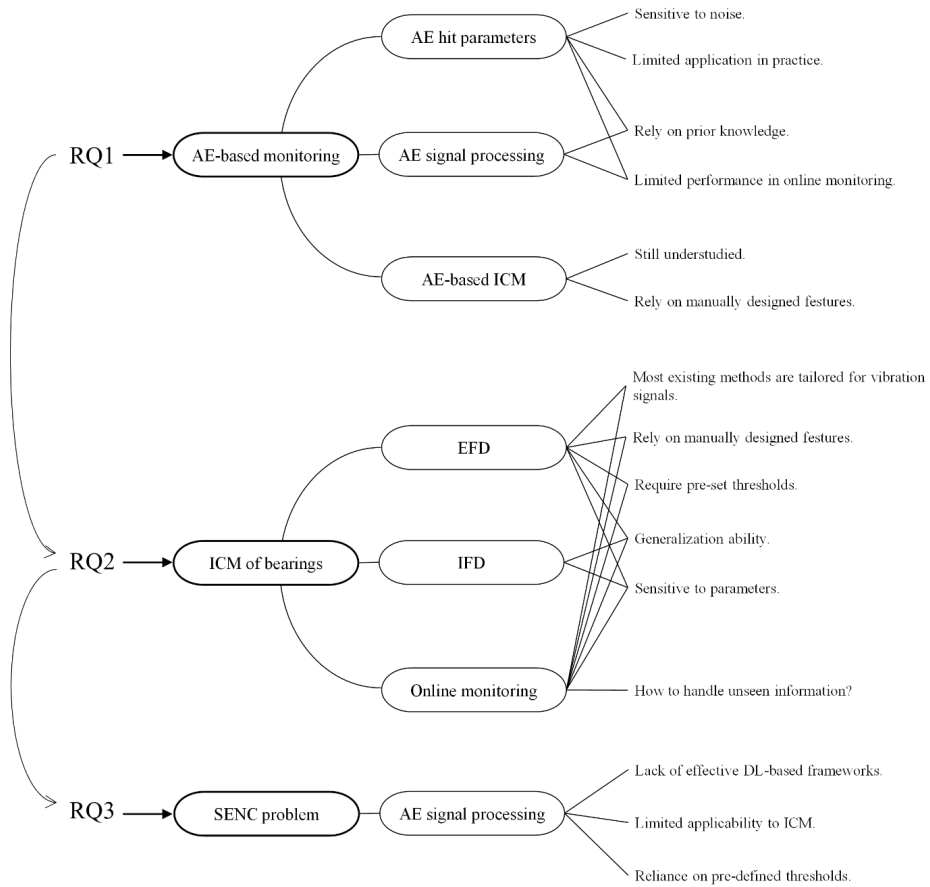
Figure 2. 5 A summary of the discussed research gaps and challenges.

Chapter 3

Contributions overview

In the previous chapter, we discussed the research gaps and limitations pertaining to the relevant topic in AE technology and ICM. The purpose of this chapter is to highlight the significant contributions made by this Ph.D. research work in addressing these gaps and establishing connections with the three RQs derived in Chapter 1. We begin by focusing on the EFD problem as a starting point and gradually delve into the investigation of conventional methods employed in this area. Through this initial exploration, we gained valuable insights that guided our research journey. Then, we critically reflected on the existing limitations and identified key research gaps that warranted further attention. The ultimate objective was to develop novel and more effective approaches for online ICM of bearings, surpassing the shortcomings of conventional methods.

Our work has led to the development of new algorithms and techniques that are capable of addressing the above-mentioned challenges including the EFD, IFD, and SENC problems. The research work has resulted in five journal papers and a conference paper. A visual representation of the interconnections between the conducted papers is provided in Figure 3.1. In the following sub-sections, the research objectives and contributions of each paper will be briefly introduced, and the connections of each work and their relevance to the target RQs will be presented as well.
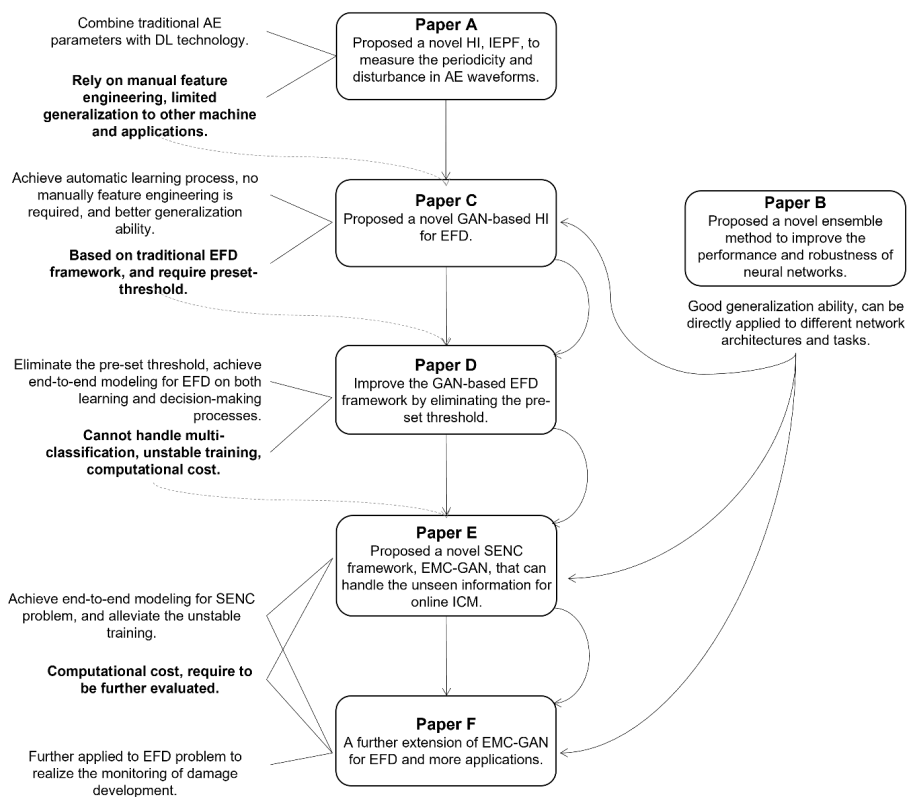
Figure 3. 1 A sketch of the connections between the papers produced in this Ph.D. research.

3.1 Paper A:

3.1.1 Objective

To begin with, we delved into the performance of conventional parameter-based methods, as they remain prevalent in the philosophy of AE waveform analysis. This provided a solid foundation for comparison with more advanced technologies utilizing intelligent algorithms. Following this, various data processing techniques, such as those based on time domain, frequency domain (FFT), and time-frequency domain (STFT, WT, and WPT) have been explored. The primary objective of this paper is to develop a sensitive HI that can effectively detect the onset of early damage in bearings during the run-to-failure process, which is crucial for predictive maintenance in many industries.

3.1.2 Relevance to the thesis

Throughout the research work presented in this thesis, the first and basic target was on answering the RQ1, *how to derive useful information from AE signals based on intelligent analytics that can reveal the emergence and development of damages in bearings?* Two keywords are highlighted: 'AE signals' and 'intelligent analytics'. Literature review in Chapter 2 uncovered a research gap, revealing that the combination of the AE technique with AI-based paradigms has received relatively less attention compared to the extensive research conducted on vibration signals. This paper marks our first attempt to address this question using DL technology.

3.1.3 Contributions

This paper addresses the existing research gap within **Section 2.1.3**, where limited studies have been conducted on the integration of AE technology with AI-based paradigms. To this end, we aim to provide a new parameter to characterize the evolution of the AE waveforms during a run-to-failure process in combination with DL method. To investigate the capacity of the AE technique in subsurface fault detection of bearings, a laboratory durability test of a roller bearing element was carried out; roller contact fatigue damage was initiated under controlled conditions, and the accompanying AE waveforms were acquired. The contributions are summarized as follows.

(1) A new qualitative parameter called Information Entropy Penalty Factor (IEPF) was proposed to monitor the evolution of the AE waveforms. The proposed parameter measures the periodicity and disturbance of the AE waveforms based on the assumption that the routine operation of the rotating machine results in periodical behavior in the AE waveforms. However, the AE signals generated by defects are paroxysmal and non-stationary, which will cause disturbance and deformation of the original waveforms. The proposed IEPF describes the AE waveforms from a different perspective, which contributes to the diversity of the existing AE parameters.

(2) The high sampling frequency of AE technology poses limitations on the application of DL methods. To overcome this challenge, the paper introduced the utilization of a Moving Variance Window (MVW) technique, which effectively reduced the dimensionality of the raw AE signals.

(3) An autoencoder-based deep neural network architecture was incorporated into the data processing pipeline. This integration effectively reduces noise from the AE signals, resulting in clearer and more prominent characteristics. We proved that reconstruction error itself is insufficient to reflect the initiation of early faults, and proposed a way of combining the reconstruction error and reconstructed signal to maintain the relatively complete information of the signal.

(4) We conducted a thorough examination of the effectiveness of conventional parameters derived from the time, frequency, and time-frequency domains. By examining the effectiveness of conventional parameters, we were able to establish a baseline for evaluating the performance of more advanced techniques and determine their potential value in developing sensitive HIs for early damage detection in bearings.

3.1.4 Results

To characterize the evolution of the AE waveforms during a run-to-failure process, a new qualitative parameter called IEPF was proposed to measure the periodicity and disturbance of in the signal. A series of experiments were carried out to assess the effectiveness of the proposed method. Our finding revealed that conventional parameters, such as RMS, skewness, crest factor, and impulse factor, were insufficient

to fully characterize the evolution of AE signals in relation to the initiation and propagation of damage. The superiority of the proposed IEPF feature against many other statistical features has been demonstrated. To be more specific, the proposed IEPF successfully characterized the waveforms changes of the recorded signal, and most importantly reflected the breakpoint between the healthy and damaged stages from the early detected crack. Besides, it was more sensitive to small disturbance of the signal in comparison with conventional statistical parameters, and thus, provided a better rate of fault alarms at the early stage the damage evolution.

3.2 Paper B:

3.2.1 Objective

Let us take a deviation from the EFD problem for a while and have a look at neural network and DL technology. While DL is revolutionizing CM in the 21st century, its application in practice still faces considerable challenges as many proposed networks are tailored for specific conditions or datasets. Neural network architectures mainly comprise four elements: (i) the number of layers; (ii) the number of neurons in each layer; (iii) the activation function of each neuron; (iv) the training algorithms. Due to the nonlinearity of the net, minor changes to any element can result in significantly different outcomes, making it difficult to guarantee consistent performance across tasks. Additionally, neural networks are highly sensitive to training parameters such as learning rate and iterations, necessitating careful consideration of these variables. The open question remains of how to improve the neural network's generalization and adaptability.

The focus of this paper is not on improving the architecture or training methods of deep networks, but rather on exploiting the potential of deep networks that has not been disclosed to a full extent. We can see that by properly utilizing these most common tools in modern neural network; the performance of deep networks can be further improved in a general sense. Additionally, the approach can also be easily applied to practical application due to its robustness and simplicity. This research is expected to benefit our following applications of neural networks across a range of tasks as well as real-time testing scenarios.

3.2.2 Relevance to the thesis

The second paper addresses RQ2 and focuses on enhancing the model's performance by increasing its robustness to hyperparameters. This improvement reduces the need for expert knowledge to fine-tune the optimal parameters, resulting in more accessible end-to-end modeling. The present paper endeavors to explore an effective and reliable approach to enhance the overall performance of deep neural networks. Specifically, the approach has the potential to (i) improve model accuracy, (ii) withstand variations in parameters, and (iii) work effectively across different network architectures. The findings derived from the present paper benefit our subsequent research in **Paper C**, **Paper E**, and **Paper F**, regarding RQ1 and RQ3, respectively.

3.2.3 Contributions

This paper proposed the ensemble technology as a viable approach to accomplish this objective. Traditional ensembled neural networks are perceived as costly in complexity as it involves intricate network architecture design, training approaches, and additional hyper-parameters to be tuned. To this end, an easy-to-implement ensemble technology was investigated in the present paper, referring to as the History-State Ensemble (HSE) method. The main contributions include:

(1) We proposed a straightforward yet efficient approach for performing ensemble learning on deep networks. The HSE-based approach offers several advantages: (i) it eliminates the need for additional training costs to generate multiple base models, making it a cost-effective solution; (ii) it can be applied to all types of neural networks without the requirement of tuning the network architecture.

(2) We conducted extensive investigation of various ensemble strategies applied to deep networks. Through comparative experiments, our proposed method demonstrated superior overall performance. The findings derived from this study offer valuable guidelines to practitioners for the rational selection of ensemble strategies.

(3) Furthermore, we presented a novel approach to explore the tradeoff between diversity and accuracy of base models within the ensemble. This unique perspective provides fresh insights into the underlying mechanisms of ensemble

learning, shedding light on how diversity and accuracy interact to influence ensemble performance.

3.2.4 Results

The proposed method was tested on eight prevalent deep network architectures and six datasets constructed from Case Western Reserve University (CWRU) Bearing Data Center and Konstruktions-und Antriebstechnik (KAt) - Bearing Data Center for bearing fault diagnostics. This paper decomposed the ensemble learning process into two components: (i) A training strategy that encourages the generation of accurate base models with diversity; and (ii) a learning strategy which combines acquired base models to form a stronger classifier. Over 20 ensemble strategies were investigated in this paper. Taken the accuracy and standard deviation as the metrics of model performance, extensive experiments demonstrated that the proposed HSE ensemble strategy exhibited higher classification accuracy and lower standard deviation, indicating the effectiveness and robustness of the methodology. These results supported the notion that the proposed approach could improve the overall performance of neural networks for a variety of datasets and deep network architectures.

3.3 Paper C:

3.3.1 Objective

Despite the successful detection of sub-surface cracks in roller bearing tests based on an effective HI described in our previous work, the approach is specific to the periodicity characteristics in AE signals, and thus, falls into the common limitations of conventional approaches as manually engineered feature. There are several reasons why we should avoid relying on manual feature engineering: (i) Such an approach is usually tailored to a specific problem or domain, which limits their applicability to other problems or domains. (ii) Manual feature engineering may not be able to capture intricate patterns in the data, especially if the data is high-dimensional or consists of non-linear relationships between the features. Moreover, (iii) this technique heavily relies on the domain expert's assumptions and prior knowledge, potentially introducing introduce human bias into the feature design. Thus, it can be particularly difficult during

real-time testing scenarios where prior knowledge about the potential damages may not be avaliable.

As a result, there is a growing need for intelligent models that can perform end-to-end learning, which involves directly learning from raw input data without the use of manually engineered features. Deep neural networks have emerged as a promising solution to this challenge. The objective of the present paper is to develop a novel end-to-end learning model that is capable of automatically learning high-level representations of AE signals. By doing so, we aim to minimize the need for expert's assumptions and prior knowledge in the learning process.

3.3.2 Relevance to the thesis

Our ongoing efforts on RQ1 for extracting useful information from AE signals are furthered in the third paper, where we reflect on the limitations of the methods probed in **paper A** and introduce a novel approach to address EFD with AE signals. In addition, the present paper also contributes to addressing RQ2, which aims to *reduce the model's dependence on expert assumptions and prior knowledge for achieving end-to-end modeling*.

3.3.3 Contributions

This paper contributes to address the research gaps in existing EFD methods that (i) Most existing methods are tailored for vibration signals, and (i) they often require manually feature engineering. To achieve the research goal of end-to-end learning, we developed a new HI that can directly process the raw AE signals, thereby reducing the need for human expertise in the modeling process. The approach follows the prevalent EFD framework by constructing a sensitive HI, and the decision-making process is realized by applying a threshold marking the boundary for the normal operating state, as depicted in Figure 2.2. The main contributions are summarized as:

(1) In this paper, a novel convolutional GAN was introduced and employed for EFD in a run-to-failure test of a roller bearing. To enhance the learning capacity of the Generator, an Autoencoder-based Generator architecture was devised. This architecture incorporated two convolutional blocks to extract local information

from the data, while two LSTM cells were integrated into the bottleneck layer to capture the time-series correlation of the signal.

(2) We proposed utilizing the output of the Discriminator as the HI to discriminate between normal and abnormal behaviors. To address the issue of unstable training in the GAN, a novel ensembled health indicator (EHI) leveraging the Discriminator score and HSE method (proposed in paper B) was constructed.

(3) Furthermore, the present paper re-evaluated the effectiveness of the HSE method on GAN and EFD task. This allows us to validate the applicability of the HSE method beyond its original implementation and provides additional evidence for its effectiveness. Overall, this paper represents a step forward in developing more effective and efficient methods for EFD with AE signals.

3.3.4 Results

The presented approach followed the prevalent EFD framework by constructing a sensitive EHI leveraging GAN for intelligent learning. The decision-making process was realized by applying a threshold marking the boundary for the normal operating state. The proposed EFD framework was applied to our roller bearing test rig, and results showed that the novel EHI provides higher accuracy (0.8732) than our previous research of IEPF (0.8583) on the same device and dataset. Besides, the present paper extended our previous efforts to study and promote the HSE method to non-destructive testing applications. Parameters analysis revealed that not only the HSE method improves the diagnostics of incipient flaws in specific rolling bearings elements under contact fatigue conditions, but it is also an efficient vehicle to enhance the performance and capacity of convolutional GAN in a general sense.

3.4 Paper D:

3.4.1 Objective

The pursuit of intelligent analytics is an ongoing progress, as there are constantly new challenges to address and opportunities for improvement. The objective of the present paper is to further improve the approach by achieving the intelligent 'decision-making

process', which refers to the ability of a model or system to identify patterns and make decisions without human intervention.

Up to now, an 'end-to-end' learning model that can process AE signals without the need for manual feature engineering has been presented in paper C, which is a significant step towards the intelligent 'learning process' of EFD. However, the contemporary approaches still have limitations in meeting the requirements for intelligent decision-making due to the need for a pre-set threshold to distinguish between normal and abnormal signals. As a standard approach used in EFD, the threshold-related issues have been seldomly discussed, they include: (i) The threshold is a hard boundary which can be easily impacted by noise; (ii) The values of the thresholds can vary for different testing conditions, leaving users with a thorny problem of rigorous threshold setting; furthermore (iii) the threshold may need to be updated as the system evolves, or new operating conditions are introduced. To this end, the strategic idea and motivation of the present paper is to develop a model with higher intelligence eliminating the threshold-related issues.

3.4.2 Relevance to the thesis

The present paper advances our previous research with the goal of developing a new technique that eliminates the need for a pre-set threshold in the EFD problem. The primary research question that this work aims to solve is '*reduces the model's dependence on expert assumptions and prior knowledge, while facilitating intelligent learning and decision-making processes?*' of RQ2 by enhancing the EFD process on the decision-making process. Additionally, the paper lays a solid foundation for future research that addresses the SENC problem corresponding to the RQ3.

3.4.3 Contributions

This paper makes a significant contribution by eliminating requirement for pre-set threshold in in existing EFD methods. The key contributions of this work are as follows.

(1) This paper introduced a novel and robust GAN-based network architecture specifically designed for EFD problems in rotating machinery. This architecture exhibits high versatility and can be applied to different condition monitoring

techniques, enhancing its applicability in various scenarios. To evaluate the performance of the proposed method, both benchmark vibration datasets and two experimental AE dataset were utilized. The experimental results obtained from these datasets demonstrated the feasibility and effectiveness of the proposed method in accurately detecting and diagnosing faults.

(2) The proposed method offers a notable advantage over traditional EFD models by not requiring a pre-set threshold. Instead, it directly produces the probability that the measured signal originates from healthy or faulty states. This approach eliminates the dependence on subjective threshold settings, providing a more robust and reliable detection framework.

In summary, this paper sets the stage for our future work by addressing the limitations of the previous model and paving the way for the development of more sophisticated and intelligent models for EFD and SENC problems.

3.4.4 results

This paper constructed a novel BC-GAN architecture for improving intelligent decision-making of EFD process. The proposed approach overcomes a significant limitation of contemporary EFD approaches that rely on manually determined pre-set thresholds. Unlike conventional methods, the proposed EFD framework is not only threshold-free, but also indicates the confidence level to its prediction. By doing so, we have improved the intelligence and automation of EFD models.

To validate the effectiveness of the presented model, the Intelligent Maintenance System (IMS) database was taken as a benchmark which consists of four run-to-failure bearing tests monitored by vibration sensors. Additionally, the method was deployed on our roller bearing and ball bearing test rigs for evaluation. We further improve the EFD accuracy on our roller bearing test rig to 0.9191, which is a significant improvement of our previous research on this dataset (0.8732 in **Paper C** and 0.8583 in **Paper A**). Extensive experiments were conducted on both vibration and AE signals, demonstrating the versatility of the proposed network in various laboratory and industrial settings. The results show that the proposed BC-GAN method outperforms conventional methods in terms of accuracy and automation.

3.5 Paper E:

3.5.1 Objective

Previous research presented in **Paper D** has established a solid foundation for the development of an end-to-end model that can achieve intelligent learning and decision-making. Building upon this work, the objective of the present paper is to further advance the field by addressing the challenges of online concept drift detection and classification under more complex scenarios, which is referred to as the SENC problem in this thesis. This is a particularly challenging task, as it requires the model to not only detect and adapt to changing data patterns over time but also handle new data points that belong to previously unseen classes in a streaming environment. The SENC problem adds another layer of complexity to this task by requiring the model to operate in a semi-supervised setting, where only a limited amount of labeled data is available for training. By addressing this problem, we aim to contribute to the development of more robust and adaptable models capable of handling the challenges of real-world data streams from the installed sensors.

3.5.2 Relevance to the thesis

The present paper builds upon the findings of our previous research in order to address the RQ3- '*how to incorporate the model trained on a closed dataset into an open environment for real-time online monitoring?*' To tackle this problem, a novel SENC framework is developed in this paper which serves as an extension of our previous work on the EFD problem (as discussed in **Paper D**) to a more generalized scenario known as the standard SENC problem. Additionally, the paper re-evaluates the effectiveness of the HSE method (**Paper B**) in a new context of the standard SENC task. The utilization of the HSE method plays a pivotal role in enhancing the stability of the proposed SENC framework. The development of the SENC framework, along with the re-examination of the HSE method, contributes significantly to the creation of more resilient and efficient models for real-time online monitoring.

3.5.3 Contributions

In this paper, we proposed that real-time online monitoring can be seen as a SENC problem. **Section 2.3.3** highlights certain limitations of current SENC frameworks, including the lack of effective DL-based frameworks, limited applicability to ICM of RM, and reliance on pre-defined thresholds. To address these research gaps, we introduced a novel SENC framework, which offers the following key contributions:

(1) We devised a novel DL-based architecture called MC-GAN. The method is notable for its capability of addressing multiple challenges with an integrated framework, including: (i) pattern recognition of already known fault types with high accuracy; (ii) timely detection of emerging new fault types; (iii) automatic model update to adapt to new fault types, and (iv) discriminating between different emerging new fault types. The framework can be applied to various practical challenges in ICM of bearings including EFD, IFD, and monitoring of the damage's progression.

(2) Compared with existing SENC frameworks, the proposed method is threshold-free, achieve intelligent learning and decision-making. To the best of our knowledge, this is the first DL-based framework for SENC problem that eliminates the need for a pre-determined threshold.

(3) The MC-GAN architecture was further enhanced by incorporating a novel ensemble technique proposed in Paper B called the HSE method. We further tailored the HSE method for GAN by introducing a model filter mechanism. This method significantly improves the performance of the model without requiring additional training resources, thereby increasing its practicality and effectiveness.

(4) To validate the proposed framework, we conducted four simulated SENC tasks using benchmark vibration signals. Through these tasks, we demonstrated the effectiveness and robustness of the framework in accurately detecting and classifying early novelty events in real-time monitoring scenarios.

3.5.4 Results

A novel SENC framework was developed in the present paper. In contrast to peer methods, the merit of the proposed method is highlighted as that it realizes intelligent

learning and decision-making simultaneously. The EMC-GAN can not only learn implicit features directly from raw signals, but also can generate automotive decision without pre-set thresholds. To evaluate the effectiveness of the proposed method in detecting and differentiating various faults in bearings, four datasets were created using the CWRU database. These datasets were specifically designed to represent different types of faults occurring in the inner race, outer race, and rolling element of bearings.

3.6 Paper F:

3.6.1 Objective

Up to now, we have successfully developed a DL-based SENC framework. Nonetheless, the validation of this method has been limited to a benchmark database, highlighting the importance of further validating its effectiveness and exploring its potential across a wider range of applications.

3.6.2 Relevance to the thesis

This paper is an extended version of our previous work, specifically **Paper E**, as we strive to evaluate and enhance the effectiveness of the EMC-GAN method for addressing the real-time online monitoring problem (RQ3) in various application scenarios. Building upon the foundation laid in **Paper E**, we have expanded the scope and explored additional applications to provide a more comprehensive analysis. Furthermore, we have revisited the effectiveness of the HSE method proposed in **Paper B**, conducting further evaluations and refining our understanding of its capabilities.

3.6.3 Contributions

Compared with the work in **Paper E**, the contribution of this paper is presented as follows:

(1) A comprehensive explanation of EMC-GAN, including its implementation details, was provided. We pointed out a prevalent issue referred to as the 'buffer contamination' problem within existing SENC frameworks. To mitigate this issue and ensure the reliability of the system, we introduced a novel safety

measure called the 'Buffer filter', drawing inspiration from Shannon's entropy concept.

(2) We introduced the EMC-GAN and model update mechanism to address the EFD problem. In contrast to the traditional EFD framework as well as our previous work in **Paper D**, the presented approach in this work is capable to achieve the following goals: (i) Identifying the onset of damage. (ii) Providing an adequate number of fault alarms related to the damage. (iii) Indicating the development of the damage. This marks a substantial advancement towards our desired intelligent modeling compared with traditional EFD methods that can only do binary classification between normal and abnormal stages.

(3) In order to provide a more comprehensive evaluation of the model performance, we conducted a thorough analysis that involved tackling more challenging tasks and performing ablation analysis.

3.6.4 Results

We conducted extensive validations on more demanding tasks to demonstrate the superiority of EMC-GAN. The results highlight how this method can effectively address the EFD problem, allowing the model to not only detect the early onset of bearing damage but also track its progression through a model update mechanism.

3.7 Summary

In conclusion, the Ph.D. research conducted in this study has resulted in six research papers that align with the proposed objectives. The work has progressively advanced from unsupervised EFD questioning to addressing the SENC problem. Throughout this process, innovative techniques and approaches were developed to tackle challenges related to EFD, IFD, and SENC.

To evaluate the effectiveness of the proposed methods in analyzing AE signals, two run-to-failure test rigs, namely the 'Roller Bearing test rig' and the 'Ball Bearing test rig', were constructed, with AE sensors used for health monitoring. The research outcomes extend beyond AE signals and show potential applicability to other sensor data, including vibration signals. The introduction of three benchmark vibration datasets
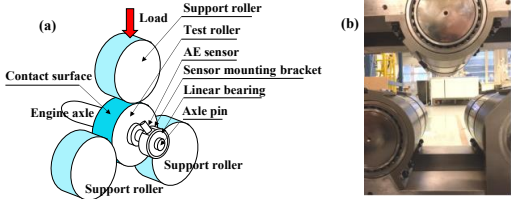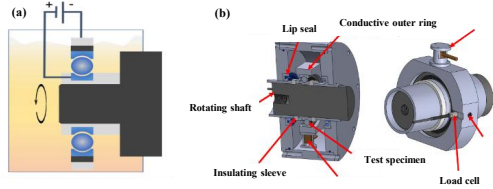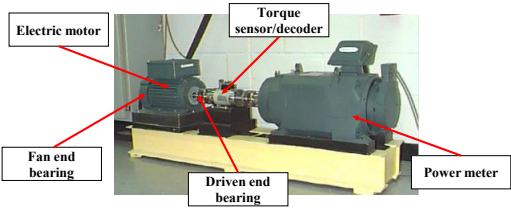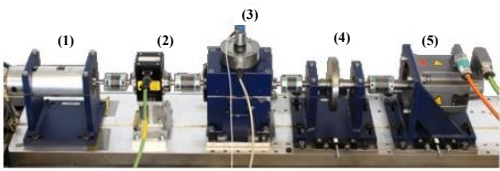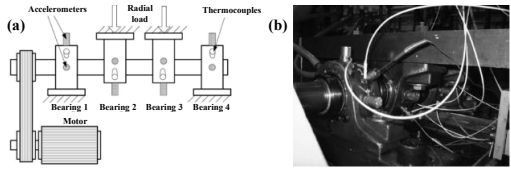
allowed for a comprehensive evaluation of the proposed approach. The results demonstrated the method's ability to detect faults in diverse scenarios, highlighting its versatility for solving problems beyond the project's scope. A summary of the utilization of these five databases can be found in Table 3.1.

Overall, this research contributes to the ICM field, offering novel insights and practical techniques for analyzing AE data. The developed methodologies have the potential to enhance fault detection and monitoring systems, benefiting various industries and applications.

Table 3.1 Connections between the utilized datasets and the research papers

Utilized datasets	A	B	C	D	E	F	Data type
1. Roller bearing fatigue test	√		√	√			AE
2. Ball bearing fatigue test		√			√		AE
3. CWRU	√			√	√		Vibration
4. Kat	√				√		Vibration
5. IMS				√			Vibration

Research papers header spans columns A B C D E F.

Chapter 4

Discussion

Let us revisit the research questions (RQs) derived in Chapter 1. The primary aim of this chapter is to thoroughly explore the connections between the six research papers conducted and the three targeted RQs, which are visually depicted in Figure 4.1. Through a comprehensive analysis of the findings presented in each paper, we will uncover how their contributions address and provide valuable insights into the corresponding RQs.
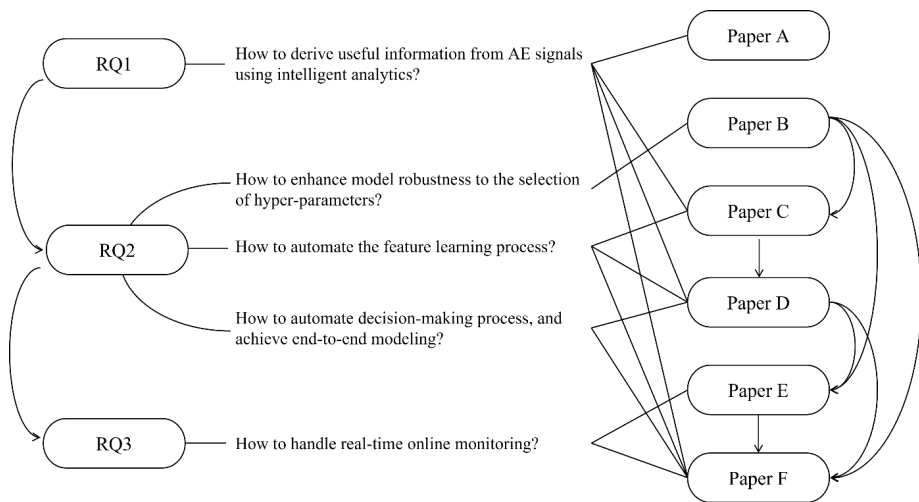
Figure 4. 1 A sketch of the connections between the conducted research papers and RQs.

4.1 RQ1: How to derive useful information from AE signals?

The entire Ph.D. project revolves around the early material damage detection by extracting useful information from AE signals. Through the literature review conducted in Chapter 2, a research gap was identified, shedding light on the relatively limited exploration of combining the AE technique with AI-based paradigms, particularly in comparison to the extensive research conducted on vibration signals. Consequently, the corresponding RQ is derived as follows:

- **RQ1**: *How to derive useful information from AE signals using intelligent analytics that can reveal the emergence and development of damages in bearings?*

This question can be approached by different methods, in this Ph.D. research, four DL-involved tools have been developed to extract useful information from AE signals. These tools are discussed in papers A, C, D, and F, as depicted in Figure 4.1. In the following, we will provide a chronological elaboration of these four tools, highlighting their respective contributions and methodologies to address the RQ1.

Paper A provided an improved manual feature engineering method by incorporating AE technology and DL. The AE signal is characterized by high dimensionality due to its high sampling frequency, typically ranging between 100 kHz and several MHz, posing computational challenges for DL methods. To this end, an efficient MVW method was leveraged to compress the AE signals. By shifting the MVW along the signal, we calculated the variance of the covered data and transform it into a numerical value. This de-dimensionalized each windowed data, producing a dimensionless number that measures the data dispersion. The method was proven to be effective and preserve useful information in AE signals.

Another contribution of **Paper A** is the introduction of an Autoencoder powered by CNN to denoise the raw AE signals that masked by heavy background noise. Autoencoders have proven to be effective in feature extraction tasks by reconstructing their own input data. This method operates on the assumption that Autoencoders can effectively capture the essential characteristics of the input data while disregarding irrelevant noise. Thus, the Autoencoder serves as a denoising mechanism, allowing the extraction and enhancement of the key features within the AE data. The benefits of this

approach can be observed in two ways: (i) the removal of redundant noise from the AE signal and (ii) the enhancement of the sensitivity of the IEPF metric to transient AE bursts. Figure 4.2 presents a comparison result that clearly demonstrates the enhanced characteristics of the reconstructed data.
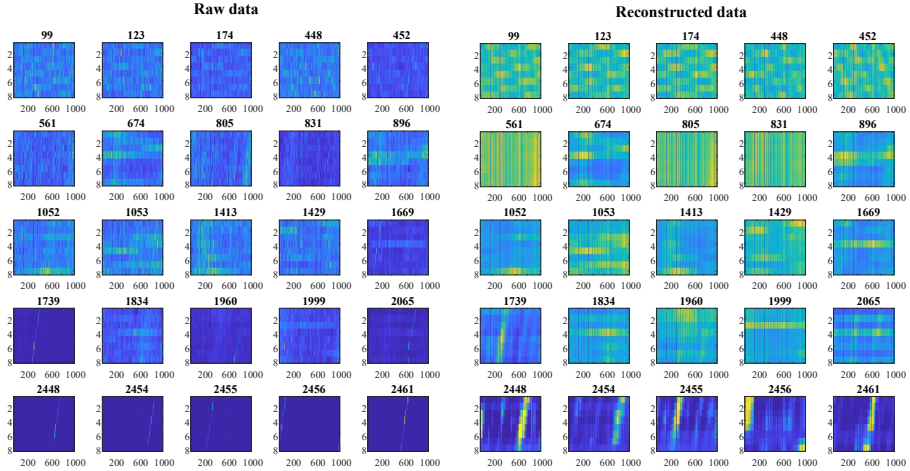
Figure 4. 2 Comparison between the raw AE data and reconstructed AE data (**Paper A** [12]).

After denoising, a novel AE parameter, referred to as IEPF in **Paper A**, was derived as $\sum_{i=1}^{n_k}\left(\widetilde{P_k}log_2\widetilde{P_k} - \frac{1}{n_k}log_2\frac{1}{n_k}\right)$ by leveraging Shannon's entropy. The parameter is aimed to characterize the periodicity and chaos within the AE data. To interpret the result, when the IEPF value is close to 0, it signifies that the signal exhibits strong periodicity. Conversely, a higher value of IEPF indicates a greater disturbance in the AE signal, as shown in Figure 4.3.
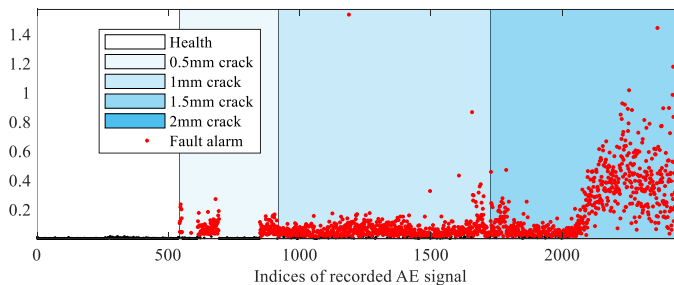
Figure 4. 3 The scores of IEPF from the durability test of the RB test.

Paper C contributes to RQ1 by introducing a novel GAN-based architecture for EFD problem. One contribution of this paper is on the development of a specialized 'Generator' that learns the data distribution of AE signals during the normal operation stage. This architecture comprises three essential components: the encoder block, the decoder block, and the bottleneck block, as depicted in the Figure 4.4. The encoder block incorporates two convolutional layers and two down-sampling layers, working synergistically to extract local information from the input AE data. Conversely, the decoder block is responsible for reconstructing the input signal and consists of two convolutional layers, one up-sampling layer, and one fully connected layer. The bottleneck block, consisting of two Long Short-Term Memory (LSTM) cells, plays a critical role in capturing the time-series correlation present in the AE signal. The primary objective of this novel architecture is to extract more valuable information from AE signals, thereby significantly enhancing the understanding and analysis of these signals for the EFD problem.
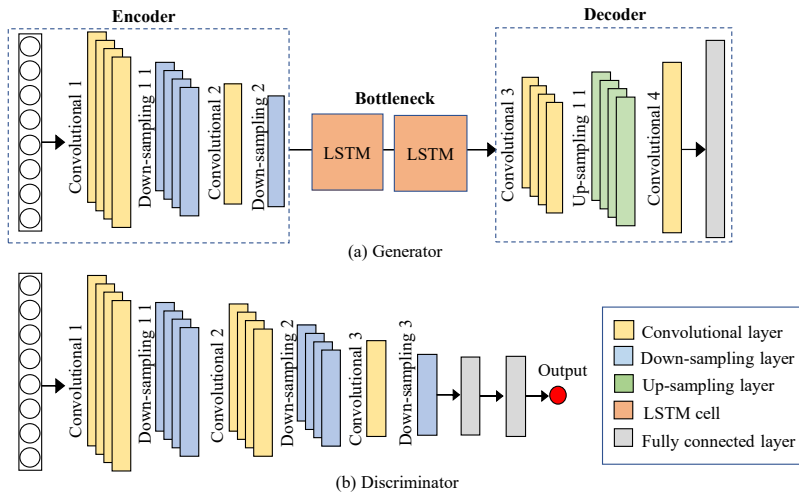
Figure 4. 4 Architectures of the proposed network architectures (**Paper C** [14]).

Paper D builds upon the basic concept of GAN but shifts its focus towards enhancing the 'Discriminator' component of the GAN framework. To address RQ1, a novel architecture called BC-GAN was proposed in **Paper D**. The re-designed Discriminator comprise two channels involving four functional modules and each of them plays a different role, as schematically depicted in Figure 4.5. This two-channel approach

allows for learning the differences between real and fake data from different perspectives, driven by distinct loss functions. By leveraging these complementary channels, the BC-GAN architecture enhances the Discriminator's ability to discern between real and fake instances, ultimately improving the overall performance of the GAN framework for early fault detection applications.
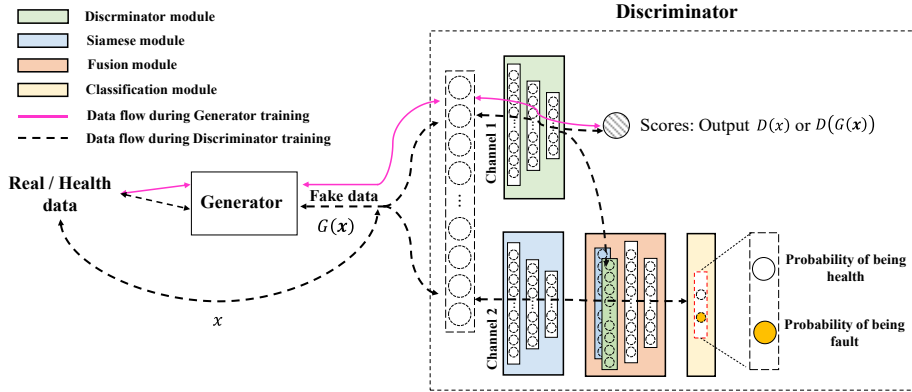
Figure 4. 5 Schematic of BC-GAN for EFD in run-to-failure process (**Paper D** [15]).

Paper F makes a significant contribution to addressing RQ1 by presenting a solution that extracts valuable information from AE signals to indicate the development of damages in bearings. In this paper, an EMT-GAN framework was introduced, which built upon the foundations of the BC-GAN architecture. By adding multiple neurons to the classification module as shown in Figure 4.5, the EMT-GAN is able to handle the multi-classification, allowing for more comprehensive analysis and detection of different types of damages in bearings.

In summary, various approaches exist for extracting valuable information from AE signals through intelligent analytics. **Paper A** focused on uncovering the periodicity and disturbance characteristics embedded within AE waveforms using a manually designed feature. On the other hand, **Papers C**, **D**, and **F** employed deep neural networks to learn tailored representations from each AE waveform, enabling more effective and generalized interpretation. Table 4.1 summarized the comparison of the four tools from the following aspects: (i) achieving end-to-end modeling, (ii) learning tailored representations, (iii) being threshold-free, and (iv) adapting to evolving learning tasks. Overall, these research papers collectively demonstrate the progression and

evolution in the field, showcasing our footprints of developing advanced techniques for extracting meaningful information from AE signals and addressing the challenges posed by RQ1.

Table 4.1. Comparison of the tools developed in different papers.

Paper	A	C	D	F
Method	IEPF	Convolutional GAN	BC-GAN	EMT-GAN
End-to-end modeling	×	√	√	√
Tailored Representation Learning	×	√	√	√
Threshold-free	×	×	√	√
Track the Damage Development	×	×	×	√

4.2 RQ2: How to perform end-to-end modeling?

End-to-end modeling is an AI-based paradigm that enables a system or model to directly process raw input data and produce the desired output without the need for explicit intermediate representations or separate modules for different processing stages. This approach brings several advantages, including flexibility, simplicity, and improved performance. However, its application in the analysis of AE signals remains relatively understudied. Therefore, the second research objective focuses on tackling the existing limitations and research gaps in the field of end-to-end analysis of AE signal. The RQ2 is presented as follows.

- **RQ2**: *How to perform end-to-end modeling that minimizes the model's dependence on expert assumptions and prior knowledge, while facilitating intelligent learning and decision-making processes?*

Paper B contributes to this question by introducing an effective but easy-to-implement ensemble technology, referring to as HSE method. The method is constructed based on the assumption that neural network can produce multiple 'local sub-optima' with diversity during training, and the combination of these local sub-optima can generate a stronger model. By doing so, the proposed method avoids the intricate process of defining qualified base models in contrast to traditional ensemble method, resulting in reduced complexity. The merits of the proposed approach are highlighted as: (i) The approach does not increase the training budget of the neural network; (ii) The approach

can be applied to all types of neural networks without twisting of network architecture; and (iii) the approach is highly accessible and straightforward, making it easy for practitioners to comprehend and implement. We analyzed the feasibility of the proposed method in terms of improving accuracy and robustness to hyper-parameters. The HSE method was further applied to our following research in **Paper C**, **Paper E**, and **Paper F**. The application of the HSE method in these domains has significantly *reduced the need for human experts to manually define the optimal hyperparameters*, such as the number of training epochs, in the DL model, which is the main contribution to RQ2.

Paper C also contributes to RQ2 by addressing *how to reduce model's dependence on expert assumptions and prior knowledge*. A remaining problem in **Paper A** pertains to the IEPF method, which relies on manually designed features tailored to specific assumptions regarding the faulty information of the damage. This approach inherently possesses limited generalization ability due to its dependency on predetermined assumptions. To this end, a novel EFD framework leveraging convolutional GAN was designed for intelligent analysis of AE signals. The key idea is to take the signals from health stage as the real input of GAN, and a Generator is trained to produce synthetic data that is close to the real data, as depicted in Figure 4.4. While a Discriminator is trained to discriminate between the synthetic and real data. The output of the Discriminator is a single value, with high values assigned to real data and low values assigned to fake data. We proposed that the scores obtained by Discriminator forms a natural HI indicating the deviation of the signals. Hence, the method is capable of learning information direct from raw input data without manually engineered features.
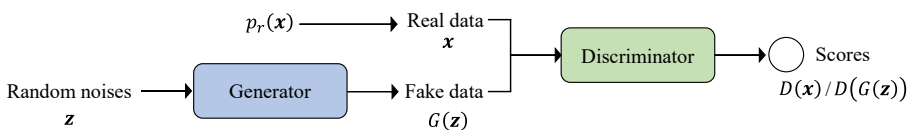
Figure 4. 6 The basic structure of GAN (**Paper C** [14]).

The convolutional GAN in **Paper C** follows the prevalent EFD framework (as shown in Figure 2.2) by extracting a sensitive HI, and the decision-making rely on a pre-set threshold. The threshold-related issues have been discussed in Section 3.4.1. **Paper D** aims to enhance the intelligence of the model by eliminating the need for pre-set thresholds in the traditional EFD approach, thereby promoting a more intelligent

decision-making process. To achieve this, a threshold-free deep architecture, referred to as BC-GAN, was constructed in the present paper. The workflow of the BC-GAN model and the traditional EFD models are compared in Figure 4.7. Unlike modern methods like the one presented in **Paper A** and **Paper C**, BC-GAN directly produces the probability of the input belonging to either the normal or abnormal state, without relying on any pre-set threshold. This results in a more concise framework for analysis. Additionally, the concept was further extended to our subsequent research papers, namely **Paper E** and **Paper F**, which explore the application of the concept to the SENC problem.

(a) Traditional EFD model	(b) BC-GAN
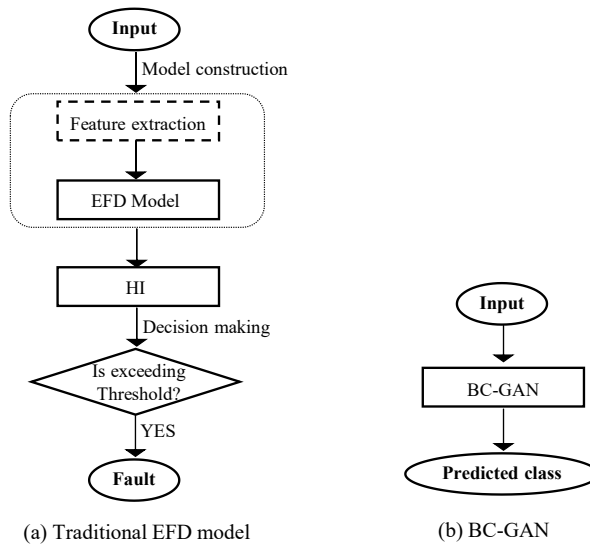
Figure 4. 7 Comparison of the framework of (a) traditional EFD models, and (b) BC-GAN, for end-to-end analysis of AE signals (**Paper D** [15]).

4.3 RQ3: How to handle real-time online monitoring?

The third research objective aligns with the project goal of designing and testing software systems to process AE signals in real-time for monitoring the emergence and development of damages. and the third question is presented as below. With this goal, the third question is proposed as below.

- **RQ3**: *How to incorporate the model trained on a closed dataset into an open environment for real-time online monitoring?*

Paper E makes a significant contribution to addressing this question by highlighting the limitations of existing supervised AI paradigms for online ICM, as discussed in **Section 2.3**. It proposed treating the online ICM of bearings as the SENC problem, enabling the model trained on a closed dataset to adapt to an open environment with new information that was not present in the training dataset. This assumption is more practical for real-world applications. To address this, a novel network architecture called EMC-GAN was introduced, which is a direct extension of the BC-GAN proposed in **Paper D**. The present paper also introduced a model update mechanism that allows the model to automatically update itself as the learning task evolves. An overview of the EMC-GAN framework for the SENC problem is illustrated in Figure 4.8.
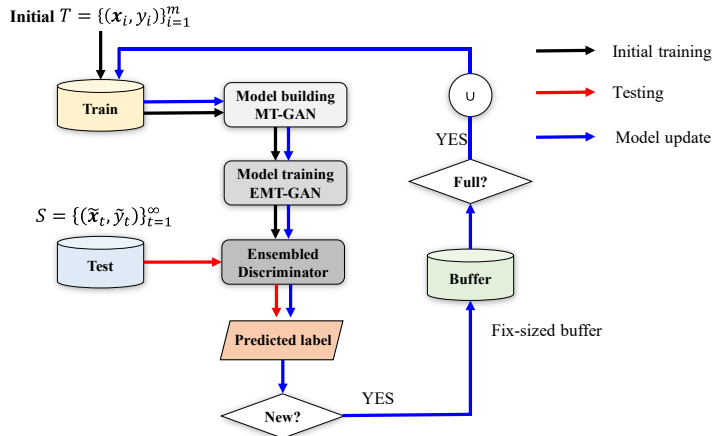
Figure 4. 8 The overall framework of EMC-GAN for classification in SENC problem (**Paper E**).

In **Paper F**, the aforementioned concept was further extended to the EFD of bearings utilizing AE technology. We treated the EFD as a specific scenario within the broader SENC problem, where $k = 1$ and $c = n$. The present paper demonstrated the effectiveness of the DL-based method in interpreting AE signals. Furthermore, it was demonstrated that AE technology could effectively track the progression of damage in bearings by categorizing them into different classes using the model update mechanism.

To summarize, the two works contribute to overcoming the limitations of existing SENC frameworks in the following ways. Firstly, a novel DL-based SENC framework is introduced, offering a fresh approach to address the problem. Secondly, the effectiveness of the proposed method is validated through its application to ICM of

bearings, utilizing both vibration and AE signals. This validation demonstrates the framework's generalization ability across different sensor data. Lastly, the proposed method is threshold-free and facilitates intelligent learning and decision-making processes, further enhancing its practicality and utility.

Chapter 5

Conclusions and future work

In this final chapter, we provide a summary of this thesis and highlight the key conclusions derived from this research. Additionally, we also offer suggestions for future work in the field.

5.1 Conclusions

The Ph.D. research revolved around early material damage detection in bearings using AE technique and intelligent analytics. Three RQs were derived in the first chapter as the guidance of this thesis, and a total of six papers were conducted to address the RQs.

The primary research objective is to extract sensitive information from AE signals based on intelligent analytics, capable of detecting the emergence and progression of damages in bearings. To accomplish this, four research papers, referred to as **Paper A**, **Paper C**, **Paper D**, and **paper F**, were devoted to investigating this issue. These works initially began with conventional methods, and building upon prior research, progressively delved deeper into the problem. As a result, four EFD frameworks were established, gradually advancing towards the goal of achieving end-to-end modeling for EFD.

After the model with the desired functionality was developed, we delved into contemplating its generalization ability for real-world applications. Consequently, the second research objective was formulated: how to enhance model performance, automate the EFD process, and minimize the model's dependence on human expertise by leveraging machine learning tools. We introduced a generalized ensemble method that enhances the overall robustness of neural networks in **Paper B**. This approach was

further validated extensively in **Paper C**, **Paper E**, and **Paper F**, showcasing its effectiveness across different tasks, ranging from EFD, IFD, and SENC problem.

Finally, we move to the online monitoring of bearings for EFD as indicated in the third research objective. We connected the online ICM with SENC problem, which, to the best of our knowledge, is a pioneering attempt in the realm of EFD. A novel SENC framework leveraging GAN was proposed in **Paper E**, and then, we further applied the method to EFD problem in **Paper F**. In contrast to traditional EFD models, the proposed method in this Ph.D. research not only detects early damage but also evolves itself to indicate the progression of the damage.

5.2 Future work

The search for better solutions is an ongoing process, and here we list some suggestions that deserve deeper investigation in the future.

- *Advanced signal processing*: AE signals exhibit high dimensionality, and in our research, we employed a simple MVW approach to compress these high-dimensional signals. However, there is ample room for future exploration and development of proper and advanced signal processing methods. Such advancements are significant for enhancing the performance of AI models.

- *Advanced model design*: A novel SENC framework based on the EMC-GAN architecture was introduced in our research papers E and F. While the current version of EMC-GAN has shown promising results for the SENC problem, there is potential for further improvement through fine-tuning the loss function, refining the training method, and exploring innovative network architectures.

- *Improved model efficiency*: As a limitation, the current version of EMC-GAN for the SENC problem is time-consuming compared to many traditional methods. Therefore, it is important to explore more effective and efficient methods that can significantly reduce the computational time required for model training and updating. This will not only improve the overall efficiency of the framework but also enhance its practical applicability in real-time or online scenarios.

- *Model interpretability*: Explainable artificial intelligence (XAI), also known as explainable artificial intelligence, has received increasing attention in recent years. DL algorithms are often perceived as 'black box' methods, lacking transparency in their decision-making processes. Therefore, the development of XAI techniques in ICM field is significant for enhancing the trustworthiness, interpretability, and overall effectiveness of AI models.

- *Broader applications*: The outcome of this research hold the potential for broader applicability beyond AE analysis and can be extrapolated to other sensor sources and industrial sectors.

Bibliography

[1] Investigation of subsurface microcracks causing premature failure in wind turbine gearbox bearings | Elsevier Enhanced Reader, (n.d.). https://doi.org/10.1016/j.rineng.2022.100667.

[2] E. Hart, B. Clarke, G. Nicholas, A. Kazemi Amiri, J. Stirling, J. Carroll, R. Dwyer-Joyce, A. McDonald, H. Long, A review of wind turbine main bearings: design, operation, modelling, damage mechanisms and fault detection, Wind Energy Science. 5 (2020) 105–124. https://doi.org/10.5194/wes-5-105-2020.

[3] A. Dhanola, H.C. Garg, Tribological challenges and advancements in wind turbine bearings: A review, Engineering Failure Analysis. 118 (2020) 104885. https://doi.org/10.1016/j.engfailanal.2020.104885.

[4] A. Choudhary, D. Goyal, S.L. Shimi, A. Akula, Condition Monitoring and Fault Diagnosis of Induction Motors: A Review, Arch Computat Methods Eng. 26 (2019) 1221–1238. https://doi.org/10.1007/s11831-018-9286-z.

[5] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mechanical Systems and Signal Processing. 138 (2020) 106587. https://doi.org/10.1016/j.ymssp.2019.106587.

[6] C.K. Tan, P. Irving, D. Mba, A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears, Mechanical Systems and Signal Processing. 21 (2007) 208–233. https://doi.org/10.1016/j.ymssp.2005.09.015.

[7] Z. Liu, B. Yang, X. Wang, L. Zhang, Acoustic Emission Analysis for Wind Turbine Blade Bearing Fault Detection Under Time-Varying Low-Speed and Heavy Blade Load Conditions, IEEE Transactions on Industry Applications. 57 (2021) 2791–2800. https://doi.org/10.1109/tia.2021.3058557.

[8] A. Choudhury, N. Tandon, Application of acoustic emission technique for the detection of defects in rolling element bearings, Tribol Int Tribol Int. 33 (2000) 39–45.

[9] A.M. Al-Ghamd, D. Mba, A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size, Mechanical Systems and Signal Processing. 20 (2006) 1537–1571.

[10] E. Caso, A. Fernandez-del-Rincon, P. Garcia, M. Iglesias, F. Viadero, Monitoring of misalignment in low speed geared shafts with acoustic emission sensors, Applied Acoustics. 159 (2020). https://doi.org/10.1016/j.apacoust.2019.107092.

[11] Y. He, X. Zhang, M.I. Friswell, Defect Diagnosis for Rolling Element Bearings Using Acoustic Emission, Journal of Vibration and Acoustics. 131 (2009). https://doi.org/10.1115/1.4000480.

[12] Y. Wang, R.H. Hestmo, A. Vinogradov, Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder, Meas. Sci. Technol. (2023). https://doi.org/10.1088/1361-6501/acc1f8.

[13] Y. Wang, A. Vinogradov, Simple is good: Investigation of history-state ensemble deep neural networks and their validation on rotating machinery fault diagnosis, Neurocomputing. 548 (2023) 126353. https://doi.org/10.1016/j.neucom.2023.126353.

[14] Y. Wang, A. Vinogradov, Improving the Performance of Convolutional GAN Using History-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Signals, Applied Sciences. 13 (2023) 3136. https://doi.org/10.3390/app13053136.

[15] Y. Wang, S. Bernat, A. Vinogradov, Bc-Gan: A Threshold-Free Framework for Unsupervised Early Fault Detection in Rotating Machinery, (2022). https://doi.org/10.2139/ssrn.4276565.

[16] Y. Wang, Q. Wang, A. Vinogradov, Semi-supervised deep architecture for classification in streaming data with emerging new classes: application in condition monitoring, (2023). https://doi.org/10.36227/techrxiv.21931476.v1.

[17]   Y. He, M. Li, Z. Meng, S. Chen, S. Huang, Y. Hu, X. Zou, An overview of acoustic emission inspection and monitoring technology in the key components of renewable energy systems, Mechanical Systems and Signal Processing. 148 (2021) 107146. https://doi.org/10.1016/j.ymssp.2020.107146.

[18]   C. James Li, S.Y. Li, Acoustic emission analysis for bearing condition monitoring, Wear. 185 (1995) 67–74. https://doi.org/10.1016/0043-1648(95)06591-1.

[19]   C.C. Tan, Application of Acoustic Emission to the Detection of Bearing Failures, in: International Tribology Conference 1990, Brisbane 2-5 December 1990: Putting Tribology to Work; Reliability and Maintainability through Lubrication and Wear Technology; Preprints of Papers, n.d.: pp. 110–114. https://doi.org/10.3316/informit.377857431358203.

[20]   S. Al-Dossary, R.I.R. Hamzah, D. Mba, Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing, Applied Acoustics. 70 (2009) 58–81. https://doi.org/10.1016/j.apacoust.2008.01.005.

[21]   A. Morhain, D. Mba, Bearing defect diagnosis and acoustic emission, Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. 217 (2003) 257–272.

[22]   C. Pachaud, R. Salvetat, C. Fray, Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces, Mechanical Systems and Signal Processing. 11 (1997) 903–916. https://doi.org/10.1006/mssp.1997.0115.

[23]   J. Miettinen, P. Pataniitty, Acoustic emission in monitoring extremely slowly rotating rolling bearing, Proceedings of COMADEM ´99. Oxford. (1999) 289–297.

[24]   W. Caesarendra, B. Kosasih, A.K. Tieu, H. Zhu, C.A.S. Moodie, Q. Zhu, Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing, Mechanical Systems and Signal Processing. 72–73 (2016) 134–159. https://doi.org/10.1016/j.ymssp.2015.10.020.

[25]   R.-P. Nikula, K. Karioja, M. Pylvänäinen, K. Leiviskä, Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features, Mechanical Systems and Signal Processing. 138 (2020) 106572. https://doi.org/10.1016/j.ymssp.2019.106572.

[26]   M. Motahari-Nezhad, S.M. Jafari, Bearing remaining useful life prediction under starved lubricating condition using time domain acoustic emission signal processing, Expert Systems with Applications. 168 (2021) 114391.

[27]   F. König, C. Sous, A. Ouald Chaib, G. Jacobs, Machine learning based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems, Tribology International. 155 (2021) 106811. https://doi.org/10.1016/j.triboint.2020.106811.

[28]   M. Elforjani, S. Shanbr, Prognosis of Bearing Acoustic Emission Signals Using Supervised Machine Learning, IEEE Transactions on Industrial Electronics. 65 (2018) 5864–5871. https://doi.org/10.1109/tie.2017.2767551.

[29]   M. Elforjani, D. Mba, Accelerated natural fault diagnosis in slow speed bearings with Acoustic Emission, Engineering Fracture Mechanics. 77 (2010) 112–127. https://doi.org/10.1016/j.engfracmech.2009.09.016.

[30]   M. Amiri, M. Modarres, E.L. Droguett, AE entropy for detection of fatigue crack initiation and growth, in: 2015 IEEE Conference on Prognostics and Health Management (PHM), 2015: pp. 1–8. https://doi.org/10.1109/ICPHM.2015.7245038.

[31]   A. Kahirdeh, C. Sauerbrunn, M. Modarres, Acoustic emission entropy as a measure of damage in materials, in: AIP Publishing LLC, 2016: p. 060007.

[32]   M. Chai, Z. Zhang, Q. Duan, A new qualitative acoustic emission parameter based on Shannon's entropy for damage monitoring, Mechanical Systems and Signal Processing. 100 (2018) 617–629.

[33]   E.L. Hidle, R.H. Hestmo, O.S. Adsen, H. Lange, A. Vinogradov, Early Detection of Subsurface Fatigue Cracks in Rolling Element Bearings by the Knowledge-Based Analysis of Acoustic Emission, Sensors. 22 (2022) 5187. https://doi.org/10.3390/s22145187.

[34]   Y. Mao, P. Borghesani, Z.Y. Chin, R.B. Randall, Z. Peng, Extraction and use of frequency-domain relationships between time-varying gear meshing properties and diagnostic measurements, Mechanical Systems and Signal Processing. 190 (2023) 110129. https://doi.org/10.1016/j.ymssp.2023.110129.

[35]   S. Gowid, R. Dixon, S. Ghani, A novel robust automated FFT-based segmentation and features selection algorithm for acoustic emission condition

based monitoring systems, Applied Acoustics. 88 (2015) 66–74.
https://doi.org/10.1016/j.apacoust.2014.08.007.

[36]   Z. Wang, F. Chegdani, N. Yalamarti, B. Takabi, B. Tai, M. El Mansori, S. Bukkapatnam, Acoustic Emission Characterization of Natural Fiber Reinforced Plastic Composite Machining Using a Random Forest Machine Learning Model, Journal of Manufacturing Science and Engineering. 142 (2020). https://doi.org/10.1115/1.4045945.

[37]   M.T. Pham, J.-M. Kim, C.H. Kim, Intelligent Fault Diagnosis Method Using Acoustic Emission Signals for Bearings under Complex Working Conditions, Applied Sciences. 10 (2020) 7068. https://doi.org/10.3390/app10207068.

[38]   Z. Gao, J. Lin, X. Wang, X. Xu, Bearing fault detection based on empirical wavelet transform and correlated kurtosis by acoustic emission, Materials. 10 (2017) 571.

[39]   D. Liu, J. Tao, A. Luo, Q. Wang, An optimized kurtogram method for early fault detection of rolling element bearings using acoustic emission, in: IEEE, 2018: pp. 365–370.

[40]   D. Bianchi, E. Mayrhofer, M. Gröschl, G. Betz, A. Vernes, Wavelet packet transform for detection of single events in acoustic emission signals, Mechanical Systems and Signal Processing. 64–65 (2015) 441–451. https://doi.org/10.1016/j.ymssp.2015.04.014.

[41]   L.-S. Law, J.H. Kim, W.Y.H. Liew, S.-K. Lee, An approach based on wavelet packet decomposition and Hilbert–Huang transform (WPD–HHT) for spindle bearings condition monitoring, Mechanical Systems and Signal Processing. 33 (2012) 197–211. https://doi.org/10.1016/j.ymssp.2012.06.004.

[42]   S.J. Kim, K. Kim, T. Hwang, J. Park, H. Jeong, T. Kim, B.D. Youn, Motor-current-based electromagnetic interference de-noising method for rolling element bearing diagnosis using acoustic emission sensors, Measurement. 193 (2022) 110912. https://doi.org/10.1016/j.measurement.2022.110912.

[43]   L. Liu, L. Chen, Z. Wang, D. Liu, Early fault detection of planetary gearbox based on acoustic emission and improved variational mode decomposition, IEEE Sensors Journal. 21 (2020) 1735–1745.

[44]   B. Kilundu, X. Chiementin, J. Duez, D. Mba, Cyclostationarity of Acoustic Emissions (AE) for monitoring bearing defects, Mechanical Systems and Signal Processing. 25 (2011) 2061–2072. https://doi.org/10.1016/j.ymssp.2011.01.020.

[45]   J. Antoni, Cyclic spectral analysis in practice, Mechanical Systems and Signal Processing. 21 (2007) 597–630. https://doi.org/10.1016/j.ymssp.2006.08.007.

[46]   Q. Ni, J.C. Ji, K. Feng, B. Halkon, A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences, Mechanical Systems and Signal Processing. 153 (2021) 107498. https://doi.org/10.1016/j.ymssp.2020.107498.

[47]   Z. Ma, M. Zhao, M. Luo, C. Gou, G. Xu, An integrated monitoring scheme for wind turbine main bearing using acoustic emission, Signal Processing. 205 (2023) 108867. https://doi.org/10.1016/j.sigpro.2022.108867.

[48]   H.O. Omoregbee, P.S. Heyns, Fault Classification of Low-Speed Bearings Based on Support Vector Machine for Regression and Genetic Algorithms Using Acoustic Emission, J. Vib. Eng. Technol. 7 (2019) 455–464. https://doi.org/10.1007/s42417-019-00143-y.

[49]   D.H. Pandya, S.H. Upadhyay, S.P. Harsha, Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN, Expert Systems with Applications. 40 (2013) 4137–4145. https://doi.org/10.1016/j.eswa.2013.01.033.

[50]   A. Meserkhani, S.M. Jafari, A. Rahi, Experimental comparison of acoustic emission sensors in the detection of outer race defect of angular contact ball bearings by artificial neural network, Measurement. 168 (2021) 108198. https://doi.org/10.1016/j.measurement.2020.108198.

[51]   M.J. Hasan, M.M.M. Islam, J.-M. Kim, Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions, Measurement. 138 (2019) 620–631. https://doi.org/10.1016/j.measurement.2019.02.075.

[52]   E. Pomponi, A. Vinogradov, A real-time approach to acoustic emission clustering, Mechanical Systems and Signal Processing. 40 (2013) 791–804. http://dx.doi.org/10.1016/j.ymssp.2013.03.017.

[53]   G. Pang, C. Shen, L. Cao, A.V.D. Hengel, Deep Learning for Anomaly
       Detection: A Review, ACM Comput. Surv. 54 (2021) 38:1-38:38.
       https://doi.org/10.1145/3439950.

[54]   X. Zhang, Y. Cong, Z. Yuan, T. Zhang, X. Bai, Early Fault Detection Method
       of Rolling Bearing Based on MCNN and GRU Network with an Attention
       Mechanism, Shock and Vibration. 2021 (2021).

[55]   C. Tutivén, Y. Vidal, A. Insuasty, L. Campoverde-Vilela, W. Achicanoy, Early
       Fault Diagnosis Strategy for WT Main Bearings Based on SCADA Data and One-
       Class SVM, Energies. 15 (2022) 4381. https://doi.org/10.3390/en15124381.

[56]   J. Saari, D. Strömbergsson, J. Lundberg, A. Thomson, Detection and
       identification of windmill bearing faults using a one-class support vector machine
       (SVM), Measurement. 137 (2019) 287–301.
       https://doi.org/10.1016/j.measurement.2019.01.020.

[57]   L.C. Brito, G.A. Susto, J.N. Brito, M.A.V. Duarte, Fault Detection of Bearing:
       An Unsupervised Machine Learning Approach Exploiting Feature Extraction and
       Dimensionality Reduction, Informatics. 8 (2021) 85.
       https://doi.org/10.3390/informatics8040085.

[58]   J. Mi, Y. Hou, W. He, C. He, H. Zhao, W. Huang, A Nonparametric
       Cumulative Sum-Based Fault Detection Method for Rolling Bearings Using High-
       Level Extended Isolated Forest, IEEE Sensors Journal. 23 (2023) 2443–2455.
       https://doi.org/10.1109/JSEN.2022.3225457.

[59]   W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, P. Zhou, Early Fault Detection
       Approach With Deep Architectures, IEEE Transactions on Instrumentation and
       Measurement. 67 (2018) 1679–1689. https://doi.org/10.1109/TIM.2018.2800978.

[60]   X. Kong, X. Li, Q. Zhou, Z. Hu, C. Shi, Attention Recurrent Autoencoder
       Hybrid Model for Early Fault Diagnosis of Rotating Machinery, IEEE
       Transactions on Instrumentation and Measurement. 70 (2021) 1–10.
       https://doi.org/10.1109/TIM.2021.3051948.

[61]   S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, W. Zhantao, Enhanced deep
       gated recurrent unit and complex wavelet packet energy moment entropy for early
       fault prognosis of bearing, Knowledge-Based Systems. 188 (2020) 105022.
       https://doi.org/10.1016/j.knosys.2019.105022.

[62]   W. Song, W. Shen, L. Gao, X. Li, An Early Fault Detection Method of Rotating Machines Based on Unsupervised Sequence Segmentation Convolutional Neural Network, IEEE Transactions on Instrumentation and Measurement. 71 (2022) 1–12. https://doi.org/10.1109/TIM.2021.3132989.

[63]   Y. Wei, Y. Li, M. Xu, W. Huang, A Review of Early Fault Diagnosis Approaches and Their Applications in Rotating Machinery, Entropy. 21 (2019) 409. https://doi.org/10.3390/e21040409.

[64]   W. Mao, J. He, Y. Li, Y. Yan, Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 231 (2016) 1560–1578. https://doi.org/10.1177/0954406216675896.

[65]   C. Lu, Z.-Y. Wang, W.-L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Processing. 130 (2017) 377–388. https://doi.org/10.1016/j.sigpro.2016.07.028.

[66]   H.O.A. Ahmed, M.L.D. Wong, A.K. Nandi, Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features, Mechanical Systems and Signal Processing. 99 (2018) 459–477. https://doi.org/10.1016/j.ymssp.2017.06.027.

[67]   A. Youcef Khodja, N. Guersi, M.N. Saadi, N. Boutasseta, Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks, The International Journal of Advanced Manufacturing Technology. 106 (2019) 1737–1751. https://doi.org/10.1007/s00170-019-04726-7.

[68]   R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, Mechanical Systems and Signal Processing. 108 (2018) 33–47. https://doi.org/10.1016/j.ymssp.2018.02.016.

[69]   C. Wu, P. Jiang, C. Ding, F. Feng, T. Chen, Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network, Computers in Industry. 108 (2019) 53–61. https://doi.org/10.1016/j.compind.2018.12.001.

[70]   J. Lu, W. Qian, S. Li, R. Cui, Enhanced K-Nearest Neighbor for Intelligent Fault Diagnosis of Rotating Machinery, Applied Sciences. 11 (2021) 919. https://doi.org/10.3390/app11030919.

[71]   A. Widodo, E.Y. Kim, J.-D. Son, B.-S. Yang, A.C.C. Tan, D.-S. Gu, B.-K. Choi, J. Mathew, Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, Expert Systems with Applications. 36 (2009) 7252–7261. https://doi.org/10.1016/j.eswa.2008.09.033.

[72]   J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, D. Huang, New imbalanced bearing fault diagnosis method based on Sample-characteristic Oversampling TechniquE (SCOTE) and multi-class LS-SVM, Applied Soft Computing. 101 (2021) 107043. https://doi.org/10.1016/j.asoc.2020.107043.

[73]   S.S. Roy, S. Dey, S. Chatterjee, Autocorrelation Aided Random Forest Classifier-Based Bearing Fault Detection Framework, IEEE Sensors Journal. 20 (2020) 10792–10800. https://doi.org/10.1109/JSEN.2020.2995109.

[74]   Q. Hu, X.-S. Si, Q.-H. Zhang, A.-S. Qin, A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests, Mechanical Systems and Signal Processing. 139 (2020) 106609. https://doi.org/10.1016/j.ymssp.2019.106609.

[75]   H. Li, G. Hu, J. Li, M. Zhou, Intelligent Fault Diagnosis for Large-Scale Rotating Machines Using Binarized Deep Neural Networks and Random Forests, IEEE Transactions on Automation Science and Engineering. 19 (2022) 1109–1119. https://doi.org/10.1109/TASE.2020.3048056.

[76]   N. Zhang, L. Wu, J. Yang, Y. Guan, Naive Bayes Bearing Fault Diagnosis Based on Enhanced Independence of Data, Sensors. 18 (2018) 463. https://doi.org/10.3390/s18020463.

[77]   S. Plakias, Y.S. Boutalis, A novel information processing method based on an ensemble of Auto-Encoders for unsupervised fault detection, Computers in Industry. 142 (2022) 103743. https://doi.org/10.1016/j.compind.2022.103743.

[78]   S. Yu, M. Wang, S. Pang, L. Song, X. Zhai, Y. Zhao, TDMSAE: A transferable decoupling multi-scale autoencoder for mechanical fault diagnosis, Mechanical Systems and Signal Processing. 185 (2023) 109789. https://doi.org/10.1016/j.ymssp.2022.109789.

[79] Z. Yang, B. Xu, W. Luo, F. Chen, Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review, Measurement. 189 (2022) 110460. https://doi.org/10.1016/j.measurement.2021.110460.

[80] H. Zhu, J. Cheng, C. Zhang, J. Wu, X. Shao, Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings, Applied Soft Computing. 88 (2020). https://doi.org/10.1016/j.asoc.2019.106060.

[81] B. Zhao, X. Zhang, H. Li, Z. Yang, Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions, Knowledge-Based Systems. 199 (2020) 105971. https://doi.org/10.1016/j.knosys.2020.105971.

[82] W. Qian, S. Li, J. Lu, Deep Sparse Topology Network for Robust Bearing Fault Diagnosis by Maximizing Prior Knowledge Functions, IEEE Transactions on Industrial Informatics. 18 (2022) 8540–8550. https://doi.org/10.1109/TII.2022.3148993.

[83] Z. Guo, M. Yang, X. Huang, Bearing fault diagnosis based on speed signal and CNN model, Energy Reports. 8 (2022) 904–913. https://doi.org/10.1016/j.egyr.2022.08.041.

[84] G. Niu, E. Liu, X. Wang, P. Ziehl, B. Zhang, Enhanced Discriminate Feature Learning Deep Residual CNN for Multitask Bearing Fault Diagnosis With Information Fusion, IEEE Transactions on Industrial Informatics. 19 (2023) 762–770. https://doi.org/10.1109/TII.2022.3179011.

[85] C. Huo, Q. Jiang, Y. Shen, C. Qian, Q. Zhang, New transfer learning fault diagnosis method of rolling bearing based on ADC-CNN and LATL under variable conditions, Measurement. 188 (2022) 110587. https://doi.org/10.1016/j.measurement.2021.110587.

[86] Z. Zhu, G. Peng, Y. Chen, H. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing. 323 (2019) 62–75. https://doi.org/10.1016/j.neucom.2018.09.050.

[87] Y. Zhang, T. Zhou, X. Huang, L. Cao, Q. Zhou, Fault diagnosis of rotating machinery based on recurrent neural networks, Measurement. 171 (2021) 108774. https://doi.org/10.1016/j.measurement.2020.108774.

[88]   X. Chen, B. Zhang, D. Gao, Bearing fault diagnosis base on multi-scale CNN and LSTM model, Journal of Intelligent Manufacturing. 32 (2021) 971–987.

[89]   J. Liang, L. Wang, J. Wu, Z. Liu, Elimination of End effects in LMD Based on LSTM Network and Applications for Rolling Bearing Fault Feature Extraction, Mathematical Problems in Engineering. 2020 (2020) 1–16. https://doi.org/10.1155/2020/7293454.

[90]   D. Gao, Y. Zhu, Z. Ren, K. Yan, W. Kang, A novel weak fault diagnosis method for rolling bearings based on LSTM considering quasi-periodicity, Knowledge-Based Systems. 231 (2021) 107413. https://doi.org/10.1016/j.knosys.2021.107413.

[91]   Y. An, K. Zhang, Q. Liu, Y. Chai, X. Huang, Rolling Bearing Fault Diagnosis Method Base on Periodic Sparse Attention and LSTM, IEEE Sensors Journal. 22 (2022) 12044–12053. https://doi.org/10.1109/JSEN.2022.3173446.

[92]   Y. Tong, P. Wu, J. He, X. Zhang, X. Zhao, Bearing fault diagnosis by combining a deep residual shrinkage network and bidirectional LSTM, Meas. Sci. Technol. 33 (2021) 034001. https://doi.org/10.1088/1361-6501/ac37eb.

[93]   X. Mao, F. Zhang, G. Wang, Y. Chu, K. Yuan, Semi-random subspace with Bi-GRU: Fusing statistical and deep representation features for bearing fault diagnosis, Measurement. 173 (2021) 108603. https://doi.org/10.1016/j.measurement.2020.108603.

[94]   Y. Cao, M. Jia, P. Ding, Y. Ding, Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network, Measurement. 178 (2021) 109287. https://doi.org/10.1016/j.measurement.2021.109287.

[95]   Z. Zhang, S. Li, J. Wang, Y. Xin, Z. An, General normalized sparse filtering: A novel unsupervised learning method for rotating machinery fault diagnosis, Mechanical Systems and Signal Processing. 124 (2019) 596–612. https://doi.org/10.1016/j.ymssp.2019.02.006.

[96]   J. Li, X. Li, D. He, Y. Qu, Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor, Journal of Intelligent Manufacturing. 31 (2020) 1899–1916. https://doi.org/10.1007/s10845-020-01543-8.

[97]   Y. Li, H. Cao, K. Tang, A general dynamic model coupled with EFEM and DBM of rolling bearing-rotor system, Mechanical Systems and Signal Processing. 134 (2019) 106322. https://doi.org/10.1016/j.ymssp.2019.106322.

[98]   Y. Zheng, Z. Xu, X. Wang, The fusion of deep learning and fuzzy systems: A state-of-the-art survey, IEEE Transactions on Fuzzy Systems. (2021).

[99]   G. Wang, J. Qiao, An efficient self-organizing deep fuzzy neural network for nonlinear system modeling, IEEE Transactions on Fuzzy Systems. (2021).

[100]   K. Su, J. Liu, H. Xiong, Hierarchical diagnosis of bearing faults using branch convolutional neural network considering noise interference and variable working conditions, Knowledge-Based Systems. 230 (2021) 107386. https://doi.org/10.1016/j.knosys.2021.107386.

[101]   C. Shen, X. Wang, D. Wang, Y. Li, J. Zhu, M. Gong, Dynamic Joint Distribution Alignment Network for Bearing Fault Diagnosis Under Variable Working Conditions, IEEE Transactions on Instrumentation and Measurement. 70 (2021) 1–13. https://doi.org/10.1109/TIM.2021.3055786.

[102]   K. Zhang, J. Wang, H. Shi, X. Zhang, Y. Tang, A fault diagnosis method based on improved convolutional neural network for bearings under variable working conditions, Measurement. 182 (2021) 109749. https://doi.org/10.1016/j.measurement.2021.109749.

[103]   H. Su, X. Yang, L. Xiang, A. Hu, Y. Xu, A novel method based on deep transfer unsupervised learning network for bearing fault diagnosis under variable working condition of unequal quantity, Knowledge-Based Systems. 242 (2022) 108381. https://doi.org/10.1016/j.knosys.2022.108381.

[104]   X. Zhao, J. Yao, W. Deng, P. Ding, J. Zhuang, Z. Liu, Multiscale Deep Graph Convolutional Networks for Intelligent Fault Diagnosis of Rotor-Bearing System Under Fluctuating Working Conditions, IEEE Transactions on Industrial Informatics. 19 (2023) 166–176. https://doi.org/10.1109/TII.2022.3161674.

[105]   T. Han, C. Liu, W. Yang, D. Jiang, Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application, ISA Trans. 97 (2020) 269–281. https://doi.org/10.1016/j.isatra.2019.08.012.

[106]   B. Zhao, X. Zhang, Z. Zhan, S. Pang, Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling

bearings under variable working conditions and domains, Neurocomputing. 407 (2020) 24–38. https://doi.org/10.1016/j.neucom.2020.04.073.

[107]   Y. Zhou, Y. Dong, H. Zhou, G. Tang, Deep Dynamic Adaptive Transfer Network for Rolling Bearing Fault Diagnosis With Considering Cross-Machine Instance, IEEE Transactions on Instrumentation and Measurement. 70 (2021) 1– 11. https://doi.org/10.1109/TIM.2021.3112800.

[108]   X. Yu, Z. Liang, Y. Wang, H. Yin, X. Liu, W. Yu, Y. Huang, A wavelet packet transform-based deep feature transfer learning method for bearing fault diagnosis under different working conditions, Measurement. 201 (2022) 111597. https://doi.org/10.1016/j.measurement.2022.111597.

[109]   X. Li, X. Jiang, Q. Wang, L. Yang, Z. Wang, C. Shen, Z. Zhu, Multi-perspective deep transfer learning model: A promising tool for bearing intelligent fault diagnosis under varying working conditions, Knowledge-Based Systems. 243 (2022) 108443. https://doi.org/10.1016/j.knosys.2022.108443.

[110]   G. Wang, M. Zhang, Lei Xiang, Z. Hu, W. Li, J. Cao, A multi-branch convolutional transfer learning diagnostic method for bearings under diverse working conditions and devices, Measurement. 182 (2021) 109627. https://doi.org/10.1016/j.measurement.2021.109627.

[111]   C. Zhao, W. Shen, Mutual-assistance semisupervised domain generalization network for intelligent fault diagnosis under unseen working conditions, Mechanical Systems and Signal Processing. 189 (2023) 110074. https://doi.org/10.1016/j.ymssp.2022.110074.

[112]   T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, E. Xu, Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions, ISA Transactions. 119 (2022) 152–171. https://doi.org/10.1016/j.isatra.2021.02.042.

[113]   Y. Wang, G. Sun, Q. Jin, Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network, Applied Soft Computing. 92 (2020) 106333. https://doi.org/10.1016/j.asoc.2020.106333.

[114]   T. Zhang, J. Chen, F. Li, T. Pan, S. He, A Small Sample Focused Intelligent Fault Diagnosis Scheme of Machines via Multi-modules Learning with Gradient

Penalized Generative Adversarial Networks, IEEE Transactions on Industrial Electronics. (2020) 1–1. https://doi.org/10.1109/tie.2020.3028821.

[115]   J. Liu, F. Qu, X. Hong, H. Zhang, A small-sample wind turbine fault detection method with synthetic fault data using generative adversarial nets, IEEE Transactions on Industrial Informatics. 15 (2018) 3877–3888. https://doi.org/10.1109/tii.2018.2885365.

[116]   Q. Li, L. Chen, C. Shen, B. Yang, Z. Zhu, Enhanced generative adversarial networks for fault diagnosis of rotating machinery with imbalanced data, Measurement Science and Technology. 30 (2019). https://doi.org/10.1088/1361-6501/ab3072.

[117]   F. Zhou, S. Yang, H. Fujita, D. Chen, C. Wen, Deep learning fault diagnosis method based on global optimization GAN for unbalanced data, Knowledge-Based Systems. 187 (2020) 104837. https://doi.org/10.1016/j.knosys.2019.07.008.

[118]   J. Fan, X. Yuan, Z. Miao, Z. Sun, X. Mei, F. Zhou, Full Attention Wasserstein GAN With Gradient Normalization for Fault Diagnosis Under Imbalanced Data, IEEE Transactions on Instrumentation and Measurement. 71 (2022) 1–16. https://doi.org/10.1109/TIM.2022.3190525.

[119]   J. Liu, C. Zhang, X. Jiang, Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet, Mechanical Systems and Signal Processing. 168 (2022) 108664. https://doi.org/10.1016/j.ymssp.2021.108664.

[120]   B. Zhao, Q. Yuan, Improved generative adversarial network for vibration-based fault diagnosis with imbalanced data, Measurement. 169 (2021) 108522. https://doi.org/10.1016/j.measurement.2020.108522.

[121]   S. Zhang, F. Ye, B. Wang, T.G. Habetler, Few-Shot Bearing Fault Diagnosis Based on Model-Agnostic Meta-Learning, IEEE Transactions on Industry Applications. 57 (2021) 4754–4764. https://doi.org/10.1109/TIA.2021.3091958.

[122]   J. Wu, Z. Zhao, C. Sun, R. Yan, X. Chen, Few-shot transfer learning for intelligent fault diagnosis of machine, Measurement. 166 (2020) 108202. https://doi.org/10.1016/j.measurement.2020.108202.

[123]   T. Zhang, J. Jiao, J. Lin, H. Li, J. Hua, D. He, Uncertainty-based contrastive prototype-matching network towards cross-domain fault diagnosis with small

data, Knowledge-Based Systems. 254 (2022) 109651.
https://doi.org/10.1016/j.knosys.2022.109651.

[124]   S. Liu, J. Chen, S. He, Z. Shi, Z. Zhou, Few-shot learning under domain shift: Attentional contrastive calibrated transformer of time series for fault diagnosis under sharp speed variation, Mechanical Systems and Signal Processing. 189 (2023) 110071. https://doi.org/10.1016/j.ymssp.2022.110071.

[125]   Y. Feng, J. Chen, T. Zhang, S. He, E. Xu, Z. Zhou, Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis, ISA Transactions. 120 (2022) 383–401. https://doi.org/10.1016/j.isatra.2021.03.013.

[126]   W. Huang, J. Cheng, Y. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, Neurocomputing. 359 (2019) 77–92. https://doi.org/10.1016/j.neucom.2019.05.052.

[127]   Janardan, S. Mehta, Concept drift in Streaming Data Classification: Algorithms, Platforms and Issues, Procedia Computer Science. 122 (2017) 804–811. https://doi.org/10.1016/j.procs.2017.11.440.

[128]   S.U. Din, J. Shao, J. Kumar, C.B. Mawuli, S.M.H. Mahmud, W. Zhang, Q. Yang, Data stream classification with novel class detection: a review, comparison and challenges, Knowl Inf Syst. 63 (2021) 2231–2276. https://doi.org/10.1007/s10115-021-01582-4.

[129]   J. Gama, R. Rocha, P. Medas, Accurate decision trees for mining high-speed data streams, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, 2003: pp. 523–528. https://doi.org/10.1145/956750.956813.

[130]   S. Jia, A VFDT algorithm optimization and application thereof in data stream classification, J. Phys.: Conf. Ser. 1629 (2020) 012027. https://doi.org/10.1088/1742-6596/1629/1/012027.

[131]   X. Mu, K.M. Ting, Z.-H. Zhou, Classification Under Streaming Emerging New Classes: A Solution Using Completely-Random Trees, IEEE Transactions on Knowledge and Data Engineering. 29 (2017) 1605–1618. https://doi.org/10.1109/TKDE.2017.2691702.

[132]   Y. Gao, S. Chandra, Y. Li, L. Khan, T. Bhavani, SACCOS: A Semi-Supervised Framework for Emerging Class Detection and Concept Drift Adaption Over Data Streams, IEEE Transactions on Knowledge and Data Engineering. 34 (2022) 1416–1426. https://doi.org/10.1109/TKDE.2020.2993193.

[133]   X.-Q. Cai, P. Zhao, K.-M. Ting, X. Mu, Y. Jiang, Nearest Neighbor Ensembles: An Effective Method for Difficult Problems in Streaming Classification with Emerging New Classes, in: 2019 IEEE International Conference on Data Mining (ICDM), 2019: pp. 970–975. https://doi.org/10.1109/ICDM.2019.00109.

[134]   J. Zhang, T. Wang, W.W.Y. Ng, W. Pedrycz, KNNENS: A k-Nearest Neighbor Ensemble-Based Method for Incremental Learning Under Data Stream With Emerging New Classes, IEEE Transactions on Neural Networks and Learning Systems. (2022) 1–8. https://doi.org/10.1109/TNNLS.2022.3149991.

[135]   Y.-N. Zhu, Y.-F. Li, Semi-Supervised Streaming Learning with Emerging New Labels, Proceedings of the AAAI Conference on Artificial Intelligence. 34 (2020) 7015–7022. https://doi.org/10.1609/aaai.v34i04.6186.

[136]   D.-W. Zhou, Y. Yang, D.-C. Zhan, Learning to Classify With Incremental New Class, IEEE Transactions on Neural Networks and Learning Systems. 33 (2022) 2429–2443. https://doi.org/10.1109/TNNLS.2021.3104882.

Appendix A

Main publications

Paper A:

Yu Wang, R. H. Hestmo, A. Vinogradov, Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder, Measurement Science and Technology, vol. 34, pp. 064008, March, 2023.

URL: http://iopscience.iop.org/article/10.1088/1361-6501/acc1f8

Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder

Yu Wang[1],* , Rune Harald Hestmo[2] and Alexei Vinogradov[1]

[1] Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology—NTNU, N-7491 Trondheim, Norway
[2] Kongsberg Maritime AS, Trondheim, Norway

E-mail: yuwa@ntnu.no

CrossMark

Abstract

Bearings are a crucial component of wind turbines. The acoustic emission (AE) technique offers the advantage of earlier detection of defects and failures of bearings in comparison to traditional vibration techniques. Parameter-based analysis is the most widely used approach to interpret AE waveforms, partly due to the challenges arising in the processing of large amounts of streaming data. In this work, the AE technique is applied to monitor a run-to-failure process of a roller bearing, and it is found that the use of multiple known parameters, such as the root mean square, skewness, crest factor, impulse factor etc, fails to characterise the evolution of the acquired AE signals, thus highlighting the long-standing necessity and significance of developing new AE indicators that are more adequate to detect the failure of rotating machines. We propose a hybrid parameter—the information entropy penalty factor (IEPF)—which uses the advantages of the entropy theory and deep learning methods. The effectiveness of the proposed method has been investigated and demonstrated for roller bearing contact fatigue experiments, and the results show that IEPF can timely and accurately detect the incipient sub-surface faults.

Keywords: sub-surface fault detection, bearings, acoustic emission signal, parameter analysis, energy entropy

(Some figures may appear in colour only in the online journal)

---

* Author to whom any correspondence should be addressed.

1. Introduction

Condition monitoring (CM) of rotating machines has been a hot topic for decades. Wind turbines are rotating machines that have evolved to become pivotal components for the generation of green energy. Wind turbines are usually installed in extreme and harsh environments and are prone to a high failure rate. Bearings are essential and highly demanding components of wind turbines. Faults in the bearings can lead to critical failures, breakdowns and consequent losses associated with the downtime of wind turbines. Therefore, CM and timely and accurate fault diagnosis offers substantial benefits to operating equipment with rolling element bearings by identifying incipient damage at as early stage as possible before the faults evolve to a critical stage. This is especially important for machines, where a fault can cause irrecoverable damage to the environment, and not least to avoid losses of human life or health.

Multiple sensing techniques have been employed in bearing CM systems used in industrial settings in general and for wind turbine CM in particular. Monitoring and trending the temperature of a bearing is a simple and cost-effective method to identify a bearing condition. However, in most applications, the temperature measurements are not sensitive enough to detect an early stage of fault development in a roller bearing. Instead, vibration analysis has become the most widespread and market-leading technology due to its simplicity, robustness and multiple uses for custom-built solutions. Various acquisition and analysis tools have been established and proven effective for vibration data. However, the vibration signals induced by tiny defects at the early stage of their development can be easily masked by the uncontrolled mechanical disturbances from the rotating machine. Moreover, the vibration acceleration signals can go undetected in heavy or slowly rotating structures until the fault increases significantly to a large (detectable) scale, by which stage it is often too late for preventive/corrective maintenance and is close to a catastrophic failure. As opposed to vibrations, acoustic emissions (AEs) reflecting the dynamics of the sources evolving under load can be generated even by microscopic flaws, such as breaks of hard non-metallic inclusions, incipient cracks, etc [1]. Moreover, the AE signal tends to increase with the growing scale of the sources. Therefore, the potential of the AE technique for early fault detection enjoys growing recognition in the industrial domain. AE methods have become an important companion of reliable monitoring systems when the impact of wear and friction of rotating components is of concern. AE is commonly defined as a phenomenon whereby transient elastic waves are spontaneously emitted by the rapid stress relaxation within localised sources in material under load. Plastic deformation and fracture associated with the nucleation and growth of cracks represent the primary mechanisms of the sources releasing the elastic strain energy associated with AE transients [2]. In contrast with the vibration signal, the sources generating AE signals are characterised by a much wider frequency range (100 kHz and 1 MHz) [3], which does not overlap significantly with low-frequency mechanical vibration signals caused by imbalance or misalignment of machine

components [4, 5]. A great deal of evidence has been accumulated, suggesting that AE parameters can reveal the faults in rotating equipment before they show up in the vibration acceleration range. Since the early work by Yoshioka *et al* [6], these results have been investigated and confirmed in abundant literature over the last 30 years; see [2, 7, 8] for examples.

With the advent of artificial intelligence, machine learning methods have become more and more extensively applied in the field of fault diagnosis. Deep learning (DL) technology is the most prominent branch of machine learning (ML) methodology, and refers explicitly to artificial neural networks with a multi-layered architecture. A number of DL architectures, such as convolutional neural networks (CNNs) [9, 10], long short-term memory (LSTM) [11] and autoencoder [12, 13], have been applied in the CM field and demonstrated outstanding potential and practicality. However, most of the relevant methods developed in this field are based on artificially seeded defects and supervised circumstances. In the reality of the run-to-failure scenario, only the data characterising the 'healthy' status of the object under inspection are accessible before the emergence of the faults. Hence, the detection of defect initiation is fundamentally an unsupervised task, and relevant studies are still scarce. Although some unsupervised DL architectures have been developed, e.g. the stacked autoencoder, deep brief network and deep Boltzmann machine, they are mainly employed with only an auxiliary role on supervised subjects; this is generally followed by a supervised model or an extra fine-tuning procedure, as proposed in [12–15]. Aiming at early fault detection, Lu *et al* proposed a DL-based architecture comprising three network blocks—a basic autoencoder, a feature extraction layer and an LSTM-based autoencoder [16]. Autoencoder is a prevalent unsupervised DL model designed to reconstruct its own input data with the learning objective to minimise the reconstruction error. It is reasonable that the reconstruction error can indicate emerging faults. Since the acquired signal may suffer serious distortion during the run-to-failure process, it can be foreseen that autoencoder will be unable to reconstruct the input correctly, thus leading to increasing reconstruction errors, which serves as a fault indicator; see [17, 18]. The above-cited works are based on vibration data. The application of AE has yet to be tested. In addition, one needs to bear in mind that the tolerance of the neural network to small variations in the data may limit the effectiveness of the entire ML-based approach. Thus, the application of DL models to early fault detection in the run-to-failure process faces serious challenges.

Up to now, the vast majority of existing studies deal with vibration signals, while attempts to pair DL methods and the AE technique are still limited. The parameter-based methods still dominate the philosophy of the AE waveform analysis. Therefore, the relevance of the involved parameters strongly affects the performance of the detectors. The conventional AE features extracted from AE waveforms include, but are not limited to, AE hit parameters such as counts, duration, rise time, counts to peak, amplitude, etc [2, 19, 20], as well as statistical parameters/features such as root mean square (RMS), kurtosis, crest factor, skewness, etc [5, 7, 21], defined in the time domain. In addition, multiple signal processing

techniques involve spectral decomposition techniques, such as Fourier transformation [22], wavelet analysis [23–25], variational mode decomposition [26], etc, to assess the AE signal in the frequency and time–frequency domains. As is commonly seen in the general statistical analysis of random data, different features characterise the AE waveforms from different angles, thus providing a great variability in features, as well as a range of strategies and options for their analysis, interpretation and decision-making.

Raw pseudo-AE waveforms harvest a wealth of mechanical interactions from rotating components, splashing oil, electrical interferences and multiple other noise-like sources of unknown origin. Therefore, the AE signal represents periodical patterns arising in response to the roller movement. To characterise the periodicity and its disturbance embedded in AE signals, a hybrid parameter that combines DL and the information entropy (IE) theory is introduced in this work. As a natural measure of uncertainty and chaos, IE provides new insight into the underlying AE process. There is no standard way to acquire IE from the AE signal. Elforjani and Mba [27] adopted the probabilities of AE events in a given AE signal to obtain the IE value. They showed that IE was more sensitive and representative than the kurtosis and crest factor. Amiri *et al* calculated the AE entropy based on counts [28]. Kahirdeh *et al* proposed three similar IE models using AE counts, accumulated counts or the estimated histogram of the AE signal [29]. However, these methods commonly suffer from shortcomings associated with the AE hit (and corresponding parameters) definition depending heavily on the present amplitude threshold, which introduces irrecoverable uncertainty in low-amplitude and/or overlapping signal detection. Thus, the early AE events are hard to identify because the AE signals caused by incipient faults are usually of low amplitude and can be masked by strong noise. This conclusion concurs with the literature review provided in [30]. Several studies have been proposed to obtain IE from the histogram of the AE signal, as documented in [29–31].

The main contributions of this work can be summarised as follows. (1) To investigate the capacity of the AE technique in sub-surface fault detection of bearings, a laboratory durability test of a roller bearing element was carried out; roller contact fatigue damage was initiated under controlled conditions, and the accompanying AE waveforms were acquired. (2) Aimed at detecting the emerging faults timely and accurately, a health indicator combining the IE theory and autoencoder was proposed to describe the evolution of AE waveforms during the run-to-failure process, which is referred to as the information entropy penalty factor (IEPF). (3) The proposed parameter is demonstrated to be more sensitive to the periodicity and disturbance in the AE signal. (4) The high sampling frequency of AE technology limits the application of DL methods; thus, a moving variance window (MVW) was utilised to reduce the dimensions of raw AE signals. Then, autoencoder was applied to denoise the signal for feature augmentation.

The rest of the paper is organised as follows. The mathematical details of the proposed method are unfolded in section 2. The test rig and the implementation details of the proposed method are introduced in section 3, along with the experimental results and discussion. Conclusions are formulated in section 4.

2. Methodology

2.1. Basic theory of autoencoder

Since it is an unsupervised task to detect the onset of early fault during the durability test to failure, a prevalent unsupervised network architecture—autoencoder—is chosen for the present work. The theoretical background of the involved neural network architectures is presented in the following sub-sections. Autoencoder aims at reconstructing its own input data. The basic form of the autoencoder is relatively simple—it is a symmetrical three-layer neural network consisting of input, hidden and output layers representing an encoder and decoder pair. For a given dataset X, the mathematical details for encoding and decoding are represented as follows:

$$\text{Encoder}: \boldsymbol{H} = \text{Activ}\left(W_{\text{e}}^T * \boldsymbol{X} + b_{\text{e}}\right) \tag{1}$$

$$\text{Decoder}: \tilde{\boldsymbol{X}} = \text{Activ}\left(W_{\text{d}}^T * \boldsymbol{H} + b_{\text{d}}\right) \tag{2}$$

where W_{e}^T and W_{d}^T stand for the weights of the encoder and decoder, respectively, and b_{e} and b_{d} are the corresponding biases. The encoder is regarded as a feature extractor, and the output $\boldsymbol{H}$ is the latent representation containing the main information of the input data. $\tilde{\boldsymbol{X}}$ is the reconstructed data that is decoded from the latent representation $\boldsymbol{H}$.

To minimise the distance between X and its reconstruction $\tilde{X}$, the mean square error (MSE) loss function J_{MSE} is generally used, which is expressed as

$$J_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} \left\| X_i - \tilde{X}_i \right\| \right)^2 \tag{3}$$

where n denotes the total number of samples.

2.2. Energy entropy (EE)

The EE is a measure of IE, which is based on the change in the energy of the signal. A moving energy window (MEW) is applied to slide over the signal to construct the probability distribution of the energy. Given a recorded AE signal X, the moving window is defined as

$$\mathbf{win}_{k,l,s}^X = X\left[x_{start}^{k,s} : x_{start}^{k,s} + l - 1\right] \tag{4}$$

$$x_{\text{start}}^{k,s} = (k-1) \times s + 1, \ k = 1, 2, \ldots n_k \tag{5}$$

where $\mathbf{win}_{k,l,s}^X$ represents the area of the signal X covered by the moving window; k, l and s are integers specifying the moving step, window length and moving stride, respectively; $x_{\text{start}}^{k,s}$ is the start point of the window on the signal X. The energy of the overlaid region is extracted at each moving step. The total

number of moving steps is $n_k = [(N - l)/s] + 1$, where N is the length of the recorded AE signal. Therefore, the partial energy of the signal E_k and its probability distribution P_k are obtained as

$$E_k = \sum_{x \in \mathbf{win}_{k,l,s}^X} |x|^2 \tag{6}$$

$$P_k = \frac{E_k}{\sum_{k=1}^{n_k} E_k}. \tag{7}$$

The MEW length is recommended to contain information about at least one entire axel revolution of the rotating machine. Thereby, it is determined by the lowest axle rotation frequency and sampling frequency. With the probability distribution, the EE is acquired based on Shannon's entropy formula:

$$H = -\sum_{k=1}^{n_k} P_k \log_b P_k. \tag{8}$$

The logarithmic base 'b' defines the unit of the measured information. The units include bits ($b = 2$), nats ($b = e$), and bans ($b = 10$) [32]. In the case of $P_k = 0$, the value of $0 \log_b 0$ is taken to be 0; therefore, the minimum value of entropy is 0.

2.3. The proposed method

In this paper, a new fault indicator combining the EE and reconstruction error of autoencoder is proposed, which is referred to as IEPF. The details are presented below.

2.3.1. Feature augmentation. To capture transient changes within the signal, an MVW, which calculates the sample variance, is applied to the original signals. For a recorded AE signal X, the procedure is formulated as follows:

$$X_{MVW} = \frac{1}{l} \sum_{x \in \mathbf{win}_{k,l,s}^{X_{MVW}}} |x - \mu|^2 \tag{9}$$

where μ is the mean of x within $\mathbf{win}_{k,l,s}^{X_{MVW}}$. The function of the MVW is to capture the transient events and highlight some important detailed information about the data. The output is a dimensionless number that measures the dispersion of the data, and thereby, the signal is de-dimensionalised. Additionally, the dimension of the original signal is largely reduced through this process, which makes it easier to be processed by the neural network.

Then, autoencoder is employed in this work to denoise and enhance the main features of the signal. With the target of reconstructing its own input data, autoencoder has been widely used for feature extraction. However, the reconstruction error is inevitable, which compels the network to outline the main features of the input data and neglect some redundant noise. To better reconstruct the input data, the CNN architecture is used to extract detailed information. The applied autoencoder architecture is shown in figure 1.
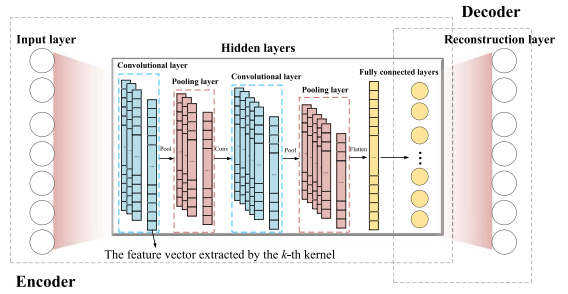
Figure 1. The neural network architecture used in the present work.

2.3.2. IEPF. During the run-to-failure process, the acquired AE signal may experience serious changes with the damage propagation through the test piece. It can be foreseen that autoencoder will eventually be unable to reconstruct the deformed signal and cause the MSE value to increase. Several researchers proposed the reconstruction error of autoencoder as an indicator of an early fault or anomaly in the mechanical behaviour of the system [16–18]. However, the neural network has a certain tolerance to waveform changes; i.e. if the discrepancy between signals is not very large, autoencoder can still fit the data with a relatively low reconstruction error. Since MSE is inevitable, the information provided by the reconstructed data itself is incomplete. Therefore, the EE is further adapted to utilise the advantages of autoencoder. To implement this coupling, equation (6) is rewritten as

$$\widetilde{E}_k = \sum_{x \in \mathbf{win}_{k,l,s}^{X_{MVW}}} \left| \widetilde{\boldsymbol{x}_{MVW}} + \delta \right|^2 \tag{10}$$

where $\widetilde{\boldsymbol{x}_{MVW}}$ stands for the reconstruction of data covered by $\mathbf{win}_{k,l,s}^{X_{MVW}}$, and δ is the corresponding MSE value. By plugging $\widetilde{E}_k$ into equation (7), a new probability distribution of the reconstructed data is adopted as $\widetilde{P}_k$.

The maximum value of Shannon's entropy is obtained when all elements are identically distributed, i.e. when the dynamic range of H is limited to $\left[0, -\sum_{n_k} \frac{1}{n_k} \log \frac{1}{n_k}\right]$. The maximum value is greater than 0, increasing monotonically over the range of n_k. Theoretically, the greater the disturbance of the AE signal, the smaller the value of Shannon's entropy. However, for the sake of convenience, one can modify the entropy calculation in such a way that it will increase with the variation in the AE signal as

$$IEPF = \sum_{i=1}^{n_k} \left(\widetilde{P}_k \log_2 \widetilde{P}_k - \frac{1}{n_k} \log_2 \frac{1}{n_k} \right). \tag{11}$$

If the IEPF value is close to 0, it indicates that the signal has strong periodicity and vice versa.

2.3.3. General procedures. The general fault detection procedures are summarised as the following five steps:

Step 1: Data acquisition. AE sensors are installed on the testing rig machine; the machine runs to failure of the rolling bearing element, and the AE waveform is recorded periodically. The acquired AE signals at the initial stage of the experiment serve as training data, and are utilised for training autoencoder and constructing a fault threshold. The rest of the data are the testing set.

Step 2: Signal pre-processing. Two moving windows are applied to the acquired AE signal.

(1) MVW is first applied to the original signal for dimension reduction and extracting detailed information.
(2) MEW is applied to the data; instead of extracting the energy feature at each moving step immediately, the covered data are prepared as the input of autoencoder.

Step 3: Data reconstruction. The processed signals are fed into autoencoder for denoising.

Step 4: Obtain health indicator. Calculate the energy probability distributions of the reconstructions. A fault alarm threshold is constructed on the basis of the IEPF values calculated from the training data.

Step 5: Decision making. The IEPF value of the testing set exceeding the threshold is considered a fault alarm.

3. Experiment and discussion

3.1. Test rig and data acquisition

To monitor the rolling contact fatigue phenomenon occurring in a roller bearing element, a run-to-failure test was carried out using an instrumented special-purpose testing rig designed at SINTEF Industry (Trondheim, Norway). The experimental setup is schematically illustrated in figure 2(a). The test specimen (central roller) is supported by three rollers, and each roller is supported by two needle bearings SKF NA 6914-zw. The wideband differential (WD) sensors (MISTRAS, USA) (only one sensor was used during the test) were connected to the data acquisition system as displayed in figure 2(b). The signal was amplified by 40 dB in the frequency band 20–1200 kHz by a 2/4/6 low-noise preamplifier (MISTRAS, USA).

The AE recording started automatically when the axle rotation frequency was greater than the threshold. After warming up to 47 ± 2 °C, the initial axle rotation frequency was set at 364 rpm at the initial load of 67.1 kN, corresponding to 1807 MPa contact stress. The test was interrupted periodically, as is indicated by the vertical lines in the test diagram represented in figure 6 for ultrasonic inspections performed with an Olympus OMNISCAN SX phase array ultrasonic scanner (PAUT). As the PAUT inspections revealed no faults after initial cycling up to approximately 3×10^6 cycles (10^6 axel rotations), the load gradually increased in a stepwise manner up to 91.3 kN (2002 MPa contact stress). The cumulative number of fatigue cycles reached 2.7×10^7 cycles. Excessive vibrations were detected in the machine at this load when running

(a) Rolling fatigue test rig

(b) Sensor's position
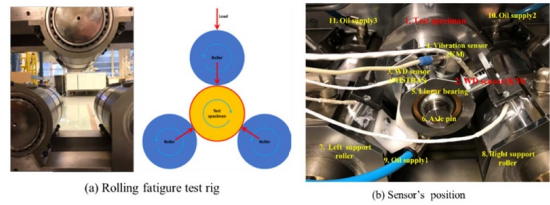
Figure 2. Rolling fatigue test rig: photographic image and schematics of the geometry of supporting rollers and the testing roller (a), and a close-up view of the setup instrumented with AE sensors (b).

Table 1. Number of AE records for different stages of damage propagation.

Health condition	Number of records	Number of fatigue cycles
No damage	542	2.8×10^7
0.5 mm crack	377	3.6×10^7
1 mm crack	809	4.8×10^7
1.5 mm crack	718	6.5×10^7
2 mm crack	25	6.6×10^7

at a rotational speed of 364 rpm. Therefore, the axle rotation frequency was reduced to 256 rpm until the end of the test. The test was continued with a 91.3 kN load, and the first sub-surface crack that was detected by the PAUT was after 2.8×10^7 fatigue cycles at approximately 4 mm below the contact surface. The smallest detected crack was estimated to be 0.5 mm long. The continued regular PAUT inspections revealed continuous slow crack growth in the longitudinal direction up to 2 mm length along the roller axis before the test was terminated. The test roller was then sectioned for metallographic inspection and verification of the PAUT results. As predicted by PAUT, three sharp fatigue cracks were observed beneath the surface.

The AE waveforms were continuously recorded at a 2 MHz sampling frequency for 2 s per record using the Kongsberg HSIO-100-A high-speed acquisition module. At the beginning of the test, AE streams were collected every 60 min. After the confirmation of the first sub-surface crack, the time interval between the successive AE acquisitions was reduced to 20 min. In total, 2471 records were qualified for the analysis. The number of records corresponding to different stages of crack growth is presented in table 1. The records are indexed from 1 to 2471 according to the time of acquisition. The recorded raw AE signals are plotted in figure 3 for illustration. An appreciable change in the AE amplitude is first observed after 4.6×10^7 fatigue cycles. Ultrasonic inspections revealed a crack of 1 mm at this stage.

Several randomly chosen AE records, which are typically observed during different stages of the damage propagation, are shown in figure 4. The evolution of the AE waveforms can be observed. At first, the AE waveforms exhibited evident periodical characteristics as shown in figure 4(1) due to the routine operation of the rotating machine. After the fatigue cycle was

Figure 3. Diagram showing the raw AE signals against the number of fatigue cycles.

Figure 4. Examples of randomly chosen representative AE records at different stages of damage propagation.

Table 2. Autoencoder architecture.

Layer	Key parameters	Output size	Activation function
Input	/	1×1024	/
C1	$1 \times 25@16$	$1 \times 1000@16$	ReLu
P1	$1 \times 5@16$	$1 \times 200@16$	/
C2	$1 \times 16@32$	$1 \times 185@32$	ReLu
P2	$1 \times 5@32$	$1 \times 37@32$	/
FC1	/	1×1184	/
FC2	1184×100	1×100	ReLu
Output	/	1×1024	Sigmoid

Note: C, P and FC denote the convolutional layer, pooling layer and fully connected layer, respectively. The notation '$a \times b@c$' describes the kernel size and the output size, where a and b represent the row and column of the matrix, and c denotes the number of channels.

3.2. Implementation details

The MEW should contain information about at least one entire axel revolution. For instance, the lowest axle rotation frequency in the present work is 254 rpm, i.e. for a 2 s recording, eight complete rotations are captured. Therefore, the moving step of MEW should be 8, and the maximum value of IEPF is calculated as 3. Based on the sampling frequency used (2 MHz), the window length and the moving stride of MVW and MEW are set at 464 and 1024, respectively.

The first 60% of healthy data (325 recorded AE fragments) were used to train the neural network and calculate the fault alarm threshold. The threshold is calculated conventionally as $\text{mean}(x) \pm 3 \times \text{std}(x)$, where x denotes the IEPF values of training data, and mean and std stand for the mean value and the standard deviation, respectively. Since each record contains eight complete rotations, after being processed by MVW and MEW, 2471 training samples were constructed to train autoencoder. The size of the training dataset is 2471×1024. The rest of the records are testing data. Both the training and testing data are normalised using the maxminmap method before feeding into autoencoder. Details of the network architecture are presented in table 2.

The reconstructed data are randomly exemplified from different stages of the experiment, as shown in figure 5. Each image represents eight stacked sub-signals covered by MEW. It is hard to identify the difference between healthy and faulty data from the raw signal by the naked eye, especially at the early damage stage featured by the 0.5 mm crack length; see, for example, indices 448, 452, 561 and 805. However, the reconstructions unveil more clear features if compared to those of the raw signal. The reconstruction error (MSE) of autoencoder is shown in figure 6. As mentioned before, autoencoder is reasonably tolerant to waveform changes; i.e. if the discrepancy between signals is not very large, the neural network can still fit the data with a low reconstruction error. As shown in figure 6, the reconstruction error of the trained network is still very low, especially at the earliest crack growth stage. The drastic increase in the reconstruction error appears only with the emergence of AE bursts. Mathematically, this is because the sigmoid activation function maps the output to the range

accumulated to 2.8×10^7, the initial periodic behaviour in the waveforms disappeared—the effect is assumed to be related to the generation of AE signals from the defect. The continuous background quasi-steady AE signal is assumed to be produced primarily by the over-rolling of roughness asperities [23]. If a defect forms on either the surface or in the bulk of the test roller, and if the sensor successfully captures it, the two random processes—the background noise and the defect-induced AE—overlap additively. The AE associated with the defect can be considered as a disturbance, distorting the waveform of the original signal. Assuming that the AE signal recorded from the healthy stage is X, the AE signal reflecting fatigue damage is denoted as $\overline{X} = X + \tau$, where τ represents the AE response to the defect that emerged in the roller. This type of signal appears as illustrated in figure 4(2–3). With the propagation of the fault, the peak amplitude of the corresponding AE burst signal clearly exceeds the noise threshold at periodic intervals, as shown in figure 4(4–5). The triple roller arrangement shown in figure 1 assumes that each point of the test roller interacts with the support rollers three times per revolution. When the test roller containing the surface (or sub-surface) faults contacts the support roller, the stress concentration in the bearing elements along the defect boundary is expected to cause an increase in the released elastic energy [2], resulting in periodical spikes in the AE waveforms.

Figure 5. Visualisation of the reconstructed data from different stages of damage propagation.
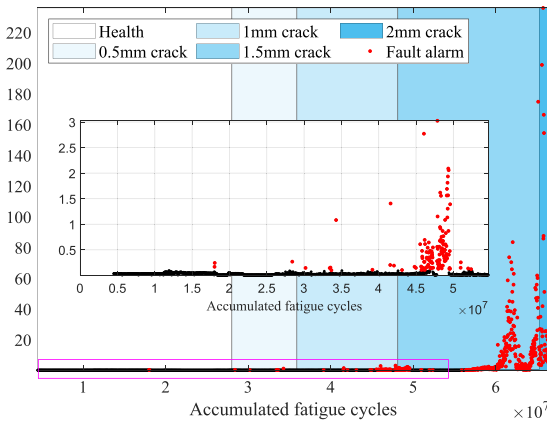
Figure 6. MSE (reconstruction error) of autoencoder; the inset shows a magnified view of MSE in the range of record indices from 0 to 2000.

Table 3. Statistical parameters extracted from time domain, frequency domain and time–frequency domain.

Domains	Parameters
Time domain	(1) Root mean square (RMS); (2) skewness; (3) kurtosis; (4) shape factor; (5) crest factor; (6) impulse factor; (7) margin factor; (8) histogram-based information entropy (IE); (9) energy entropy (EE); (10) power.
Frequency domain	(11) Mean frequency (MF); (12) RMS of frequency (RMSF); (13) standard deviation of frequency (SF).
Time–frequency domain	(14) STFT + power entropy; (15) STFT + MF entropy; (16) STFT + RMSF entropy; (17) STFT + SF entropy; (18) wavelet packet energy entropy (WPEE); (19) wavelet packet singular entropy (WPSE).

Note: (14–17) short-time Fourier transform (STFT) was implemented with a hamming window with a length equal to 4096 readings. The window slid over the original data to calculate the discrete Fourier transform of the windowed data, and the overlap of each moving step was 512. (18–19) Wavelet packet transform was applied to perform three-layer decomposition of the original AE signal using the 'dmey' wavelet, and results in eight decomposed frequency bands.

of (0, 1); however, the peak of the AE burst will exceed the upper boundary of the sigmoid function without the corresponding training data. The results manifest that the MSE indicator taken alone is not sufficient to identify early faults.

3.3. Evaluation methods

In the following analysis, the performance of conventional statistical parameters is investigated and compared with the one proposed in this work. The quality of the probed parameters is assessed from two aspects: (1) timely and accurate detection of emerging faults and (2) better description of the AE waveform evolution. The 19 statistical parameters

listed in table 3 were extracted from the time domain, frequency domain and time–frequency domain and were probed for the sake of comparison. To quantify the performance of all these indicators, the data were categorised into two classes as 'healthy' and 'faulty', and three evaluation indicators— Accuracy, Specificity and F_1-score were measured, as defined below [33]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (12)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (13)$$
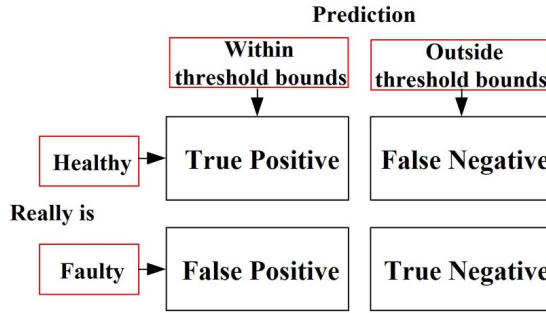
Figure 7. Definition of the confusion matrix.
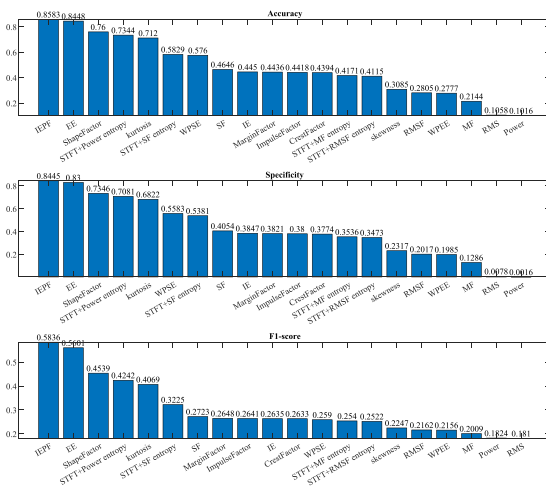
Figure 8. Accuracy, specificity and F_1-score of the probed parameters sorted in descending order.

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (14)$$

where TP, FN, FP and TN are abbreviations of true positive, false negative, false positive and true negative, respectively, as described in figure 7. Accuracy measures all correctly classified samples. Specificity quantifies the ratio of negative class predictions of all negative samples. The F_1-score provides a single score that balances both the concerns of precision and recall. Precision and recall are defined as $TP/TP + FP$ and $TP/TP + FN$, and quantify the number of correct positive results divided by all positive results and relevant samples, respectively.

Accuracy, Specificity and F_1-score of the probed parameters are compared in figure 8. Based on these three quality indicators, the proposed parameter exhibits the highest scores. IEPF generates fewer false fault alarms and more true fault alarms compared with other parameters tested. Although parameters such as RMS, skewness, crest factor, impulse factor,

etc, have been used with greater or lesser success by many researchers, in the present settings, they perform quite unsatisfactory. This prompted us to seek new reliable parameters.

The top eight parameters performing better in Accuracy (excluding IEPF) were selected for further comparison. Table 4 presents the values of accuracy of the selected parameters at different experimental stages. The accuracy at each stage was obtained using the formula F/N, where N denotes the total number of records at a specific experimental stage, and F represents the true fault alarms at this stage. Most of the parameters indicated the AE waveform changed substantially when the crack was propagated to the mature stage with a final length of up to 2 mm. Although parameters like wavelet packet singular entropy (WPSE) and standard deviation of frequency show relatively high accuracy at the 0.5 mm crack stage, WPSE exhibits a higher rate of false alarms at the healthy stage, and standard deviation of frequency (SF) fails to detect the propagation of the fault. Since failure is an irreversible and progressively propagating process, the indicator is expected to be continuous and monotonic. Compared with other parameters, the IEPF generates notably fewer false alarms and more true fault alarms.

Scores of IEPF are plotted in figure 9 against the cumulative number of fatigue cycles. The red dots represent the events with IEPF values exceeding the threshold, which are denoted as fault alarms. One can see that IEPF transparently characterises the evolution of the recorded AE waveforms from the following aspects. First, the IEPF value corresponding to the initial healthy stage is approximately 0 (the average IEPF value of the recorded AE signal at the healthy stage is 0.0026), which indicates that the recorded AE signals present strong periodical patterns. The IEPF increases steeply in the second stage when the first 0.5 mm crack is detected. Thus, a breakpoint between the healthy and faulty stages can be easily identified. Second, the IEPF value captures the initiation of the persistent AE bursts at the intersection of the 1 mm crack and 1.5 mm crack. Additionally, the results successfully characterise the increase in the AE bursts after the 2000th record while maintaining the general trend towards higher values.

To further compare the performance of IEPF with other parameters, figure 10 shows the variation of the selected parameters from the beginning of the test to failure. Although EE

Table 4. The accuracy of selected parameters at different stages of damage propagation.

Parameter	Fault conditions				
	No damage	0.5 mm crack	1 mm crack	1.5 mm crack	2 mm crack
IEPF	0.9742	0.4259	0.9148	0.9791	1
EE	0.9723	0.4233	0.8938	0.9652	1
Shape factor	0.9815	0.3836	0.6642	0.9889	1
STFT + power entropy	0.9705	0.4233	0.5889	0.9819	1
Kurtosis	0.9760	0.3307	0.5691	0.9833	1
STFT + SF entropy	0.9797	0.3598	0.3370	0.8428	1
WPSE	0.8930	0.6614	0.1037	1.0000	1
SF	0.9779	0.6111	0.1938	0.5369	0.3462
IE	0.9742	0.0397	0.1889	0.7650	1
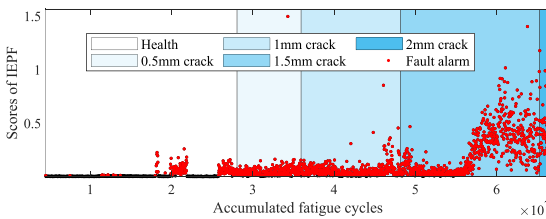
Figure 9. The scores of IEPF from the durability test of the roller bearing element.
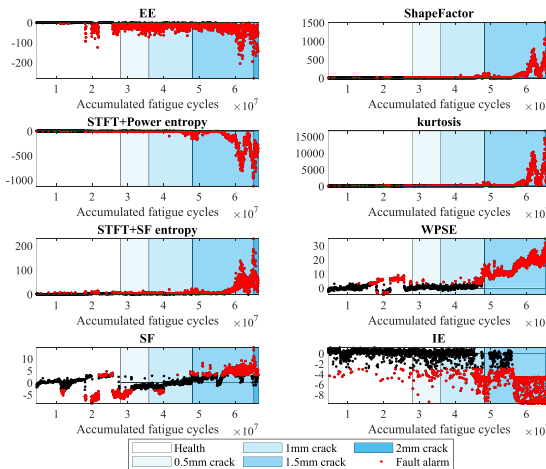
Figure 10. The scores of selected parameters of the AE signals from the durability test of the roller bearing element.

also presents relatively high accuracy in figure 8, it fails to characterise the AE behaviour in response to the crack growth up to 1 mm and further to 1.5 mm length. Compared to other parameters, the IEPF shows excellent sensitivity to the emergence of periodical AE impulses and exhibits a clearer description of the waveform evolution corresponding to the propagation of internal fatigue cracks.

4. Conclusion

In this paper, a durability test of a roller bearing element was carried out to investigate the application of the AE technique to sub-surface fault detection in a roller. The experimental results show that many known parameters, such as RMS, skewness, crest factor, impulse factor etc, fail to characterise the evolution of AE signals in relation to the damage initiation and propagation. Therefore, a hybrid parameter called IEPF is proposed to assess the fault behaviour through the evolution of AE waveforms. The proposed method combines the advantages of information theory and autoencoder to achieve a high sensitivity to the periodicity and its disturbance in AE signals. Comparative tests were carried out to assess the quality of the health status indicators from two aspects: (i) timely and accurate detection of emerging faults and (ii) a more elucidative description of the AE waveform evolution in response to the emerging and propagating fatigue damage. The experimental results verify the effectiveness of the proposed data processing scheme for fault monitoring and possible diagnostics in roller bearings. The proposed methodology can be reasonably easily adapted to the CM of other rotating machines since it is driven primarily by data and does not rely on specific knowledge of the mechanical features of the system under control.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

ORCID iD

Yu Wang https://orcid.org/0000-0003-0854-6975

References

[1] Wadley H and Mehrabian R 1984 Acoustic emission for materials processing: a review *Mater. Sci. Eng.* **65** 245–63

[2] Choudhury A and Tandon N 2000 Application of acoustic emission technique for the detection of defects in rolling element bearings *Tribol. Int.* **33** 39–45

[3] Liu Z, Yang B, Wang X and Zhang L 2021 Acoustic emission analysis for wind turbine blade bearing fault detection under time-varying low-speed and heavy blade load conditions *IEEE Trans. Ind. Appl.* **57** 2791–800

[4] He Y and Zhang X 2012 Approximate entropy analysis of the acoustic emission from defects in rolling element bearings *J. Vib. Acoust.* **134** 061012

[5] Hemmati F, Alqaradawi M and Gadala M S 2016 Optimized statistical parameters of acoustic emission signals for monitoring of rolling element bearings *Proc. Inst. Mech. Eng. J* **230** 897–906

[6] Yoshioka T and Fujiwara T 1982 A new acoustic emission source locating system for the study of rolling contact fatigue *Wear* **81** 183–6

[7] Al-Ghamd A M and Mba D 2006 A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size *Mech. Syst. Signal Process.* **20** 1537–71

[8] Caso E, Fernandez-del-rincon A, Garcia P, Iglesias M and Viadero F 2020 Monitoring of misalignment in low speed geared shafts with acoustic emission sensors *Appl. Acoust.* **159** 107092

[9] Prosvirin A, Kim J and Kim J-M 2018 Bearing fault diagnosis based on convolutional neural networks with kurtogram representation of acoustic emission signals *Advances in Computer Science and Ubiquitous Computing* (*Singapore*) vol 474 pp 21–6

[10] Hasan M J, Manjurul Islam M M and Kim J-M 2019 Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions *Measurement* **138** 620–31

[11] Shi H, Guo L, Tan S and Bai X 2019 Rolling bearing initial fault detection using long short-term memory recurrent network *IEEE Access* **7** 171559–69

[12] Li X, Li J, He D and Qu Y 2019 Gear pitting fault diagnosis using raw acoustic emission signal based on deep learning *Eksploat. Niezawodn.* **21** 403–10

[13] Zhiyi H, Haidong S, Lin J, Junsheng C and Yu Y 2020 Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder *Measurement* **152** 107393

[14] He Z, Shao H, Wang P, Lin J, Cheng J and Yang Y 2020 Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples *Knowl.-Based Syst.* **191** 105313

[15] Zhu H, Cheng J, Zhang C, Wu J and Shao X 2020 Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings *Appl. Soft Comput.* **88** 106060

[16] Lu W, Li Y, Cheng Y, Meng D, Liang B and Zhou P 2018 Early fault detection approach with deep architectures *IEEE Trans. Instrum. Meas.* **67** 1679–89

[17] Dwiputranto T H , Setiawan N A and Aji T B (eds) 2017 Machinery equipment early fault detection using artificial neural network based autoencoder *2017 3rd Int. Conf. on Science and Technology-Computer (ICST)* (IEEE) (https://doi.org/10.1109/icstc.2017.8011854)

[18] König F, Sous C, Chaib A O and Jacobs G 2021 Machine learning based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems *Tribol. Int.* **155** 106811

[19] Caesarendra W, Kosasih B, Tieu A K, Zhu H, Moodie C A and Zhu Q 2016 Acoustic emission-based condition monitoring methods: review and application for low speed slew bearing *Mech. Syst. Signal Process.* **72** 134–59

[20] Mano R, Yoshioka T, Miti A K and Yamamoto T 2000 Relationship between growth of rolling contact fatigue cracks and load distribution *Tribol. Trans.* **43** 367–76

[21] Morhain A and Mba D 2003 Bearing defect diagnosis and acoustic emission *Proc. Inst. Mech. Eng. J.* **217** 257–72

[22] Wang Z, Chegdani F, Yalamarti N, Takabi B, Tai B, El Mansori M and Bukkapatnam S 2020 Acoustic emission characterization of natural fiber reinforced plastic composite machining using a random forest machine learning model *J. Manuf. Sci. Eng.* **142** 031003

[23] Bianchi D, Mayrhofer E, Gröschl M, Betz G and Vernes A 2015 Wavelet packet transform for detection of single events in acoustic emission signals *Mech. Syst. Signal Process.* **64** 441–51

[24] Gao Z, Lin J, Wang X and Xu X 2017 Bearing fault detection based on empirical wavelet transform and correlated kurtosis by acoustic emission *Materials* **10** 571

[25] Liu D, Tao J, Luo A and Wang Q (eds) 2018 An optimized kurtogram method for early fault detection of rolling element bearings using acoustic emission *2018 Int. Conf. on Information Systems and Computer Aided Education (ICISCAE)* (IEEE) (https://doi.org/10.1109/iciscae.2018.8666929)

[26] Liu L, Chen L, Wang Z and Liu D 2020 Early fault detection of planetary gearbox based on acoustic emission and improved variational mode decomposition *IEEE Sens. J.* **21** 1735–45

[27] Elforjani M and Mba D 2010 Accelerated natural fault diagnosis in slow speed bearings with acoustic emission *Eng. Fract. Mech.* **77** 112–27

[28] Amiri M, Modarres M and Droguett E L (eds) 2015 AE entropy for detection of fatigue crack initiation and growth *Austin 2015 IEEE Conf. on Prognostics and Health Management (PHM)* (IEEE) (https://doi.org/10.1109/ICPHM.2015.7245038)

[29] Kahirdeh A, Sauerbrunn C and Modarres M 2016 Acoustic emission entropy as a measure of damage in materials *AIP Conf. Proc.* vol 1757 060007

[30] Chai M, Zhang Z and Duan Q 2018 A new qualitative acoustic emission parameter based on Shannon's entropy for damage monitoring *Mech. Syst. Signal Process.* **100** 617–29

[31] Karimian S F, Modarres M and Bruck H A 2020 A new method for detecting fatigue crack initiation in aluminum alloy using acoustic emission waveform information entropy *Eng. Fract. Mech.* **223** 106771

[32] Shannon C E 1948 A mathematical theory of communication *Bell Syst.Tech. J.* **27** 379–423

[33] Amin H U, Mumtaz W, Subhani A R, Saad M N M and Malik A S 2017 Classification of EEG signals based on pattern recognition approach *Front. Comput. Neurosci.* **11** 103

Paper B:
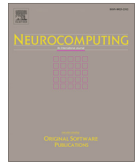
Simple is good: Investigation of history-state ensemble deep neural networks and their validation on rotating machinery fault diagnosis

Yu Wang *, Alexey Vinogradov

Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology – NTNU, Trondheim 7491, Norway

ARTICLE INFO

ABSTRACT

The present work is motivated by the desire to find an efficient approach that can improve the performance of deep neural networks in a general sense. To this end, an easy-to-implement ensemble approach is proposed in this paper leveraging the 'local sub-optima' of deep networks, which is referred as to history-state ensemble (HSE) method. We demonstrated that neural networks can naturally generate multiple 'local sub-optima' with diversity during training process, and their combination can effectively improve the accuracy and stability of the single network. The merits of HSE are twofold: (1) It does not require additional training cost in order to acquire multiple base models, which is one of the main drawbacks limiting the generalization of ensemble techniques in deep learning. (2) It can be easily applied to any types of deep networks without tuning of network architectures. We proposed the simplest way to perform HSE and investigated more than 20 ensemble strategies for HSE as comparison. Experiments are conducted on six datasets and eight popular network architectures for the case of rotating machinery fault diagnosis. It is demonstrated that the stability and accuracy of neural networks can be generally improved through the simplest ensemble strategy proposed in this paper.

AV	Average voting
BMA	Base model accuracy
CCALR	Cyclic cosine annealing learning rate
CLR	Constant learning rate
CNN	Convolutional Neural Network
CWRU	Case Western Reserve University
DBN	Deep Belief Network
DL	Deep Learning
FCNN	Fully connected neural network
GBDT	Gradient-Boosted Decision Trees
GRU	Gate Recurrent Unit Network
HSE	History-state ensemble
IFD	Intelligent fault diagnostics
KAt	Konstruktions-und Antriebstechnik
KNN	k-Nearest Neighbor
LR	Learning rate
LRD	Learning rate with decay
MBGD	Mini-batch gradient descent
ML	Machine Learning
PSO	Particle swarm optimization
SAE	Stacked Autoencoder
SVM	Support Vector Machine

* Corresponding author.
  E-mail address: yuwa@ntnu.no (Y. Wang).

1. Introduction

With the recent rapid advent of artificial intelligence, intelligent fault diagnostics (IFD) techniques are burgeoning in new dimensions. IFD generally refers to the application of Machine Learning (ML) algorithms in fault diagnostics to reduce human labour demand and cost [1]. Among all branches of ML, Deep Learning (DL) technology has attracted the most attention for its capacity to extract implicit features automatically from training data through multi-layered hidden neurons. Moreover, the procedures of feature extraction and fault recognition in DL are integrated, which makes it suitable to deal with the raw signal directly without any pre-processing. Since 2015, the area of DL applications has expanded rapidly; thus, the DL-based machine fault diagnostics framework has become the mainstream of IFD *de facto* [2]. Up to now, hundreds of deep networks have been designed and applied to IFD of bearing to take advantage of DL philosophy. Just to name some of them: back propagation neural network [3,4], convolutional neural network (CNN) [5–9], deep Boltzmann machine [10], deep belief network (DBN) [11], stacked autoencoder [12–15], long short-term memory network [16–18] and their modifications are among the most popular examples.

Deep network architectures mainly comprise four elements: (i) the number of layers, (ii) the number of neurons in each layer, (iii) the activation function of each neuron, and (iv) the training algorithms [19]. Due to the nonlinear nature of the network, a slight change in one element may lead to a significantly different result. Therefore, substantial research efforts have been devoted to the design of network architectures. Increasingly, people are looking to improve network performance by increasing network complexity or adopting more complicated approaches. Indeed, this has achieved remarkable success in many fields, including rotating machinery fault diagnostics. However, the performance of a carefully designed neural network may decline when applied to tasks different from what they had been originally intended for. For example, for the multi-scale cascade convolutional neural network proposed in [10], the classification accuracy varied between 99.7% and 96.9% when the authors tested the same network with different scales of the convolutional kernel. Besides, the neural network is sensitive to training parameters such as the learning rate (LR) and the training epochs.

To this end, we endeavor to develop an efficient and robust approach that can improve the performance of deep networks in a general sense. The ensemble strategy is a promising technique for improving the performance of a single model [20]. Generally, it comprises a great number of weak classifiers like Decision Tree and combines them to form a stronger model. Classical ensemble models are known as Random Forest [21], Adaboost [22], Xgboost [23] and so on. Thomas et al. demonstrated in their investigation that an ensemble of different types of classifiers leads to an increase in accuracy [24]. In recent years, ensemble strategy has gained increasing attention in DL. In [25], an ensemble strategy that combines a convolutional residual network, deep belief network and deep autoencoder was proved to be more effective than a single model. Cruz et al. proposed an evolutionary way to ensemble a fixed number of CNNs [5]. A multiobjective deep belief networks ensemble method was proposed in [26] for the remaining useful life estimation of engineering systems. Zhang et al. proposed an ensemble deep network architecture based on sparse deep autoencoder, denoising deep autoencoder and contractive deep autoencoder in [27] for the rotating machinery fault diagnostics. At the same time, Yang et al. [28] proposed another ensembled fault diagnostics scheme based on Sparse Autoencoder [29] and Denoising Autoencoder. The bootstrap sampling and plurality voting were employed for the ensemble in this paper. In ref. [30], a new ensemble deep network was developed to combine the result generated by fifteen different activation functions. To make use of the advantages offered by different neural networks, Ma et al. applied a multiobjective optimization algorithm to integrate CNN, DBN and deep autoencoder [25]. Zhang et al. proposed an ensemble learning model based on convolutional neural network [31]. Their method is implemented by adding multiple classification layers to generate a 'poll matrix' before the majority voting is used to generate the ensembled classification result. Li et al. proposed an optimal ensemble deep transfer network [32], and parameter transfer learning was used to initialize the start points of several base models with different kernels of maximum mean discrepancy. However, these ensemble strategies involve an intricate network architecture design, training approaches, and additional hyperparameters to be tuned.

In this paper, we do not concern with improving the structure or training of deep networks; instead, our goal is to exploit the potential of deep networks that have been overlooked. We can see that by virtue of adequately organized and together-tuned commonly accessible tools, the performance of deep networks can be greatly improved in a general sense. In this connection, a simple yet efficient ensemble strategy referred to as *history-state ensemble*

(HSE) method is proposed in the present paper. The 'history-state' in the method is represented by the network weights after the network update in each training cycle. We hypothesize that network performance can be further improved in terms of stability and accuracy by incorporating these history-states that should be covered after each model update. Details will be elaborated on in the following section. In comparison with the ensembled deep networks introduced above, the advantages of the proposed method are twofold: (1) it offers an improvement in the efficiency of the network by acquiring multiple base models without increasing training costs, and (2) the method is versatile and can be easily applied to different neural networks without a re-design or tuning of network architectures.

Some related works of HSE are to be mentioned. Back in 2013, Xie et al. proposed a series of ensemble methods for DL, including vertical voting, horizontal voting, and horizontal stacked ensemble, of which horizontal voting has a similar concept to the proposed HSE [33]. However, there is not enough experimental verification in this work, the approach has not received considerable attention. To encourage the deep networks to produce diverse base models, Huang et al. have proposed that cyclic cosine annealing learning rate (CCALR) helps the deep networks to attain multiple local minima, and a Snapshot ensemble method was introduced in their work [34]. There have been two applications of the Snapshot ensemble in the field of machinery fault diagnostics. Wen et al. [35] improved the original CCALR of the Snapshot ensemble and applied it to a CNN-based model for the fault diagnostics of bearings. Another Improve Snapshot Ensemble CNN was proposed in [36] using diversity regularization to encourage the diversity of the training history-states. Zhang et al. [37] proposed the Snapshot boosting to improve the snapshot ensemble.

These works are similar to our proposed method in that they all assume that the deep networks can generate multiple base models during the training process for ensemble learning; in this sense, they all refer to HSE methods. However, the difference lies in how the ensemble strategy is designed. There are several significant gaps in the current research landscape that need to be filled. (1) While the ensembled deep networks have been compared with a single deep network, the generalization of the ensemble method has yet to be done to adapt the concept to various network architectures. (2) Different ensemble strategies have to be systematically compared. (3) The HSE methods remain to be only scarcely studied, and their effectiveness needs to be examined and documented in a broader range of applications. This motivated us to address these specific issues on the examples of case studies relevant to practically significant problems of condition monitoring and fault diagnostics in rotating machinery. To this end, we investigated the efficiency and robustness of different HSE methods with various deep network architectures and proposed the ensemble strategy with the best generalization ability. Compared to previous works, our contributions are summarised as follows.

(1) We propose the most straightforward yet efficient way to perform ensemble learning on deep networks. It has been experimentally confirmed that deep networks can generate multiple local sub-optima during the training process, and the combination of them improves the network performance on average without increasing training costs.

(2) We conduct an extensive investigation of various ensemble strategies on deep networks in order to address the aforementioned existing issues. Through the comparative experiments, our proposed method demonstrates the best overall performance. The conclusions derived from the study provide practitioners with the guidelines for the rational selection of the ensemble strategy.

(3) We present a novel approach to investigate the tradeoff between the diversity and accuracy of base models, which provides new insights into the underlying mechanisms of ensemble learning.

The rest of the paper is organized as follows. The basic theory of HSE method is described in Section 2. The experimental setup and contrastive methods are introduced in Section 3. Experimental results of the probed methods are discussed in Section 4. Conclusions are drawn in Section 5.

2. Methodology

The training of neural network consists of two stages: (i) the forward propagation of input data from the first layer to the last layer of the network structure for feature extraction; (ii) the backward propagation of the prediction error from the opposite direction for network optimization. A closed loop is formed through forward and backward propagation, and the updated weights after each cycle are referred to as one history-state of the network in this paper. In general, a single network model is trained on fixed training cycles or iterations. However, it is hard to choose a 'magic' training budget to build a reliable model [34], as the acquired local optimum based on training dataset cannot represent the true local optimum of the whole dataset especially in practical applications. Local optimum denotes a solution that is optimal within a neighboring set of candidate solutions. Loss value is usually served as a measure of how well or poorly the network behaves after each optimization cycle. If the network is properly trained, the evolutionary trend in the loss value has the 'elbow-like' shape, as shown in Fig. 1: it drops sharply at the first iterations and then decays slowly before converging finally at very low values. At the same time, the network accuracy shows the opposite trend.

We assume that (1) a neural network can produce multiple 'local sub-optima' with diversity during training; (2) the combination of these local sub-optima improves the network performance in terms of stability and accuracy. We use the term' local sub-optima' to indicate that the solution is merely approximate to the true 'local optima'. In this paper, they refer to all the history-states of neural networks when the training process enters a stable phase, i.e., the loss value converges to a small value. Therefore, to obtain base models for ensemble learning using a neural network, one only needs to preserve these history states or local sub-optima of the network that should be covered after each training cycle, and the time cost of this is negligible. Since there is no requirement for the network structure, the method can be applied to all neural networks.
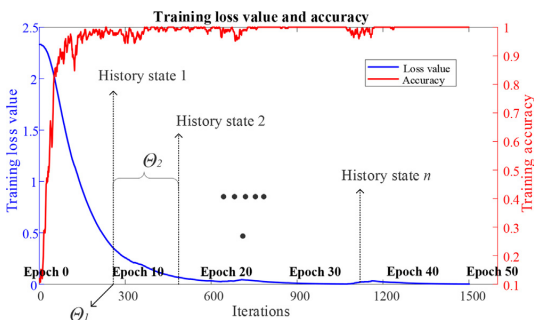
2.1. Feasibility analysis

The feasibility of the proposed ensemble method is analyzed below.

2.1.1. Improve network stability
The network stability is evaluated by the variability of performance under randomly initialized weights. The variance or standard deviation is utilized as the metric. With the previous hypothesis on local sub-optima, we assume the acquired base models are (i) unbiased, (ii) with the same variance, and (iii) uncorrelated with each other. The accuracy of the i-th base models is modeled as: $a_i = A + \omega$, where A stands for the accuracy of the true local optimum to be estimated, ω is random noise caused by the diversity of base models. Based on the minimum variance unbiased estimator, a reasonable estimator of a can be expressed as $\tilde{a} = \sum_{i=1}^{n} a_i/n = E(A)$, assuming the ω is Gaussian noise, and n represents the base model number. Then, the variance of the estimator $\tilde{a}$ is expressed as:

$$var\left(\tilde{a}\right) = \frac{\sum_{i=1}^{n} var(a_i)}{n^2} = \frac{var(a)}{n} \qquad (1)$$

where $var(a)$ denotes the variance of a single base model. Since $var(a)/n < var(a)$, i.e., the variance of the ensemble network is smaller than the single network. Hence, it can be inferred that the ensemble method reduces the performance variability.

2.1.2. Improve network accuracy
For better illustration, we enumerate possible scenarios when comparing the single network with the ensemble network by introducing a simple example of binary classification, as displayed in Fig. 2. In case 1–1 and 1–2, most base models produce the true labels, thus underlying the efficacy of the ensemble. However, the single network might fail. And cases 2–1 and 2–2 are the opposite. The ensemble method will fail if most base models produce false labels, whereas the single network might succeed. If the two scenarios have the same frequency, then the ensemble and single network should have the same performance on average. Since deep networks are generally considered to be a strong classifier, we hypothesize that the first scenario is more frequent than the other one, and the ensemble method has a higher chance to output a true label, which remains to be proven in the following experiments.

2.2. The proposed ensemble strategy

Similarly to conventional ensemble techniques, the implementation of the HSE method faces two challenges: (1) a training strategy that encourages the generation of accurate base models with diversity [5,24]; and (2) a learning strategy which combines ac-

Fig. 1. Illustration showing the typical behaviour of the training loss value and the accuracy as a function of the number of iterations; the definition of parameters θ_1 and θ_2 is illustrated (see the text for details)

Fig. 2. Graphical illustration of possible scenarios when comparing the single network with the ensemble network. Assuming there are 10 base models, and the prediction of the ensemble is given by majority voting of all base models, whereas the prediction of a single network is given by the last base model.

quired base models to form a stronger classifier [38]. In this paper, we are not concerned with improving the training of neural networks; instead, our goal is to show the potential of current neural networks with the most common tools we already have.

2.2.1. Training strategy

To improve the ensemble performance, it is pivotal to balance the diversity and accuracy of the base models. We propose that the simple combination of the widely used MBGD and CLR in the network training process is sufficient to produce the required base models.

MBGD is a variant of the gradient descent algorithm whereby the whole training dataset is divided into multiple small batches, and only one batch is used to calculate the gradient at each iteration. The iteration refers to the number of batches or steps through partitioned packets of the training data. One epoch is counted after all batches of training data are fed into the neural network. For example, if the dataset is divided into m batches with k training epochs, the total number of iterations i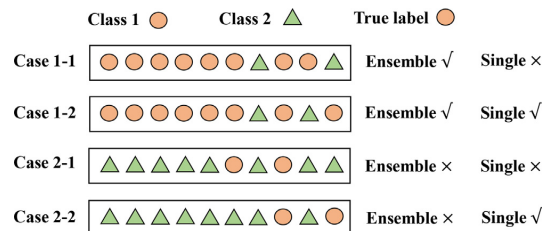s $m \times k$. The application of MBGD increases the model update frequency, thus giving rise to a larger number of base models generated for the ensemble classification.

CLR is a *de facto* tool for training modern neural networks, which is advocated in HSE for the following reasons: (1) it approaches local minima in a noisy manner, which helps to encourage the diversity of base models; (2) as a rule of thumb, it will not deviate far from the local minimum, which guarantees the accuracy of the acquired base models to a certain extent. Nevertheless, the hypothesis is based on experience, which needs to be proven experimentally. To demonstrate this, different LR schedulers are investigated in the following experiments, and analysis of the diversity-accuracy tradeoff has shown that CLR exhibits better balance in the acquired base models.

2.2.2. Learning strategy

A SoftMax layer is generally attached after the last hidden layer of a neural network to normalize the output of a network to a probability distribution over the predicted output classes. The AV method takes the average score of all the recorded base models, which is mathematically described as follows:

$$E_j = \frac{1}{N} \times \sum_{i=1}^{N} \frac{exp(\boldsymbol{b}^i + \boldsymbol{W}_j^i * \boldsymbol{H})}{\sum_{j=1}^{c} exp(\boldsymbol{b}^i + \boldsymbol{W}_j^i * \boldsymbol{H})} \tag{2}$$

where N denotes the number of base models, and c represents the number of categories. $\boldsymbol{H}$ stands for the input of the Softmax layer, $\boldsymbol{W}$ and $\boldsymbol{b}$ are weights and bias, respectively, which connect the Softmax layer and the last hidden layer of the neural network. E_j is the predicted probability that the input data is classified as the category j, and the predicted class of HSE is given by $max(E_j)$.

2.2.3. Model selection

We do not propose strict selection criteria for base models in HSE for the following reasons: (1) to avoid overfitting; (2) to be consistent with the previous assumptions of base models, namely, that they are unbiased and have the same variance. Nevertheless, the ensemble model performance is affected by three factors: the number, the accuracy, and the diversity of base models, which are closely correlated with the following parameters: (i) the S represents the applied training cycles; (ii) the θ_1 represents the required training cycles for model warm-up; and (iii) the θ_2 denotes the update frequency between two adjacent base models. Generally, the larger θ_2 can reduce redundancy and increase the diversity of base models. The total number of the acquired base models is calculated as:

$$N = int(\frac{S - \theta_1}{\theta_2}) \tag{3}$$

where $int(\cdot)$ returns an integer number. The key to improving ensemble performance is the balance between the diversity and accuracy of base models, which needs to be investigated experimentally.

3. Experimental setup

3.1. Description of datasets

The utilized datasets are acquired from two bearing data centers: Case Western Reserve University (CWRU) Bearing Data Center and Konstruktions-und Antriebstechnik (KAt) – Bearing Data Center [39]. Details of the used data are described in Table 2. The CWRU data centre provides vibration signals of bearing with various artificial damages. It has been widely used as benchmark data to evaluate the effectiveness of the proposed models. Table 1 summarises the use of CWRU data in recent publications. The vibration signals database comprises samples collected at both 12 k and 48 k samples/second rates. The highest accuracy for the signals from the 12 kHz dataset was observed to be 99.94%, while for signals from the 48 kHz dataset, it was 98.95%. Ten categories of bearing under different fault levels are used in this study, as presented in Table 2 (Index 1–10). Each category of vibration data is collected from 3 different loads (1, 2, 3 hp) at a sampling frequency of 48 kHz. The KAt data centre provides real bearing damage signals caused by the accelerated lifetime test. The test rig and experimental details can be found in [39]. Ten categories of faulty bearings with different fault levels in different locations are gathered in Table 2 (Index 11–20). Each bearing was tested under four different operating conditions, which are denoted as N1, N2, N3, and N4 in Table 3.

In this study, six datasets are constructed from the above-gathered bearing vibration data to validate the proposed method, as described in Table 3. Among them, dataset A and dataset E are collected from CWRU and KAt datasets, respectively, with the same operating condition on training and test data. It can be concluded from the previous study that it is more difficult to classify the datasets with different operating conditions on training and test data. Therefore, datasets B-D and dataset E are constructed to evaluate the performance of the proposed method under various operating conditions.

3.2. Evaluation method

Each model runs ten times with random initial conditions, and the average classification accuracy and standard deviation are used to evaluate the probed models, which are defined as follows:

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}$$

$$Std = \frac{1}{n}(Acc_i - \mu)^2 \tag{5}$$

where TP, FN, FP and TN abbreviate the True Positive, False Negative, False Positive and True Negative, respectively, n denotes the number of trials, Acc_i is the accuracy of the i-th trial, μ is the mean: $\mu = \frac{1}{n}\sum_{i=1}^{n} Acc_i$.

3.3. Implementation details and comparison methods

3.3.1. Comparison with single deep networks

Various deep neural network architectures with increasing depths are proposed in this work for comparison, which are detailed as follows:

Table 1
Summary of the use of CWRU dataset in recent publications.

Refs	Load (HP)	Frequency (kHz)	Class	Training/test samples	Accuracy (%)
[4]	*	12	12	4200/600	98.47
[8]	3	48	10	1680/720	98.46
[9]	$0 \rightarrow 0/0 \rightarrow 1/0 \rightarrow 2$	12	12	2400/1200	98.5/97.1667/95.8333
[11]	0/1/2/3	12	10	1000/1000	99.57/99.32/99.54/99.43
[14]	0–3	12	8	*	99.94
[15]	0–3	*	12	1800/900	96.44
[18]	0–3	48	10	*	98.95

Table 2
Description of the gathered vibration signals from CWRU and KAt data centers.

Data base	Index	Fault location	Fault level	Data base	Index	Fault location	Bearing code	Fault type	Fault level
CWRU	1	–	0	KAt	11	–	K001	–	0
	2	IR	0.007		12	IR	KI16	Fatigue pitting	3
	3	IR	0.014		13	IR	KI17	Fatigue pitting	1
	4	IR	0.021		14	IR	KI18	Fatigue pitting	2
	5	Ball	0.007		15	OR	KA16	Fatigue pitting	2
	6	Ball	0.014		16	OR	KA22	Fatigue pitting	1
	7	Ball	0.021		17	OR	KA15	Indentations	1
	8	OR	0.007		18	OR + IR	KB23	Fatigue pitting	2
	9	OR	0.014		19	OR + IR	KB24	Fatigue pitting	3
	10	OR	0.021		20	OR + IR	KB27	Indentations	2

Table 3
Description of the extracted six datasets under various operating conditions.

Datasets	Class number	indices	Training set		Test set	
			Operating condition	Sample number	Operating condition	Sample number
A	10	1–10	L1, L2, L3	4500	L1, L2, L3	4500
B			L1	3000	L2, L3	6000
C			L2	3000	L1, L3	6000
D			L3	3000	L1, L2	6000
E		11–20	N1, N2, N3, N4	8000	N1, N2, N3, N4	8000
F			N1, N2	8000	N3, N4	8000

(1) Fully connected neural network (FCNN).
(2) Stacked Autoencoder with supervised fine-tuning (SAE).
(3) Gaussian-Bernoulli Deep Belief Network with supervised fine-tuning (DBN).
(4) Gate Recurrent Unit Network for fault diagnostics (GRU) [40].
(5) One-dimensional LeNet5 (1D-LeNet5).
(6) One-dimensional AlexNet (1D-AlexNet).
(7) One-dimensional Deep Residual Network (1D-ResNet).
(8) One-dimensional Deep Densely Connected Network (1D-DenseNet).

Among the above eight architectures, (1)-(3) are fully connected networks with the shared architecture configuration of $[1024, 512, 256, 128]$. SAE and DBN are employed to learn unsupervised features from data which are followed by a supervised fine-tuning process for fault diagnostics. The unsupervised learning is trained for 30 epochs. (4) is a fault diagnostics-based recurrent neural network architecture proposed in [40], which consists of a linear layer, a GRU layer and a classification module with a multilayer perceptron. To maintain the consistence of the input of all probed networks, the data length is set as 1024, which is converted to a $[16 \times 64]$ image, i.e., the sequential length of GRU cell is 16. The linear layer maps the dimension of the raw image to $[16 \times 1024]$, and the output is fed into GRU cell with a hidden size of 1024. The classification module consists of two hidden layers with 1024 neurons and an output layer. (5)-(8) are convolutional neural networks with increasing depth of hidden layers, the configuration of their network architectures is detailed in Tables 4 and 5.

To probe their performance under a relative fair circumstance, all networks are trained with same hyperparameters. The Adam stochastic optimization algorithm is applied to update the network weights [55]. Epoch number and batch size are set as 50. The initial LR is set as 0.001 and decreases with a decay rate of 0.001 for each iteration after 20 training epochs. The LR with decay is denoted as LRD in this work. The deep network without HSE is denoted as a single network. Single networks are served as references to the ensembled neural networks.

3.3.2. Comparison with different ensemble strategies

Several ensemble approaches designed for neural networks are investigated. They are roughly divided into two categories in this study: training strategy and learning strategy. The training strategy aims at encouraging the diversity of the acquired base models, while the learning strategy provides optimal solutions to combine these base models. The implementation details are presented below:

3.3.2.1. Training strategies:.
(1) MBGD + CLR (the proposed).
(2) CCALR in Snapshot ensemble [34].
(3) Boosted framework.
(4) Snapshot boosting ensemble [37].

In (2), the models are warmed up for 20 epochs with initial LR of 0.001, and then the LR is scheduled with CCALR, and a snapshot or history-state of the model is taken when the LR reaches its minimum at each cycle. The Snapshot number of 5 and 30 are investi-

Table 4
The proposed 1D-CNN architectures with increasing network depths.

1D-LeNet-5		1D-AlexNet		1D-ResNet		1D-DenseNet	
Layers	Output size	Layers	Output size	Layers	Output size	Layers	Output size
Input	1024@1	*Input*	1024@1	*Input*	1024@1	*Input*	1024@1
$\begin{bmatrix} C125,S1,P0 \\ A12,S2,P0 \end{bmatrix}$	445@16	$\begin{bmatrix} C125,S1,P0 \\ A12,S2,P0 \end{bmatrix}$	445@16	$\begin{bmatrix} C125,S1,P0 \\ A12,S2,P0 \end{bmatrix}$	445@16	$\begin{bmatrix} C125,S1,P0 \\ A12,S2,P0 \end{bmatrix}$	445@16
$[C7,S2,P0]$	220@16	$\begin{bmatrix} C7,S1,P3 \\ A7,S2,P0 \end{bmatrix}$	220@16	$RL \times 2$	445@16	$DL \times 2$	445@16
		$[C5,S1,P0] \times 3$	220@32	$[C5,S2,P0]$	221@16	TL	221@16
				$RL \times 2$	221@32	$DL \times 2$	221@32
				$[C5,S2,P0]$	109@32	TL	109@32
				$RL \times 2$	109@64	$DL \times 2$	109@64
				$[C5,S2,P0]$	53@64	TL	53@64
				$RL \times 2$	53@64	$DL \times 2$	53@64
$[A5,S5,P0]$	44@32	$[A5,S5,P0]$	44@32	$\begin{bmatrix} C5,S2,P0 \\ A4,S1,P0 \end{bmatrix}$	22@64	$\begin{bmatrix} C5,S2,P0 \\ A4,S1,P0 \end{bmatrix}$	22@64
Flatten	1408@1	*Flatten*	1408@1	*Flatten*	1408@1	*Flatten*	1408@1

Note: *C* denotes the convolutional kernel, *A* denotes the average pooling kernel, *S* and *P* are the stride and padding number of each kernel, respectively. *Flatten* stands for the concatenated layer. The output size is denoted by the notation '*a*@*b*', where *a* represents the length of the output vector, and *b* is the number of output channels.

Table 5
Description of *RL*, *DL*, and *TL*.

Layers	Description	Parameters
RL	Residual block	$\begin{bmatrix} C15,S1,P7 \\ C15,S1,P7 \end{bmatrix}$
DL	Densely connected block	$\begin{bmatrix} C1,S1,P0 \\ C15,S1,P7 \end{bmatrix}$
TL	Transition layer	$\begin{bmatrix} C1,S1,P0 \\ A5,S2,P0 \end{bmatrix}$

gated in this work, i.e., 5 and 30 base models are acquired, which are denoted as 'Snap5' and 'Snap30', respectively. (3) The models are warmed up for 20 epochs and then trained with the boosted framework as described in subsection 3.2.2. The sample weights are updated every epoch after the first 20 epochs, and a total of 30 history states are recorded. (4) The Snapshot boosting ensemble is introduced in [37]; the original method is based on an image processing task. The models are warmed up for 20 epochs, and then both boosted framework and CCALR are applied to train the models. The Snapshot number is 30.

3.3.2.2. Learning strategies:.
(5) AV.
(6) Ranking voting.
(7) Weighted voting + PSO.
(8) Selective voting + PSO.

In (6), a validation set is used to evaluate the performance of the recorded base models, and only the base models with the top *n* classification accuracy on the validation set participate in the voting. In this work, 10% of training samples are randomly extracted out as a validation set, and *n* is set as 3. The PSO algorithms are served as *meta*-learner to implement (7) and (8). The swarm size

and the number of iterations are set as 100. Acceleration coefficients, c_1 and c_2, are set as 0.5. The inertia weight is 0.9. In (7), particle swarm optimization (PSO) is utilized to adaptively learn the weights of all base models based on their performance on the validation set, and the prediction is given by the weighted base models. While in (8), only the base models with positive weights participate in the voting, and the prediction is given by the averaged output of the selected base models.

3.3.3. Comparison with shallow learning algorithms
The performance of several types of classical shallow learning and ensemble algorithms are probed as comparison methods, as listed below.

(1) *k*-nearest neighbors (KNN).
(2) Support Vector Machine (SVM).
(3) Random forest.
(4) AdaBoost.
(5) Gradient-Boosted Decision Trees (GBDT).
(6) XGBoost.

Common time- and frequency-dependent features are extracted from the raw signal as the input of each model as Root Mean Square, Skewness, Kurtosis, Shape factor, Crest factor, Impulse factor, Margin factor, Power, Mean frequency, RMSF, RVF. In model (1), the number of neighbors is set as 1. (2) The Gaussian kernel is employed, and the kernel coefficient is defined by $1/(L \times var(\boldsymbol{X}))$, where *L* denotes the number of classes, $var(\boldsymbol{X})$ represents the variance of input $\boldsymbol{X}$. The regularization parameter is 1. DT is served as the base model for (3–6), and the number of base models is set as 50. In (4), the LR value is 0.5, and the maximum depth of the individual model is 3. (5) LR is 0.5, and the maximum depth of each individual tree is 5. (6) The LR value is 0.9, L1 regu-

Table 6
Fault diagnostics accuracy of shallow learning algorithms based on time and frequency features.

Models	Datasets					
	A	B	C	D	E	F
KNN	0.8658	0.8085	0.8085	0.7395	0.7059	0.6609
SVM	0.9106	0.8362	0.8642	0.8053	0.7592	0.7035
RF	0.9251	0.8303	0.8360	0.7891	0.8084	0.7047
Adaboost	0.8802	0.7945	0.8080	0.7332	0.7374	0.6500
GBDT	0.9195	0.8271	0.8336	0.7716	0.7916	0.6874
XGboost	0.8716	0.7728	0.7901	0.7492	0.6910	0.6193

larization term is 0.1, and the maximum depth of each individual tree is 5.

4. Results and discussion

We first look at the influence of different LR schedulers on the loss value of deep networks, as shown in Fig. 3. It can be observed from Fig. 3a that the loss value converges with the decay of LR, which indicates that the model gradually reaches a local minimum. Fig. 3c shows the loss value of the model with the CCALR scheduler. There are obvious ups and downs of loss value with the cyclic evaluation of LR, which manifests the CCALR encourages the model to attain multiple local minimums. Fig. 3b and d show the loss value of the same network architecture, which has been trained by CLR. Compared with Fig. 3a, their loss value converges at a higher level and exhibits more significant fluctuations. The observation of LR and its corresponding loss value indicates that the models have been adequately trained.

Then, each probed method runs ten times on every network with the randomly generated training dataset and weights, and the average classification accuracy is measured. Detailed experimental results on datasets A-F of each method are shown in Tables 8–13 in the appendix. When comparing with recent publications cited in Table 1, which employed the same datasets sampled at 48 kHz (data A), it was observed that 1D-DenseNet using the proposed ensemble method (**MBGD + CLR + AV**) achieved the highest average accuracy of 99.19% on dataset A. This result is superior to what has been reported earlier, c.f., Table 1.

To facilitate the observation of the results, the average accuracy of 1D-LeNet5, 1D-AlexNet, 1D-ResNet and 1D-DenseNet are plotted in Fig. 4. To measure the effectiveness and robustness of DHSE on different network architectures and datasets, the accuracy and standard deviation of the probed eight deep networks on datasets A-F are averaged and shown in Fig. 5.

4.1. Display of comparison experiments

4.1.1. Ensemble networks compared to single networks

LR scheduler is crucial for network training. We probed the single networks in two cases: trained by CLR and LRD schedulers, respectively. LRD is a widely used tool, which encourages the network to converge to a local minimum. It can be seen in Fig. 5 that the average accuracy of single networks with CLR is 0.7584, which is improved to 0.7861 through the utilization of LRD. In addition, the average standard deviation of single networks with LRD is smaller than CLR. The result reveals that for single networks, the use of LRD helps to improve the performance and stability of the model. For comparison, the ensemble method was applied to single networks, which are denoted as 'MBGD + CLR + AV' and 'MBGD + LRD + AV', respectively. Among them, the 'MBGD + CLR + AV' is the strategy advocated in this work. Through HSE, the performance of 'single + LRD' is further improved to 0.7883 in Fig. 5. More notably, the performance of 'single + CLR' is largely improved up to 0.7994 average accuracy by 'MBGD + CLR + AV'. A conclusion can be drawn that for HSE the CLR scheduler works better than LRD, which is the opposite of the case of single networks. And this is not just a coincidence; the same behaviour can be observed in all probed network architectures on each dataset, as shown in Fig. 4 (more details can be found in Tables 8–13 in the appendix). The reason is presumed to be the balance between the diversity and accuracy of the base models, which will be discussed later.

Besides, it is worth noting from Fig. 4 that although 1D-DenseNet exhibits the best performance on datasets A, E and F, its performance decreases on datasets B, C and D compared with other networks. The result confirms the above statement that the

performance of a carefully designed network architecture might decrease on other tasks. However, the model performance can be generally improved by HSE regardless of the network architectures and datasets, demonstrating the general capability of the proposed method to be self-adapted to different network structures.

4.1.2. Comparison of different ensemble strategies

Through different combinations of training and learning strategies, a total of 21 ensemble strategies for HSE are probed in this work. Taking the performance of 'single + CLR' and 'single + LRD' as baselines, Figs. 4 and 5 show that most of the ensemble strategies outperform the 'single + CLR' model. Besides, this model also exhibits the highest average standard deviation. However, only seven probed ensemble strategies work better than the 'single + LRD' model. Among them, the proposed 'MBGD + CLR + AV' strategy exhibits the highest average accuracy and lowest standard deviation, thus demonstrating the relative effectiveness and robustness of the method.

4.1.3. Comparison with shallow learning methods

The performance of shallowing learning methods is presented in Table 6. It can be concluded that deep networks like 1D-LeNet5, 1D-AlexNet, 1D-ResNet and 1D-DenseNet are much superior to shallow learning methods. Thus, we demonstrated that ensemble learning combined with strong classifiers could make full use of the advantages of both methods.

4.2. Discussion on the comparison results

The ensemble performance is directly affected by the diversity and accuracy of the acquired base models. To uncover their inner working mechanism, the diversity and base model accuracy (BMA) acquired by the probed ensemble strategies are quantified as:

$$Diversity = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\sum_{k=1}^{m} 1\left(f_i^k \neq f_j^k\right)}{m} \right)/N^2 \qquad (6)$$

$$BMA = \left(\sum_{j=1}^{N} \frac{\sum_{k=1}^{m} 1\left(f_i^m = y^k\right)}{m} \right)/N \qquad (7)$$

where $1(\cdot)$ is an indicator function, which output is 1 when the result of the Boolean operation is true, otherwise it is 0. N and m represent the number of base models and test samples. f_i^m stands for the predicted label of the i-th base model and the k-th test sample. y^m represents the true label of the k-th test sample. Therefore, $Diversity$ measures the ratio of the two base models generating different labels for the same sample, and BMA measures the mean accuracy of all base models. The average $Diversity$ and BMA of all probed network architectures and datasets are plotted in Fig. 6. One can observe that the diversity of base models is roughly inversely proportional to the accuracy of base models. The training strategies and learning strategies are discussed separately. More details will be u below.

4.2.1. Discussion on the training strategies

Then, we discuss the experimental results with diversity-accuracy tradeoff.

4.2.1.1. Discussion of LRD and CCALR.
As concluded above, although LRD effectively increases the accuracy of single networks, its performance is inferior to CLR in HSE. This is because LRD encourages the network to converge to one local minimum; thus, the acquired base models have low diversity. The same reason ap-
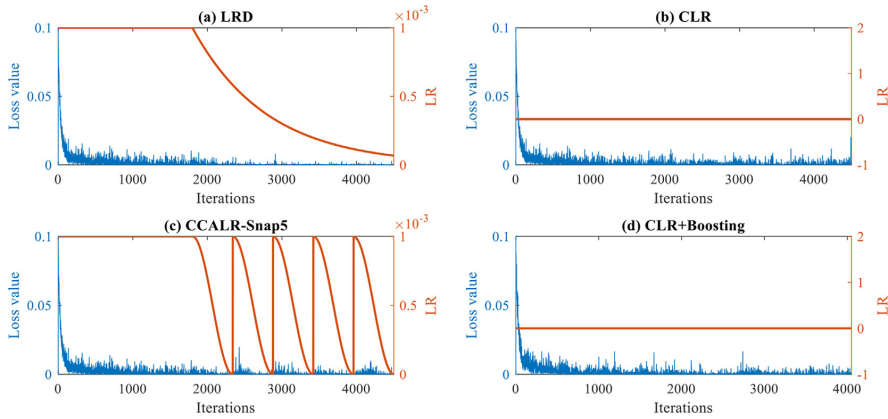
Fig. 3. Loss value of different training methods - exemplified from 1D-LeNet5 of trial 1 on dataset A. (a) The LRD is initially set as 0.001 and then starts to decrease with a decay rate of 0.001 for each iteration after 20 epochs. (b) CLR+MBGD with a constant value of 0.001 (the proposed method). (c) CCALR with Snapshot number of 5. (d) Boosting +CLR with a constant value of 0.001, 10% of training samples are randomly extracted out as validation set.

Table 7
Training and test cost time of 1D-LeNet5 trained by 'MBDG + CLR' methods on dataset A with different learning strategies.

Learning strategy	Number of base models	Training time (s)/ Sample number	Test time(s)/ Sample number
Single	1	25.56/4500	0.1/4500
AV	30	25.75/4500	2.96/4500
Ranking voting	3	24.26/4500	0.38/4500
Weighted voting + PSO	30	29.29/4500	3.17/4500
Selective voting + PSO	14	29.37/4500	1.43/4500

plies to the CCALR scheduler in Snapshot ensemble. However, CLR usually approaches the local minima in a noisy manner which encourages the diversity of base models. It can be seen that the 'Snap5′, 'MBGD + LRD' and 'Snap30′ occupy the top three positions of *BMA* across all methods. But their diversity

is low. Hence, the performance increment of these methods to single networks is limited. Although, theoretically, the Snapshot ensemble provides the likelihood to jump out of the current local minimum in some cases, it is practically difficult to guarantee a successful result. Instead, the 'MBGD + CLR' is easier to implement and more robust.

4.2.1.2. Discussion on the boosted framework. Contrary to LRD and CCALR, both 'Boosting' and 'MBGD + CLR' training strategies generate high diversity in base models; however, the *BMA* of the 'Boosting' method is much lower than 'MBGD + CLR'. The reason presumably is that the 'Boosting' method forces each base model to focus on these currently difficult-to-classify samples; consequently, the total training sample of each base model is reduced and biased. Although the diversity of base models has increased, the accuracy has also decreased. The combination of 'Boosting' and CCALR in Snapshot ensemble still cannot overcome the above-mentioned problems.
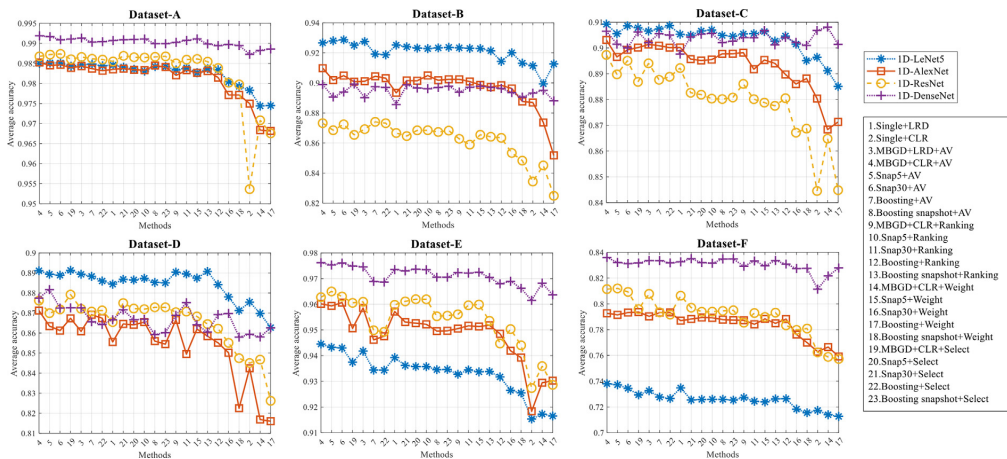
Fig. 4. The average accuracy of the selected four network architectures on datasets A-F. Plots are sorted by the descending order of the average accuracy of the four models on each dataset.

Fig. 5. The averaged classification accuracy and standard deviation of DHSEs with various ensemble strategies and single model on datasets A-E and different network architectures. (a) Average classification accuracy in descending order. (b) Average standard deviation in descending order.

Fig. 6. Measure of diversity and mean accuracy of base models. The picture is sorted by the descending order of the average diversity of the probed network architectures and datasets.
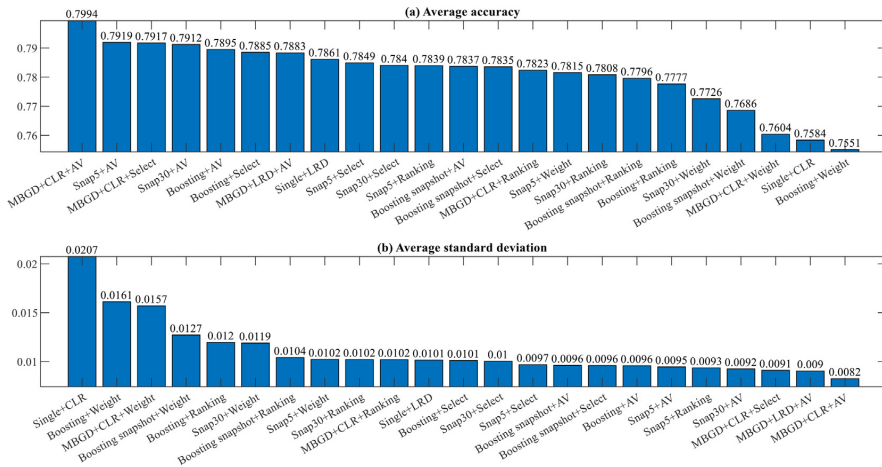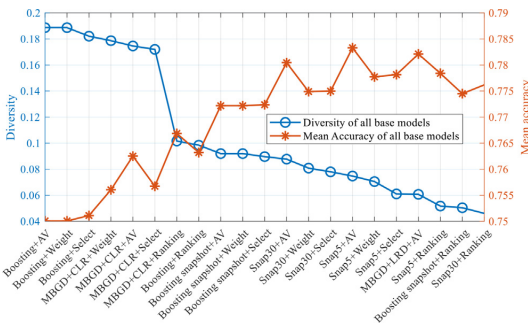

Fig. 7. Average accuracy of MBGD+CLR+AV with different selection of parameters θ_1 and θ_2.

4.2.2. Discussion on the learning strategies

Three learning strategies are introduced as a comparison, including 'Ranking voting', 'Weighted voting' and 'Selective voting'. The probed methods apply different selection criteria to the ensemble method, endowing different degrees of freedom to the obtained base model, among which the proposed 'AV' gives the highest degree of freedom to the base models, followed by the 'selective voting', 'Ranking Voting', and 'Weighted Voting'. As can be observed from Fig. 5, the ensemble performance is highly correlated with the selection criteria, and the methods with a higher degree of freedom tend to perform better. The reason presumably is that the strict selection criteria formulated on training or validation datasets are likely to cause overfitting.

4.3. Parameter analysis

4.3.1. Ensemble parameters

The selection of the parameters θ_1 and θ_2 directly affects the number of base models. Fig. 7 shows the average accuracy of 'MBGD + CLR + AV' with different selection of parameters θ_1 and θ_2 as well as corresponding base model number. Take the performance of 'Snap5' and 'Single + LRD' as two base lines, it can be concluded from the experimental results that: (1) the ensemble
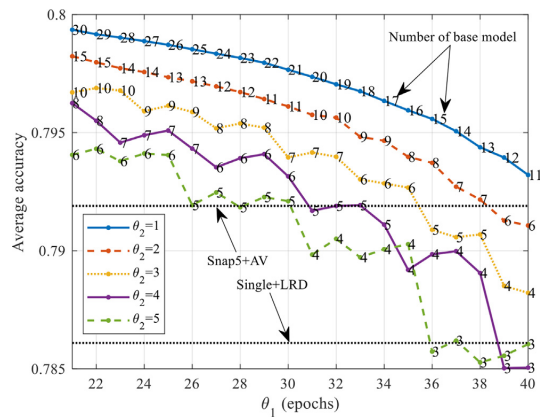
performance is highly related with base model number; (2) with the same number of base models, larger θ_2 tends to have higher accuracy, this is presumably because each base model is updated more often with larger θ_2 which encourages the diversity of base models. The two rules can be regarded as the selection criteria of the proposed method. However, the specific setting of θ_1 and θ_2 depends on the specific task. Nevertheless, the performance of Snapshot ensemble does not show a direct relationship with the base model number through the comparison between 'Snap5' and 'Snap30'.

4.3.2. Training epochs

We investigated the influence of training budget to the model performance. To facilitate analysis, the ensemble parameter θ_2 is fixed as 1, and the number of base models is fixed as 10; the warm-up parameter θ_1 is defined as $N - 10$, where N denotes the number of training epochs. The other hyper-parameters remain unchanged. The performance of the proposed HSE method is compared with the corresponding single model under different training
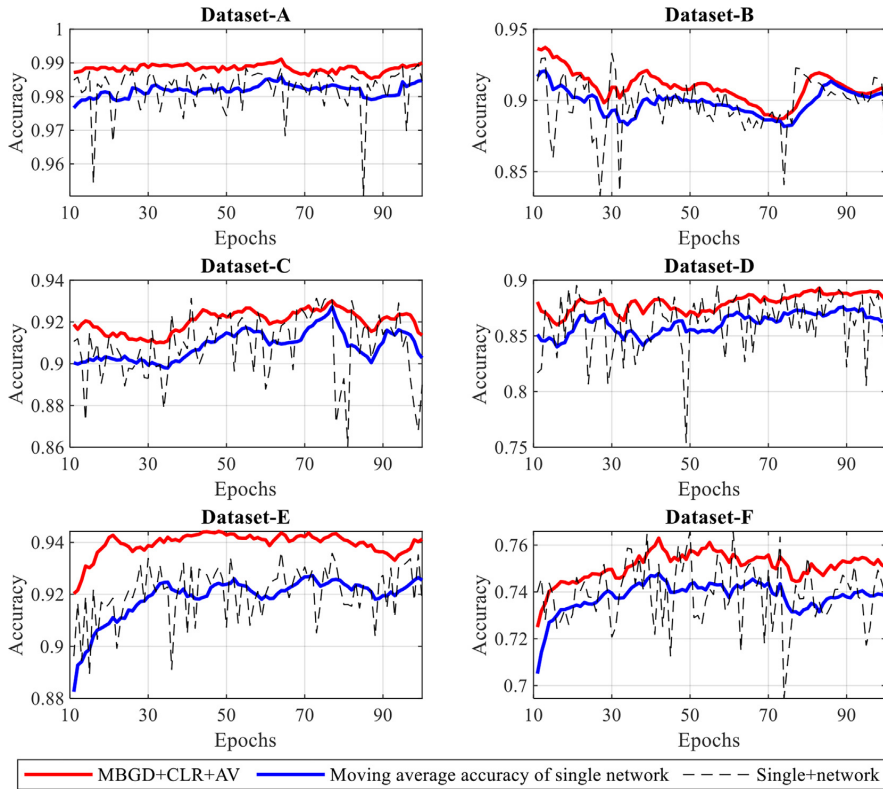
Fig. 8. Accuracy of the probed methods with different training epochs.

epochs ranging from 10 to 100, as displayed in Fig. 8. The example is taken from 1D-LeNet5. The red line represents the accuracy of the proposed method under different training epochs. The black line represents the accuracy of a single network corresponding to training epochs, while the blue line is the moving average accuracy of the single network with a window length of 10. One can observe that the ensemble network shows less fluctuation in accuracy with different training epochs. The blue line and red line show similar performance variability; however, the accuracy of the proposed method is maintained on the upper envelope of the single network.

4.4. Computational cost

As has been noted, the DHSE does not increase the training time of the model. However, the test time is inevitably increased to $t \times N$, where t denotes the test time of a single network, and N represents the number of base models. Table 7 presents the training and test cost time of 1D-LeNet5 using the 'MBDG + CLR' training method on dataset A. The training cost of 'Weight Voting' and 'Selective Voting' is increased because of the use of PSO algorithms. Although the test time is increased, it is small for every single sample.

5. Conclusion

In this paper, the ensemble techniques combining the 'history-states' generated during network training are denoted as HSE methods. The proposed methodology aims at enhancing the relia-

bility of bearing fault diagnostics by introducing an effective and user-friendly HSE strategy, which is evaluated across various network architectures. The experimental results reveal that deep networks can produce multiple base models for ensemble learning using a combination of MBGD, CLR, and AV methods once the training process reaches a stable phase. Compared with peer methods, the proposed ensemble strategy benefits from simplicity, which is evident from the following aspects:

(1) Ensemble strategy: this work integrates existing concepts in an enhanced accuracy-improving workflow. The use of *de facto* accepted tools enhances the method's accessibility and applicability, making it an easy-to-implement option for practitioners seeking to improve the model performance.
(2) Selection of base models: the proposed method differs from many peer methods like Snapshot ensemble struggling to encourage neural networks to reach different local optima. Instead, the proposed method focuses on utilizing 'local sub-optima' to achieve its goals. By doing so, the proposed method avoids the intricate task of defining qualified base models, resulting in reduced complexity.
(3) Hyper-parameters: the proposed method requires fewer hyperparameters to be tuned and exhibits robustness to the selection of hyperparameters through parameter analysis.

Comparing the experimental findings with peer methods demonstrates that the proposed 'MBGD + CLR + AV' exhibits a better classification accuracy while having a lower standard deviation,

thus showing the relative effectiveness and robustness of the proposed methodology.

Nevertheless, the efficiency of HSE methods still needs to be investigated in a broader range of applications, which is in the focus of our further study. Up to now, we have successfully tested the proposed methodology against several other applications, including unsupervised early fault detection and the problem of streaming data with emerging new classes, the readers can find more details in [41,42].

CRediT authorship contribution statement

Yu Wang: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Alexey Vinogradov:** Writing – review & editing, Supervision, Funding acquisition.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix

1. Cyclic cosine annealing learning rate

Snapshot ensemble adopted a cyclic cosine annealing learning rate (CCALR) schedule to encourage the model to reach multiple local minima during training [34]. The LR is lowered at a very fast pace at first, encouraging the neural networks to converge towards its local minimum. Then the optimization continues at the initial LR, and the procedure repeats several times. The shifted cosine function is used to obtain the LR at each iteration, which is mathematically described below.

$$\alpha(t) = \frac{\alpha_0}{2}\left(cos\left(\frac{\pi \cdot mod(t-1,[T/M])}{[T/M]}\right) + 1\right) \tag{8}$$

where $\alpha(t)$ denotes the LR at the iteration number of t, α_0 is the initial LR, T represents the total number of training iterations and M is the number of cycles the procedure is repeated. A 'snapshot' of the model is taken when the LR reaches its minimum at each cycle; thus, a total of M models are acquired. The snapshot of the model is also referred as the history-state of the model in this work. There-

fore, Snapshot ensemble can be regarded as an implementation of DHSE with scheduled LR, and CCALR is a type of training strategy.

2. Boosted training strategy

Boosting technique trains a number of weak learners sequentially through an iterative arrangement of training samples to form a stronger model. The technique gives larger weights to those samples that were misclassified by the previous weak learners; in this way, the classifiers are supposed to have less overlap in the set of samples they misclassify. A boosted framework for the neural network was proposed in [37], the data distribution of each classifier is arranged as follows:

$$W_t(i) = \frac{1/n}{Z_{t-1}}e^{-\beta_{t-1}\cdot\alpha}, i = 1,2,\cdots n. \tag{9}$$

$$\beta_t = \frac{1}{2}log\frac{1-\epsilon_t}{\epsilon_t} + \frac{1}{10}log(k-1) \tag{10}$$

$$\epsilon_t = \frac{1}{n}\sum_{i=1}^{n}I(f_t(x_i)\neq y_i) \tag{11}$$

$$\alpha = \begin{cases} 1, iff_t(x_i) = y_i \\ -1, iff_t(x_i)\neq y_i \end{cases} \tag{12}$$

where $W_t(i)$ denotes the weight of i-th sample for the t-th classifier, ϵ_t counts the number of misclassified samples. A higher value of ϵ_t yeilds a smaller value of the coefficient β_t, which will be further reflected in a higher value of weight. Z_t is a normalization factor, which is defined as $\sum_{i=1}^{n}e^{-\beta_{t-1}y_if_t(x_i)}$. The difference between the boosted framework for a deep network with AdaBoost is: (1) A validation set is used to count the ϵ_t, while conventionally, it is calculated by all training samples. (2) The weights W_t is updated from the uniform distribution of $1/n$ for each classifier: note that in the conventional procedure, it is updated from W_{t-1} sequentially. (3) The boosted framework used a *meta*-learner to combine the weak classifiers with a validation set.

3. Weighted learning strategy

The weighted ensemble output is expressed as:

$$E_i = \boldsymbol{w}_i \times \sum_{i=1}^{N}\frac{exp(\boldsymbol{b}^i + \boldsymbol{W}_j^i * \boldsymbol{H})}{\sum_{j=1}^{c}exp(\boldsymbol{b}^i + \boldsymbol{W}_j^i * \boldsymbol{H})} \tag{13}$$

Compared with AV, the weighted voting applies a *meta*-leaner to adaptively learn the weight of each base model $\boldsymbol{w}_i$.

4. The detailed experimental results on dataset A-F are presented in Tables 8–13.

Table 8
Average accuracy of the probed methods on dataset A.

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Single + LRD	0.8407	0.8402	0.8377	0.9494	0.9846	0.9835	0.9855	0.9907
Single + CLR	0.7861	0.7827	0.7786	0.9094	0.9783	0.9750	0.9536	0.9873
MBGD + LRD + AV	0.8433	0.8418	0.8393	0.9501	0.9846	0.9843	0.9866	0.9913
MBGD + CLR + AV	**0.8698**	**0.8700**	**0.8689**	**0.9543**	**0.9850**	**0.9852**	**0.9868**	**0.9919**
Snap5 + AV	0.8573	0.8509	0.8540	0.9516	**0.9851**	0.9845	**0.9872**	0.9917
Snap30 + AV	0.8516	0.8495	0.8497	0.9497	**0.9851**	0.9847	**0.9874**	0.9909
Boosting + AV	0.8366	0.8388	0.8402	0.9475	0.9848	0.9837	0.9862	0.9903
Boosting + Snapshot + AV	0.8265	0.8300	0.8263	0.9458	0.9842	0.9845	0.9866	0.9899
MBGD + CLR + Ranking	0.8265	0.8246	0.8310	0.9416	0.9834	0.9820	0.9850	0.9903

Table 8 (*continued*)

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Snap5 + Ranking	0.8317	0.8346	0.8316	0.9450	0.9831	0.9833	0.9865	0.9911
Snap30 + Ranking	0.8213	0.8229	0.8241	0.9444	0.9839	0.9833	0.9859	0.9907
Boosting + Ranking	0.8082	0.8116	0.8120	0.9391	0.9834	0.9815	0.9838	0.9894
Boosting + Snapshot + Ranking	0.8144	0.8176	0.8194	0.9437	0.9837	0.9831	0.9854	0.9899
MBGD + CLR + Weight	0.7897	0.7969	0.7834	0.9111	0.9744	0.9684	0.9708	0.9883
Snap5 + Weight	0.8247	0.8295	0.8272	0.9434	0.9827	0.9827	0.9861	0.9911
Snap30 + Weight	0.8028	0.8031	0.8048	0.9390	0.9804	0.9771	0.9802	0.9897
Boosting + Weight	0.7793	0.7769	0.7830	0.9171	0.9745	0.9682	0.9676	0.9886
Boosting + Snapshot + Weight	0.7978	0.7999	0.8041	0.9308	0.9793	0.9771	0.9798	0.9895
MBGD + CLR + Select	0.8485	0.8499	0.8500	0.9486	0.9843	0.9839	0.9860	0.9911
Snap5 + Select	0.8335	0.8375	0.8329	0.9451	0.9835	0.9834	0.9865	0.9910
Snap30 + Select	0.8272	0.8292	0.8312	0.9453	0.9841	0.9837	0.9869	0.9909
Boosting + Select	0.8341	0.8362	0.8366	0.9463	0.9844	0.9832	0.9860	0.9904
Boosting + Snapshot + Ranking	0.8254	0.8295	0.8259	0.9456	0.9842	0.9841	0.9868	0.9899

Table 9
Average accuracy of the probed methods on dataset B.

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Single + LRD	0.5719	0.5754	0.5777	0.8608	0.9252	0.8934	0.8667	0.8856
Single + CLR	0.5316	0.5533	0.5435	0.8190	0.9115	0.8869	0.8344	0.8932
MBGD + LRD + AV	0.5777	0.5799	0.5809	0.8609	0.9275	0.9011	0.8692	0.8901
MBGD + CLR + AV	**0.5968**	**0.6041**	**0.6050**	**0.8635**	**0.9266**	**0.9098**	**0.8732**	**0.8989**
Snap5 + AV	0.5803	0.5839	0.5851	0.8594	0.9280	0.9018	0.8685	0.8906
Snap30 + AV	0.5804	0.5827	0.5837	0.8584	0.9287	0.9049	0.8724	0.8940
Boosting + AV	0.5874	0.5942	0.5868	0.8685	0.9192	0.9044	0.8741	0.8974
Boosting + Snapshot + AV	0.5667	0.5698	0.5717	0.8653	0.9233	0.9017	0.8673	0.8970
MBGD + CLR + Ranking	0.5734	0.5769	0.5696	0.8572	0.9233	0.9024	0.8628	0.8939
Snap5 + Ranking	0.5686	0.5721	0.5678	0.8631	0.9228	0.9049	0.8685	0.8961
Snap30 + Ranking	0.5675	0.5709	0.5666	0.8665	0.9230	0.9009	0.8589	0.8972
Boosting + Ranking	0.5661	0.5747	0.5643	0.8602	0.9143	0.8984	0.8635	0.8963
Boosting + Snapshot + Ranking	0.5651	0.5665	0.5692	0.8659	0.9214	0.8973	0.8642	0.8971
MBGD + CLR + Weight	0.5482	0.5572	0.5481	0.8287	0.8994	0.8737	0.8451	0.8951
Snap5 + Weight	0.5649	0.5718	0.5697	0.8631	0.9229	0.8986	0.8655	0.8978
Snap30 + Weight	0.5622	0.5679	0.5582	0.8592	0.9201	0.8962	0.8534	0.8935
Boosting + Weight	0.5477	0.5459	0.5369	0.8141	0.9126	0.8518	0.8249	0.8881
Boosting + Snapshot + Weight	0.5566	0.5581	0.5624	0.8564	0.9131	0.8877	0.8483	0.8906
MBGD + CLR + Select	0.5892	0.5954	0.5915	**0.8654**	0.9250	0.9008	0.8654	0.8993
Snap5 + Select	0.5698	0.5731	0.5708	0.8619	0.9232	0.9014	0.8685	0.8966
Snap30 + Select	0.5677	0.5744	0.5684	0.8656	0.9240	0.9016	0.8647	0.8987
Boosting + Select	0.5858	0.5921	0.5849	0.8662	0.9187	0.9031	**0.8732**	0.8970
Boosting + Snapshot + Ranking	0.5668	0.5698	0.5716	0.8651	0.9237	0.9024	0.8683	0.8978

Table 10
Average accuracy of the probed methods on dataset C.

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Single + LRD	0.5375	0.5383	0.5389	0.8357	0.9053	0.9002	0.8922	0.8976
Single + CLR	0.5150	0.5106	0.5096	0.8096	0.8964	0.8804	0.8445	**0.9068**
MBGD + LRD + AV	0.5403	0.5402	0.5421	0.8362	0.9066	0.9014	0.8940	0.9021
MBGD + CLR + AV	**0.5606**	**0.5599**	**0.5596**	0.8457	**0.9092**	**0.9031**	**0.8973**	0.9065
Snap5 + AV	0.5455	0.5472	0.5438	0.8362	0.9055	0.8965	0.8898	0.9015
Snap30 + AV	0.5424	0.5408	0.5438	0.8334	0.9086	0.8993	0.8951	0.9000
Boosting + AV	0.5440	0.5464	0.5466	**0.8478**	0.9074	0.9009	0.8877	0.9055
Boosting + Snapshot + AV	0.5294	0.5339	0.5326	0.8352	0.9049	0.8978	0.8802	0.9021
MBGD + CLR + Ranking	0.5338	0.5378	0.5372	0.8345	0.9054	0.8982	0.8861	0.9041
Snap5 + Ranking	0.5279	0.5349	0.5317	0.8363	0.9070	0.8955	0.8803	0.9057
Snap30 + Ranking	0.5264	0.5293	0.5265	0.8357	0.9056	0.8918	0.8801	0.9040
Boosting + Ranking	0.5255	0.5251	0.5276	0.8374	0.9047	0.8897	0.8805	0.9041
Boosting + Snapshot + Ranking	0.5239	0.5279	0.5311	0.8319	0.9027	0.8940	0.8776	0.9012
MBGD + CLR + Weight	0.5144	0.5198	0.5214	0.8187	0.8912	0.8684	0.8648	**0.9081**
Snap5 + Weight	0.5256	0.5316	0.5299	0.8337	0.9064	0.8954	0.8788	**0.9068**
Snap30 + Weight	0.5211	0.5221	0.5251	0.8320	0.9015	0.8860	0.8672	0.9022
Boosting + Weight	0.5114	0.5083	0.5147	0.8055	0.8851	0.8713	0.8448	0.9014
Boosting + Snapshot + Weight	0.5181	0.5243	0.5255	0.8290	0.8951	0.8882	0.8687	0.9010
MBGD + CLR + Select	0.5456	0.5486	0.5472	0.8449	0.9079	0.9002	0.8869	0.9062
Snap5 + Select	0.5284	0.5354	0.5336	0.8368	0.9066	0.8952	0.8819	0.9054
Snap30 + Select	0.5297	0.5311	0.5298	0.8352	0.9050	0.8957	0.8826	0.9042
Boosting + Select	0.5423	0.5448	0.5441	**0.8466**	0.9087	0.9001	0.8888	0.9050
Boosting + Snapshot + Ranking	0.5287	0.5330	0.5324	0.8349	0.9046	0.8978	0.8808	0.9027

Table 11
Average accuracy of the probed methods on dataset D.

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Single + LRD	0.5283	0.5351	0.5314	0.7932	0.8844	0.8556	0.8655	0.8666
Single + CLR	0.4991	0.4983	0.5006	0.7773	0.8755	0.8424	0.8451	0.8594
MBGD + LRD + AV	0.5324	0.5380	0.5344	0.7959	0.8893	0.8609	0.8721	0.8726
MBGD + CLR + AV	**0.5499**	**0.5479**	**0.5519**	**0.8036**	**0.8911**	**0.8712**	**0.8762**	**0.8774**
Snap5 + AV	0.5281	0.5339	0.5323	0.7949	0.8894	0.8635	0.8699	**0.8817**
Snap30 + AV	0.5332	0.5349	0.5336	0.7947	0.8888	0.8612	0.8718	0.8725
Boosting + AV	0.5332	0.5308	0.5365	0.7966	0.8883	0.8692	0.8707	0.8656
Boosting + Snapshot + AV	0.5241	0.5202	0.5194	0.7931	0.8853	0.8560	0.8728	0.8592
MBGD + CLR + Ranking	0.5214	0.5226	0.5266	0.7901	0.8904	0.8666	0.8704	0.8689
Snap5 + Ranking	0.5268	0.5201	0.5233	0.7897	0.8873	0.8655	0.8719	0.8669
Snap30 + Ranking	0.5224	0.5168	0.5172	0.7919	0.8895	0.8495	0.8707	0.8752
Boosting + Ranking	0.5209	0.5143	0.5176	0.7959	0.8842	0.8552	0.8622	0.8692
Boosting + Snapshot + Ranking	0.5219	0.5172	0.5160	0.7918	0.8907	0.8585	0.8645	0.8605
MBGD + CLR + Weight	0.5020	0.4978	0.5036	0.7709	0.8698	0.8169	0.8468	0.8581
Snap5 + Weight	0.5227	0.5198	0.5217	0.7902	0.8874	0.8620	0.8683	0.8642
Snap30 + Weight	0.5166	0.5124	0.5135	0.7958	0.8780	0.8502	0.8551	0.8698
Boosting + Weight	0.4944	0.4908	0.4994	0.7518	0.8627	0.8160	0.8262	0.8626
Boosting + Snapshot + Weight	0.5133	0.5073	0.5149	0.7905	0.8712	0.8225	0.8475	0.8580
MBGD + CLR + Select	0.5367	0.5362	0.5397	0.7943	**0.8913**	0.8674	0.8792	0.8726
Snap5 + Select	0.5261	0.5206	0.5233	0.7887	0.8865	0.8641	0.8722	0.8667
Snap30 + Select	0.5251	0.5185	0.5210	0.7924	0.8868	0.8645	0.8750	0.8716
Boosting + Select	0.5339	0.5311	0.5339	0.7960	0.8860	0.8674	0.8713	0.8643
Boosting + Snapshot + Ranking	0.5239	0.5204	0.5194	0.7928	0.8851	0.8545	0.8728	0.8602

Table 12
Average accuracy of the probed methods on dataset E.

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Single + LRD	0.6306	0.6293	0.6286	0.9093	0.9392	0.9572	0.9598	0.9735
Single + CLR	0.5860	0.5874	0.5916	0.8588	0.9153	0.9183	0.9274	0.9615
MBGD + LRD + AV	0.6305	0.6296	0.6295	0.9100	0.9418	0.9585	0.9610	0.9746
MBGD + CLR + AV	**0.6555**	**0.6554**	**0.6545**	**0.9213**	**0.9445**	**0.9602**	**0.9628**	**0.9762**
Snap5 + AV	0.6475	0.6456	0.6442	0.9202	0.9433	0.9594	**0.9650**	0.9753
Snap30 + AV	0.6412	0.6418	0.6415	0.9185	0.9430	**0.9606**	0.9630	0.9760
Boosting + AV	0.6363	0.6383	0.6388	0.9132	0.9344	0.9462	0.9499	0.9689
Boosting + Snapshot + AV	0.6292	0.6309	0.6337	0.9122	0.9345	0.9496	0.9553	0.9705
MBGD + CLR + Ranking	0.6216	0.6199	0.6207	0.9000	0.9328	0.9505	0.9561	0.9723
Snap5 + Ranking	0.6236	0.6257	0.6252	0.9086	0.9357	0.9522	0.9619	0.9734
Snap30 + Ranking	0.6172	0.6191	0.6192	0.9028	0.9344	0.9516	0.9596	0.9721
Boosting + Ranking	0.6110	0.6163	0.6106	0.8977	0.9317	0.9485	0.9447	0.9680
Boosting + Snapshot + Ranking	0.6125	0.6169	0.6199	0.9014	0.9338	0.9518	0.9536	0.9704
MBGD + CLR + Weight	0.5900	0.5927	0.5928	0.8517	0.9172	0.9294	0.9359	0.9682
Snap5 + Weight	0.6196	0.6212	0.6197	0.8994	0.9337	0.9514	0.9599	0.9725
Snap30 + Weight	0.6085	0.6051	0.6054	0.8758	0.9265	0.9419	0.9503	0.9689
Boosting + Weight	0.5837	0.5902	0.5861	0.8543	0.9165	0.9303	0.9286	0.9637
Boosting + Snapshot + Weight	0.5968	0.6011	0.6004	0.8718	0.9255	0.9392	0.9440	0.9662
MBGD + CLR + Select	0.6405	0.6401	0.6408	0.9131	0.9374	0.9506	0.9606	0.9749
Snap5 + Select	0.6287	0.6319	0.6308	0.9126	0.9358	0.9527	0.9620	0.9737
Snap30 + Select	0.6254	0.6245	0.6270	0.9106	0.9362	0.9531	0.9611	0.9730
Boosting + Select	0.6347	0.6369	0.6368	0.9116	0.9343	0.9476	0.9493	0.9686
Boosting + Snapshot + Ranking	0.6278	0.6307	0.6324	0.9113	0.9346	0.9497	0.9556	0.9705

Table 13
Average accuracy of the probed methods on dataset F.

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Single + LRD	0.5755	0.5752	0.5769	0.7430	0.7348	0.7870	0.8063	0.8326
Single + CLR	0.5345	0.5353	0.5375	0.6999	0.7173	0.7629	0.7624	0.8114
MBGD + LRD + AV	0.5759	0.5760	0.5774	0.7434	0.7326	0.7903	0.8078	0.8336
MBGD + CLR + AV	**0.6033**	**0.6022**	**0.6031**	**0.7513**	**0.7380**	**0.7927**	**0.8114**	**0.8360**
Snap5 + AV	0.5930	0.5920	0.5900	0.7446	0.7372	0.7915	0.8120	0.8321
Snap30 + AV	0.5906	0.5868	0.5901	0.7483	0.7345	0.7932	0.8093	0.8312
Boosting + AV	0.5874	0.5852	0.5854	0.7456	0.7276	0.7932	0.7934	0.8336
Boosting + Snapshot + AV	0.5773	0.5808	0.5776	0.7406	0.7259	0.7878	0.7945	0.8348
MBGD + CLR + Ranking	0.5671	0.5690	0.5689	0.7272	0.7273	0.7873	0.7854	0.8292
Snap5 + Ranking	0.5750	0.5767	0.5744	0.7343	0.7259	0.7892	0.7940	0.8316
Snap30 + Ranking	0.5656	0.5642	0.5675	0.7304	0.7244	0.7842	0.7929	0.8333
Boosting + Ranking	0.5568	0.5641	0.5592	0.7296	0.7264	0.7882	0.7833	0.8309
Boosting + Snapshot + Ranking	0.5629	0.5661	0.5670	0.7309	0.7262	0.7850	0.7932	0.8336
MBGD + CLR + Weight	0.5413	0.5377	0.5482	0.6815	0.7140	0.7665	0.7588	0.8217

(*continued on next page*)

Table 13 (*continued*)

Methods	BPNN	SAE	DBN	GRU	1D-LeNet5	1D-AlexNet	1D-ResNet	1D-DenseNet
Snap5 + Weight	0.5688	0.5702	0.5717	0.7285	0.7239	0.7883	0.7898	0.8296
Snap30 + Weight	0.5496	0.5490	0.5528	0.7108	0.7183	0.7762	0.7798	0.8274
Boosting + Weight	0.5417	0.5384	0.5406	0.6908	0.7127	0.7592	0.7573	0.8279
Boosting + Snapshot + Weight	0.5423	0.5481	0.5498	0.7072	0.7155	0.7699	0.7809	0.8276
MBGD + CLR + Select	0.5886	0.5889	0.5885	0.7480	0.7294	0.7937	0.7959	0.8317
Snap5 + Select	0.5791	0.5819	0.5781	0.7358	0.7258	0.7894	0.7939	0.8320
Snap30 + Select	0.5738	0.5783	0.5762	0.7348	0.7254	0.7882	0.7970	0.8350
Boosting + Select	0.5857	0.5848	0.5863	0.7456	0.7266	0.7934	0.7917	0.8318
Boosting + Snapshot + Ranking	0.5761	0.5802	0.5779	0.7395	0.7253	0.7874	0.7950	0.8349

References

[1] H.Y. Sim, R. Ramli, A. Saifizul, M.F. Soong, Detection and estimation of valve leakage losses in reciprocating compressor using acoustic emission technique, Measurement 152 (2020) 107315.

[2] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech. Syst. Sig. Process. 138 (2020) 106587.

[3] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, Mech. Syst. Sig. Process. 108 (2018) 33–47, https://doi.org/10.1016/j.ymssp.2018.02.016.

[4] R. Wang, H. Jiang, X. Li, S. Liu, A reinforcement neural architecture search method for rolling bearing fault diagnosis, Measurement 154 (2020) 107417.

[5] Y.J. Cruz, M. Rivas, R. Quiza, A. Villalonga, R.E. Haber, G. Beruvides, Ensemble of convolutional neural networks based on an evolutionary algorithm applied to an industrial welding process, Comput. Ind. 133 (2021), https://doi.org/10.1016/j.compind.2021.103530.

[6] H. Wang, J. Xu, R. Yan, R.X. Gao, A New Intelligent Bearing Fault Diagnosis Method Using SDP Representation and SE-CNN, IEEE Trans. Instrum. Meas. 69 (2020) 2377–2389, https://doi.org/10.1109/TIM.2019.2956332.

[7] C. Wu, P. Jiang, C. Ding, F. Feng, T. Chen, Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network, Comput. Ind. 108 (2019) 53–61, https://doi.org/10.1016/j.compind.2018.12.001.

[8] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Ind. Electron. 65 (2018) 5990–5998, https://doi.org/10.1109/tie.2017.2774777.

[9] B. Zhao, X. Zhang, H. Li, Z. Yang, Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions, Knowl.-Based Syst. 199 (2020), https://doi.org/10.1016/j.knosys.2020.105971.

[10] Y. Li, H. Cao, K. Tang, A general dynamic model coupled with EFEM and DBM of rolling bearing-rotor system, Mech. Syst. Sig. Process. 134 (2019), https://doi.org/10.1016/j.ymssp.2019.106322.

[11] X. Yan, Y. Liu, M. Jia, Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions, Knowl.-Based Syst. 193 (2020) 105484.

[12] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Sig. Process. 95 (2017) 187–204, https://doi.org/10.1016/j.ymssp.2017.03.034.

[13] S. Plakias, Y.S. Boutalis, A novel information processing method based on an ensemble of Auto-Encoders for unsupervised fault detection, Comput. Ind. 142 (2022), https://doi.org/10.1016/j.compind.2022.103743.

[14] H. Zhu, J. Cheng, C. Zhang, J. Wu, X. Shao, Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings, Appl. Soft Comput. 88 (2020) 106060.

[15] X. Kong, G. Mao, Q. Wang, H. Ma, W. Yang, A multi-ensemble method based on deep auto-encoders for fault diagnosis of rolling bearings, Measurement 151 (2020) 107132.

[16] S. Xiang, Y.i. Qin, C. Zhu, Y. Wang, H. Chen, Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction, Eng. Appl. Artif. Intel. 91 (2020) 103587.

[17] X. Chen, B. Zhang, D. Gao, Bearing fault diagnosis base on multi-scale CNN and LSTM model, J. Intell. Manuf. 32 (4) (2021) 971–987.

[18] Y. Wang, L. Cheng, A combination of residual and long–short-term memory networks for bearing fault diagnosis based on time-series model analysis, Meas. Sci. Technol. 32 (2020).

[19] P.G. Benardos, G.-C. Vosniakos, Optimizing feedforward artificial neural network architecture, Eng. Appl. Artif. Intel. 20 (2007) 365–382, https://doi.org/10.1016/j.engappai.2006.01.005.

[20] W. Huang, J. Cheng, Y. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, Neurocomputing 359 (2019) 77–92, https://doi.org/10.1016/j.neucom.2019.05.052.

[21] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, C. Li, Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Sig. Process. 70–71 (2016) 87–103, https://doi.org/10.1016/j.ymssp.2015.08.030.

[22] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, R. de J. Romero-Troncoso, Early fault detection in induction motors using adaboost with imbalanced small data and optimized sampling, IEEE Trans. Ind. Appl. 53 (2017) 3066–3075, https://doi.org/10.1109/TIA.2016.2618756.

[23] R. Zhang, B. Li, B. Jiao, Application of XGboost algorithm in bearing fault diagnosis, IOP Conf. Ser.: Mater. Sci. Eng. 490 (2019), https://doi.org/10.1088/1757-899X/490/7/072062.

[24] P. Thomas, H. Bril El Haouzi, M.-C. Suhner, A. Thomas, E. Zimmermann, M. Noyel, Using a classifier ensemble for proactive quality monitoring and control: the impact of the choice of classifiers types, selection criterion, and fusion process, Comput. Ind. 99 (2018) 193–204, https://doi.org/10.1016/j.compind.2018.03.038.

[25] S. Ma, F. Chu, Ensemble deep learning-based fault diagnosis of rotor bearing systems, Comput. Ind. 105 (2019) 143–152, https://doi.org/10.1016/j.compind.2018.12.012.

[26] C. Zhang, P. Lim, A.K. Qin, K.C. Tan, Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics, IEEE Trans. Neural Networks Learn. Syst. 28 (10) (2017) 2306–2318.

[27] Y. Zhang, X. Li, L. Gao, W. Chen, P. Li, Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method, Measurement 151 (2020) 107232.

[28] J. Yang, G. Xie, Y. Yang, An improved ensemble fusion autoencoder model for fault diagnosis from imbalanced and incomplete data, Control Eng. Pract. 98 (2020) 104358.

[29] H.A. Saeed, M.-J. Peng, H. Wang, B.-W. Zhang, Novel fault diagnosis scheme utilizing deep learning networks, Prog. Nucl. Energy 118 (2020) 13066.

[30] Y. Zhang, X. Li, L. Gao, W. Chen, P. Li, Ensemble deep contractive auto-encoders for intelligent fault diagnosis of machines under noisy environment, Knowl.-Based Syst. 196 (2020) 105764.

[31] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Sig. Process. 100 (2018) 439–453, https://doi.org/10.1016/j.ymssp.2017.06.022.

[32] X. Li, H. Jiang, R. Wang, M. Niu, Rolling bearing fault diagnosis using optimal ensemble deep transfer network, Knowl.-Based Syst. 213 (2021), https://doi.org/10.1016/j.knosys.2020.106695.

[33] J. Xie, B. Xu, Z. Chuang, Horizontal and Vertical Ensemble with Deep Representation for Classification, (2013). https://doi.org/10.48550/arXiv.1306.2759.

[34] G. Huang, Y. Li, G. Pleiss, Z. Liu, J.E. Hopcroft, K.Q. Weinberger, Snapshot Ensembles: Train 1, get M for free, (2017). https://doi.org/10.48550/arXiv.1704.00109.

[35] L. Wen, L. Gao, X. Li, A new snapshot ensemble convolutional neural network for fault diagnosis, IEEE Access 7 (2019) 32037–32047, https://doi.org/10.1109/ACCESS.2019.2903295.

[36] L. Wen, X. Xie, X. Li, L. Gao, A new ensemble convolutional neural network with diversity regularization, J. Manuf. Syst. (2020).

[37] W. Zhang, J. Jiang, Y. Shao, B. Cui, Snapshot boosting: a fast ensemble framework for deep neural networks, Sci. China Inf. Sci. 63 (2019), https://doi.org/10.1007/s11432-018-9944-x.

[38] J. Yang, X. Zeng, S. Zhong, S. Wu, Effective neural network ensemble approach for improving generalization performance, IEEE Trans. Neural Networks Learn. Syst. 24 (2013) 878–887, https://doi.org/10.1109/TNNLS.2013.2246578.

[39] C. Lessmeier, J.K. Kimotho, D. Zimmer, 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, PHM Society European Conference. 3 (2016). https://doi.org/10.36001/phme.2016.v3i1.1577.

[40] Y. Zhang, T. Zhou, X. Huang, L. Cao, Q. Zhou, Fault diagnosis of rotating machinery based on recurrent neural networks, Measurement 171 (2021), https://doi.org/10.1016/j.measurement.2020.108774.

[41] Y. Wang, Q. Wang, A. Vinogradov, Semi-supervised deep architecture for classification in streaming data with emerging new classes: application in condition monitoring, (2023). https://doi.org/10.36227/techrxiv.21931476.v1.

[42] Y. Wang, A. Vinogradov, Improving the Performance of Convolutional GAN Using History-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Signals, Applied Sciences. 13 (2023) 3136. https://doi.org/10.3390/app13053136.
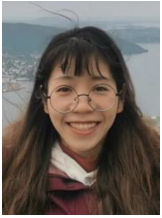
Yu Wang received the MM. degree in Management Science and Engineering from the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing, China in 2020. She is currently working towards the Ph.D. degree in the Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway. She has worked on the intelligent fault diagnosis of rotating machinery with machine learning since 2017.

Alexey Vinogradov majored in acoustic emission signal processing, data mining, and failure analsyis and pre-diction. He received his PhD degree in physics and mathematics from A.F. Ioffe Physical-Technical Institute of St- Petersburg, USSR, in 1988. Since 1992 he has taken several academic positions in Japan and Norway. Since 2023 AV serves as a distinguished professor of the Magnesium Research Center of Kumamoto University, Japan.

Paper C:

Article

Improving the Performance of Convolutional GAN Using History-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Signals

Yu Wang [1] [ID] and Alexey Vinogradov [2],*

1   Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, 7034 Trondheim, Norway
2   Magnesium Research Center, Kumamoto University, Kumamoto 860-8555, Japan
*   Correspondence: vinogradov@kumamto-u.ac.jp

Abstract: Early fault detection (EFD) in run-to-failure processes plays a crucial role in the condition monitoring of modern industrial rotating facilities, which entail increasing demands for safety, energy and ecological savings and efficiency. To enable effective protection measures, the evolving faults have to be recognized and identified as early as possible. The major challenge is to distil discriminative features on the basis of only the 'health' signal, which is uniquely available from various possible sensors before damage sets in and before the signatures of incipient damage become obvious and well-understood in the signal. Acoustic emission (AE) signals have been frequently reported to be able to deliver early diagnostic information due to their inherently high sensitivity to the incipient fault activities, highlighting the great potential of the AE technique for EFD, which may outperform the traditional vibration-based analysis in many situations. To date, the 'feature-based' multivariate analysis dominates the interpretation of AE waveforms. In this way, the decision-making relies heavily on experts' knowledge and experience, which is often a weak link in the entire EFD chain. With the advent of artificial intelligence, practitioners seek an intelligent method capable of tackling this challenge. In the present paper, we introduce a versatile approach towards intelligent data analysis adapted to AE signals streaming from the sensors used for the continuous monitoring of rotating machinery. A new architecture with a convolutional generative adversarial network (GAN) is designed to extract the deep information embedded in the AE waveforms. In order to improve the robustness of the proposed EFD framework, a novel ensemble technique referred to as 'history-state ensemble' (HSE) is introduced and paired with GAN. The primary merits of HSE are twofold: (1) it does not require extra computing time to obtain the base models, and (2) it does not require a special design of the network architecture and can be applied to different networks. To evaluate the proposed method, a durability rolling contact fatigue test was performed with the use of AE monitoring. The experimental results have demonstrated that the proposed ensemble method largely improves the robustness of GAN.

Keywords: early fault detection; acoustic emission signal; unsupervised learning; ensembled method; convolutional GAN

1. Introduction

A rolling bearing is the core component in many rotating machines. Any failure in rolling bearings can lead to a chain reaction of faults in the whole mechanical system, causing the rapid and unexpected breakdown of the machine. Being an essential part of condition monitoring, the early and accurate identification of an imminent failure is deemed effective in reducing property loss and even possible casualties caused by catastrophic industrial breakdowns. Recent years have seen the rapid development of innovative artificial intelligence (AI) algorithms, including two major groups: machine learning (ML) and deep learning (DL). Inspired by the progress in this field on the one hand, and informed by the

long-standing unaddressed challenges faced by traditional 'feature-based' approaches, practitioners strive to find new solvers—intelligent methods capable of detecting the emerging faults early, reliably and seamlessly, without heavy reliance on human labor and expert experience [1,2]. To date, numerous intelligent fault diagnosis schemes have been proposed. Lei et al. have analyzed the relevant bibliometric data in this field [2]. The application of DL started to thrive after 2015 and it gradually surpassed traditional ML models. At present, the DL-based machine fault diagnosis framework has become the mainstream of intelligent fault diagnostics. Compared with traditional ML models, the procedures of feature extraction and fault recognition are integrated within the DL approach. The DL approach is unique in that it is capable of extracting features automatically from the input data through multiple layers comprising processing units called hidden neurons. This makes DL suitable to process the raw signal straightforwardly, without any signal pre-processing. Classical DL architectures include the Back Propagation Neuron Network (BPNN) [3], Convolutional Neuron Network (CNN) [4,5], Deep Boltzmann Machine (DBM), Deep Brief Network (DBN), Autoencoder (AEN) [6,7], Long Short-Term Memory (LSTM) network [8,9] and their variants.

The emerging damage in the machine can be communicated through different signal sources. Among them, the vibration-based technique is the most widely used one owing to its simplicity, the transparency of the analysis based on spectral features and cost advantages. However, there are shortcomings of vibration signals, which are to be mentioned. Firstly, the vibration acceleration signals can hardly be detectable until the damage develops significantly to a mature stage, corresponding to large-scale faults causing vibrations in heavy or slowly rotating structures. It is often too late to use this information for preventive maintenance [10]. Secondly, the vibration signals induced by early defects are easily masked by the routine background mechanical vibration of the rotating machine. It has been frequently reported that acoustic emission (AE) signals can detect the incipient crack earlier than traditional vibration signals [11–14]. Even if this claim is not always justified, the modern AE technique provides a promising means for EFD in roller bearings. The AE is referred to as a phenomenon of transient elastic wave generation by a sudden local drop in internal stress within the material. Compared to the vibration signal, AE has a much wider frequency range (20 kHz to 10 MHz), and, thus, it does not overlap or interfere with low-frequency mechanical vibration signals. To date, the 'feature-based' parametric analysis prevails in interpreting information derived from AE waveforms [15–17]. However, these hand-designed features are inherently linked through AE to the specific signal processing techniques in the time, frequency or time–frequency domains [15,18,19]. These features need to be carefully extracted and analyzed by experts, and there is no guarantee that the features tailored to a specific fault diagnosis condition are applicable for other tasks.

To reduce the risk of biased opinions, we leverage the DL technique to explore the implicit fault information embedded in AE signals. Among all types of DL models, the generative adversarial network (GAN) has shown a remarkable capacity to perform distribution fitting. GAN is a powerful generative model that was first proposed by Goodfellow et al. in 2014 [20]. Unlike conventional neural networks, GAN implements generative modeling as a game between two separated networks: a Generator is trained to produce synthetic data that are close to the real data, while a Discriminator is trained to discriminate between the synthetic and real data. During this training process, the probability distribution of the real data can be learned by the Generator. There are many successful applications of GAN in the fault diagnosis field. Existing research mainly focuses on the problem of 'unbalanced data', i.e., the sample size of anomal data is much smaller than that of regular data [21–26]. GAN is utilized to generate synthetic abnormal data to assist model training. These studies have demonstrated that GAN has an excellent ability to learn representative features from mechanical signals. Xia et al. summarized the applications of GAN to anomaly detection in a number of fields [27]. It can be concluded that GAN is suitable for early fault detection (EFD) problems in two ways [28]: (1) it shows superiority in fitting the distribution of health signals; the Generator can be trained to learn rich and hierarchical information from the data

GAN is suitable for early fault detection (EFD) problems in two ways [28]: (1) its superiority in fitting the distribution of health signals; the Generator can be trained to learn rich and hierarchical information from the data characteristic of the normal operating state of the machine; thus, these abnormal data should be poorly reconstructed; (2) the Discriminator forms a health indicator (HI) to indicate the abnormal signals.

Although GAN is a promising approach for EFD, it faces a well-known issue of unstable training. In contrast to traditional supervised networks, whose performance can be well reflected by the loss value, GAN consists of two networks' 'fighting' with each other, so the loss values of the two networks show a relationship of 'as one falls, the other rises'. The balance between the Generator and Discriminator is subtle. Therefore, it is difficult to determine when the Discriminator is well-trained, leading to the problem of instability. In order to improve the robustness of GAN in the EFD problem, we introduced a new ensemble technique, referred to as the 'history state ensemble' (HSE) [29] method. HSE assumes that neural networks can generate multiple local optima during the training history. Our previous experimental results [29] have demonstrated that these local optima are diverse, and their combination improves the accuracy and stability of the single network. The ensemble method is generally perceived to be time-consuming. However, the benefit of the HSE is that it does not need extra training; we only need to record the critical training models when using the back-propagation algorithm to adjust the training model weights, which are operational algorithm states just the models. Therefore, it does not require a specially designed network architecture. These training model weights are denoted as 'history states' or base models.

In a brief summary, the main contributions of this paper are as follows.

(1) We proposed a novel architecture of GAN consisting of convolutional blocks and LSTM for EFD in the run-to-failure process.
(2) A new ensemble health indicator (EHI) is constructed by integrating GAN and a novel ensemble technique called the HSE method.
(3) A laboratory durability test of a roller bearing element monitored by the acoustic emission technique was carried out to evaluate the effectiveness of the proposed method.

2. The Proposed Architecture of Convolutional GAN

We first introduce the basic theory of traditional GAN, and the architecture of the proposed convolutional GAN is elaborated.

2.1. Basic Theory of GAN

The basic structure of GAN consists of two networks, as illustrated in Figure 1. The main idea is to construct a neural network model, known as the 'Generator', in order to map the random noises z into a new data space $G(z)$. The goal is to minimize the discrepancy between the 'fake data' from the mapped space $G(z)$ and the 'real data' from the target space $p_r(x)$. In contrast to traditional neural networks such as the Autoencoder, which directly minimizes the distance through the mean square error (MSE), one more neural network is introduced in GAN, referred to as the 'Discriminator', which is aimed to distinguish $G(z)$ from $p_r(x)$.
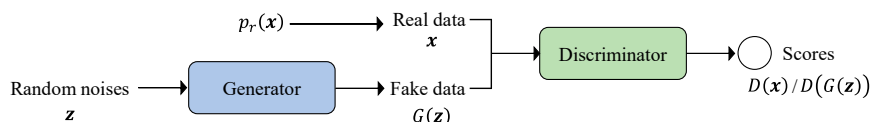
Figure 1. The basic structure of GAN.

The above stated goal can be achieved through the joint training of the two networks, and the original loss functions for the Generator and Discriminator are express as follows:

$$\max_D L(D) = E_{x\sim p_r(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))] \qquad (1)$$

$$\min_G L(G) = E_{z\sim p_z(z)}[log(1-D(G(z)))] \qquad (2)$$

where G and D represent the Generator and Discriminator, respectively. The D outputs a score ranging from 0 to 1 for each sample. In order to distinguish between the 'fake data' and 'real data', D tries to assign a value close to 1 for the sample x from 'real data', and a value close to 0 for the sample $G(z)$ from 'fake data'. Then, the loss function for D is maximized. On the other hand, G attempts to produce 'fake data' to fool D; thus, it will be adjusted to produce 'fake data' that are close to 'real data', allowing $D(G(z))$ to be close to 1; therefore, the loss function of G is minimized. If both networks have sufficient capacity, they will reach a point at which both models cannot be improved anymore because the generated distribution approximates the real distribution well enough.

The original loss function of GAN has been reported with challenges, such as unstable training and the poor quality of generated data. The key problem stems from the embedded Jensen–Shannon divergence (JSD) as a measure of the distance between real and generated distribution; the details can be found in [30]. To overcome the training challenges, Arjovsky et al., proposed to replace JSD with the Wasserstein distance (WD) defined as

$$W(p_r, p_\theta) = \frac{1}{K} sup_{\|f\|_L \leq 1} E_{x \sim p_r}[f(x)] - E_{\tilde{x} \sim p_g}[f(\tilde{x})] \tag{3}$$

where $f : x \rightarrow R$ is a set of Lipschitz functions satisfying the condition $\frac{|f(x_1)-f(x_2)|}{|x_1-x_2|} \leq K$, and K is the Lipschitz constant. It can be observed that the absolute value of the derivative of f does not exceed K. By applying this distance metric in GAN, the function f is referred to as a 'critic' in the original paper, which can be approximated by D, and p_g is the model distribution implicitly defined by $\tilde{x} = G(z)$, $z \sim p_z(z)$. The variant of GAN with WD is known as Wasserstein GAN (WGAN). Compared with JSD, WD has a smoother change rate when measuring the distance between two distributions; thus, it can provide meaningful gradient information to G.

The original WGAN applies a 'weight clipping' method to enforce a Lipschitz constraint by clamping the network weights to a fixed range $[-c, c]$ after each gradient update. This method still leads to optimization difficulties such as gradient vanishing [23,31]. Therefore, Gulrajani et al. proposed an alternative solution by enforcing a soft version of the Lipschitz constraint with a penalty on the gradient norm of random samples $\check{x} \sim p_{\check{x}}$, which is expressed as

$$\delta = E_{\tilde{x} \sim p_{\check{x}}} \left[(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - K)^2 \right] \tag{4}$$

where $\check{x} = \varepsilon x + (1 - \varepsilon \tilde{x})$, with $x \sim p_r$, $\tilde{x} \sim p_g$, and $\varepsilon \sim N[0, 1]$. The K is generally set as 1.

With the gradient penalty, a new objective is proposed as

$$\max_D L(D) = E_{\tilde{x} \sim p_\theta}[D(\tilde{x})] - E_{x \sim p_r}[D(x)] + \beta E_{\tilde{x} \sim p_{\check{x}}} \left[(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2 \right] \tag{5}$$

$$\min_G L(G) = -E_{\tilde{x} \sim p_{\check{x}}}[D(\tilde{x})] \tag{6}$$

where β is the penalty coefficient, and the WGAN with a gradient penalty is referred to as gp-WGAN.

2.2. Design of Generator and Discriminator

It is well-known that GAN is difficult to train, even with the use of WD and gradient penalty methods. One of the prime challenges is the diversity of the generated data, which is needed to cover the data distribution sufficiently. Otherwise, the Generator may become 'lazy', therefore producing the homogeneous data fooling the Discriminator. However, a good point in early fault detection, i.e., in early anomaly detection in the streaming of diagnostic data, is that the primary focus should be the performance of the Discriminator. This is the difference between our task and many other tasks, such as image generation or imbalanced data problems in fault diagnostics. In our work, it is not necessary that the Generator's input is a random signal. Hence, the Autoencoder-based Generator architecture is adopted in this work, i.e., both the input and output of the Generator are real data. The

ever, a good point in early fault detection, i.e., in early anomaly detection in the streaming of diagnostic data, is that the primary focus should be the performance of the Discriminator. This is the difference between our task and many other tasks, such as image generation or imbalanced data problems in fault diagnostics. In our work, it is not necessary that the Generator's input is a random signal. Hence, the Autoencoder-based Generator architecture is adopted in this work, i.e., both the input and output of the Generator are real data. The Autoencoder is known as a powerful and versatile non-linear dimensionality reduction technique employing neural networks for which the target output is the same as the input.

The proposed architectures of the Generator and Discriminator are shown in Figure 2. The Generator comprises an encoder block and a decoder block with a bottleneck block between them. The encoder is built of two convolutional layers to learn hierarchical representations of the real input, and each convolutional layer is followed by a down-sampling layer to reduce the feature size. Moreover, the down-sampling layer helps to boost the model's robustness to noise and variations in input data. Then, the real input is compressed into a compact representation called the bottleneck. In this paper, the LSTM cell is used in the bottleneck layer to capture the time-series correlation in data. By stacking two layers of LSTM cells, the network can learn more complex patterns in the input data and have better long-term memory retention. Next, the compressed representation is fed into the decoder, which consists of two convolutional layers, one up-sampling layer and one fully connected layer that finally reconstructs the input. In the competition between the Generator and the Discriminator, if the Discriminator is too powerful, it will quickly converge before the Generator can learn useful information from the input. Therefore, a more concise structure is used in the Discriminator, which is composed of three convolutional layers, three down-sampling layers, two fully connected layers and a one-dimensional output layer.
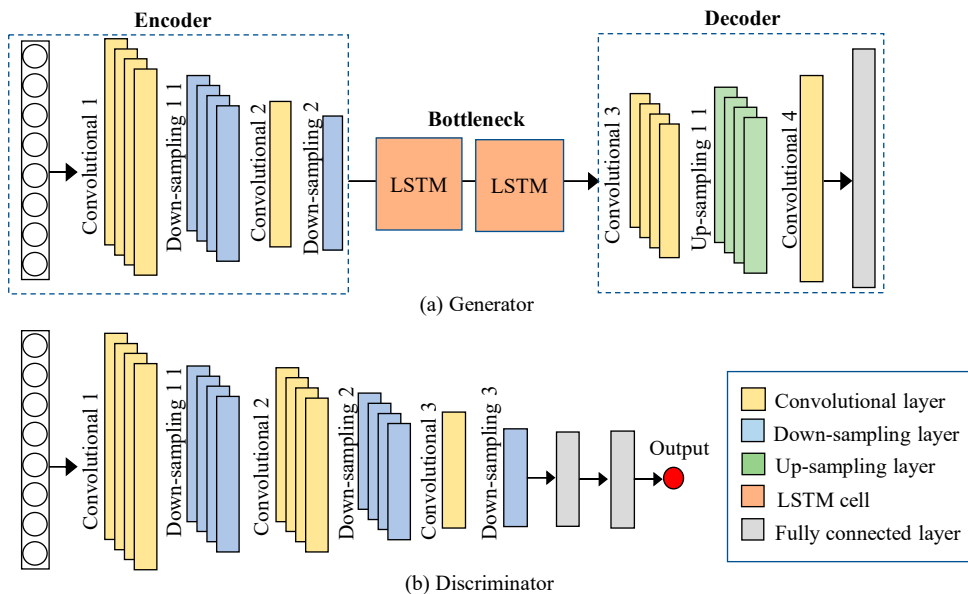

(a) Generator

(b) Discriminator

Figure 2. Architectures of the proposed Generator and Discriminator.

The most important mathematical details related to the above structure are presented below.

(1) **Convolutional layer.** The convolution process refers to a specialized linear operation where a small window called a kernel or filter overlays and slides through the entire input with a preset stride, which is mathematically expressed as

$$H_k = Activ\left(\sum_{i=1}^{inMap} Conv(W_k, X_i) + b_k\right) \tag{7}$$

where $Conv(\cdot)$ denotes the convolution window, W_k is the kernel that slides through the data X and b_k is the corresponding bias. $Activ(\cdot)$ represents the activation function, H_k

stands for the feature vector extracted by the *k*-th kernel and the subscript *k* defines the number of kernels.

(2) **Average down-sampling layer**. The down-sampling layer is also referred to as the pooling layer, which is commonly applied after a convolution layer to reduce the dimension of the feature maps. It refers to a special type of convolution whereby the kernel slides through the entire input map, and, generally, instead of creating the element-wise product, the average value of the overlaid input region is extracted; this quantity is the so-called 'average pooling'.

(3) **Up-sampling layer**. In contrast to down-sampling, up-sampling is generally used after the encoder to restore the resolution of the original data. The most common up-sampling techniques include Nearest Neighbor, Bilinear and Bicubic [32]. The Bilinear method is adopted in the present work.

(4) **Fully connected layer**. The fully connected layer refers to the type of neural network where all the input from the previous layer is connected to every neural node of the next layer:

$$\hat{H} = \text{Activ}(W * H + b) \tag{8}$$

where *H* stands for the neural nodes of the previous layer; *W* and *b* represent the weights and bias.

(5) **LSTM cell**. LSTM is a variant of Recurrent Neural Network (RNN), which has the advantage of exploiting the information of time-series signals. LSTM alleviates the vanishing gradient problem in the original RNN by introducing a memory cell, as described in Figure 3. The memory cell consists of a forget gate, input gate, output gate and state gate, which are mathematically described as follows:

$$f_t = \sigma\left(W_f \cdot [x_t, \ h_{t-1}] + b_f\right) \tag{9}$$

$$i_t = \sigma(W_i \cdot [x_t, \ h_{t-1}] + b_i) \tag{10}$$

$$o_t = \sigma(W_o \cdot [x_t, \ h_{t-1}] + b_o) \tag{11}$$

$$\widetilde{C}_t = tanh(W_c \cdot [x_t, \ h_{t-1}] + b_c) \tag{12}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{13}$$

$$h_t = o_t \cdot tanh(C_{t-1}) \tag{14}$$

where the f_t, i_t, o_t and C_t represent the forget gate, input gate, output gate and state gate, respectively. x_t is the input vector of the LSTM unit at the current time *t*, and h_{t-1} is the hidden vector of the previous time $t - 1$. Therefore, LSTM considers the information at the current time and the previous time. σ and *tanh* represent the 'sigmoid' and 'tanh' activation functions, respectively.
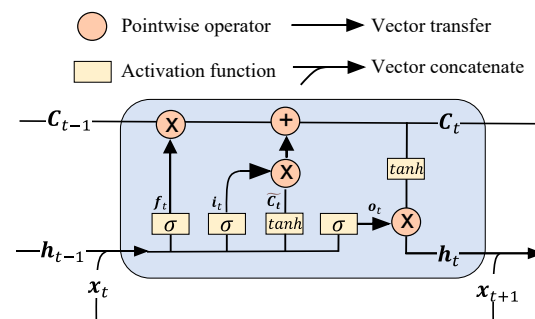
Figure 3. The structure of the LSTM memory cell.

It is worth noting that the input of LSTM should be a matrix, but the AE signal is a vector. Thus, the original 1D data should be converted to two-dimensional space. To this end, we simply divide the signal into multiple segments along the time axis. Thus, the input is reshaped into a $N \times M$ matrix, where *N* denotes the number of segments, and
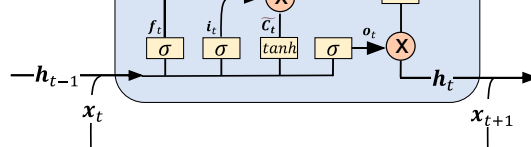
Figure 3. The structure of the LSTM memory cell.

It is worth noting that the input of LSTM should be a matrix, but the AE signal is a vector. Thus, the original 1D data should be converted to two-dimensional space. To this end, we simply divide the signal into multiple segments along the time axis. Thus, the input is reshaped into a $N \times M$ matrix, where N denotes the number of segments, and M stands for the length of each segment. The matrix is firstly processed by the encoder as M separated samples to capture detailed information about each segment, and the extracted features are stacked as a matrix and fed into the bottleneck layer. The output of the bottleneck layer is again processed by the decoder as a separate dataset. The reconstructed data are concatenated by the last fully connected layer. The data flow in the Generator is illustrated in Figure 4.
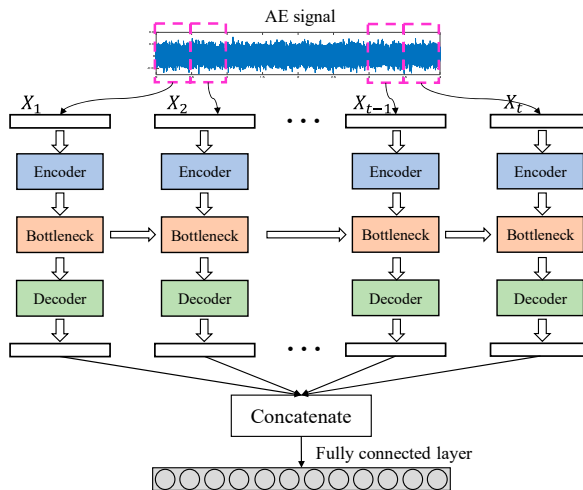

Figure 4. Illustration of the data flow within the Generator.

3. The Proposed EFD Framework Based on Convolutional GAN and History-State Ensemble

3.1. Definition of a Health Indicator (HI) Based on GAN

The output of the Discriminator is a single value that distinguishes the fake data from real data. If the input is identified as real, the Discriminator will ascribe a high value to it, whereas a low value will be set otherwise. If we apply this to the early fault detection problem with the training data from only the normal operating state, the Discriminator will ascribe the faulty data to the low value based on the assumption that the defect has distorted the normal AE waveform, which is successfully captured by the sensors. Therefore, the Discriminator naturally determines the health indicator HI, as presented below:

$$HI = D(\boldsymbol{X}) \tag{15}$$

where $\boldsymbol{X}$ denotes the evaluated data.

3.2. Ensembled Health Indicator

3.2.1. Motivation

The remaining question is how to determine whether the Discriminator has been well-trained or not. Differing from traditional supervised networks, whose performance can be assessed by the loss value, GAN benefits from the competition between the Generator and Discriminator. Therefore, the loss values of the two networks exhibit a relation of 'as one falls, the other rises'. To overcome this problem, we introduce a simple yet effective ensemble technique referred to as the history-state ensemble (HSE) method, as described in our dedicated study [29]. The advantages of the HSE method are twofold: (1) it does not

well trained or not. Differing from traditional supervised networks, whose performance can be assessed by the loss value, GAN benefits from the competition between the Generator and Discriminator. Therefore, the loss values of the two networks exhibit a relation of 'as one falls, the other rises'. To overcome this problem, we introduce a simple yet effective ensemble technique referred to as the history-state ensemble (HSE) method, as described in our dedicated study [29]. The advantages of the HSE method are twofold: (1) it does not require extra computing time to obtain the base models, and (2) it is versatile enough to be seamlessly applied to plain neural networks without readjusting the network architecture. Similarly to traditional ensemble techniques, the implementation of HSE methods assumes (i) encouraging the model to generate accurate base models with high diversity, and (ii) assembling these models to create a more robust classifier.

3.2.2. Base Model Generation

To obtain the base models, HSE is based on the assumption that the neural networks can generate multiple local optima, also referred to as 'history states', during the training process, and these local optima can be taken as base models for ensemble learning. Therefore, to generate multiple base models, one only needs to preserve the historical weights of the network, as illustrated in Figure 5. Hence, the time cost of this procedure is negligible.
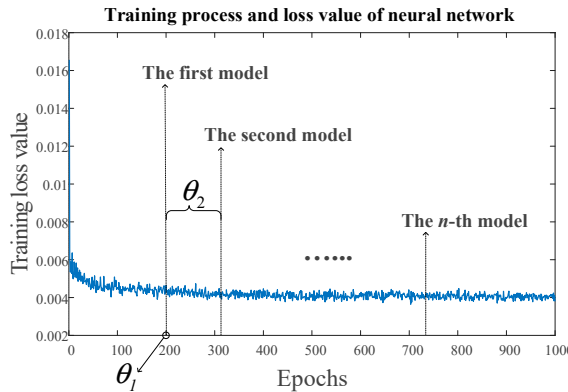
Figure 5. Illustration showing the typical behavior of the training loss value as a function of the number of training epochs; the definition of parameters θ_1 and θ_2 is clarified graphically (see the text for details).

The obtained base models have a direct impact on the model performance, which is affected chiefly by three factors: (1) the number of base models, (2) the accuracy of each base model and (3) the diversity within all base models. In order to encourage the diversity of base models, the Mini-Batch Gradient Descent (MBGD) is recommended in the training process. MBGD is a variant of the gradient descent algorithm whereby the whole training dataset is divided into multiple MBGD batches, and only one batch is used to calculate the gradient at each iteration. The application of MBGD increases the model update frequency, which helps to generate more models and encourages their diversity. Moreover, two extra parameters need to be defined, as illustrated in Figure 5. Here, θ_1 denotes the number of training epochs at which the first base model is acquired, and θ_2 indicates the model update frequency. The total number of acquired base models is calculated as

$$N = int\left(\frac{S - \theta_1}{\theta_2}\right) \qquad (16)$$

where *S* is the total number of training epochs.

3.2.3. Ensemble Results

Average voting (AV) is used to integrate the results of all base models, and, for the case of this paper, a new ensembled HI (EHI) can be constructed:

$$EHI = \sum_{i=1}^{N} D_i(\hat{X})/N \qquad (17)$$

where $D_i(\hat{X})$ represents the score given by the *i*-th base model, and *N* is the total number of the base models.

To define the fault alarm, a threshold *th* is set as

$$th = min(EHI_{train}) - C * std(EHI_{train}) \qquad (18)$$

where HI_{train} represents the HI values obtained by training signals. $min(\cdot)$ selects the minimum value of all the HI_{train} values, and *std* represents the standard deviation. *C* is the constant that reflects the confidence of the result. In this paper, the *C* is set as 3.

3.3. Dimension Reduction of the Raw AE signals

The high sampling frequency of AE signals results in a large amount of acquired data, which increases the computational burden of the model. Therefore, a Moving Variance Window (MVW) is applied to the raw AE signals for dimension reduction, as illustrated in Figure 6. With the step-wise shifting of the window, the variance of the covered signal is calculated, i.e., each windowed data segment is transformed into a single value of the variance. The output is a dimensionless number, which measures the dispersion of the data; thereby, the sub-signal is de-dimensionalized. The function of MVW is to capture the transient events and highlight some essential detailed features of the data. Moreover, since the signal dimension is vastly reduced, processing by the neural network is faster and easier. The MVW is mathematically described as follows:

$$\hat{X} = \frac{1}{l} \sum_{x \in win_{k,l,s}^{X}} |x - \mu|^2 \qquad (19)$$

$$win_{k,l,s}^{X} = X\left[x_{start}^{k,s} : x_{start}^{k,s} + l - 1\right] \qquad (20)$$

$$x_{start}^{k,s} = (k-1) \times s + 1, \ \ k = 1, 2, \ldots n_k \qquad (21)$$

where *X* denotes the raw AE signal, $win_{k,l,s}^{X}$ represents the area of signal *X* covered by the moving window, and *k*, *l* and *s* are integers specifying the moving step, window length and moving stride, as illustrated in Figure 6. $x_{start}^{k,s}$ is the start point of the window on signal *X*. The total number of moving steps is computed as $n_k = [(N - l)/s] + 1$, where *N* is the length of the recorded AE signal; μ in definition (19) denotes the mean of $win_{k,l,s}^{X}$. The MVW applies a moving window slide over the original AE signal to extract the variance; thus, the signal dimension can be largely reduced, which makes it easier to be processed by the neural network. Additionally, the MVW helps to capture the transient events and highlight some important detailed information in the data.
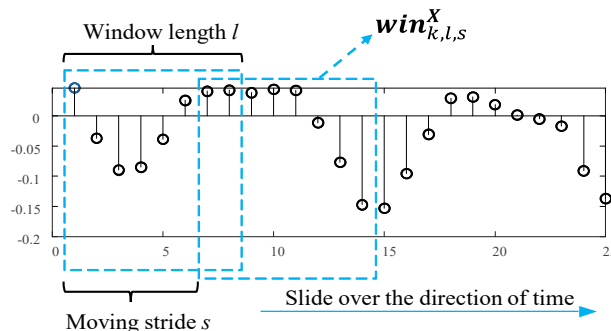
Figure 6. Illustration of the moving window $win_{k,l,s}^{X}$.

3.4. Overall Framework

With the pre-processing of dimensional reduction, the original high-dimensional AE signals can be processed by the proposed ensembled convolutional GAN. The general procedures of the proposed EFD framework are summarized as follows.

Step 1: Data acquisition. AE signals are acquired at fixed time intervals from sensors mounted on the test machine. The signals received at the initial health stage of the exper-

3.4. Overall Framework

With the pre-processing of dimensional reduction, the original high-dimensional AE signals can be processed by the proposed ensembled convolutional GAN. The general procedures of the proposed EFD framework are summarized as follows.

Step 1: Data acquisition. AE signals are acquired at fixed time intervals from sensors mounted on the test machine. The signals received at the initial health stage of the experiment are treated as the training set, and the remaining serve as the test set.

Step 2: Dimension reduction The MVW is first applied to the raw signals to reduce the dimension.

Step 3: Model set-up.

(1) Offline training stage: the pre-processed training data are fed into the convolutional GAN. During the training phase, the history states are recorded at fixed training epochs. Thus, N base models are obtained.

(2) Setup threshold: the training set is fed into the Discriminator only, and N scores are generated by the base models for each sample. The *EHI* and the threshold are calculated by Equations (17) and (18).

Step 4: Online test stage. The samples in the test set are sequentially fed into the Discriminator to calculate the *EHI*. The *EHI* values exceeding the threshold are considered a fault alarm.
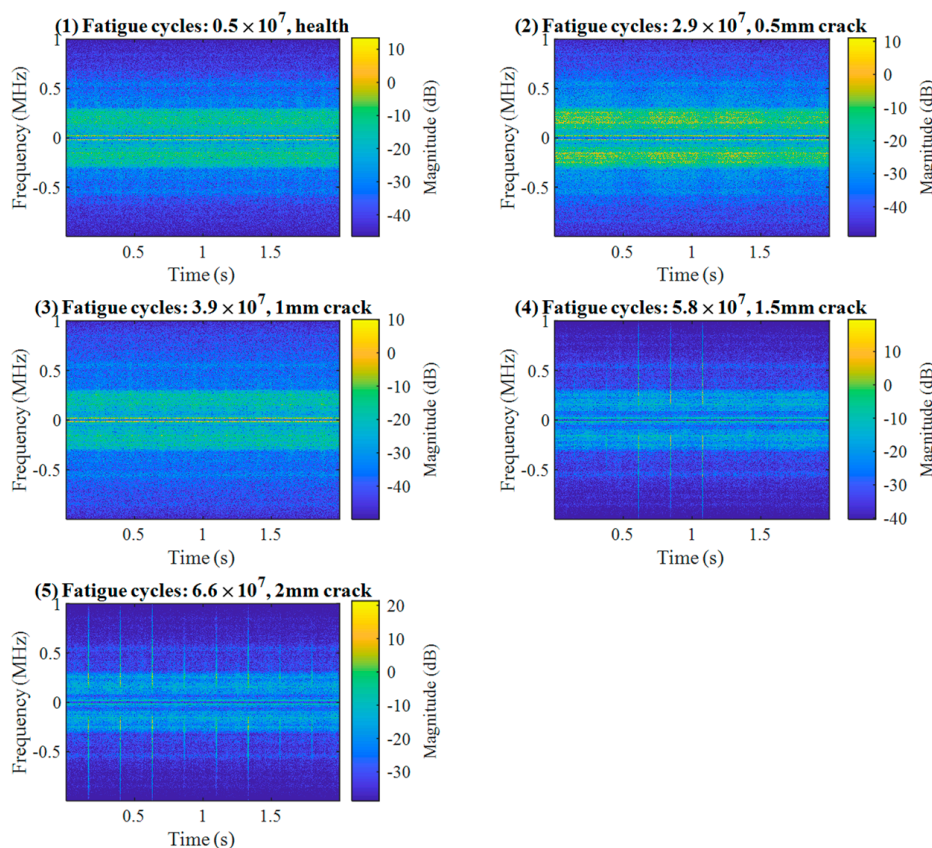
4. Experimental Validation

4.1. Test Rig and AE Data Acquisition

One can find more details of the experimental setup and durability test in [17]. To evaluate the performance of the proposed method, a rolling contact fatigue test was carried out in this section. The test rig, designed at SINTEF Industry (Trondheim, Norway), consists of four roller bearings, as illustrated in Figure 7a. The test specimen is in the central position, supported by another three rollers. Each roller is supported by two needle bearings (type SKF NA 6914-zw). To monitor damage associated with the rolling contact fatigue, the broadband WD (MISTRAS, Princeton, NJ, USA) sensors were mounted on the housing of the needle bearing supporting the test roller. A close-up view of the sensors and their location on the rig is presented in Figure 7b. The streaming AE signals were recorded periodically at fixed time intervals, and each data file contains 2 s of streaming AE waveforms sampled at 2 MHz using the Kongsberg HSIO-100-A (Kongsberg Maritime, Trondheim, Norway) high-speed acquisition module. At the beginning of the test, the recording time interval was set at 60 min. When the first damage was confirmed by periodic ultrasonic inspections of the test roller, the recording time interval was reduced to 20 min to obtain more AE realizations containing information about the faults. At the end of the experiment, 2471 AE records were qualified for the analysis. Figure 8 displays the amplitude of the raw AE signal against contact fatigue cycles. An appreciable change in the AE amplitude is observed for the first time after 4.6×10^7 fatigue cycles.

(a)   **(b)**

Figure 7. Display of the rolling fatigue test rig. (**a**) Photographic image and schematics of the geometry of supporting rollers and the testing roller, and (**b**) a close-up view of the setup instrumented with sensors.

Figure 8. Display the raw AE signals according to the fatigue cycles.

To monitor and visualize the development of incipient damage, ultrasonic inspections were performed periodically during the test using the Olympus OMNISCAN SX (Olympus, Tokyo, Japan) and a stationary ultrasonic phased array testing unit (PAUT). According to the PAUT results, the health condition of the test spot then was split into five stages, as indicated by different colors in Figure 8. The typical AE records that were present at these five stages were presented in Figure 9. It can be seen that at the early stage of the damage (corresponding to 0.5 mm and 1 mm length of the integral crack), are still invisible to the naked eye. As the faults grow up to 1.5 mm, some high-frequency components induced by AE bursts emerge gradually, and the number of AE bursts increases with the damage development.

Table 1. The number of AE records for different stages of damage propagation.

Health Condition	Number of Records	Number of Fatigue Cycles
No damage	542	2.8×10^7
0.5 mm crack	377	3.6×10^7
1 mm crack	809	4.8×10^7
1.5 mm crack	718	6.5×10^7
2 mm crack	25	6.6×10^7

Figure 9. Representative time–frequency spectral decompositions of AE signals corresponding to different stages of damage propagation.

4.2. Results and Discussion

4.2.1. Data Preprocessing

Original high-dimensional AE signals are firstly processed by MVW with the moving step k of 8192. The window length l and moving stride s are set as 464. Therefore, each AE file is downsized to a shorter vector with a length of 8192. Let us recall the structure of the Generator, where two LSTM cells are utilized to capture the time-series correlations in the data. To fit the Generator, each input datum is divided into N segments with length M, as described in Figure 4. We recommend that each segment should contain information about at least one entire axel revolution. In this way, LSTMs can capture the correlation of AE signals generated in continuous axel revolutions. For instance, the lowest axel rotation frequency in the present work is 254 rpm, i.e., for a 2-s recording, 8 complete rotations are captured. Therefore, the segment parameters N and M are defined as 8 and 1024, respectively.

4.2.2. Network Training

The proposed method was implemented with the open-source PyTorch machine learning framework. The detailed architectures of the proposed Generator and Discriminator are described in Table 2. The first 60% of healthy data (325 recorded AE signals) are used for training the convolutional GAN, and the EHI of each training sample is calculated by Equation (17). Then, a fault alarm threshold can be obtained according to Equation (18).

Table 2. The detailed parameters of the proposed Generator and Discriminator.

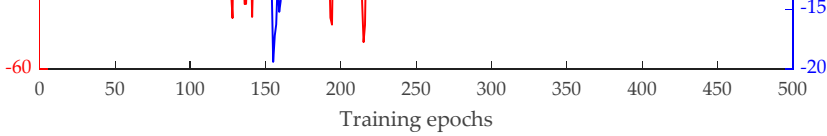| Generator		Discriminator	
Layers	Output size	Layers	Output size
Input	1024@1	*Input*	8192@1
$[C:25, S:1, P:12]$	1024@16	$[C:125, S:1, P:63]$	8192@8
$\begin{bmatrix} A:2, S:2, P:0 \\ Batchnorm \end{bmatrix}$	512@16	$[A:2, S:2, P:0]$	4096@8
$[C:5, S:15, P:2]$	512@1	$[C:25, S:1, P:12]$	4096@16
$\begin{bmatrix} A:2, S:2, P:0 \\ Batchnorm \end{bmatrix}$	256@1	$[A:2, S:2, P:0]$	2048@16
$LSTM \times 2$	256@1	$[C:25, S:1, P:12]$	2048@1
$[C:5, S:1, P:2]$	256@16	$[A:2, S:2, P:0]$	1024@1
$\begin{bmatrix} Batchnorm \\ Up:2 \end{bmatrix}$	512@16	FC	512@1
$[C:5, S:1, P:2]$	512@1	$FC:1$	
$[Batchnorm]$	1024@1		
$FC:1024$	1024@1		

Note: C denotes the convolutional kernel, A denotes the average pooling kernel, S and P are the stride and padding of each kernel, respectively. FC stands for the fully connected layer. The output size is denoted by '$a@b$', where a represents the length of the output vector, and b is the number of output channels.

The network was trained for 500 epochs, and the loss values of the Generator and Discriminator are shown in Figure 10. It can be observed that these loss values oscillate, indicating that both the Generator and Discriminator concurrently attempt to improve their individual capacity during training. To implement the HSE method, two ensemble parameters H and J need to be preset. In this section, the values for H and J are set at 300 and 20, respectively, and a total of 10 base models are obtained. The output layers of each base model are all fully connected layers (16). The averaging voting is applied to the ensemble and the results according to the Equation (17). The output generated by the proposed method is shown in Figure 11.
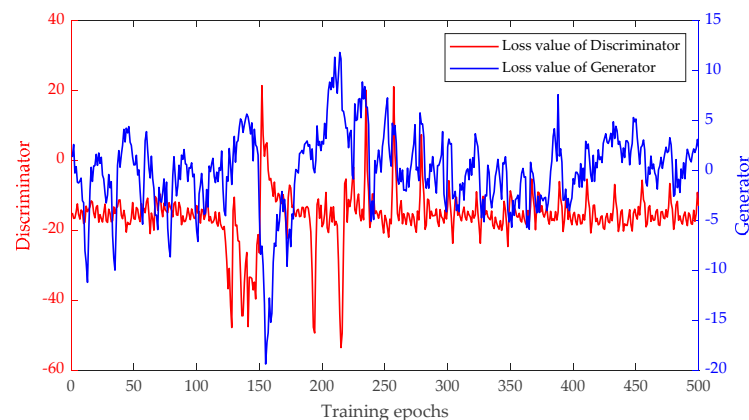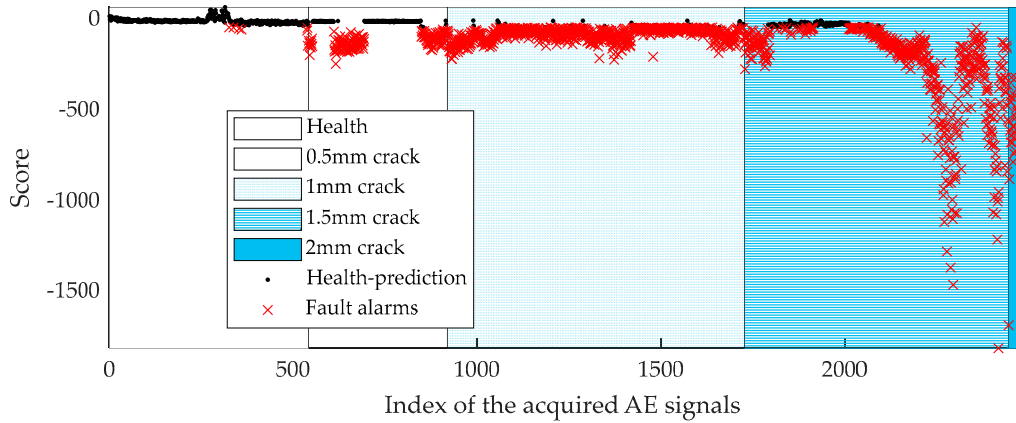

Figure 10. Loss values of the Generator and Discriminator during training.
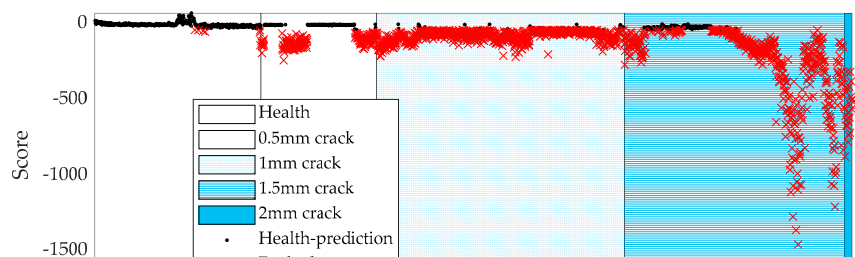
Figure 10. Loss values of the Generator and Discriminator during training.

Figure 11. The generated *EHI* of the acquired AE signals.

We assess the performance of HIs from the following aspects: (1) the ability to reflect the breakpoint between the healthy stage and the onset of defects; (2) the ability to characterize the waveform change of the signal; since failure is an irreversible process, (3) HI is expected to be continuous and monotonic. Figure 11 shows the obtained *EHI* of all of the 2471 AE data files. Observations show that the *EHI* successfully characterizes the evolution of the recorded AE waveforms from the following aspects. Firstly, the *EHI* value shows rapid growth at the stage of the 0.5 mm crack, and a breakpoint between the healthy and fault stage is easily observed. Secondly, the IEPF value captures the initiation of the continuous AE transient bursts at the intersection of the 1 mm crack and 1.5 mm crack. Additionally, the *EHI* presents excellent monotonicity.

In order to evaluate the proposed method and highlight its superiority over existing conventional procedures, the following techniques are introduced and compared with each other.

Statistical parameters: (1) Mean; (2) Variance; (3) Root Mean Square (RMS); (4) Skewness; (5) Kurtosis; (6) Shape Factor; (7) Crest Factor; (8) Impulse Factor; (9) Margin Factor; (10) Information Entropy (IE); (11) Energy Entropy; (12) Mean Frequency (MeanFreq); (13) RMS Frequency (RMSF); (14) Root Variance Frequency (RVF); (15) Median Frequency (MedFreq).

Machine learning models: (16) One-Class SVM (17) Local Outlier Factor (LOF); (18) Isolation Forest (iForest); (19) Autoencoder.

The samples are fed into the probed models sequentially. The streaming accuracy (SA) is used as a metric to quantify the performance of each model, which is expressed as

$$SA_t = n_t / N_t \tag{22}$$

where N_t denotes the total number of samples from the start to time t; n_t denotes the number of the correctly classified samples until time t. SA_t shows the performance of the probed methods against the acquisition time of each AE signal, and their results are compared in Figure 12, while the average accuracy is plotted in Figure 13. It can be seen that the superiority of the proposed method becomes gradually more and more obvious during the 1 mm crack growth stage, where the developed ensembled convolutional GAN shows the highest average accuracy among all probed contenders.

where N_t denotes the total number of samples from the start to time t; n_t denotes the number of the correctly classified samples until time t. SA_t shows the performance of the probed methods against the acquisition time of each AE signal, and their results are compared in Figure 12, while the average accuracy is plotted in Figure 13. It can be seen that the superiority of the proposed method becomes gradually more and more obvious during the 1 mm crack growth stage, where the developed ensembled convolutional GAN shows the highest average accuracy among all probed contenders.
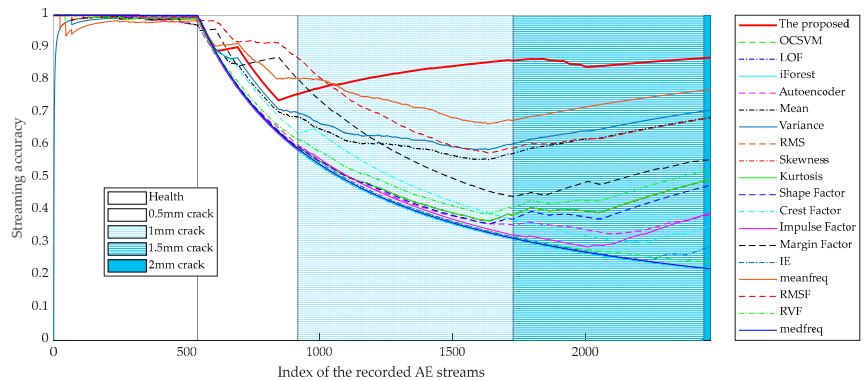
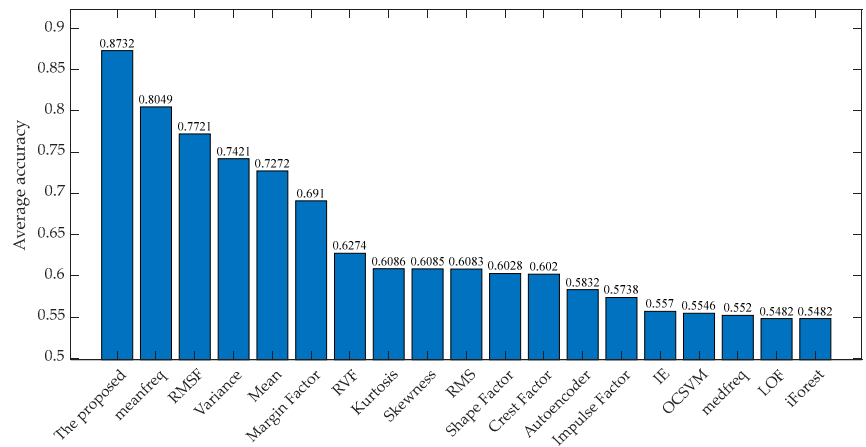**Figure 12.** The streaming accuracy of the probed methods.

Figure 13. The average accuracy of the probed methods.

Finally, we investigate the influence of the ensemble parameters on the model performance. The ensemble model performance is impacted by the two ensemble parameters θ_1 and θ_2, and the model training epoch S. To facilitate the analysis, we set the parameter θ_2 at 1, and examine the influence of θ_1 and the training epochs at fixed base model numbers. With fixed θ_2 and training epochs, the ensemble parameter θ_1 is defined as $S - n$, where S denotes the current training epoch, and n represents the base model number. Figure 14 shows the experimental results with different training epochs ranging from 50 to 500. The black line denotes the average accuracy of the convolutional GAN, while the red lines show the ensembled model accuracy with different base model numbers ranging from 5 to 30. Figure 15 represents the result of the quantitative analysis of the accuracy improvement of the model after applying the HSE procedure. Based on the observation of Figures 14 and 15, the following conclusions can be drawn.

(1) The HSE effectively reduces the accuracy fluctuation of a single GAN under different training epochs.

(2) One can observe that the accuracy of the ensembled model remains at the upper bound of the single GAN, indicating that the proposed method can improve the performance of convolutional GAN in a general sense.

(3) With a smaller number of training epochs ranging from 50 to 300, the HSE procedure can effectively improve the model accuracy by an average of 10% or more, as shown in Figure 15. It also indicates that the proposed ensemble method can improve
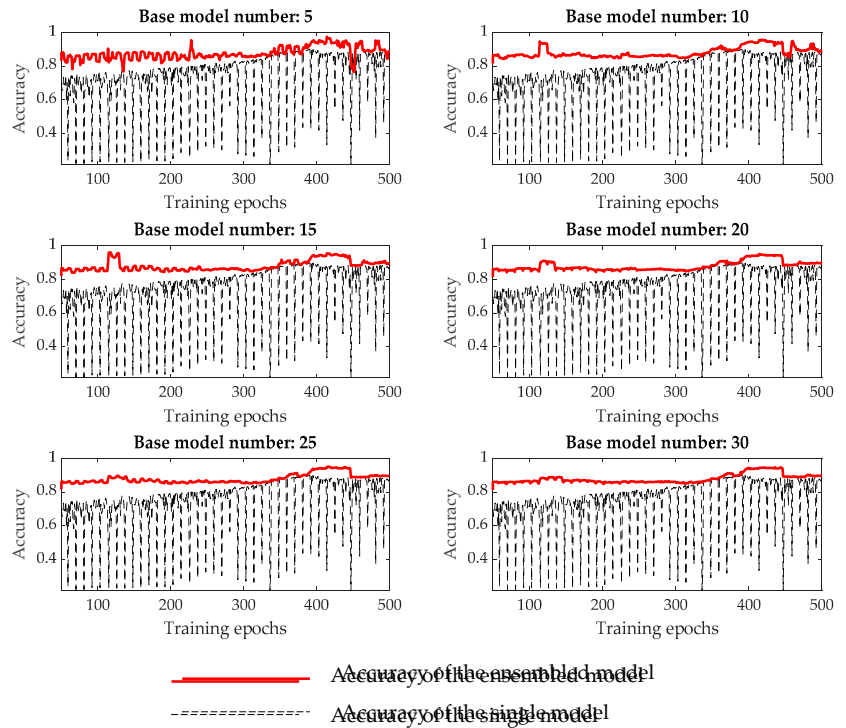
Figure 14. Comparison analysis of the ensembled and single convolutional GAN.
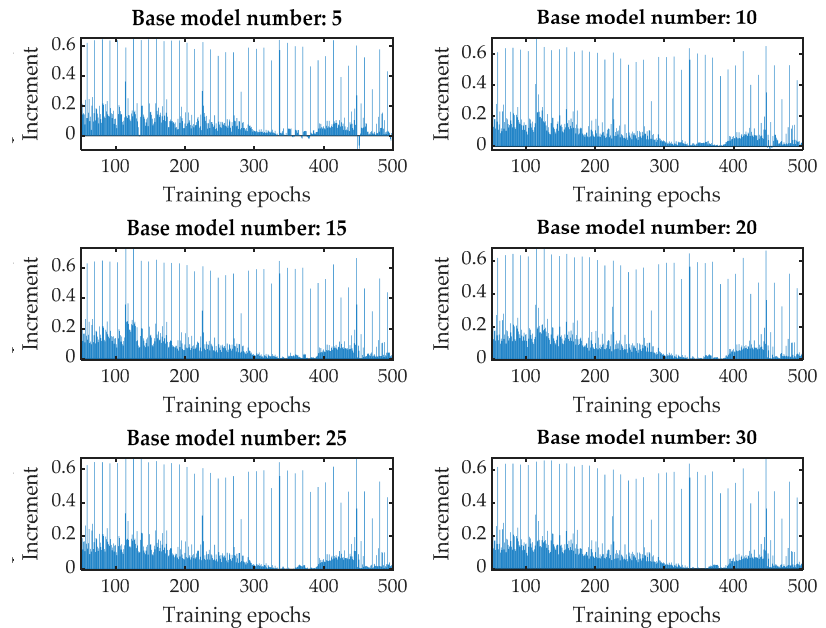
Figure 15. Quantitative analysis of increased accuracy after application of HSE.

(1) The HSE effectively reduces the accuracy fluctuation of a single GAN under different training epochs.

(2) One can observe that the accuracy of the ensembled model remains at the upper bound of the single GAN, indicating that the proposed method can improve the performance of convolutional GAN in a general sense.

(3) With a smaller number of training epochs ranging from 50 to 300, the HSE method can effectively improve the model accuracy by an average of 10% or more, as shown in Figure 15. It also indicates that the proposed ensemble method can improve the model efficiency with a smaller training budget

(4) The model performance is directly related to the number of base models, i.e., the robustness of the ensembled model increases with the number of base models.

In summary, we have demonstrated that HSE improves the overall performance of the single GAN in terms of model stability and accuracy.

5. Conclusions

The present paper extends our previous efforts to study and promote the HSE method to non-destructive testing applications: here, we tested its effectiveness for early fault diagnostics in rotating machinery. The main findings can be summarized as follows.

(1) A new convolutional GAN is designed in this paper and applied for EFD in the run-to-failure test of a roller bearing. To boost the learning capacity of the Generator, an Autoencoder-based Generator architecture is designed in this work. Two convolutional blocks are used to extract the local information of the data, and the Long Short-Term Memory (LSTM) cells are embedded in the bottleneck layer to extract the time-series correlation of the signal.

(2) A novel HSE method is introduced in the designed convolutional GAN to establish an ensembled health indicator (EHI). The proposed ensembled convolutional GAN is combined with the AE technique. There have been limitations in the use of AE technology for condition monitoring, partly due to challenges with processing a large amount of data; thus, a smoothing Moving Variance Window (MVW) is used in this work to reduce the dimensions of the raw AE signal.

(3) We demonstrate the effectiveness of the HSE method when applied to GAN and EFD problems. Roller fatigue test monitoring by AE sensors was performed to evaluate the proposed method. Experimental results demonstrate the effectiveness of the proposed method.

(4) The HSE-based approach benefits from the fact that (i) it does not require extra training costs to generate multiple base models, and (ii) it can be applied to all types of neural networks without tuning the network architecture. Experimental results indicate that not only does the HSE method improve the diagnostics of incipient flaws in specific rolling bearing elements under contact fatigue conditions, but it is also an efficient vehicle to enhance the performance and capacity of convolutional GAN in a general sense.

Author Contributions: Conceptualization, methodology, code, and writing—original draft preparation, Y.W.; supervision, experiment and data acquisition, writing—review, A.V. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this paper can be requested from the corresponding author.

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

References

1. Sun, Y.; Wang, J.; Wang, X. Fault Diagnosis of Mechanical Equipment in High Energy Consumption Industries in China: A Review. *Mech. Syst. Signal Process.* **2023**, *186*, 109833. [CrossRef]
2. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
3. Zhou, L.; Wang, P.; Zhang, C.; Qu, X.; Gao, C.; Xie, Y. Multi-Mode Fusion BP Neural Network Model with Vibration and Acoustic Emission Signals for Process Pipeline Crack Location. *Ocean Eng.* **2022**, *264*, 112384. [CrossRef]
4. Zhang, Y.; Xing, K.; Bai, R.; Sun, D.; Meng, Z. An Enhanced Convolutional Neural Network for Bearing Fault Diagnosis Based on Time–Frequency Image. *Measurement* **2020**, *157*, 107667. [CrossRef]
5. Yu, J.; Zhou, X. One-Dimensional Residual Convolutional Autoencoder Based Feature Learning for Gearbox Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6347–6358. [CrossRef]
6. Zhang, Y.; Li, X.; Gao, L.; Chen, W.; Li, P. Intelligent Fault Diagnosis of Rotating Machinery Using a New Ensemble Deep Auto-Encoder Method. *Measurement* **2020**, *151*, 107232. [CrossRef]
7. Yang, S.; Wang, Y.; Li, C. Wind Turbine Gearbox Fault Diagnosis Based on an Improved Supervised Autoencoder Using Vibration and Motor Current Signals. *Meas. Sci. Technol.* **2021**, *32*, 114003. [CrossRef]
8. Wang, Y.; Du, X.; Lu, Z.; Duan, Q.; Wu, J. Improved LSTM-Based Time-Series Anomaly Detection in Rail Transit Operation Environments. *IEEE Trans. Ind. Inform.* **2022**, *18*, 9027–9036. [CrossRef]
9. Ding, Y.; Jia, M.; Miao, Q.; Cao, Y. A Novel Time–Frequency Transformer Based on Self–Attention Mechanism and Its Application in Fault Diagnosis of Rolling Bearings. *Mech. Syst. Signal Process.* **2022**, *168*, 108616. [CrossRef]
10. Tan, C.K.; Irving, P.; Mba, D. A Comparative Experimental Study on the Diagnostic and Prognostic Capabilities of Acoustics Emission, Vibration and Spectrometric Oil Analysis for Spur Gears. *Mech. Syst. Signal Process.* **2007**, *21*, 208–233. [CrossRef]
11. Yoshioka, T.; Fujiwara, T. Application of Acoustic Emission Technique to Detection of Rolling Bearing Failure. *Am. Soc. Mech. Eng.* **1984**, *14*, 55–76.
12. Al-Ghamd, A.M.; Mba, D. A Comparative Experimental Study on the Use of Acoustic Emission and Vibration Analysis for Bearing Defect Identification and Estimation of Defect Size. *Mech. Syst. Signal Process.* **2006**, *20*, 1537–1571. [CrossRef]
13. Caso, E.; Fernandez-del-Rincon, A.; Garcia, P.; Iglesias, M.; Viadero, F. Monitoring of Misalignment in Low Speed Geared Shafts with Acoustic Emission Sensors. *Appl. Acoust.* **2020**, *159*, 107092. [CrossRef]
14. Motahari-Nezhad, M.; Jafari, S.M. Bearing Remaining Useful Life Prediction under Starved Lubricating Condition Using Time Domain Acoustic Emission Signal Processing. *Expert Syst. Appl.* **2021**, *168*, 114391. [CrossRef]
15. AlShorman, O.; Alkahatni, F.; Masadeh, M.; Irfan, M.; Glowacz, A.; Althobiani, F.; Kozik, J.; Glowacz, W. Sounds and Acoustic Emission-Based Early Fault Diagnosis of Induction Motor: A Review Study. *Adv. Mech. Eng.* **2021**, *13*, 1687814021996915. [CrossRef]
16. Hou, D.; Qi, H.; Li, D.; Wang, C.; Han, D.; Luo, H.; Peng, C. High-Speed Train Wheel Set Bearing Fault Diagnosis and Prognostics: Research on Acoustic Emission Detection Mechanism. *Mech. Syst. Signal Process.* **2022**, *179*, 109325. [CrossRef]
17. Hidle, E.L.; Hestmo, R.H.; Adsen, O.S.; Lange, H.; Vinogradov, A. Early Detection of Subsurface Fatigue Cracks in Rolling Element Bearings by the Knowledge-Based Analysis of Acoustic Emission. *Sensors* **2022**, *22*, 5187. [CrossRef] [PubMed]
18. Ma, Z.; Zhao, M.; Luo, M.; Gou, C.; Xu, G. An Integrated Monitoring Scheme for Wind Turbine Main Bearing Using Acoustic Emission. *Signal Process.* **2023**, *205*, 108867. [CrossRef]
19. Liu, W.; Rong, Y.; Zhang, G.; Huang, Y. A Novel Method for Extracting Mutation Points of Acoustic Emission Signals Based on Cosine Similarity. *Mech. Syst. Signal Process.* **2023**, *184*, 109724. [CrossRef]
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., Eds.; Curran: Red Hook, NY, USA, 2014; Volume 27, Available online: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b1 22f61f8f06494c97b1afccf3-Paper.pdf (accessed on 8 February 2023).
21. Zhang, W.; Li, X.; Jia, X.-D.; Ma, H.; Luo, Z.; Li, X. Machinery Fault Diagnosis with Imbalanced Data Using Deep Generative Adversarial Networks. *Measurement* **2020**, *152*, 107377. [CrossRef]
22. Wang, Y.; Sun, G.; Jin, Q. Imbalanced Sample Fault Diagnosis of Rotating Machinery Using Conditional Variational Auto-Encoder Generative Adversarial Network. *Appl. Soft Comput.* **2020**, *92*, 106333. [CrossRef]
23. Zhang, T.; Chen, J.; Li, F.; Pan, T.; He, S. A Small Sample Focused Intelligent Fault Diagnosis Scheme of Machines via Multi-Modules Learning with Gradient Penalized Generative Adversarial Networks. *IEEE Trans. Ind. Electron.* **2020**, *68*, 10130–10141. [CrossRef]
24. Yin, H.; Li, Z.; Zuo, J.; Liu, H.; Yang, K.; Li, F. Wasserstein Generative Adversarial Network and Convolutional Neural Network (WG-CNN) for Bearing Fault Diagnosis. *Math. Probl. Eng.* **2020**, *2020*, 2604191. [CrossRef]

25. Luo, J.; Huang, J.; Li, H. A Case Study of Conditional Deep Convolutional Generative Adversarial Networks in Machine Fault Diagnosis. *J. Intell. Manuf.* **2020**, *32*, 407–425. [CrossRef]

26. Liang, P.; Deng, C.; Wu, J.; Yang, Z. Intelligent Fault Diagnosis of Rotating Machinery via Wavelet Transform, Generative Adversarial Nets and Convolutional Neural Network. *Measurement* **2020**, *159*, 107768. [CrossRef]

27. Xia, X.; Pan, X.; Li, N.; He, X.; Ma, L.; Zhang, X.; Ding, N. GAN-Based Anomaly Detection: A Review. *Neurocomputing* **2022**, *493*, 497–535. [CrossRef]

28. Mao, J.; Wang, H.; Spencer, B.F. Toward Data Anomaly Detection for Automated Structural Health Monitoring: Exploiting Generative Adversarial Nets and Autoencoders. *Struct. Health Monit.* **2021**, *20*, 1609–1626. [CrossRef]

29. Wang, Y.; Vinogradov, A. Simple Is Good: Investigation of History-State Ensemble Deep Neural Networks and Their Validation on Rotating Machinery Fault Diagnosis. 2022. Available online: https://ssrn.com/abstract=4278481 (accessed on 8 February 2023).

30. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Generative Adversarial Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017.

31. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved Training of Wasserstein GANs. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017.

32. Thévenaz, P.; Blu, T.; Unser, M. Image Interpolation and Resampling. Ch. 28. In *Handbook of Medical Image Processing and Analysis*; Bankman, I.N., Ed.; Academic Press: Cambridge, MA, USA, 2000; pp. 465–493. [CrossRef]

Paper D:

Yu Wang, S. Bernat, A. Vinogradov, BC-GAN: A Threshold-Free Framework for Unsupervised Early Fault Detection in Rotating Machinery, 2023. (Under review).

BC-GAN: a threshold-free framework for unsupervised early fault detection in rotating machinery

Yu Wang[1*], Szymon Bernat [2], Alexey Vinogradov[1]

[1]Department of Mechanical and Industrial Engineering,

Norwegian University of Science and Technology – NTNU, Trondheim, 7491, Norway. yuwa@ntnu.no (Y.W.); alexei.vinogradov@ntnu.no (A.V.)

[2]SINTEF Industry, Trondheim, 7465, Norway. Szymon.Bernat@sintef.no (S.B.).

Abstract: Early fault detection (EFD) is a crucial component of proactive maintenance that can prevent expensive downtime, enhance safety, and optimize equipment performance and longevity. Existing limitations of the contemporary EFD approaches frequently include: (1) manually designed features relying on expert skills and knowledge, (2) the vaguely determined pre-set thresholds to distinguish between the faulty and healthy state of the testing object (however, the threshold value can vary notably from task to task), and (3) most methods, which have been originally designed for vibration data, cannot be generalised to other techniques, for example acoustic emission (AE) signals. To address these issues, a novel binary-classification generative adversarial network (BC-GAN) is designed for general EFD problems and applied to several specific datasets acquired from different run-to-failure tests of rotating components. Compared with conventional methods, unsupervised BC-GAN directly outputs the probability that an input belongs to either the "health" or "fault" state of the rotating machine without a priory threshold setting. Experimental results demonstrate the high versatility of the proposed network, which can be applied in various laboratory and industrial settings to both vibration and AE signals.

Key words: Binary-Classification Generative Adversarial Network (BC-GAN), early fault detection, run-to-failure process, unsupervised learning, threshold-free.

1. Introduction

Early diagnosis of defects in rotating machinery is an effective way to avoid downtime and loss of revenue as well as to protect both assets and employees, which has been a hotspot for decades. Fault detection strategies are roughly categorized as model-driven and data-driven methods in abundant literature. The first one is based on physical insight into the machine system, which is represented by a dynamic mathematical model with a set of descriptive parameters. Alternatively, data-driven fault diagnosis models depend on mathematical models that explore the statistical features hidden in the data acquired by the array of monitoring sensors. With the development of artificial intelligence, especially the advent of deep learning, modern industries are seeking techniques that rely less on human experience and prior knowledge. Under this circumstance, data-driven methods have gained increasing attention and become the mainstream in the field (Lei et al., 2020).

Up to date, the vibration-based analysis is the most commonly used technique to monitor the condition of rotating machinery due to its simplicity, robustness and wide availability. However, vibration signals ranging from several Hz to several kHz, which are induced by small surface flaws, especially at the very early stage of their propagation, can be easily masked by the mechanical vibration of the system or may simply not exist in heavy structures until the fault grows up significantly to a scale large enough to cause a measurable vibration. There are some alternative techniques that can compensate the limitations of vibration technique. For example, acoustic emission (AE) signals have different dynamics and are characterized by a much wider frequency range (typically from 20 kHz to 2 MHz) (Z. Liu et al., 2021), which does not overlap significantly with low-frequency mechanical vibration signals caused by imbalance or misalignment (He & Zhang, 2012; Hemmati et al., 2016, 2016). However, there is still a huge black gap in the combination of AE technique with the prevalent deep learning techniques, and parameter-based analysis still plays a dominant role in this field. For these approaches, which work well in vibration data, there may be degradation in the AE signal. We need a general method that is feasible and immune to data source, it helps us to study and compare different condition monitoring techniques under the same standard. Although it is out of the scope of this work, in this paper, we aim to propose a threshold-free early fault detection (EFD) model and verify the superiority of the method.

EFD from the run-to-failure cyclic loading process is a typical unsupervised problem because only the data from the healthy stage are available for the training phase. When tackling such a problem, the crucially important task is to identify unexpected events as early as possible. The problem is therefore generally considered as novelty detection or anomaly detection. A traditional way is to extract a health indicator (HI) that reveals the trend of the equipment degradation process. Thus, the emerging faults can be observed by the changes in HI values, as illustrated in Fig.1(a). The decision is made by applying a threshold marking the boundary for the normal operating state, and the signal is used as a fault indicator if its HI value exceeds the pre-set threshold.
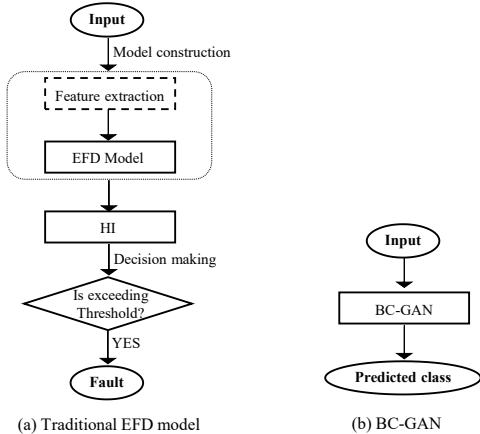
(a) Traditional EFD model   (b) BC-GAN

Fig. 1. Comparison of the framework of traditional EFD models and the proposed BC-GAN.

Although machine learning algorithms have been widely used in EFD problems, the limitations of contemporary approaches still exist, and these are listed as follows. (1) Most of the methods require manually constructed features, and the model performance is heavily dependent on the quality of extracted features (C. Liu & Gryllias, 2020; H. Liu et al., 2018; Xu et al., 2021). Thus, the entire approach relies on the vaguely defined expert's skills, and the risks of misinterpretation are exacerbated when a novice is performing the EFD procedure. (2) The methods that can directly process the raw mechanical signals still require the pre-set thresholds delineating healthy and faulty states (Hidle et al., 2022; C. Liu & Gryllias, 2020; Lu et al., 2018; Luo et al., 2018; Mao et al., 2022). The values of the thresholds can vary for different testing conditions, leaving users with a thorny problem of rigorous threshold setting. (3) Most existing methods are originally designed specifically for the vibration signal and cannot be generalised to other techniques like AE time series (C. Liu & Gryllias, 2020; Lu et al., 2018; Luo et al., 2018; Mao et al., 2022; Song et al., 2022). The motivation of this work is to construct an intelligent unsupervised model that is effective on different data sources, and we specifically validate on vibration and AE signals.

To this end, a novel Binary-Classification Generative Adversarial Network (BC-GAN) is proposed in this paper, aiming to address the EFD problem of the cyclic run-to-failure process. Generative Adversarial Network (GAN) has been shining for addressing the fault diagnosis issues by supplementing the unbalanced dataset with generated data (Gao et al., 2020; Li et al., 2021; J. Liu et al., 2019; S. Liu et al., 2022; Wang et al., 2020; T. Zhang et al., 2020). Although promising results have been reported in the cited works, the approaches used are essentially built upon a supervised learning strategy, whereas the unsupervised approaches are still scarcely probed. To make the method easy to implement, we maintain the architecture of standard GANs to the maximum extent possible: Generator learns the probability distribution of data from the normal operating stage to produce synthetic data, while a specially designed

Discriminator is applied to determine whether an instance belongs to either a synthetic or real distribution. Binary classification is performed between the health and fault states using an unsupervised strategy, i.e., only data from the normal (health) state are required. The keys to the successful training of BC-GAN will be detailed in the next section.

The main contributions of this work are as follows.

(1) A novel and stable GAN-based network architecture has been proposed in this work for EFD problems in rotating machinery, which has high versatility on the application to different condition monitoring techniques.

(2) The proposed method does not require a pre-set threshold. It can directly produce the probability that the measured signal is from the healthy or faulty states, as illustrated in Fig.1(b), which is the main advantage over traditional EFD models.

(3) Validation experiments on benchmark vibration datasets have demonstrated the feasibility and effectiveness of the proposed method. Besides, two durability tests of roller element bearings have been performed using AE technique as the monitor of the failure process. Comparison experiments have demonstrated that the proposed BC-GAN can accurately and timely detect the emergence of faults. The generality of the proposed method to different data source has also been verified by its application to both vibration and acoustic emission signals.

The rest of the paper is organized as follows. The details of the proposed BC-GAN are unfolded in Section 2, and benchmark vibration datasets are used to validate the approach and assess its effectiveness. Two instrumented bearing contact fatigue testing rigs are introduced in Section 3 and the experimental results exemplifying the proposed method with AE signals are presented and discussed. Section 4 summarizes all findings and concludes the present study.
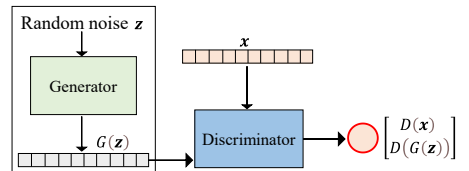

Fig. 2. Schematics of general GAN.

2. Details of the proposed method

Before introducing the proposed method, let us first look at the basic structure of GAN, which comprises two neural networks - a generative model and a discriminative model (Goodfellow et al., 2014), referring to as Generator and Discriminator, respectively, as illustrated in Fig. 2. The Generator takes random noise z as input and attempts to produce the synthetic or fake data with the same dimensionality as the real training samples. On the other hand, the Discriminator takes both real data and the fake data as input, which is trained to judge whether an input originates from the real distribution or the fake distribution
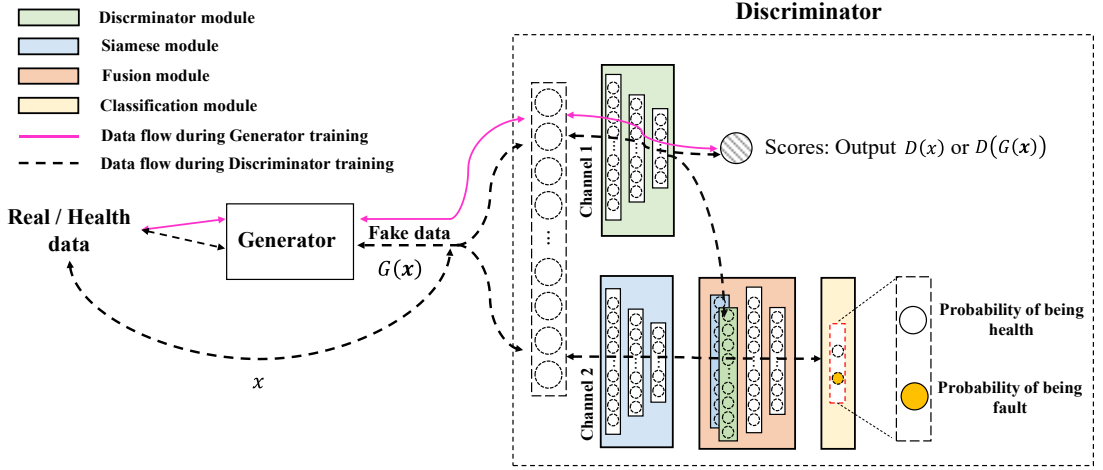
Fig. 3. Schematic of BC-GAN for EFD in run-to-failure process.

produced by the Generator. Both networks are trained simultaneously like a two-player minimax game aiming at obtaining a generative model that can approximate the real data distribution. The loss function is usually expressed as:

$$\min_G \max_D L(D, G) = E_{x \sim p_r(x)}[log D(\boldsymbol{x})]$$
$$+ E_{\boldsymbol{z} \sim p_z(\boldsymbol{z})}\left[log\left(1 - D\big(G(\boldsymbol{z})\big)\right)\right] \tag{1}$$

where $G(\cdot)$ and $D(\cdot)$ represent the Generator and the Discriminator, respectively. $\boldsymbol{x}$ denotes the real dta, $p_r(\boldsymbol{x})$ stands for the real probability distribution of data $\boldsymbol{x}$. $p_z(\boldsymbol{z})$ is generally a normal distribution, and the noise $\boldsymbol{z}$ is randomly sampled as the input of G. The output of $G(\boldsymbol{z})$ is the fake data with the same dimension as $\boldsymbol{x}$ generated by G, and the $D(\boldsymbol{x})$ and $D\big(G(\boldsymbol{z})\big)$ output a score for the real and fake data, respectively.

Based on the traditional EFD scheme in Figure 1(a), the standard GAN can be applied to EFD by introducing the output of Discriminator $D(\boldsymbol{x})$ as HI, as used in (Wang & Vinogradov, 2023), which is employed to indicate the health condition, for example, of the rolling element bearings, by applying a threshold. Although it is theoretically feasible, the uncertainty in thresholding it is an apparent "classic" limitation of conventional EFD models. Hence, the strategic idea and motivation of BC-GAN is to develop a model with higher intelligence eliminating the threshold-related issues. In contrast to many modern methods, the proposed BC-GAN directly produces the probability that the input belongs to either the normal or abnormal state. The details of the proposed framework are elaborated in the following sub-sections.

2.1 Model establishment

The basic architecture of the proposed network configuration is presented in Fig 3. For EFD in the run-to-failure process, only the signals from the healthy stage are available to construct the training dataset. The Generator remains to be used for learning the distribution of the training

data. However, this task differs from traditional generative objectives, such as image generation or tackling imbalanced data issues in fault diagnostics, which focus on developing a strong Generator whereas our goal here is to obtain a 'well-trained' Discriminator. Therefore, to reduce the training difficulty, the Generator takes the health signals as input, which also referred to as real data. The remaining data flow is similar to that in traditional GAN applications, and both real and fake data are sent to the Discriminator for judgment.

The design of the Discriminator is pivotal for the proposed method. As illustrated in Fig 3, the re-designed Discriminator features two channels on its input layer and comprises four neural network modules. It is worth noting that the first channel maintains the original function of traditional GAN, which produces a score for each input sample as described before. By utilising the second channel, the probability of an input instance belonging to either the real or fake class can be acquired. In EFD problems, incorporating the two channels prompts the Discriminator to distinguish between real and fake data through multiple learning objectives, thereby enabling a comprehensive understanding of the differences between them. Detailed network configurations are quite flexible as long as the following four functional modules are included; each module is a multi-layer neural network block.

1) *Discriminator module* $\leftarrow D_M$: receives input data in the first channel. The module is designed to maintain the original function of traditional GAN, and scores are calculated based on the extracted features of its last hidden layer. A linear fully connected layer is attached at the end of D_M to compute the scores as:

$$score(\boldsymbol{x}) = \omega_{score} D_M(\boldsymbol{x}) + b_{score} \tag{3}$$

where $\boldsymbol{x}$ is the input instance of the Discriminator, ω_{score} and b_{score} represent the weight and bias of the linear layer, respectively. D_M assists the training of the Generator to produce the fake data that is close to the real data.

2) *Siamese module* $\leftarrow S_M$: receives input data in the second channel, which is employed to learn the implicit features that distinguish the real and fake data, albeit from a

distinct aspect driven by a different learning objective. It is worth noting that S_M is not involved in the training of Generator and only focuses on classification.

3) *Fusion module* $\leftarrow F_M(D_M, S_M)$: is employed to fuse the features extracted from D_M and S_M, which is designed as a two-channel convolutional neural network block corresponding to D_M and S_M:

$$F_M = f\big(D_M(\boldsymbol{x}), S_M(\boldsymbol{x})\big) \qquad (4)$$

4) *Classification module* $\leftarrow C_M$: receives the fused feature from F_M, and outputs the predicted probability that the input $\boldsymbol{x}$ belongs to either normal or abnormal states. A SoftMax layer is utilized as the classifier, which is expressed as

$$C_M(y_i|\boldsymbol{x}) = \frac{exp\big(\omega_{C_M}^k F_M\big)}{\sum_{j=0}^{1} exp\big(\omega_{C_M}^j F_M\big)} \qquad (5)$$

where $y_i \in [0,1]$ denotes the corresponding label of $\boldsymbol{x}$. Since our target is to differentiate between the normal and abnormal states, which is a binary classification problem. As shown in Fig. 3, C_M comprises two output neurons, which indicate the likelihood of the Discriminator's input instances belonging to healthy and faulty categories, respectively.

Novelty detection: Recall the assumption of the EFD task is that only data from the healthy stage are available at the training phase, which is labelled as 0. To perform the binary classification, the remaining problem is to define the negative training dataset representing abnormal states. The core idea of BC-GAN is to take the fake data produced by Generator as the negative training dataset to train the entire network. Therefore, the network is enforced to dig the implicit features of the data from the normal state and distinguish it from abnormal data. It is worth noting that although the final decision does not depend on the discriminator scores given by D_M, it still plays an important role in keeping the training of the entire network stable. A well-known fact is that GAN is hard to train, by preserving the original function of traditional GAN, it helps us to apply

---

Algorithm 1 Training of BC-GAN using RMSprop stochastic gradient descent. We use default value of RMSprop hyper-parameters: $lr = 0.0001, \alpha = 0.99$.

Input: Training dataset $T = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{c}$; Initial Discriminator weights ω^0; Initial Generator weights φ^0; Initial training hyper-parameters, including, epoch number N, batch size m, critic number n_{critic}, penalty coefficient β, and balance coefficient ρ.
Output: Well-trained ω^N and φ^N.
1: for $epoch = 1, \dots, N$ do
2:     for $batch = 1, \dots, m$ do
3:         **Train Discriminator**
4:         Sample m real data $\boldsymbol{x} \sim T$
5:         Generate fake data $\boldsymbol{x}' \leftarrow G(\boldsymbol{x})$
6:         $L_{score} \leftarrow Eq.(8)$
7:         $L_{classifier} \leftarrow Eq.(9)$
8:         $L_D \leftarrow Eq.(7)$
9:         $\omega \leftarrow \text{RMSprop}(\nabla_\omega E(L_D), \omega, lr, \alpha)$
10:       **Train Generator**
11:       if $rem(batch, n_{critic}) = 1, \dots, m$ do
12:         $L_G \leftarrow Eq.(6)$
13:         $\varphi \leftarrow \text{RMSprop}(\nabla_\varphi E(L_D), \varphi, lr, \alpha)$

---

these new findings that can improve the stability of the training. In addition, the scores can also be served as an auxiliary tool to observe the evolution of streaming data.

2.2 Model training

Up to now, the feedforward process of BC-GAN for EFD problem has been revealed, this sub-section will discuss the optimization of the model. The training details of the BC-GAN is described in **Algorithm 1**. Fig. 3. displays the data flow of the Generator and Discriminator throughout the optimization process. During the feed-forward phase, the Generator creates the fake data, which is then fed along with the real data into the re-designed Discriminator. The Discriminator generates both a score and predicted probabilities for them concurrently. However, it is an important detail that the $L_{classifier}$ shall not be involved during the optimization of the Generator as illustrated by the magenta line in Fig. 3, thus, the loss function of the Generator is expressed as:

$$min\ L_G = -E_{\boldsymbol{x}' \sim p_{\boldsymbol{x}'}}[score(\boldsymbol{x}')] \qquad (6)$$

where $\boldsymbol{x}' = G(\boldsymbol{x})$ denotes the generated fake data, and $\boldsymbol{x}$ represents the real data. The $score(\boldsymbol{x}')$ is defined in equation (3). The objective of training the Generator is to maximize the $score(\boldsymbol{x}')$, which is analogous to minimizing L_G.

As illustrated by the dotted line in Fig. 3, the optimization of the Discriminator consists of two tasks, which are reflected in a combined loss function as:

$$min\ L_D = \rho \times L_{score} + (1-\rho) \times L_{classifier} \qquad (7)$$

where L_{score} denotes the loss function of D_M, which has no difference with general GAN, and the $L_{classifier}$ is the loss function of C_M. The $\rho \in (0,1)$ is a trade-off parameter that balances the discriminator score and classifier. The optimization of BC-GAN is realized by minimizing the combined loss function L. If $\rho = 1$, the model can be regarded as a general GAN.

Let us first define the L_{score}. The original loss function of GAN in Eq. (1) has challenges such as unstable training and poor quality of generated data. The central problem stems from the embedded Jensen-Shannon divergence (JSD) serving as a measure of the dissimilarity between real and generated distributions (see (Arjovsky et al., 2017; Gulrajani et al., 2017) for details). To overcome the training challenges, Arjovsky et al. proposed to replace JSD with Wasserstein distance (WD) (Arjovsky et al., 2017). Introducing the gradient penalty method (Gulrajani et al., 2017), a new objective is expressed as:

$$\begin{aligned} L_{score} &= E_{\boldsymbol{x}' \sim p_{\boldsymbol{x}'}}[score(\boldsymbol{x}')] - E_{\boldsymbol{x} \sim p_r}[score(\boldsymbol{x})] \\ &+ \beta E_{\boldsymbol{x}' \sim p_{\boldsymbol{x}'}}[(\|\nabla_{\boldsymbol{x}'} score(\boldsymbol{x}')\|_2 - 1)^2] \end{aligned} \qquad (8)$$

where $\boldsymbol{x}'$ and $\boldsymbol{x}$ stand for the fake data and real, respectively. β is the penalty coefficient, and GAN with Wasserstein distance and gradient penalty is referred to as gp-WGAN.

The $L_{classifier}$ is presented by the prevalent Mean Square Error (MSE).

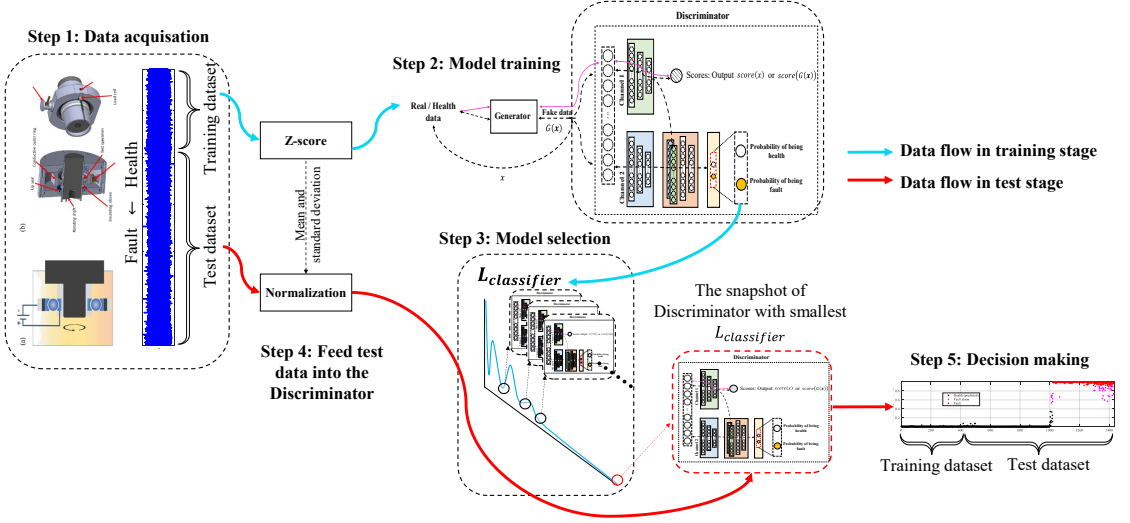$$L_{classifier} = E[(C_M(y_i|\boldsymbol{x}) - y_i)^2] \qquad (9)$$

Fig. 4. Workflow the proposed EFD method.

where $C_M(y_i|x)$ represents the predicted probability that x belongs to class y, and y is a two-dimensional one-hot vector.

Model selection: Determining whether the Discriminator has been adequately trained is a tricky problem. We propose the $L_{classifier}$ as a metrics to evaluate the performance of the Discriminator. A snapshot of the Discriminator with the smallest $L_{classifier}$ is captured as the optimized model for testing which will be elaborated in the following experiments.

2.3 Overall framework

When applying BC-GAN to EFD of bearing elements experiencing contact fatigue damage, the overall framework of the proposed method is illustrated in Fig. 4, which is summarized below.

Step 1: Data acquisition.
(1) Sensors are installed on the test machine, and signal snapshots are recorded at specific time intervals.
(2) The signals acquired in the early stage of the experiment under a normal operating state form the training set, and the remaining signals are for testing.

Step 2: Model training. The acquired training data are normalized using 'z-score' method, and then, feed into the BC-GAN. The model is trained according to **Algorithm 1**.

Step 3: **Model selection**. The Discriminator's snapshot with the smallest $L_{classifier}$ is chosen.

Step 4: **Model test**. The test data are fed into the selected Discriminator, and the predicted probabilities are obtained directly from equation. (5).

Step 5: Decision making. The predicted probabilities represent the level of the predictive confidence of the model. To reveal more information, the condition of an instance is divided into three stages directly based on the predicted

probability given by BC-GAN, as explained in the equation (10):

$$\begin{cases} Health, & P \leq 0.5 \\ Fault\ alarm, & 0.5 < P \leq 0.9 \\ Fault, & P > 0.9 \end{cases} \quad (10)$$

where P represents the predicted probability of an instance being fault. It is worth noting that the division of the model's confidence levels is based on human intervention and does not impact the outcome of the model.

2.4 Validation on benchmark vibration dataset

To validate the effectiveness of BC-GAN, the Intelligent Maintenance System (IMS) database is taken as a benchmark to probe the proposed method. The database refers to four bearings on the shaft. Three run-to-failure tests were performed, and the results were tabulated. At the end of the three experiments, four tested bearings were found to be differently damaged. Fig. 5 shows the original vibration
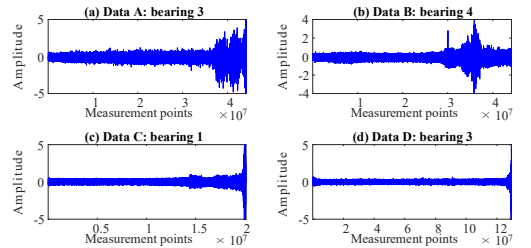

Fig. 5. The raw vibration signal of the tested four bearings. (a) Bearing 3 of test 1: run-to-failure signals ending with an inner race defect. (b) Bearing 4 of test 1: run-to-failure signals ending with a roller element defect. (c) Bearing 1 of test 2: run-to-failure signals ending with an outer race defect. (d) Bearing 3 of test 3: run-to-failure signals ending with an outer race defect.
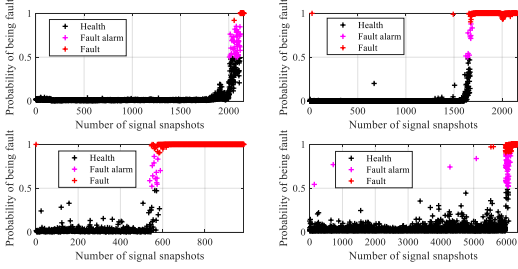
Fig. 6. The classification results using BC-GAN. (a) Bearing 3 of test 1, with training epoch 114. (b) Bearing 4 of test 1, with training epochs 500. (c) Bearing 1 of test 2, with training epochs 450. (d) Bearing 3 of test 3, with training epochs 203.

signals of the four bearings evolving from the healthy to the faulty stage. Each data stream is divided into individual realizations of 1-second vibration signal snapshots recorded at specific time intervals. Thus, each file represents a snapshot containing 20480 readings, and all files relevant to one bearing constitute a pseudo-data stream reflecting the evolution of the vibration signal during the run-to-failure test.

In this work, we use only 8192 points of each file as input to train the model. The first 30% fragment of each data stream is used to train BC-GAN without any pre-processing, and the rest represents the test set. Fig 6 displays the predicted probability given by BC-GAN. The number of training epochs is determined when the loss value $L_{classifier}$ converges to a low level, which will be further discussed in Section 4. The y-axis denoted the probability of being fault. The magenta markers are used to highlight these snapshots with probabilities between 0.5 and 0.9, which means they are highly likely to be faults. When the probability reaches 0.9, the model is confident that this is a fault. The formation of fault is a gradually evolving process, which BC-GAN can reveal by observing the predicted probability. The juxtaposition of our results with the peer studies (Haidong et al., 2020; C. Liu & Gryllias, 2020; Lu et al., 2018; Song et al., 2022) shows that the proposed method successfully detects the occurrence of faults.

3   Experimental verification

In this section, the effectiveness of BC-GAN is validated in bearing elements using AE signals, and two run-to-failure tests of roller element bearings and ball bearings were performed. We introduced two different dilemmas that may be encountered in the EFD problem. In the first case, the breakpoint between health and fault is difficult to be defined, while in the second case, a breakpoint can be easily observed, but the threshold cannot be well-defined using training data as we have mentioned before. We have shown that the threshold-free BC-GAN can easily tackle the EFD task in these cases and achieve the best performance among the probed contrastive methods.

3.1 Dimension reduction of the original AE signals

The high-sampling frequency of AE signal results in a huge amount of data, which increases the computational burden of the model. Therefore, a Moving Variance Window (MVW) is utilized to reduce the dimension of raw AE signals (Wang & Vinogradov, 2023), as defined below:

$$\widehat{X} = \frac{1}{l} \sum_{x \in win_{k,l,s}^X} |x - \mu|^2 \qquad (10)$$

where X denotes the raw AE signal, $win_{k,l,s}^X$ represents the area of the signal X covered by the moving window; k, l, and s are integers specifying the moving step, window length and moving stride, respectively. The MVW procedure applies a moving window sliding over the original AE signal to extract the variance. In this way, the signal dimension can be largely reduced, which makes it easier to be processed by a neural network. Additionally, MVW helps to capture the transient events and highlight important information in the data.

3.2 Contrastive methods

To evaluate the effectiveness of the proposed method, the performance of several conventional HIs of different types are employed as baselines. These HIs are listed below:

Statistical parameters (Wang, Hestmo, et al., 2023): (1) Mean; (2) Variance; (3) Root Mean Square (RMS); (4) Skewness; (5) Kurtosis; (6) Shape factor; (7) Crest factor; (8) Impulse factor; (9) Margin factor; (10) Information Entropy (IE); (11) Energy Entropy; (12) Mean frequency (MeanFreq); (13) RMS Frequency (RMSF); (14) Root Variance Frequency (RVF); (15) Median frequency (MedFreq).

Machine learning methods: (16) one-class SVM (Saari et al., 2019); (17) Local Outlier Factor (LOF); (18) Isolation Forest (iForest) (Mao et al., 2022); (19) SENCForest (Mu et al., 2017); (20) SENNE (Cai et al., 2019); (21) KNNENS (J. Zhang et al., 2022).

Deep learning methods: (22) Deep autoencoder: The HI is constructed from reconstruction error; (23) GAN: The HI is constructed from the output of traditonal Discriminator, which is expressed as $D(X)$ (Wang & Vinogradov, 2023), as illustrated in Fig. 2.

For these models that cannot process the raw mechanical signals directly, the statistical parameters are served as their input. Traditional EFD frameworks identify the occurrence of defects based on a threshold, which should have different values from task to task. Therefore, a heuristic approach is used to define the threshold level for these comparison methods, by adjusting its value to ensure at least 98% of training data can be recognized as "health".

3.3 Roller bearing (RB) test

To monitor the rolling contact fatigue (RCF) phenomenon occurring a roller bearing element, a run-to-failure test was carried out using an instrumented special purpose testing rig
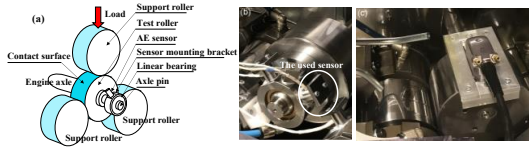
Fig. 7. Rolling fatigue test rig: photographic image and schematics of the geometry of supporting rollers and the testing roller (a), and a close-up view of the setup instrumented with sensors (b), and a view of PAUT inspection performed periodically on the test roller (c) (see (Hidle et al., 2022)).
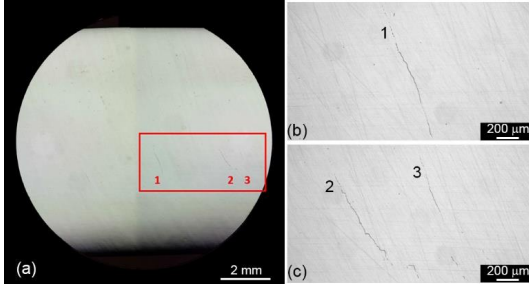

Fig. 8. Subsurface cracks observed in transverse metallographic sections of the test roller (a) and magnified views of three major subsurface contact-fatigue induced cracks marked 1-3 (b,c).

designed at SIN TEF Industry (Trondheim, Norway). The experimental setup is schematically illustrated in Fig 7 (a). The test specimen (central roller) is supported by three rollers, and each roller is supported by two needle bearings. Two broadband WD sensors (MISTRAS, USA) (only one sensor was used for the analysis) were connected to the data acquisition system, as displayed in Fig 7 (b). The signal from the transducer was amplified by 40 dB in the frequency band 20-1200 kHz by the 2/4/6 low-noise preamplifier (MISTRAS, USA).

The test was interrupted periodically for non-destructive inspections performed with the Olympus OMNISCAN SX phase array ultrasonic tester (PAUT) with the Dual Linear Array 7.50L32-REX1-IHC sensor and MX2 control box. The AE waveforms were continuously recorded at 2 MHz sampling frequency for 2s per snapshot using the Kongsberg HSIO-100-A high-speed acquisition module. For details of the experimental setup and the roller durability test itself, the readers are encouraged to read the recent publication (Hidle et al., 2022).

The first sub-surface crack of approximately 0.5 mm was detected by the PAUT at 2.8×10^7 fatigue cycles (see (Hidle,

et al., 2022) for details of PAUT inspections and metallographic observations). The crack then slowly expanded up to 2 mm before the test was terminated, and the appearance of internal cracks was confirmed by the post-mortem destructive metallographic examination unveiling several macroscopic cracks in the transverse cross-section of the tested roller as shown in Fig. 8 illustrating three major cracks in the subsurface area and propagating towards the rolling contact surface. No surface damage has been detected on the test roller of support rollers after the test.

Totally, 2471 AE records were qualified for the analysis and indexed from 1 to 2741 according to their acquisition time. Raw streaming AE data and two typically extracted representative features - root mean square voltage (RMS) and kurtosis of the AE waveform - are presented in Fig. 9 (a-c), respectively (note here that these features are shown as an example; many more statistical features have been extracted and analysed, albeit without much success for potential use of health indicators which motivated us turn to artificial intelligence in data analysis in this case). The appreciable change in the original AE amplitude has been firstly noticed after 4.6×10^7 (or even more reliably only after 5.5×10^7 fatigue cycles or 2100 AE snapshots, when the crack reached the terminal size of 2 mm according to PAUT observations). No apparent breakpoint can be determined by means of traditional statistical features around 2.8×10^7 fatigue cycles corresponding to the 543rd snapshot in Fig. 9 (b) when actual incipient subsurface damage has been unveiled by ultrasonic inspections.

The predicted probability of each data fragment being identified as "fault" is displayed Fig. 10 (a). In the sharp
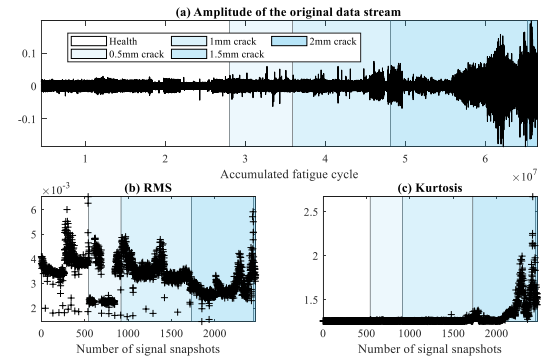

Fig. 9. The acquired run-to-failure AE signals of RB test ending with a sub-surface crack. (a) Original AE signals according to fatigue cycle, and selective features (b) RMS, and (c) kurtosis.

Table 1

Comprehensive classification outcomes of BC-BAN on roller bearing experiment, as well as the chosen comparative models.

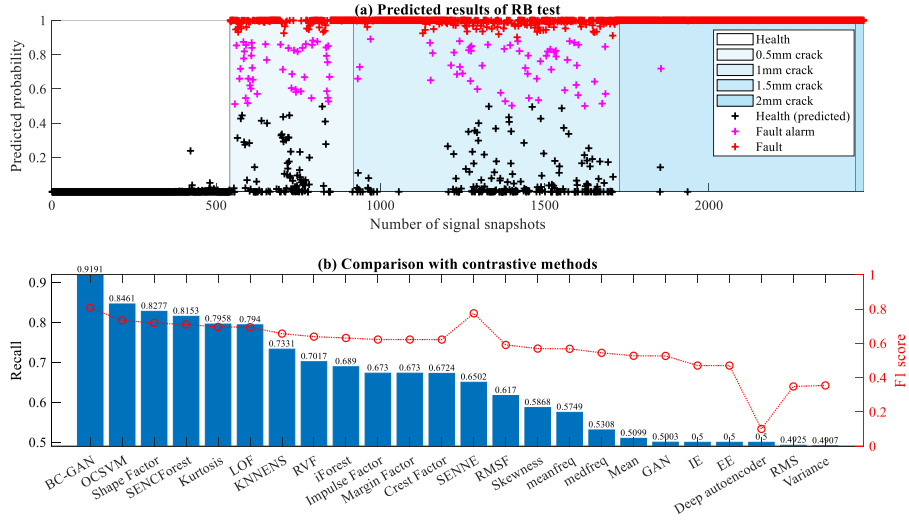| Method	Threshold	Health conditions					False alarm
		Health	0.5mm crack	1mm crack	1.5mm crack	2mm crack	
BC-GAN	/	542	273	604	715	25	0
OCSVM	0.77	531	210	428	712	25	11
Shape factor	1.26	532	124	458	697	25	2
SENCForest	1.66	528	159	387	693	25	14
Kurtosis	0.12	533	99	369	680	25	9
LOF	-2.15	539	260	222	643	25	3
KNNENS	1.6	537	50	187	653	25	5

Fig. 10. Results of RB test. (a) Visualization of the classification result using BC-GAN with 2000 training epochs; (b) Measure of F1 score and recall of the probed methods. Plots are sorted by the descending order of recall.

contrast with the conventional data representation shown above, BC-GAN classified the first input as fault alarm at the 543rd snapshot, i.e., pretty much at the same time as the incipient subsurface crack was detected by PAUT.

The F1 score and recall were used as the performance metrics for the diagnosis system. Using the 543rd snapshot as a breakpoint, the dataset is divided into two classes – 'health' and 'fault' (including fault alarms and faults). The results of the comparison of results obtained by different probed methods are presented in Fig.10 (b). Recall measures the percentage of actual positve/negative signals that are labeled as positive/negative among all corresponding samples. The recall of the proposed method reaches 0.9191 followed by 0.8461 of OCSVM. A higher recall indicates that the proposed BC-GAN is effective for identifying both healthy and fault instances. F1-score balances the importance of precision and recall. The proposed method attains the highest F1-score among all contrastive methods. While SENNE demonstrates the high F1-score for this dataset, its recall is only 0.6502, indicating a considerable number of healthy signals being misidentified as faults.

Table 1 presents the comprehensive classification outcomes of BC-BAN for the roller bearing experiment, as well as the chosen comparative models, based on the top 6 recall values. It provides the number of the accurately classified samples for each bearing health condition, and the thresholds of the comparison methods are also presented. A competent HI is anticipated to yield a higher number of faults and fewer false alarms. Observing the results, it is apparent that BC-GAN provides the highest faults of 273 during the initial falting stage corresponding to the smallest observabale crack of 0.5 mm, followed by LOF at 260. However, the LOF's performance deteriorates as the crack grows to 1 mm. In contrast to comparison methods, BC-GAN delivers the best performance throughout all stages of the fault progression without generating any false alarms.

3.4 Ball bearing (BB) test

The second experiment was designed to enable accelerated subsurface white-etching cracking (WEC) in running ball bearings. The setup is illustrated schematically in Fig.11. The base unit of the test setup comprises an AC variable-speed gear-motor, sealed test chamber, electrically insulated rotating shaft on which the test bearing is mounted, DC power, load cell and AE sensor. The load to the bearing was applied by a screw directly located on the top of the bearing and measured by a load cell. All the metallic parts of the equipment were electrically isolated by using Teflon rings or Teflon sleeves as illustrated in Fig. 11 (b), while the
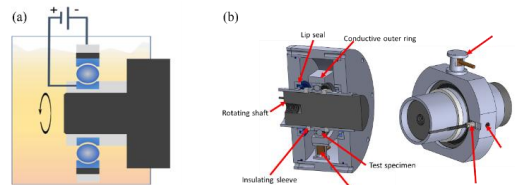

Fig. 11. Schematics of BB test: (a) illustration of the test rig promoting the white etched crack formation, and (b) close-up view of the bearing assembly.
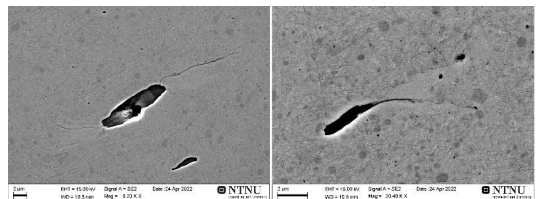

Fig.12. Examples of incipient white etched cracks (WECs) initiated at non-metallic inclusions on the outer raceway of the ball bearing tested (the courtesy of T. A. Nymark and A. B. Haagen).
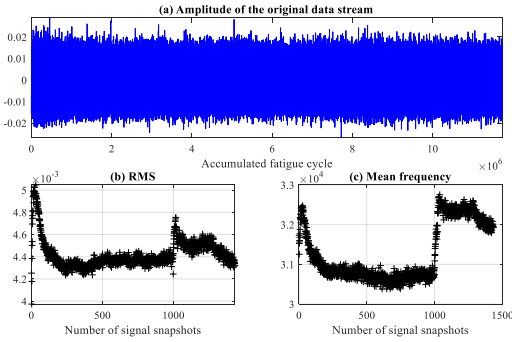
Fig. 13. The acquired run-to-failure AE signals of BB test ending with a crack. (a) Original AE signals over recording time, and selective features (b) RMS, (c) mean frequency.

Table 2
Comprehensive classification outcomes of BC-BAN on ball bearing experiment, as well as the chosen comparative models.

Method	Threshold	Health conditions		False alarm
		Health	Fault	
BC-GAN	/	1008	419	1
LOF	-1.51	969	420	41
SENCForest	1.13	1007	331	3
SENNE	0.1	787	423	223
Medfreq	0.48	1006	261	4
iForest	-0.58	1001	237	9
OCSVM	1.89	940	257	70

bearing was separated by a metallic ring that allowed electricity to flow between the outer and inner race. To promote the formation of white-etched cracks, a plasma discharge was initiated in the lubricating film through the DC current applied between the two bearings. The negative pole was attached to the outer race, while the positive pole was connected to the inner race. LD 75W-80 oil with ZDDP

anti-wear additive was used as the lubricant. The tests were carried out under contact pressures up to 1.9 GPa and 2000 rpm at 50 °C.

When the test was terminated, the inner and outer race of the test bearing were sectioned in the circumferential plane in the middle of the ball track for crack observations by means of the scanning electron microscopy (SEM) in the secondary electron contrast (see (Nymark, 2022) for experimental details and examination results). Multiple microscopic cracks (identified as WECs), like those shown in Fig.12, were readily observed on the sections at the depth of 30-150 μm from the rolling contact surface. The microcracks up to 15 mm initiate primarily from non-metallic inclusions and are surrounded by the so-called white etched matter, which has been investigated in detail in (Nymark, 2022).

The AE waveforms were continuously recorded at 2 MHz sampling frequency for about 1s per record (snapshot), and a total of 1433 records were acquired. The raw noise-like continuous AE signal and a couple of statistical features chosen example are presented in Fig.13 (a-c), respectively. No appreciable change in the AE amplitude can be observed from the original AE signals throughout the test. However, a sharp increase in the RMS value and mean frequency of the AE power spectral density is observed around the 1010th snapshot, which is presumably caused by the emergence of the internal damage development associated with the microcracks illustrated in Fig. 12. Since ultrasonic inspection was not possible in this experiment, the exact time of the first occurrence of the crack is unknown. Based on the analysis of the data stream, the 1010th snapshot is thought as the approximate watershed dividing the healthy and faulty stages of the bearing and the corresponding data for assessing the model performance. The experimental result shown in this example differs from the EFD problem in RB test. In the second case, a breakpoint can be likely observed from traditional parameters, However, it is still challenging
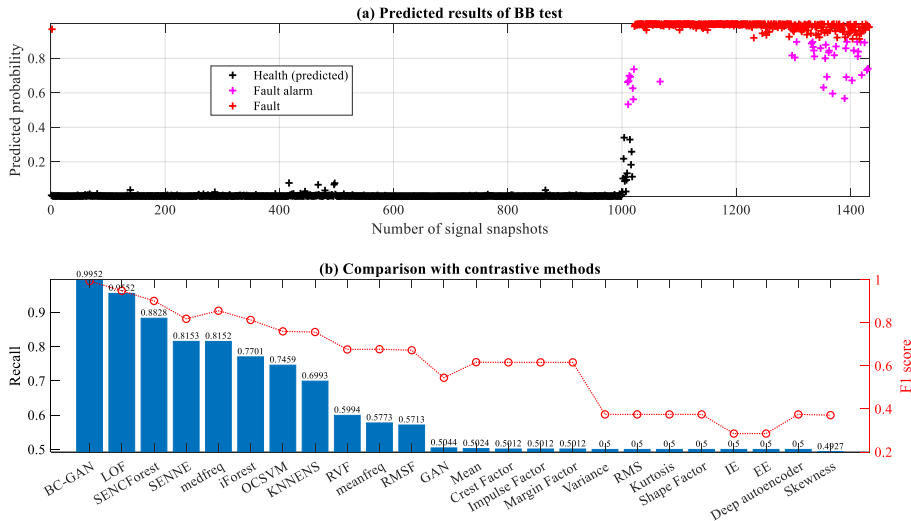

Fig. 14. Results of BB test. (a) Visualization of the predicted result using BC-GAN with 300 training epochs; (b) Measure of F1 score and recall of the probed methods. Plots are sorted by the descending order of recall.
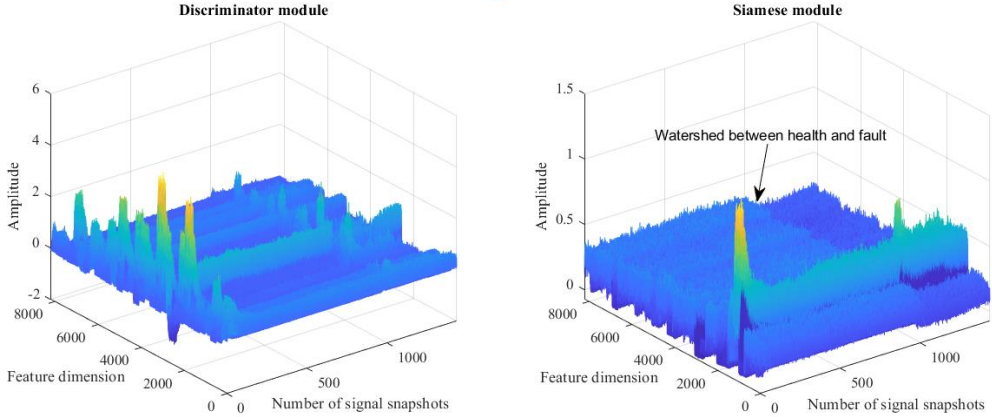
Fig. 16. Visualization of the hidden features extracted from D_M and S_M. Experiment 2 is taken as an example.

to define a proper threshold that can delineate the healthy stage from the onset of the faulty stage in the ball bearing behaviour. Most importantly, we want a roust model that can be generalized and adapted flexibly to notably different situations.

Fig. 14 (a) shows the obtained probability using BC-GAN. The first instance that is classified as "fault alarm" is reported as the 1010th snapshot, and the first "fault" is reported as the 1022nd snapshot which is in excellent agreement with the breakpoint anticipated from the observations of the RMS and AE mean frequency behaviour shown in Fig.13 (b). The recall and F1-score of the probed methods are compared in Fig. 14 (b). The comparison of results reveals that C-GAN exhibits the highest recall of 0.9952, while LOF comes second with the recall value of 0.9552. Furthermore, the BC-GAN has the highest F1-score compared to all other methods.

Table 2 outlines the comprehensive classification outcomes of BC-BAN on the ball bearing experiment, as well as the chosen comparative models based on the top 6 recall values. Despite LOF achieving largest fault alarms of 420, it is also associated with a significantly higher number of false alarms. The BC-GAN achieves a better balance between fault alarms and false fault alarms. Comparing Table 1 and Table 2, it is evident that while the probed methods may exhibit good performance in one task, they tend to experience a decline in performance in another task, which is hard to predict a priory. Furthermore, their optimal threshold settings vary across different tasks. In contrast, BC-GAN delivers exceptional performance in both experiments, with the added advantage of being threshold-free.

4   Discussion and conclusion

In summary, the proposed method has been tested against four vibration benchmark datasets and two original experimental datasets. The discussion and conclusions of the experimental results are presented in this section.
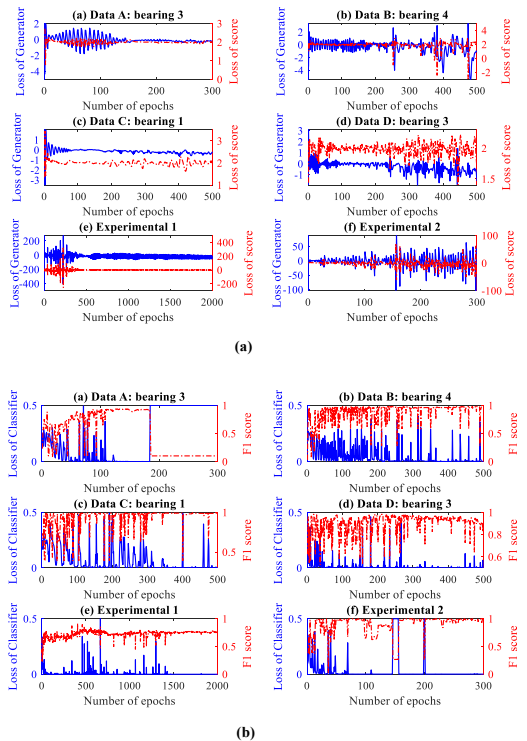
Fig. 15. Observation of loss values. (a) Visualization of L_G and L_{score} over training epochs for different datasets; (b) Visualization of $L_{classifier}$ and accuracy over training epochs for different datasets.

(1) **Loss value and model selection**. The relationship between the model performance and the training loss value is investigated. We first have a look at the L_G and L_{score} utilized in general GAN. Traditionally, it is difficult to judge if the network has been well-trained by observing the loss values of Generator and Discriminator. Unlike traditional regression or classification models, L_G and L_{score} will not

(a) Dataset B

(b) Experiment 2

Fig. 18. Display of the $L_{classifier}$ and F1 score under different values of trade-off parameter ρ ranging from 0-0.9. Dataset B and experiment 2 are taken as an example.

converge to zero with the increasing number of training epochs, as displayed in Fig.15 (a). The oscillating behavior observed in their loss values 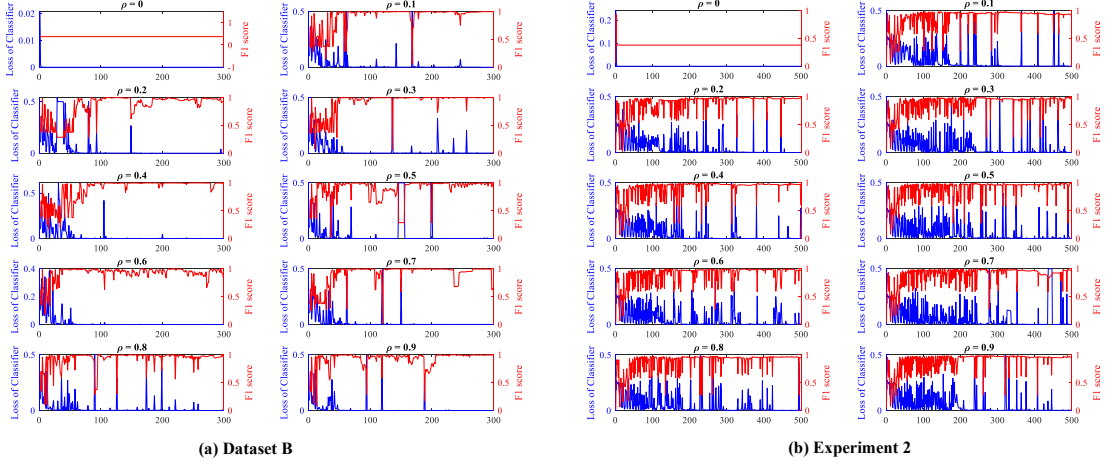corresponds to the fact that the two networks are battling with each other during training. Therefore, determining whether GAN has been well-trained on not is a thorny problem for general GAN. However, BC-GAN provides an indicator that helps to evaluate the model performance during training by introducing the $L_{classifier}$. Fig.15 (b) shows the evolution of $L_{classifier}$ as well as test accuracy over training epochs. One can see that the prediction accuracy is highly correlated with the value of $L_{classifier}$. The training of the model can stop at a sub-optimal point when $L_{classifier}$ converges to a low value.

(2) **Visualisation of the features extracted from D_M and S_M**. In the proposed architecture, the Discriminator module D_M and the Siamese module S_M are two separated neural network blocks. They, however, share essentially the same network architecture. These modules have different initial network weights but receive the same input simultaneously. This design encourages the network to learn

Fig. 17. Comparison of real data and generated signals under different epochs; dataset A is taken as an example.

implicit features from different perspectives. The D_M aids the training of Generator to produce the fake data that is close to the real data, in contrast, S_M is not involved in the Generator training and focuses on classification. The features learned by the two modules are displayed in Fig 16. Observing the results, it is apparent that the two channels extract different features from the input data. However, the features learned by the S_M module appear to be more effective in distinguishing between different health and fault stages.

(3) **Generated data**. In contrast to traditional GAN, BC-GAN is aimed at training Discriminator that captures the implicit features from health data for the EFD problem. Therefore, it does not require that Generator to produce very authentic 'fake data' that can entirely fool Discriminator. In fact, a slight difference between the generated data and the real data helps to produce a more powerful Discriminator in our problem. Fig.17 displays the generated signals at different training epochs. One can see that the generated signals gradually approach the real signals as the number of training epochs increases. However, the generated signals are not completely consistent with the real data.

(4) **Trade-off parameter**. The influence of the trade-off parameter ρ is studied in this sub-section. Fig.18 displays the F1 score of BC-GAN with different ρ values ranging from 0 to 0.9. When ρ=0, L_{score} has no influence on the model. Although $L_{classifier}$ rapidly converges to zero, in this case, Discriminator fails to set the fault alarms correctly. The reason presumably is that L_{score} and $L_{classifier}$ have different functions in the training of the networks. The L_{score} value is essential for Generator training, while $L_{classifier}$ focuses mainly on the classification between real and generated data. According to the parametric analysis, the performance of BC-GAN is relatively stable with the choice of the trade-off parameter. Nevertheless, ρ =0.5 can be recommended in general applications.
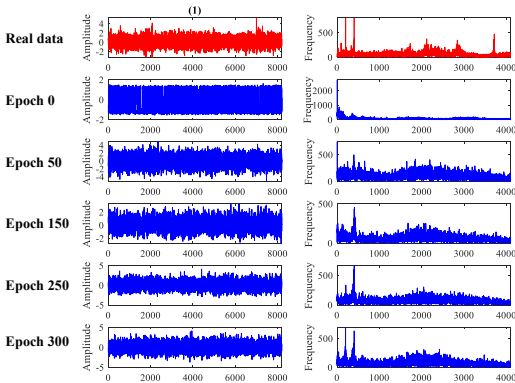
(5) **Advantages and disadvantages**. The main advantages of the proposed EFD method are summarized as follows: (i) BC-GAN directly processes the raw mechanical signals without manually extracted features; experimental results have demonstrated that the method is effective on both vibration and AE signals; (ii) the method directly produces the predicted probability that the captured input signal belongs to either a "health" or "fault" class without any pre-set threshold; (iii) by introducing the $L_{classifier}$, the model performance of BC-GAN can be monitored by observing the loss value, which helps to find a local optimum of the model. (iv) The BC-GAN model is versatile, and its application is not limited to the EFD problem. It can be further extended to address a wide class of problems dealing with streaming data with emerging new classes (SENC), the readers can find more details in our further study with this (Wang, Wang, et al., 2023).

Finally, the limitations of the proposed method are to be mentioned. The training phase is relatively time-consuming if compared with traditional machine learning methods. Besides, BC-GAN still suffers from unstable training, and the balance between Generator and Discriminator is crucial for the overall performance. These issues will also be the focus of our further study.

Acknowledgments

CRediT authorship contribution statement

Yu Wang: Conceptualization, Methodology, Code, Analysis, Writing - original draft. **Szymon Bernat**: Test rig, Data collection, writing - review. **Alexey Vinogradov**: Supervision, Writing - review.

Declaration of Competing Interes

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

We propose a general framework for EFD problem on the basis of deep learning principles. The proposed model is implemented with pytorch. RMSprop stochastic gradient descent is utilized for training the model with the learning rate of 0.0001. The batch size m is 50, critic number n_{critic} is 3, penalty coefficient β is 2, and balance coefficient ρ is set at 0.5 for all experiments. Instead of feeding random noise into the generator, an autoencoder-structured generator is used in this work, and the input is formed by training examples. We construct D_M, F_M and C_M using a convolutional block consisting of three convolutional layers and average pooling layers, the detailed architectures of the utilized Generator and Discriminator are displayed in **Appendix-A**, where C denotes the convolutional kernel, A denotes the average pooling kernel, S and P are the stride and padding number of each kernel, respectively. FC stands for the concatenated layer. For simplicity, we adopted the same network structure in the three main modules of the Discriminator, a straightforward improvement of this work is to construct a more powerful network structure. The number after the symbol @ stands for the number of channels. 'LeakyReLU' is employed as the nonlinear transfer function with negative slope of 0.2, however, it is worth noting that linear mapping is utilized in the output layer of Generator and D_M.

References

Arjovsky, M., Chintala, S., & Bottou, L. (2017). *Wasserstein GAN* (arXiv:1701.07875). arXiv. https://doi.org/10.48550/arXiv.1701.07875

Cai, X.-Q., Zhao, P., Ting, K.-M., Mu, X., & Jiang, Y. (2019). Nearest Neighbor Ensembles: An Effective Method for

Appendix-A – Detailed architecture of the utilized Generator and Discriminator

Generator		Discriminator					
		Discriminator module		Siamese module		Fusion module	
Layers	Output size	Layers	Output size	Layers	Output size	Layers	Output size
Input	8192@1	*Input*	8192@1	*Input*	8192@1	*Input*	8192@2
$\begin{bmatrix} C10,S1,P5 \\ A4,S4,P0 \end{bmatrix}$	2048@8	$\begin{bmatrix} C10,S1,P5 \\ A4,S4,P0 \end{bmatrix}$	2048@8	$\begin{bmatrix} C10,S1,P5 \\ A4,S4,P0 \end{bmatrix}$	2048@8	$\begin{bmatrix} C10,S1,P5 \\ A4,S4,P0 \end{bmatrix}$	2048@8
$\begin{bmatrix} C25,S1,P12 \\ A4,S4,P0 \end{bmatrix}$	512@16	$\begin{bmatrix} C25,S1,P12 \\ A4,S4,P0 \end{bmatrix}$	512@16	$\begin{bmatrix} C25,S1,P12 \\ A4,S4,P0 \end{bmatrix}$	512@16	$\begin{bmatrix} C25,S1,P12 \\ A4,S4,P0 \end{bmatrix}$	512@16
$\begin{bmatrix} C25,S1,P12 \\ A4,S4,P0 \end{bmatrix}$	128@32	$\begin{bmatrix} C25,S1,P12 \\ A2,S2,P0 \end{bmatrix}$	256@32	$\begin{bmatrix} C25,S1,P12 \\ A2,S2,P0 \end{bmatrix}$	256@32	$\begin{bmatrix} C25,S1,P12 \\ A2,S2,P0 \end{bmatrix}$	256@32
FC	4096	*FC*	8192	*FC*	8192	*FC*	8192
output	8192	*output*	8192	*output*	8192	*output*	8192

Difficult Problems in Streaming Classification with Emerging New Classes. *2019 IEEE International Conference on Data Mining (ICDM)*, 970–975. https://doi.org/10.1109/ICDM.2019.00109

Gao, X., Deng, F., & Yue, X. (2020). Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty. *Neurocomputing*, *396*, 487–494. https://doi.org/10.1016/j.neucom.2018.10.109

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*. https://doi.org/10.48550/arXiv.1406.2661

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). *Improved Training of Wasserstein GANs* (arXiv:1704.00028). arXiv. https://doi.org/10.48550/arXiv.1704.00028

Haidong, S., Junsheng, C., Hongkai, J., Yu, Y., & Zhantao, W. (2020). Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. *Knowledge-Based Systems*, *188*, 105022. https://doi.org/10.1016/j.knosys.2019.105022

He, Y., & Zhang, X. (2012). Approximate entropy analysis of the acoustic emission from defects in rolling element bearings. *Journal of Vibration and Acoustics*, *134*(6). https://doi.org/10.1115/1.4007240

Hemmati, F., Alqaradawi, M., & Gadala, M. S. (2016). Optimized statistical parameters of acoustic emission signals for monitoring of rolling element bearings. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *230*(8), 897–906. https://doi.org/10.1177/1350650115619611

Hidle, E. L., Hestmo, R. H., Adsen, O. S., Lange, H., & Vinogradov, A. (2022). Early Detection of Subsurface Fatigue Cracks in Rolling Element Bearings by the Knowledge-Based Analysis of Acoustic Emission. *Sensors*, *22*(14), Article 14. https://doi.org/10.3390/s22145187

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

Li, Z., Zheng, T., Wang, Y., Cao, Z., Guo, Z., & Fu, H. (2021). A Novel Method for Imbalanced Fault Diagnosis of Rotating Machinery Based on Generative Adversarial Networks. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–17. https://doi.org/10.1109/tim.2020.3009343

Liu, C., & Gryllias, K. (2020). A semi-supervised Support Vector Data Description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mechanical Systems and Signal Processing*, *140*, 106682. https://doi.org/10.1016/j.ymssp.2020.106682

Liu, H., Zhou, J., Xu, Y., Zheng, Y., Peng, X., & Jiang, W. (2018). Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. *Neurocomputing*, *315*, 412–424. https://doi.org/10.1016/j.neucom.2018.07.034

Liu, J., Qu, F., Hong, X., & Zhang, H. (2019). A Small-Sample Wind Turbine Fault Detection Method With Synthetic Fault Data Using Generative Adversarial Nets. *IEEE Transactions on Industrial Informatics*, *15*(7), 3877–3888. https://doi.org/10.1109/tii.2018.2885365

Liu, S., Jiang, H., Wu, Z., & Li, X. (2022). Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. *Mechanical Systems and Signal Processing*, *163*, 108139. https://doi.org/10.1016/j.ymssp.2021.108139

Liu, Z., Yang, B., Wang, X., & Zhang, L. (2021). Acoustic Emission Analysis for Wind Turbine Blade Bearing Fault Detection Under Time-Varying Low-Speed and Heavy Blade Load Conditions. *IEEE Transactions on Industry Applications*, *57*(3), 2791–2800. https://doi.org/10.1109/tia.2021.3058557

Lu, W., Li, Y., Cheng, Y., Meng, D., Liang, B., & Zhou, P. (2018). Early fault detection approach with deep architectures. *IEEE Transactions on Instrumentation and Measurement*, *67*(7), 1679–1689. https://doi.org/10.1109/TIM.2018.2800978

Luo, B., Wang, H., Liu, H., Li, B., & Peng, F. (2018). Early fault detection of machine tools based on deep learning and dynamic identification. *IEEE Transactions on Industrial Electronics*, *66*(1), 509–518. https://doi.org/10.1109/tie.2018.2807414

Mao, W., Ding, L., Liu, Y., Afshari, S. S., & Liang, X. (2022). A new deep domain adaptation method with joint adversarial training for online detection of bearing early fault. *ISA Transactions*, *122*, 444–458. https://doi.org/10.1016/j.isatra.2021.04.026

Mu, X., Ting, K. M., & Zhou, Z.-H. (2017). Classification Under Streaming Emerging New Classes: A Solution Using Completely-Random Trees. *IEEE Transactions on Knowledge and Data Engineering*, *29*(8), 1605–1618. https://doi.org/10.1109/TKDE.2017.2691702

Nymark, T. A. (2022). *Micro cracks in wind turbine bearings Investigation of microstructural characteristics and nanomechanical properties in White Etching Cracks (WEC)* [Master thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3017025

Saari, J., Strömbergsson, D., Lundberg, J., & Thomson, A. (2019). Detection and identification of windmill bearing faults using a one-class support vector machine (SVM). *Measurement*, *137*, 287–301. https://doi.org/10.1016/j.measurement.2019.01.020

Song, W., Shen, W., Gao, L., & Li, X. (2022). An Early Fault Detection Method of Rotating Machines Based on Unsupervised Sequence Segmentation Convolutional Neural Network. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–12. https://doi.org/10.1109/TIM.2021.3132989

Wang, Y., Hestmo, R. H., & Vinogradov, A. (2023). Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder. *Measurement Science and Technology*. https://doi.org/10.1088/1361-6501/acc1f8

Wang, Y., Sun, G., & Jin, Q. (2020). Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network. *Applied Soft Computing*, *92*, 106333. https://doi.org/10.1016/j.asoc.2020.106333

Wang, Y., & Vinogradov, A. (2023). Improving the Performance of Convolutional GAN Using History-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Signals. *Applied Sciences*, *13*(5), Article 5. https://doi.org/10.3390/app13053136

Wang, Y., Wang, Q., & Vinogradov, A. (2023). *Semi-supervised deep architecture for classification in streaming data with emerging new classes: Application in condition monitoring*. TechRxiv. https://doi.org/10.36227/techrxiv.21931476.v1

Xu, Y., Zhen, D., Gu, J. X., Rabeyee, K., Chu, F., Gu, F., & Ball, A. D. (2021). Autocorrelated Envelopes for early fault detection of rolling bearings. *Mechanical Systems and Signal Processing*, *146*, 106990.

Zhang, J., Wang, T., Ng, W. W. Y., & Pedrycz, W. (2022). KNNENS: A k-Nearest Neighbor Ensemble-Based Method for Incremental Learning Under Data Stream With Emerging New Classes. *IEEE Transactions on Neural Networks and Learning Systems*, 1–8. https://doi.org/10.1109/TNNLS.2022.3149991

Zhang, T., Chen, J., Li, F., Pan, T., & He, S. (2020). A Small Sample Focused Intelligent Fault Diagnosis Scheme of Machines via Multi-modules Learning with Gradient Penalized Generative Adversarial Networks. *IEEE Transactions on Industrial Electronics*, 1–1. https://doi.org/10.1109/tie.2020.3028821

Paper E:

Ensembled multi-classification generative adversarial network for condition monitoring in streaming data with emerging new classes

Yu Wang[1], Qingbo Wang[1] and Alexey Vinogradov [2]

[1] Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway
[2] Institute of Light Alloys, Kumamoto University, Kumamoto, Japan

Abstract. Condition monitoring (CM) process can be viewed as the problem of streaming classification with emerging new classes (SENC). There are four fundamental challenges we are faced with: (1) timely detection of emerging new classes; (2) model update to adapt to new classes; (3) pattern recognition of these already known classes with high accuracy and (4) distinguishing between different new classes. Although intelligent surrogate models like deep learning have achieved remarkable success in the field of CM, SENC still remains a thorny challenge that has seldom been studied. This paper presents an approach to address the challenge of SENC using an ensembled multi-classification generative adversarial network (EMC-GAN). The proposed method includes a novel deep network architecture called MC-GAN that integrates the tasks of novelty detection and multi-classification into a single framework. To address issues of model stability, an efficient history-state ensemble (HSE) method that does not require additional training costs to generate multiple base models is introduced. Experimental validation is conducted on four simulated SENC tasks using benchmark data, and the results have shown the effectiveness of the proposed approach.

Keywords: Streaming data with emerging new classes (SENC), condition monitoring (CM), history-state ensemble (HSE), generative adversarial network, novelty detection.

1   Introduction

Condition monitoring (CM) of rotating machinery is a continuous process of identifying changes that are indicative of the developing failures. The main objectives include: (1) timely detection of the breakpoint when the monitored machine deviates from its normal operating condition, (2) accurate pattern recognition for multiple failure types, and (3) quantification of the damage development rate. With the improvement in computational ability, intelligent surrogate models based on deep learning have achieved remarkable success in multiple fault diagnosis [1–4], early fault detection [5–7] and

remaining useful life prediction [8], etc. Traditional CM frameworks attempt to tackle these tasks separately, and their limitations are standing out from the surroundings. Most existing surrogate models are built on an ideal assumption that when we perform CM on a particular machine, failures encountered in the real streaming data must be included in the categories of training data. Nevertheless, this requirement is hard to meet in real-world applications because the signals acquired from sensors are fed into the CM system over time, one cannot access the whole dataset to train the surrogate models. A common solution is to utilize historical datasets to train the models, however in practice, it is rather difficult to obtain a representative set of actual fault data, not to mention the problem complexity due to numerous possible failure types and their locations. It is impossible to create an all-round dataset that contains all failure signals. Therefore, the surrogate models must have the ability to handle these unseen failure types in streaming data while identifying these already-known faults, a scenario known as streaming classification with emerging new classes (SENC), where most state-of-art approaches fail.

The challenges faced in traditional SENC problem are decomposed as follows [9]: (1) timely detection of emerging new classes that are not included in the training dataset; (2) automatically model update in order to adapt to new classes; (3) multi-classification of the already known classes, and (4) discriminating different new classes. The surrogate models are required to preserve the historical health condition memories of the machine, which contributes to maintenance efficiency. Due to the growing need to reduce reliance on expert knowledge and manpower, it is required that (5) the surrogate models are capable to extract features and make the data-driven decisions automatically. To this end, a novel SENC framework is proposed to address the above challenges. The main contributions of this paper are briefly summarized as:

(1) A new architecture called MC-GAN is proposed to tackle the SENC problem by integrating novelty detection into the multi-classification framework.

(2) The MC-GAN architecture is further improved by incorporating a novel ensemble technique known as the "history-state ensemble" (HSE) method, which significantly enhances the model's performance without requiring additional training resources.

(3) The effectiveness of the proposed framework is validated through four simulated SENC tasks using benchmark vibration signals.

2 Methodology

2.1 Notations

Given a training dataset $T = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^m$, where $\boldsymbol{x}_i \in R^d$ is the i-th training instance and y_i represents the corresponding label or class which belongs to $Y = \{1, 2, \ldots c\}$, c is the total number of known classes. The task is to build an initial multi-classification model based on the given dataset T so as to process the stream data $S = \{(\widetilde{\boldsymbol{x}}_t, \widetilde{y}_t)\}_{t=1}^{\infty}$, where $\widetilde{\boldsymbol{x}}_t \in R^d$ is a streaming instance at time t, and the corresponding class $\widetilde{y}_t \in \widetilde{Y} = \{1, 2, \ldots c, c+1, c+2, \ldots c+K\}$, K is the number of all new classes.

2.2 Basic structure of the proposed MC-GAN

Fig. 1 illustrates the main structure of the proposed model, comprising a Generator and a Discriminator that are inherited from traditional GAN. Similarly, the Generator is employed to produce the fake data that are close to the input real data, and the Discriminator is trained to distinguish between the real and fake data. As the original Discriminator in traditional GAN is designed as a binary classifier, its network architecture is required to be modified to accommodate multi-classification tasks. As depicted in Fig. 1, the designed Discriminator's architecture consists of two distinct channels, which are employed to learn implicit features from different aspects of the signals.
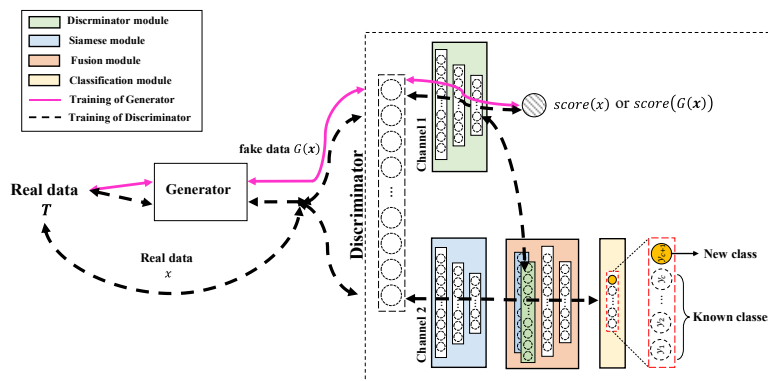

Fig. 1. Basic structure of the proposed MC-GAN.

The 'channel 1' contains a 'Discriminator module', which is used to learn the similarities of all real data from different categories that can distinguish them from the fake data. Therefore, the module does not distinguish between different classes of real data. It takes in all input instances and outputs a 'score' for each of them that indicates whether the instance belongs to fake or real. A linear fully connected layer is employed to compute the score, expressed as:

$$score(x) = \omega_s D_M(x) + b_s \tag{1}$$

where $x \in T$, D_M represents the Discriminator module, which is a neural network block consisting of multiple convolutional layers and pooling layers.

The 'channel 2' comprises three models including:

(1) A Siamese module, S_M, that also takes in all input instances. But different from the previous D_M, it is designed to learn the difference between different classes of real data.

(2) A Fusion module, $F_M(D_M, S_M)$, that receives the extracted features, generally the last hidden layer of D_M and S_M.

(3) A Classification module, C_M, that receives the fused feature from F_M. A SoftMax layer is utilized with the number of hidden neurons of $c + 1$, where c is the number of

all known classes of training instances, and an additional hidden neuron is added as the detector of emerging new classes.

$$C_M(y_i|\boldsymbol{x}) = \frac{exp\left(\omega_{C_M}^k F_M\right)}{\sum_{j=1}^{c+1} exp\left(\omega_{C_M}^j F_M\right)} \tag{2}$$

where $y_i \in Y \cup \{c+1\}$. The equation outputs the probability of $\boldsymbol{x}$ belongs to label y_i, the $c+1$ represents the label of a new class that does not belong to the training dataset T.

Detection of a new class. The basic idea of MC-GAN is to take the fake data produced by Generator as a new class to train the entire network. Given a training dataset T with c known classes, in the feed-forward phase, the input real data that are randomly sampled from T are fed into the Generator to produce fake data. The fake data $G(x)$ is considered as a new class labeled as '$c+1$'. Then, both fake and real data are fed into the modified Discriminator. The MC-GAN can directly output the predicted probability that an instance belongs to each known class or a new class. The training of MC-GAN is detailed in **Algorithm 1**.

Training of Generator. The data flow during the training of Generator is depicted by magenta line in Fig. 1. The Generator is optimized by minimizing the following loss function:

$$L(G) = -E[score(G(\boldsymbol{x}))] + \gamma \cdot \|\boldsymbol{x} - G(\boldsymbol{x})\|_1 \tag{3}$$

where γ is a penalty coefficient, the second term measures the L1 distance between the real data $\boldsymbol{x}$ and generated data $G(\boldsymbol{x})$ [10].

Training of Discriminator. The discriminator is optimized by minimizing the following combined loss function:

$$L(D) = \rho \times L_d + (1 - \rho) \times L_c \tag{4}$$

where $\rho \in (0,1)$ in formula (4) is a trade-off parameter, L_d and L_c represents the loss functions of the 'channel 1' and 'channel 2', respectively, which are defined as:

$$L_d = E[score(G(\boldsymbol{x}))] - E[score(\boldsymbol{x})] + \beta E[(\|\nabla_{\boldsymbol{x}'} score(\boldsymbol{x}')\|_2 - 1)^2] \tag{5}$$

$$L_c = E[(C_M(\tilde{\boldsymbol{x}}) - y)^2] \tag{6}$$

where the L_d adopts the Wasserstein distance (WD) as a measure of the distance between real and generated data distribution; the details of gp-WGAN are presented in [11,12]. $\boldsymbol{x}' = \alpha \cdot \boldsymbol{x} + (1 - \alpha) \cdot G(\boldsymbol{x})$ in formula (3) represents the interpolation of $\boldsymbol{x}$ and $G(\boldsymbol{x})$, where $\alpha \in (0,1)$. $\tilde{\boldsymbol{x}} = F_M(D_M(\boldsymbol{x}), S_M(\boldsymbol{x}))$ represents the output of Fusion module.

Algorithm 1 Training of MC-GAN using RMSprop stochastic gradient descent. We use default value of Adam hyper-parameters: $lr = 0.0001, \alpha = 0.99$.

Input: Training dataset $T = \{(x_i, y_i)\}_{i=1}^c$; Initial Discriminator weights ω^0; Initial Generator weights φ^0; Initial training hyper-parameters, including, epoch number N, batch size m, critic number n_{critic}, penalty coefficient β, and balance coefficient ρ.

Output: Well-trained ω^N and φ^N.

1: **for** $epoch = 1, \dots, N$ **do**
2:   **for** $critic = 1, \dots, n_{critic}$ **do**
3:     **for** $batch = 1, \dots, m$ **do**
4:       Sample a batch of real data $x \sim T$
5:       Generate fake data $G(x)$
6:       $L_d \leftarrow Eq.(5)$
7:       $L_c \leftarrow Eq.(6)$
8:       $L(D) \leftarrow Eq.(4)$
9:       $\omega \leftarrow \text{RMSprop}(\nabla_\omega E(L(D)), \omega, lr, \alpha)$, optimize the Discriminator.
10:     **End for**
11:   **End for**
12:   Sample a batch of real data $x \sim T$
13:   Generate fake data $G(x)$
14:   $L(G) \leftarrow Eq.(3)$
15:   $\varphi \leftarrow \text{RMSprop}(\nabla_\varphi E(L(G)), \varphi, lr, \alpha)$, optimize the Generator.
16: **End for**

2.3 History-state ensemble (HSE)

As the Generator and the Discriminator are competing with each other during training, their performance tends to fluctuate, which makes it challenging to assess whether the Discriminator has been sufficiently trained for the SENC task. To address the challenge, a novel ensemble technique referred to as history-state ensemble (HSE) method is introduced in this work [13]. The HSE method operates on the assumption that neural networks can generate several local optima or 'history-states' during training, which can serve as base models for ensemble learning. To generate multiple base models, only the historical weights of the network need to be retained, as shown in Fig. 2. As a result, the time cost of this procedure is negligible.

In order to minimize redundancy among the base models, we introduce two parameters: θ_1 and θ_2. The former parameter, $\theta_1 1$, specifies the number of epochs required for model warm-up, while the latter parameter, θ_2, sets the update frequency between two consecutive base models. Besides, to effectively leverage the HSE, an important conclusion drawn from our previous work is applied, which pointed out that the performance of MT-GAN is strongly related to the loss value L_c, as presented in formula (6). Therefore, the model selection is crucial. A simple model selection approach is adopted in this paper. By sorting the L_c values corresponding to each base models, these base models with smaller L_c are selected. With the selected base models,

$$Y_i = \frac{1}{m} \sum_{j=1}^{m} P_j(y = i | \boldsymbol{x}) \tag{7}$$

where $P(y = i | \boldsymbol{x})$ denotes the predicted probability that a given instance $\boldsymbol{x}$ belongs to label $i \in \{1,2, \dots c, c + 1\}$. m represents the number of base models, and P_j denotes the predicted probability given by the j-th base model.
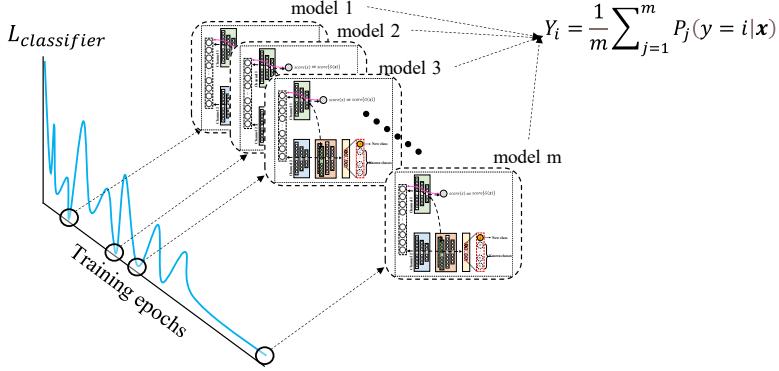
Fig. 2. Illustration of ensembled MC-GAN based on HSE.

2.4 The proposed SENC framework

The proposed workflow is shown in Fig 3. In order to detect multiple new classes, the model is required to update itself over time. Like traditional SENC models, we apply a fix-sized buffer to store the instances from the detected new class. Given an initial training dataset $T = \{(x_i, y_i)\}_{i=1}^{m}$ with c known classes, a streaming dataset $S = \{(\tilde{x}_t, \tilde{y}_t)\}_{t=1}^{\infty}$ with K new classes and a fix-sized buffer B, the primary procedures of the proposed workflow are summarized as follows:

(1) The initial training dataset T is normalized to train the initial MC-GAN.
(2) Choose fix-sized Discriminator snapshots with the smallest L_c as base models to construct the EMC-GAN.
(3) Each instance $\tilde{x}_t$ from the streaming dataset S is fed into the ensembled Discriminator sequentially. If the predicted label of $\tilde{x}_t$ is $c + 1$, then feed the instance into the buffer B.
(4) If the buffer is full, then merge the instances in buffer into the original T, and re-normalized the new training dataset.
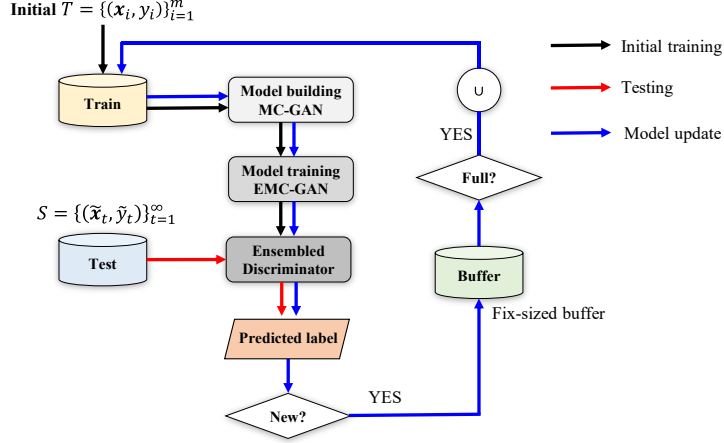(5) Re-train the MC-GAN with updated T and repeat steps (2-4) until the last instances in S.

Initial $T = \{(x_i, y_i)\}_{i=1}^m$

$S = \{(\tilde{x}_t, \tilde{y}_t)\}_{t=1}^\infty$

Fig. 3. The workflow of classification in SENC with EMC-GAN.

# 3   Experimental validation

3.1   Dataset description

We used a prevalent benchmark dataset collected by Case Western Reserve University (CWRU) Bearing Data Center [14] to evaluate the proposed method. The CWRU Bearing dataset contains a collection of vibration signals acquired from ball bearings mounted on a motor. Different levels of artificial damages were seeded on the three different locations of the bearings, namely the inner raceway (I), outer raceway (O), and rolling elements or balls (B). The seeded fault sizes were 0.007, 0.014, and 0.021 inches in diameter. Together with health bearing, a total of ten categories are used in this work, including Health (H), inner race faults (I7, I14, I21), outer race faults (O7, O14, O21) and rolling element faults (B7, B14, B21). The vibration signals of the utilized faulted bearings were recorded for motor loads of 1 to 3 horsepower. The sampling frequency is 12kHz. Each category contains 750 samples with dimension of 1024. Please note that the first paragraph of a section or subsection is not indented.

Table 1 illustrates the four datasets that were generated to evaluate the proposed method. The Data A-C solely comprise of one known class, namely 'Health,' in the initial training dataset T, while the test set S encompasses four fault conditions, including three new classes that were not present in the training dataset. Data A-C were employed to investigate the model's ability to detect and differentiate between faults in the inner race, outer race, and rolling elements of the bearings, respectively. On the other hand, Data set D further increases the complexity of the problem by introducing four known classes and six emerging new classes, featuring distinct fault locations and sizes in the training and test datasets.

Table 1. The utilized datasets.

| Datasets | Training set | | Test set | |
	Fault condition	Number	Fault condition	Number
A	H	250	H, I07, I14, I21	500*1+750*3
B	H	250	H, O07, O14, O21	500*1+750*3
C	H	250	H, B07, B14, B21	500*1+750*3
D	H, I07, I14, I21	250*4	H, I07, I14, I21 O07, O14, O21, B07, B14, B21	500*4+750*6

The simulated SENC tasks using the utilized datasets are visualized in Fig. 4. The health conditions designated by the color black indicate the initial known classes, while several new classes emerge at different times, represented by different colors. As shown in Fig. 3, instances will be fed into the model one by one, and to evaluate the performance of the model on 'known classes', samples from known classes or previously detected new classes will reappear with a certain probability.
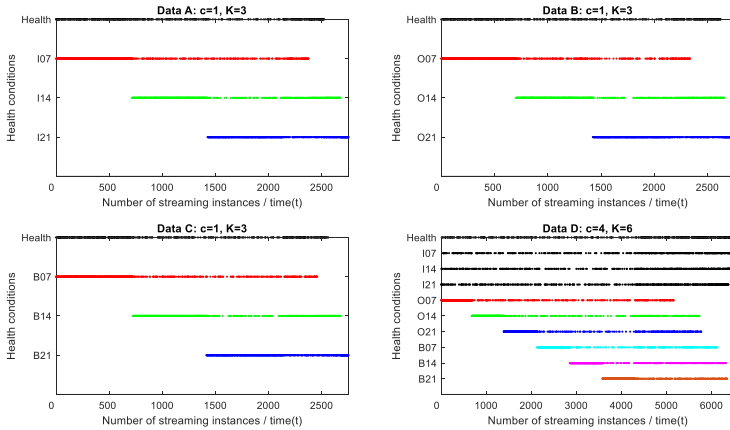
Fig. 4. Visualization of the utilized datasets.

3.2 Results and discussion

The proposed method is implemented using Pytorch. The RMSprop stochastic gradient descent is utilized for training the model. The learning rate is set at 0.0001, and the model is trained for 300 epochs on dataset A-C and 500 epochs on dataset D. The batch size is set at 20, and the trade-off parameter ρ is set to 0.5. The buffer size for model updates is set at 200. For ensemble learning, the warm-up epochs (θ_1) is set to 100,

stride (θ_2) is set to 1, and the number of base models is set at 5 for dataset A-C and 15 for dataset D.

Fig. 5 displays the comprehensive outcomes of the probed methods using confusion matrix. Based on the observations, the following points can be concluded: (1) The proposed method proved to be highly effective in detecting all types of emerging unseen instances, exhibiting the excellent classification accuracy on the four datasets that were evaluated. (2) The proposed method also demonstrated remarkable performance in differentiating between various emerging new classes, with an impressive number of instances being correctly classified. This indicates that the proposed method has the potential to perform reliably in diverse real-world scenarios where emerging new classes are present.

Data A: c=1, K=3

True Class \ Predicted Class	H	I07	I14	I21	NaN
H	500				
I07		750			
I14			750		
I21		23		725	2
NaN					

Data B: c=1, K=3

True Class \ Predicted Class	H	O07	O14	O21	NaN
H	497	3			
O07		750			
O14			738	10	2
O21		32		718	
NaN					

Data C: c=1, K=3

True Class \ Predicted Class	H	B07	B14	B21	NaN
H	500				
B07		749	1		
B14		130	620		
B21		12	25	694	19
NaN					

Data D: c=4, K=6

True Class \ Predicted Class	H	I07	I14	I21	O07	O14	O21	B07	B14	B21	NaN
H	500										
I07		500									
I14			498								2
I21				500							
O07					750						
O14				3		747					
O21	73				277		399				1
B07				43			45	661	1		
B14	2			26			3	8	710	1	
B21								13	2	717	18
NaN											

Fig. 5. The confusion matrix of the proposed method.

For comparison, three state-of-art SENC approaches are introduced:

(1) SENNE [15], a cluster-based ensemble method which is built based on k-Nearest neighbor ensembles and can be used for both new class detection and known class classification after training.

(2) KNNENS [9], another cluster-based ensemble method derived from SENNE and incorporates a reliable model retirement mechanism.

(3) SENCForest [16], which incorporates a completely-random tree technique for identifying newly emerging classes by generating an 'outlying' anomaly sub-region. Additionally, to establish a comprehensive framework for SENC, multi-classification is performed by capturing the class labels within the leaves of the previously trained trees.

As comparative methods are based on shallow learning algorithms that exhibit inadequate performance on raw mechanical signals, 15 features extracted from both the time and frequency domains were utilized as their input. These features encompass several

statistics such as mean, variance, root mean square (RMS), skewness, kurtosis, shape factor, crest factor, impulse factor, margin factor, entropy, energy entropy, power, mean frequency, root mean square frequency, and the standard deviation of frequency. A grid search approach is employed to identify the optimal combination of key parameters that yield the best performance of the competing methods. Determining the decision boundaries is heavily dependent on the threshold value [9]. Nevertheless, the selection of suitable thresholds is contingent on the particularities of each dataset. Therefore, the threshold value is calibrated to guarantee that no less than 98% of the training data can be accurately identified as known.

Initially, we examine the collective performance of the evaluated methods over time compared with the above competitor. By utilizing classification accuracy as the metric, we compute the streaming accuracy (SA) as follows:

$$SA(t) = n_t/N_t \tag{8}$$

where N_t denotes the total number of instances from the start to time t, n_t^j represents the number of instances being correctly classified as the corresponding labels.

Fig. 6 presents the SA of the probed method on the employed four datasets, the best performance of the probed methods is chosen for exhibition. The following observations are worth noting: (1) Initially, all methods perform well on data A-C during testing, but around the 250th instance, the performance of SENNE and SENCForest begins to decline. Meanwhile, KNNENS and the proposed EMC-GAN maintain excellent performance on dataset A and B. However, (2) KNNENS's performance starts to decrease around the 1500th instance on datasets C and D, and the proposed EMC-GAN maintains the highest accuracy at the end of the test. (3) Although the proposed EMC-GAN also experiences a decrease in performance around the 800th instance on data C and 1500th instance on data D, indicating some incorrect predictions at that time, it subsequently recovers and exhibits an increase in performance.
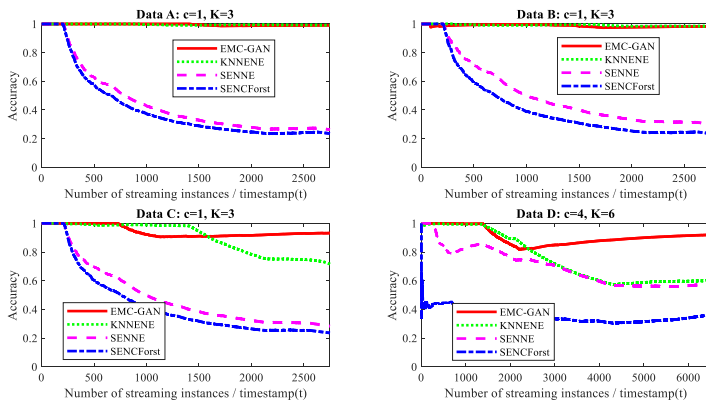

Fig. 6. The streaming accuracy (SA) of the probed methods.

The model update information of the probed methods is illustrated in Fig. 7. The baseline signifies the precise time when the model should be updated to detect all new classes. If the probed methods update the model for a new class before the baseline, it implies that some known instances are being erroneously classified as new. On the other hand, if the probed methods update the model for a new class after the baseline, it suggests that some new instances are being inaccurately classified as known.

The following points are noteworthy: (1) The frequent updates of SENNE and SENCForest, as compared to the baselines, suggest that these methods have detected numerous false new classes. This observation elucidates why these methods exhibit a decline in performance after the 250th instance on data A-C. (2) The KNNENS only update twice for dataset C, however, there are three new classes, which means that the model has poor performance to detect the third new classes. The same behavior can also be observed from data D, the KNNENS only update three times for this dataset given six new classes. In contrast, the proposed EMC-GAN successfully detected and distinguished all the emerging new classes of the four datasets.

Table 2 presents the average accuracy and standard deviation of the probed methods. It can be seen that the EMC-GAN achieves over 90% accuracy on the four datasets. However, the SENNE and SENCForest perform poor because they tend to provide many false new classes after model update.
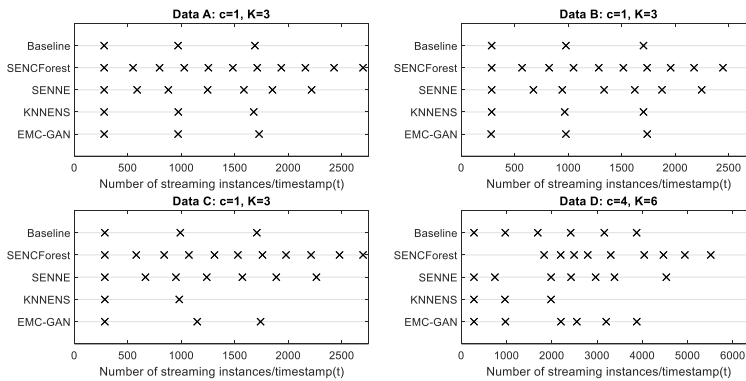
Fig. 7. The comprehensive illustration of the model update timestamp and frequency of the probed methods.

Table 2. Comparison of the average accuracy/standard deviation among probed methods.

Methods	Datasets			
	A	B	C	D
EMC-GAN	**95.12% /0.06**	**94.21% /0.0574**	**92.21% /0.0086**	**91.72% /0.0069**
KNNENS	94.49% /0.0974	89.66% /0.1178	79.45% /0.0653	55.17% /0.0479
SENNE	18.81% /0.0818	21.52% /0.0938	19.81% /0.0831	39.74% /0.2295
SENCForest	22.44% /0.021	22.47% /0.0207	23.17% /0.0257	33.17% /0.0114

4 Conclusion

In this paper, a novel framework, namely EMC-GAN, is proposed to tackle the challenges associated with SENC. The effectiveness of the proposed method is investigated through four simulated SENC tasks based on well-established benchmark datasets. The proposed method offers several advantages over traditional SENC models, including: (1) being threshold-free, the EMC-GAN approach integrates novelty detection procedures with the multi-classification task and can accurately predict whether a signal belongs to one of the known classes or a new class; (2) the proposed EMC-GAN method can process raw signals directly without requiring any manual feature extraction which is in line with the modern pursuit of intelligent model; and (3) the EMC-GAN model achieves high classification accuracy between different emerging new classes. Overall, the proposed EMC-GAN approach has the potential to significantly enhance the accuracy and efficiency of SENC systems, enabling more reliable and effective monitoring of mechanical systems.

References

1. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, doi:10.1016/j.ymssp.2019.106587.
2. Zhang, Y.; Ji, J.C.; Ren, Z.; Ni, Q.; Wen, B. Multi-Sensor Open-Set Cross-Domain Intelligent Diagnostics for Rotating Machinery under Variable Operating Conditions. *Mech. Syst. Signal Process.* **2023**, *191*, 110172, doi:10.1016/j.ymssp.2023.110172.
3. Zhao, X.; Yao, J.; Deng, W.; Ding, P.; Zhuang, J.; Liu, Z. Multiscale Deep Graph Convolutional Networks for Intelligent Fault Diagnosis of Rotor-Bearing System Under Fluctuating Working Conditions. *IEEE Trans. Ind. Inform.* **2023**, *19*, 166–176, doi:10.1109/TII.2022.3161674.
4. Li, X.; Jiang, H.; Wang, R.; Niu, M. Rolling Bearing Fault Diagnosis Using Optimal Ensemble Deep Transfer Network. *Knowl.-Based Syst.* **2021**, *213*, 106695, doi:10.1016/j.knosys.2020.106695.

5.   Xia, X.; Pan, X.; Li, N.; He, X.; Ma, L.; Zhang, X.; Ding, N. GAN-Based Anomaly De-
        tection: A Review. *Neurocomputing* **2022**, *493*, 497–535, doi:10.1016/j.neu-
        com.2021.12.093.

6.   Song, W.; Shen, W.; Gao, L.; Li, X. An Early Fault Detection Method of Rotating Ma-
        chines Based on Unsupervised Sequence Segmentation Convolutional Neural Network.
        IEEE Trans. Instrum. Meas. **2022**, *71*, 1–12, doi:10.1109/TIM.2021.3132989.

7.   Zhou, H.; Lei, Z.; Zio, E.; Wen, G.; Liu, Z.; Su, Y.; Chen, X. Conditional Feature Disen-
        tanglement Learning for Anomaly Detection in Machines Operating under Time-Varying
        Conditions. *Mech. Syst. Signal Process.* **2023**, *191*, 110139,
        doi:10.1016/j.ymssp.2023.110139.

8.   Pan, Y.; Wu, T.; Jing, Y.; Han, Z.; Lei, Y. Remaining Useful Life Prediction of Lubrica-
        tion Oil by Integrating Multi-Source Knowledge and Multi-Indicator Data. *Mech. Syst.
        Signal Process.* **2023**, *191*, 110174, doi:10.1016/j.ymssp.2023.110174.

9.   Zhang, J.; Wang, T.; Ng, W.W.Y.; Pedrycz, W. KNNENS: A k-Nearest Neighbor Ensem-
        ble-Based Method for Incremental Learning Under Data Stream With Emerging New
        Classes. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–8,
        doi:10.1109/TNNLS.2022.3149991.

10.   Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional
        Adversarial Networks 2018.

11.   Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN 2017.

12.   Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved Training
        of Wasserstein Gans. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, doi:Improved training of
        wasserstein gans.

13.   Wang, Y.; Vinogradov, A. Improve the Performance of Convolutional GAN Using His-
        tory-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Sig-
        nals 2023.

14.   Zhao, Z.; Li, T.; Wu, J.; Sun, C.; Wang, S.; Yan, R.; Chen, X. Deep Learning Algorithms
        for Rotating Machinery Intelligent Diagnosis: An Open Source Benchmark Study. *ISA
        Trans.* **2020**, *107*, 224–255, doi:10.1016/j.isatra.2020.08.010.

15.   Cai, X.-Q.; Zhao, P.; Ting, K.-M.; Mu, X.; Jiang, Y. Nearest Neighbor Ensembles: An
        Effective Method for Difficult Problems in Streaming Classification with Emerging New
        Classes. In Proceedings of the 2019 IEEE International Conference on Data Mining
        (ICDM); 2019; pp. 970–975.

16.   Mu, X.; Ting, K.M.; Zhou, Z.-H. Classification Under Streaming Emerging New Classes:
        A Solution Using Completely-Random Trees. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*,
        1605–1618, doi:10.1109/TKDE.2017.2691702.

Paper F:

Yu Wang, Wang, Q, S. Bernat, Vinogradov, A. Ensembled multi-task generative adversarial network (EMT-GAN): a deep architecture for classification in streaming data with emerging new classes and its application to condition monitoring of rotating machinery, 2023. (Under review).

Appendix B

GitHub repository

This repository contains the source code of the algorithms developed in Paper E and Paper F. The repository can be found at:

https://github.com/YuWa1994/EMC-GAN

NTNU

Norwegian University of
Science and Technology