ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

MECCANICA E SCIENZE AVANZATE DELL'INGEGNERIA

Ciclo 38

Settore Concorsuale: 09/C1 - MACCHINE E SISTEMI PER L'ENERGIA E L'AMBIENTE

Settore Scientifico Disciplinare: ING-IND/09 - SISTEMI PER L'ENERGIA E L'AMBIENTE

APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO CONDITION MONITORING AND PREDICTIVE MAINTENANCE OF MECHANICAL COMPONENTS

Presentata da: Francesco Bellucci

Coordinatore Dottorato

Lorenzo Donati

Supervisore

Enrico Corti

Co-supervisore

Paolo Bovicelli

Esame finale anno 2026

Abstract

In the era of Industry 4.0, the reliability of complex mechanical systems is paramount, yet unexpected equipment failures remain a critical challenge, leading to significant economic losses and operational disruptions. This thesis presents the development and validation of a systematic, multi-stage framework for early fault diagnosis by applying Artificial Intelligence (AI) techniques to Condition Monitoring (CM) data.

The core of the proposed methodology is a hybrid approach that integrates physics-informed feature engineering with unsupervised anomaly detection models. A key contribution is the novel strategy of using one sensing modality, such as online oil debris analysis, to supervise the definition of a robust 'healthy' baseline for training the unsupervised models. The framework processes multi-sensor data—primarily vibration and ultrasound—to generate both a General Health Index (GHI) for an overall assessment of system health and component-level Specific Health Indices (SHIs) for targeted diagnostics.

The effectiveness of this framework is validated across two distinct, large-scale industrial case studies: a run-to-failure durability test on a heavy-duty mechanical gearbox, and a series of developmental tests on automotive electric powertrains, including Accelerated Durability and End-of-Line testing. In the gearbox case study, the methodology successfully identified an incipient bearing fault, providing over 800 hours of lead time before catastrophic failure. In the electric powertrain applications, the framework demonstrated its adaptability in detecting a range of anomalies, from lubrication system degradation to the onset of severe mechanical faults, and its viability for real-time, onboard implementation was verified.

Ultimately, this research contributes a validated, adaptable, and scalable framework for implementing data-driven predictive maintenance. It demonstrates a practical and effective methodology that bridges the gap between theoretical AI models and the demands of real-world industrial machinery, offering a clear pathway to enhanced operational reliability and efficiency.

Table of Contents

Contents

List of Table

List of Figure

List of Symbols

The following tables summarize the main symbols, acronyms, and abbreviations used throughout this thesis.

Acronyms and Abbreviations

AI	Artificial Intelligence
B2B	Back-to-Back (Test Configuration)
CAGR	Compound Annual Growth Rate
CAN	Controller Area Network
CM	Condition Monitoring
CPM	Cycles Per Minute
CUS	Contact Ultrasound Sensor
DL	Deep Learning
DUT	Device Under Test
EFD	Early Fault Diagnosis
EIF	Extended Isolation Forest
EOL	End-of-Line
ESA	Electrical Signature Analysis
FFT	Fast Fourier Transform
FSD	Full-Scale Deflection
GHI	General Health Index
GMF	Gear Mesh Frequency
GMM	Gaussian Mixture Models
GPU	Graphics Processing Unit
HI	Health Index
IEPE	Integrated Electronics Piezo-Electric
IIoT	Industrial Internet of Things
IQR	Interquartile Range
ISO	International Organization for Standardization
LOF	Local Outlier Factor
LRD	Local Reachability Density
MA	Moving Average
MAD	Median Absolute Deviation
MCSA	Motor Current Signature Analysis
MCD	Minimum Covariance Determinant

ML	Machine Learning
MM	Moving Median
NDE	Non-Drive End
NOC	Normal Operating Condition
OC-SVM	One-Class Support Vector Machine
PCA	Principal Component Analysis
PdM	Predictive Maintenance
PHM	Prognostic and Health Management
PM	Preventive Maintenance
P-F	Potential-to-Failure (Curve)
Q1, Q3	First and Third Quartile
RM	Reactive Maintenance
RMS	Root Mean Square
ROI	Return on Investment
RUL	Remaining Useful Life
RxM	Prescriptive Maintenance
SHI	Specific Health Index
SK	Spectral Kurtosis
SNR	Signal-to-Noise Ratio
STB	System Test Bench
STFT	Short-Time Fourier Transform
STL	Seasonal-Trend decomposition using LOESS
SVM	Support Vector Machine
TAN	Total Acid Number
TBN	Total Base Number
TPU	Tensor Processing Unit
t-SNE	t-Distributed Stochastic Neighbor Embedding
WT	Wavelet Transform

Introduction

In the contemporary industrial landscape, often termed Industry 4.0, the reliance on complex machinery and automated systems is more pronounced than ever. The efficiency, reliability, and safety of these industrial operations are intrinsically linked to the health of their mechanical components. However, unexpected equipment failure remains a significant challenge, often leading to costly unplanned downtime, reduced productivity, and potential safety hazards. To mitigate these risks, maintenance strategies have undergone a significant evolution, moving from reactive fixes to proactive and intelligent interventions.

This thesis enters discourse at a pivotal moment where the convergence of industrial machinery and artificial intelligence (AI) is redefining the boundaries of maintenance. The research presented herein explores the application of advanced AI and machine learning (ML) techniques to the fields of Condition Monitoring (CM) and Predictive Maintenance (PdM). The primary objective is to develop and validate intelligent models capable of moving beyond traditional maintenance paradigms. Specifically, this work focuses on enhancing the ability to automatically discriminate between healthy and faulty operational states of various mechanical components.

This research was conducted as part of an industrial PhD program funded by **Alma Automotive s.r.l.**, and its findings are grounded in extensive practical experience from multiple industrial collaborations. The case studies presented herein draw upon data from durability testing on gearboxes, performed in partnership with **S.T.M. Riduttori**, and development tests on electric powertrains, based on data acquired during a six-month industrial internship at **Rimac Technology**. Additional experience was gained from monitoring internal combustion engines and high-productivity die casting activities. This thesis aims to provide a robust framework for implementing data-driven maintenance strategies by leveraging the knowledge acquired from both a strong mechanical engineering background and specialized training in machine learning, bridging the gap between theoretical models and real-world industrial applications.

This work will demonstrate how machine learning models can be effectively utilized for anomaly detection, damage classification, and the prediction of Remaining Useful Life (RUL), thereby offering a pathway to more efficient, reliable, and cost-effective industrial operations.

The structure of this thesis is organized to guide the reader from foundational theory through to practical application and validation. **Chapter 1** provides a comprehensive overview of the evolution of maintenance strategies, the physical principles of equipment failure, and the industrial context for this research. **Chapters 2 and 3** establish the analytical foundation, first by introducing the fundamental principles of machine learning, and then by detailing the specific algorithms for dimensionality reduction and anomaly detection used in this research. **Chapter 4** shifts the focus to physical data, providing a detailed examination of the key signals and sensors relevant to condition monitoring, such as vibration and oil analysis. **Chapter 5** synthesizes these elements by outlining the overarching research framework and the specific, multi-stage methodological workflow applied throughout the case studies. The practical validation of this framework is presented in the subsequent experimental chapters: **Chapter 6** details the early fault diagnosis in a heavy-duty mechanical gearbox, while **Chapter 7** presents the results from several test campaigns on electric vehicle powertrains. Finally, the thesis concludes with a synthesis of the key findings, a discussion of the contributions and limitations of the research, and an outlook on future work.

1. The Evolution of Maintenance: From Reactive to Predictive Strategies

1.1 Introduction to Industrial Maintenance

Industrial maintenance encompasses the combination of all technical, administrative, and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function. A well-structured maintenance strategy is a critical investment for any industrial enterprise, directly impacting production efficiency, equipment longevity, and overall operational costs. The primary objective is to ensure the availability and reliability of physical assets, preventing expensive production stoppages and catastrophic failures that can have significant financial and safety implications.[1]

The selection and implementation of an appropriate maintenance strategy are therefore crucial decisions. European standard EN 13306:2017, "Maintenance — Maintenance terminology," provides a standardized framework for understanding the fundamental concepts and terms within this discipline. This standard serves as a foundational reference, defining the types of maintenance and the relationships between them, which helps in establishing a clear and consistent approach to maintenance management. According to this standard, maintenance is broadly divided into two main categories: preventive maintenance and corrective maintenance.

1.2 Understanding Equipment Failure

To develop an effective maintenance strategy, it is first essential to understand the patterns of equipment failure over its lifecycle. The reliability of a mechanical component is not static; it changes over its operational life. A widely accepted conceptual model used in reliability engineering to represent this behavior is the Bathtub Curve. This curve plots the failure rate of a component or a population of components as a function of time. As illustrated in Figure 1, the curve is named for its characteristic "bathtub" shape, which is composed of three distinct phases.

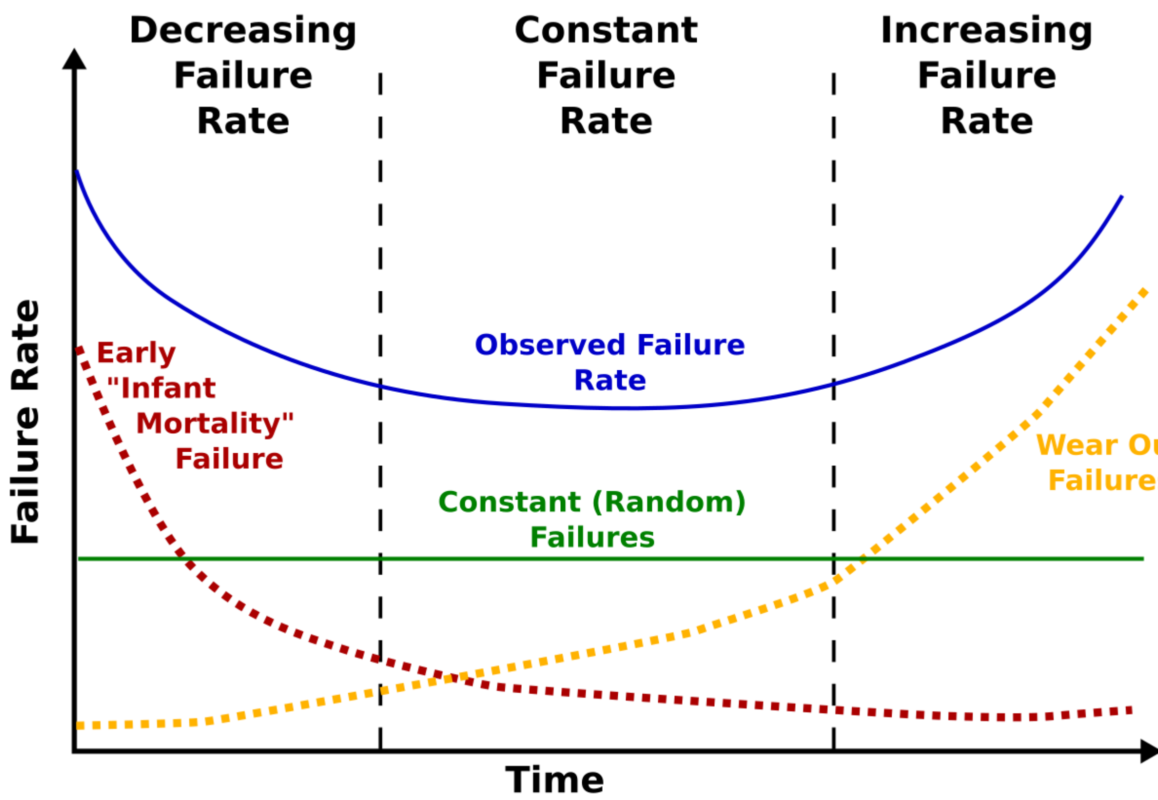


Figure 1 - Bathtub curve, illustrating the three phases of equipment failure rate over its lifecycle. Based on the model presented in Ebeling [2].

1. Phase 1: Infant Mortality (Early-Life Failures): This initial period is characterized by a high but decreasing failure rate. Failures in this phase are typically premature and can be attributed to latent defects from the manufacturing process, poor quality control, damage during shipping, or errors made during installation and commissioning. These are not wear-related failures but rather inherent weaknesses that are quickly exposed once the equipment is put into service.
2. Phase 2: Useful Life (Random Failures): After the initial period, the failure rate drops to a low and relatively constant level. Failures during this longest phase of an asset's life are considered random and unpredictable. They are typically caused by sudden stress events, operator error, or other unforeseeable external factors, rather than by age or wear. Because these failures are random, traditional time-based preventive maintenance is often ineffective in this phase. This is the period where Condition Monitoring and Predictive Maintenance

provide the greatest value, as they aim to detect the incipient stages of these random faults before they lead to a catastrophic failure.

3. Phase 3: Wear-Out (Age-Related Failures): In the final phase of the component's lifecycle, the failure rate begins to increase rapidly. These failures are a direct result of age and accumulated wear. Components have reached the end of their design life, and failures occur due to material fatigue, corrosion, abrasion, or other cumulative degradation mechanisms. This is the phase where traditional, time-based preventive maintenance (e.g., replacing a bearing after a set number of operating hours) is most effective, as the failures are directly correlated with operational age.

Understanding these three distinct failure periods is fundamental to maintenance management. It highlights that no single maintenance strategy is optimal for the entire lifecycle of a machine. A comprehensive strategy must therefore be a dynamic blend of approaches, designed to mitigate the specific risks associated with each phase of the bathtub curve. For example, O’Connor & Kleyner (2012) define the bathtub curve as showing “an initial decreasing hazard rate or infant mortality period, an intermediate useful life period and a final wearout period”. [3] Likewise, the NIST reliability handbook describes an early “Infant Mortality” phase (with origins in actuarial life tables), a long constant-failure “Intrinsic” period, and a late “Wearout” period. [4] Klutke *et al.* (2003) review notes that the bathtub shape dates back to 17th-century actuarial studies (as early as 1693) and is treated in “nearly every standard reliability text”. [5] These foundational sources establish the classical three-phase failure model and stress its prevalence (and limits) in reliability theory.

While the Bathtub Curve provides a high-level view of failure probability over a component's lifecycle, the Potential-to-Failure (P-F) Curve offers a more granular model for understanding the progression of a single, specific failure mode. More recent and comprehensive interpretations have expanded this model into the D-I-P-F Curve to provide a holistic view of an asset's lifecycle, from its design inception to its functional failure. This expanded model, illustrated in Figure 2, clarifies how decisions made long before a fault is detectable can profoundly impact an asset's lifespan.[6]

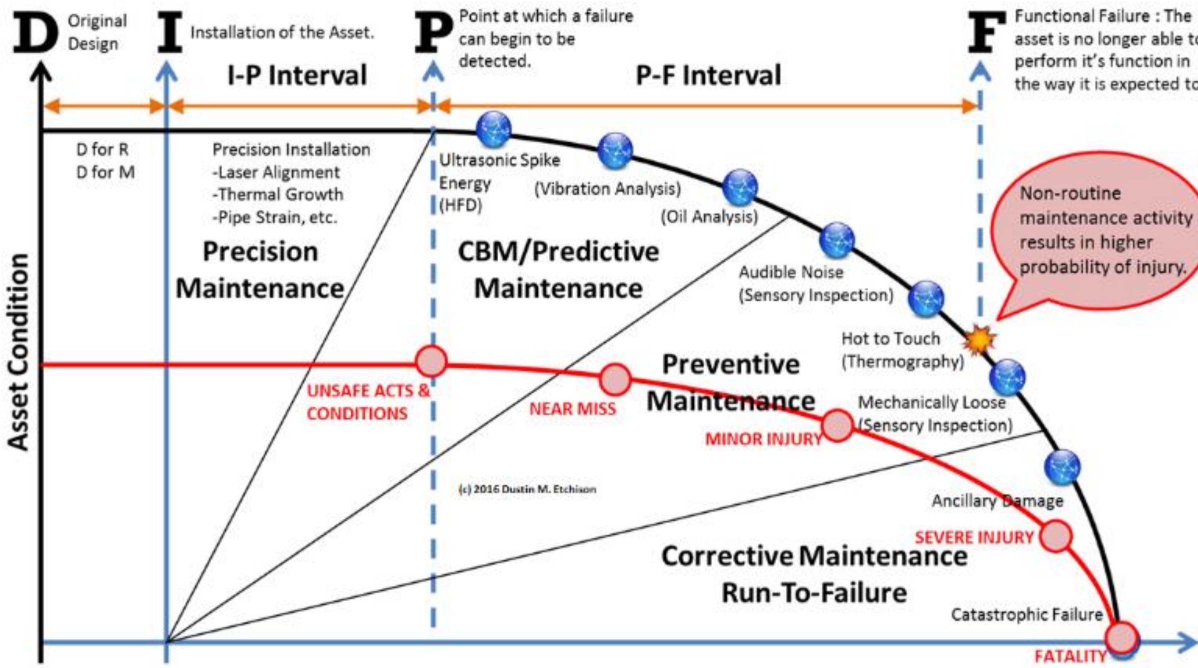


Figure 2 – The D-I-P-F Curve, illustrating the stages of asset degradation. Source: Aladon [6].

The curve is defined by several critical points and intervals:

- D (Design): The curve begins at point D, which represents the asset's maximum achievable lifespan and inherent reliability as determined by its engineering, materials, and manufacturing specifications.
- I (Installation): This point represents the actual condition of the asset immediately after commissioning. The gap between D and I reflect the quality of installation; poor practices like misalignment or improper fastening can mean the asset starts its life already in a degraded state. The interval between these points, the D-I Interval, is where Precision Maintenance practices during installation and commissioning are crucial to preserving the asset's designed reliability.
- P (Potential Failure): This is the earliest point at which a developing fault can be detected by a given condition monitoring technology (e.g., ultrasound, vibration analysis).
- F (Functional Failure): This is the point where the asset no longer meets its specified performance requirements and has failed.

This expanded model introduces two critical time windows. The I-P Interval represents the period of normal operation where proactive maintenance can extend the asset's life before any fault is detectable. The classic P-F Interval remains the critical window of opportunity for predictive maintenance. It is the lead time available, after a fault has been detected at point P, to plan, schedule, and execute a repair before the functional failure at point F occurs. The entire philosophy of Predictive Maintenance is built upon deploying condition monitoring techniques (such as vibration analysis, thermography, or oil analysis), that can identify point P as early as possible, thereby maximizing the P-F interval and providing the longest possible warning period for a planned, proactive response.

1.3 A Spectrum of Maintenance Strategies

Maintenance philosophies have undergone significant evolution, driven by technological advancements, the increasing complexity of industrial machinery, and a growing emphasis on operational efficiency. This evolution can be categorized into several distinct approaches, which are clearly structured within the EN 13306 standard. As depicted in Figure 3- Maintenance Strategies, the fundamental distinction lies between actions taken before a failure (preventive) and actions taken after a failure (corrective).

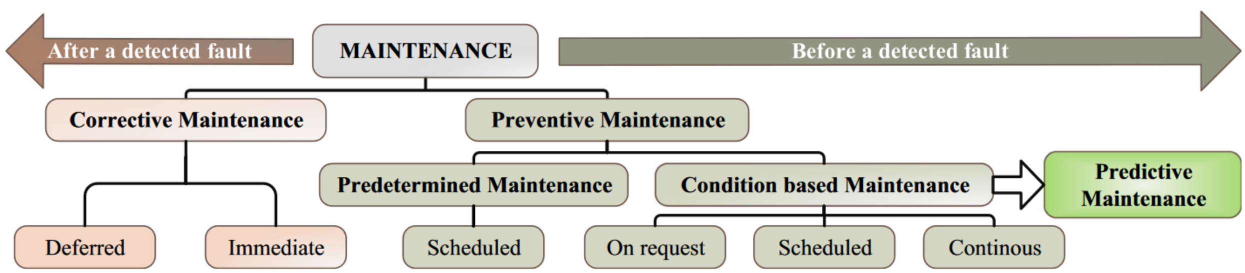


Figure 3 - A classification of maintenance strategies, based on the framework defined in the European Standard EN 13306:2017 [7].

1.3.1 Reactive Maintenance (Corrective)

The most traditional approach, reactive maintenance, is performed only after a piece of equipment has broken down or a malfunction has occurred[1], [8]. Often referred to as a "run-to-failure" strategy, its primary goal is to restore the equipment to its functional state as quickly as possible to minimize downtime [1]. This method is generally the most expensive due to its unplanned nature, which can lead to significant delays in production and higher costs for urgent repairs[9]. It is typically only suitable for non-critical equipment where the cost of preventive measures would exceed the cost of failure [1].

This strategy is often characterized as a "run-to-failure" approach. As defined by EN 13306, corrective maintenance can be further subdivided into:

- Immediate/Emergency Corrective Maintenance: Carried out without delay after a fault is detected to avoid unacceptable consequences, such as safety hazards or major production losses. This is the most disruptive and costly form of maintenance due to its unplanned nature.
- Deferred Corrective Maintenance: Maintenance that is delayed for a period in accordance with given maintenance rules, but before a complete failure occurs.

While simple to implement, a purely reactive strategy is generally the most expensive due to its association with unplanned downtime, collateral damage to equipment, and higher costs for urgent repairs and overtime labor. It is typically only suitable for non-critical, redundant, or low-cost equipment where the cost and consequences of failure are minimal.

1.3.2 Preventive Maintenance

In contrast to the reactive approach, preventive maintenance is proactive and is carried out at predetermined intervals or according to prescribed criteria, aimed at reducing the probability of failure or degradation of the functioning of an item. The primary goal is to perform maintenance before failure occurs. The EN 13306 standard outlines two main sub-categories of preventive maintenance, as shown in Figure 3:

- Predetermined Maintenance: This is the most traditional form of preventive maintenance, where tasks are performed at fixed intervals of time, usage, or distance (e.g., inspecting a machine every 1,000 operating hours). While it is effective in preventing many age-related failures, its key limitation is that it does not account for the actual condition of the equipment. This can lead to inefficiency, as maintenance may be performed unnecessarily—leading to the premature replacement of perfectly functional components—or may not be performed soon enough if a component degrades faster than anticipated.

- Condition-Based Maintenance (CBM): This represents a more advanced and efficient preventive strategy. CBM involves monitoring the physical condition of an asset to decide when maintenance should be performed. Maintenance is only carried out when the monitored data indicates that a fault is imminent. This approach relies on the real-time or periodic measurement of specific parameters (e.g., vibration, temperature) and comparing them against known limits. By basing maintenance decisions on actual equipment health, CBM optimizes resource allocation, minimizes unnecessary interventions, and provides an early warning of potential failures.

1.3.3 Predictive Maintenance (PdM)

Predictive Maintenance (PdM) can be considered the most advanced evolution of Condition-Based Maintenance. While CBM focuses on detecting the onset of a fault, PdM aims to predict the future trend of that fault. The core principle of PdM is to use condition-monitoring data in conjunction with predictive algorithms to forecast when a failure is likely to occur. This allows maintenance to be scheduled with a lead time, just before the predicted failure, thereby maximizing the useful life of the component while minimizing the risk of unplanned downtime.

A properly implemented PdM strategy has been shown to yield significant benefits, including:

- Increased Equipment Uptime: By predicting failures in advance, maintenance can be scheduled during planned shutdowns.
- Reduced Maintenance Costs: Eliminates unnecessary preventive tasks and reduces the high costs associated with emergency repairs.
- Improved Safety: Identifies potentially hazardous conditions before they can lead to accidents.

The success of a PdM program is intrinsically linked to the quality of the data collected and the sophistication of the analytical models used to interpret it.

Figure 4 shows a simplified graphical representation of the evolution of the types of maintenance described above.

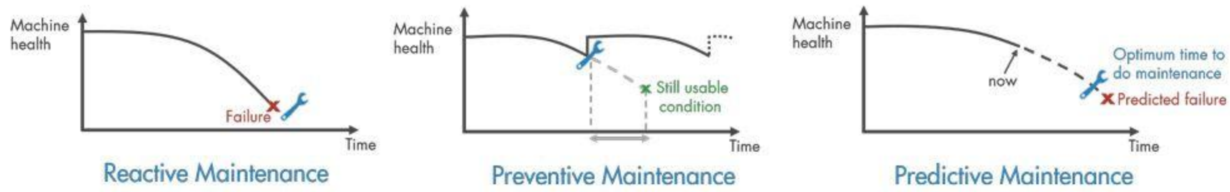


Figure 4 - Evolution of Type of Maintenance, Source: MathWorks [10].

Beyond the operational benefits of increased uptime and reliability, a primary driver for adopting a predictive strategy is its significant economic advantage. The relationship between maintenance costs and the frequency of intervention varies dramatically across different strategies. This economic trade-off is effectively conceptualized in Figure 5, which compares the cost profiles of reactive, preventive, and predictive maintenance. Reactive Maintenance (RM) consistently incurs a high cost, as expenses are driven by unplanned downtime, secondary equipment damage, and emergency repairs. Preventive Maintenance (PM) exhibits a characteristic U-shaped cost curve; performing maintenance too infrequently leads to increased failure-related costs, while performing it too frequently results in unnecessary expenditure on labor and components. The goal of a PM strategy is to find the optimal point at the bottom of this curve. Predictive Maintenance (PdM), however, fundamentally breaks this trade-off. By leveraging real-time condition data to intervene only, when necessary, PdM aims to consistently operate at a near-optimal cost, avoiding both the excessive costs of failure and the waste of premature maintenance.[11]

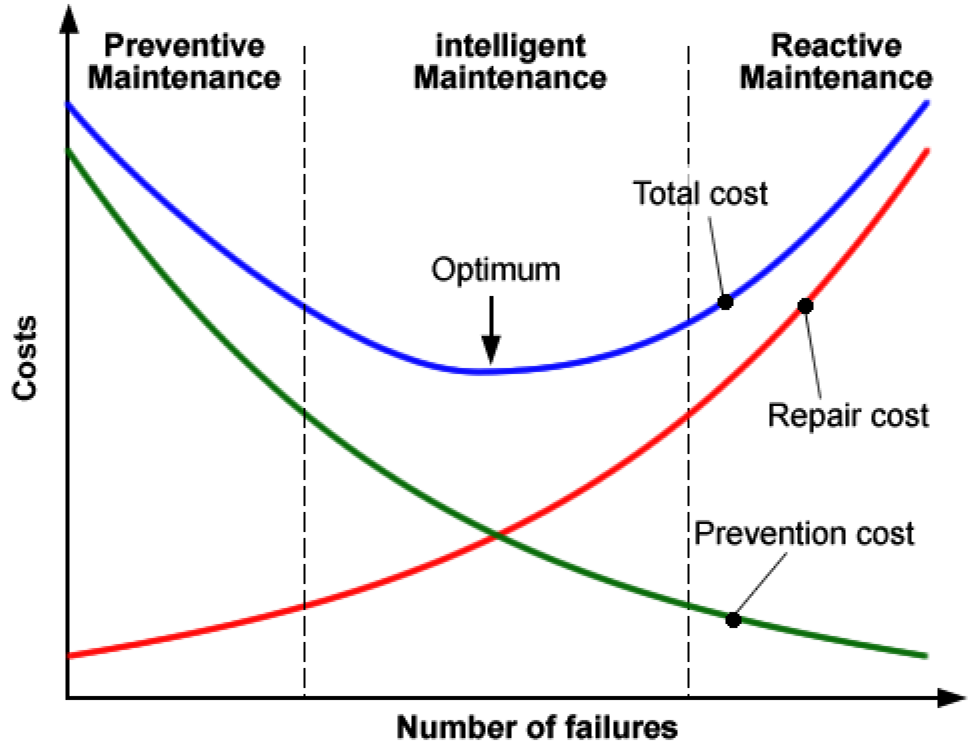


Figure 5 - Cost associated with traditional maintenance strategies, illustrating the relationship between intervention frequency and cost for different approaches. Adapted from Tchakoua et al. [11].

1.3.4 Prescriptive Maintenance (RxM)

Prescriptive Maintenance (RxM) represents the current apex of maintenance strategy evolution, moving beyond prediction to actively recommend optimal solutions. As illustrated in Figure 2, it is the logical next step after predictive maintenance, answering not just "When will it fail?" but also "What should we do about it?". This advanced strategy fundamentally changes the role of maintenance from a necessary operational activity to a dynamic, profit-driving component of the business.

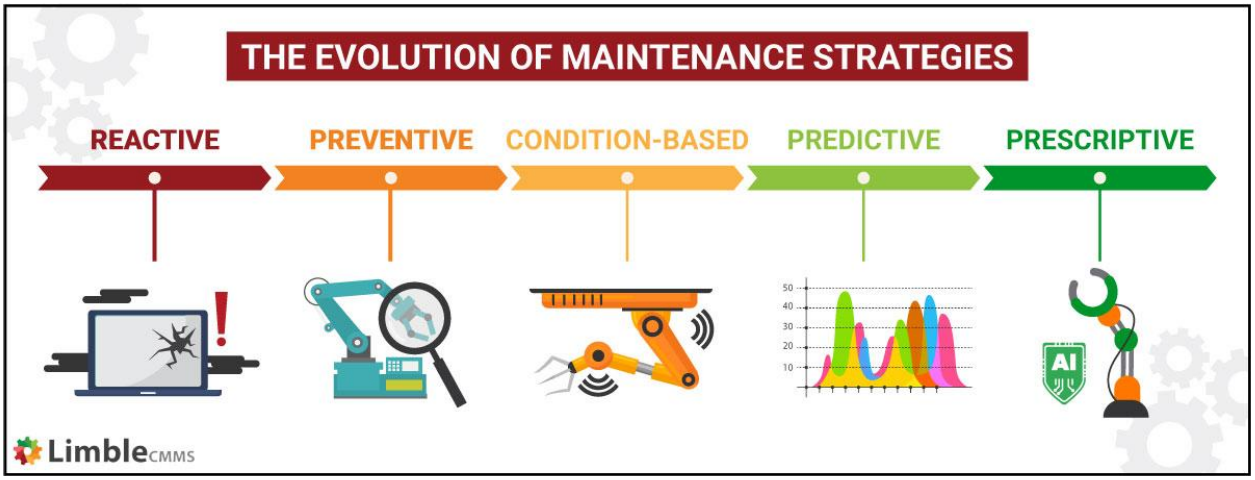


Figure 6 - An expanded view of the evolution of maintenance strategies. Adapted from Limble CMMS [12].

While predictive maintenance focuses on forecasting a future failure event, prescriptive maintenance integrates that prediction with a broader operational context. It leverages artificial intelligence, particularly machine learning and AI-driven simulations, to analyze a variety of potential scenarios and their outcomes. For a predicted upcoming failure, a prescriptive system might evaluate several possible responses:

- Immediate repair: what would be the cost in terms of parts, labor, and immediate downtime?
- Continued operation at reduced capacity: could the machine run at a lower speed or load, to extend its life until a scheduled maintenance window, and what would be the impact on production targets?
- Deferring maintenance: What is the increasing risk of catastrophic failure if the repair is postponed, and what are the potential costs associated with that risk?

The system then weighs these options against business-level objectives, such as production goals, energy costs, supply chain logistics, and maintenance budgets. Based on this comprehensive analysis, it provides a set of clear, actionable recommendations—or "prescriptions" to human operators. For example, it might suggest, "Reduce the machine's operating speed by 15% to safely extend its life by 72 hours, which will align the repair with the planned plant-wide shutdown on Saturday, avoiding 8 hours of lost production."

By not only identifying future problems but also prescribing the most efficient and cost-effective solutions, RxM moves maintenance towards a highly autonomous and optimized process, minimizing human intervention for complex decision-making and maximizing the overall profitability and reliability of the industrial operation.[12]

1.4 Condition Monitoring: The Cornerstone of Predictive Maintenance

As outlined in the hierarchy of maintenance strategies, the successful implementation of both Condition-Based and Predictive Maintenance is fundamentally dependent on Condition Monitoring (CM). CM is the process of systematically collecting and analyzing data to track the health of machinery over time.[13], [14] It provides the essential real-time information required to detect changes that could indicate a developing fault.[13]

Historically, condition monitoring was a manual and often subjective process, relying on the experience of engineers to inspect machinery and assess its operational state based on sensory inputs like unusual sounds, vibrations, or heat. Modern CM, however, leverages a wide array of sensors and non-destructive testing technologies to continuously collect objective data on key performance indicators. Some of the most common and effective techniques include:

1. **Vibration Analysis:** A widely used method for rotating machinery, as changes in vibration patterns can indicate issues like imbalance, misalignment, or bearing wear.[15], [16]
2. **Thermal Imaging (Infrared Thermography):** Used to detect abnormal temperature profiles, which can signify electrical problems (e.g., loose connections), lubrication issues, or excessive friction.[17], [18]
3. **Oil Analysis:** Involves analyzing the physical and chemical properties of lubricants to detect the presence of wear particles from internal components or contamination from external sources.[17], [18]
4. **Acoustic Analysis:** Utilizes microphones and specialized sensors to capture the sound patterns of machinery, identifying anomalies such as leaks in compressed air systems or the early stages of bearing faults.[18]
5. **Ultrasound Analysis:** This technique focuses on detecting high-frequency sound waves (typically above 20 kHz) that are generated by machinery during operation and are inaudible to the human ear. These ultrasonic waves are produced by phenomena such as friction, turbulence, or impacts. Technology is highly effective for early fault detection in mechanical and electrical assets, as changes in ultrasound signatures can provide the first indication of

a problem. Common applications include detecting bearing faults (e.g., lack of lubrication, early-stage spalling), identifying leaks in compressed gas or vacuum systems, and pinpointing electrical issues like arcing, tracking, or corona discharge. The directional nature of high-frequency sound allows for precise source location even in noisy industrial environments.[19]

6. **Motor Current Signature Analysis (MCSA):** Analyzes the electrical current drawn by a motor to detect electrical or mechanical faults within the motor or the equipment driven.

By continuously monitoring these and other relevant parameters, CM provides a rich, high-velocity data stream that serves as the essential input for making informed, data-driven maintenance decisions.[15]

By monitoring these and other relevant parameters, CM provides a rich, high-velocity data stream that serves as the essential input for making informed, data-driven maintenance decisions. The successful implementation of these various CM techniques translates directly into a wide spectrum of operational and financial benefits, as summarized in Figure 7. The foundational advantage is the Early Detection of incipient faults, which enables proactive intervention. This capability initiates a cascade of positive outcomes: Reliability Improvement is achieved by minimizing unexpected failures, which in turn leads to Safety Enhancement by mitigating the risks of catastrophic breakdowns. From a financial standpoint, this proactive stance drives significant cost reductions by avoiding the high expenses of unplanned downtime and secondary damage. Ultimately, the integration of CM elevates maintenance from a reactive necessity to a strategic, Data-Driven Decision-making process, directly resulting in an Improved Return on Investment (ROI) for the organization's physical assets [20].

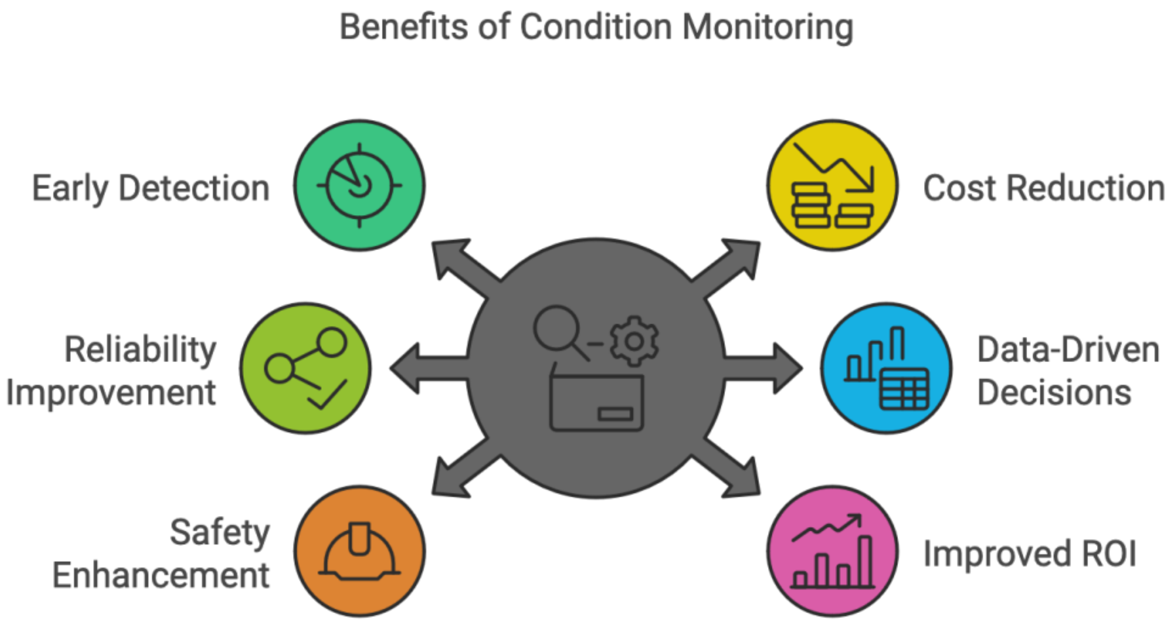


Figure 7 - Benefits of Condition Monitoring. Source: [20].

1.5 The Global Predictive Maintenance Market: Trends and Forecasts

The evolution from traditional to predictive maintenance is not only a technological shift but also a significant economic and industrial trend. The concepts discussed in this chapter are now at the core of a rapidly expanding global market, which is transitioning from a niche application to a mainstream competitive necessity.

Market analysis shows that the predictive maintenance market is valued at USD 14.09 billion in 2025 and is forecast to reach USD 63.64 billion by 2030, advancing at a remarkable Compound Annual Growth Rate (CAGR) of 35.2% [21]. This substantial growth is propelled by several key drivers, including declining sensor costs, the convergence of edge and cloud computing, and wider industrial digitization across asset-intensive sectors. Enterprises now view advanced maintenance as essential, as AI models can flag potential failures weeks or months in advance, enabling precise scheduling of repairs and optimal resource allocation. The scalability of cloud platforms removes traditional infrastructure barriers, while edge analytics lowers latency, making solutions viable for remote sites. While global supply-chain volatility has impacted hardware prices, these pressures also spur innovation in more efficient, on-device processing architectures.

However, this rapid expansion is not without its challenges. The industry faces two significant hurdles that can restrain the pace of adoption: data security vulnerabilities and a shortage of skilled talent.

- **Data-Security and Privacy Gaps**: The proliferation of connected sensors in operational technology (OT) environments significantly enlarges the attack surface, exposing critical systems to cyberthreats. Many legacy industrial protocols lack native encryption, necessitating the retrofitting of secure tunneling and authentication layers to comply with regulations such as GDPR and NERC CIP. A persistent challenge remains in balancing the need for real-time data flows with the implementation of strict security policies, a critical concern in sectors like healthcare and energy where safety and stability are paramount. Consequently, enterprises are increasingly prioritizing platforms that feature device-level encryption, zero-trust architectures, and audit-ready data-handling practices.
- **Skilled-Talent Shortage**: Successful predictive maintenance programs require hybrid expertise in mechanical engineering, data science, and network security; however, the labor market supplies few professionals with this triple skill set. A skills gap often exists where traditional technicians require upskilling to interpret model outputs, while data scientists may lack the specific domain context to recognize genuine equipment degradation signatures. This shortage is exacerbated by a pipeline lag, as academic institutions are only now beginning to integrate curricula covering OT analytics. While managed-service providers and emerging AI-enabled copilots can help bridge this gap, they do not fully offset the need for deep institutional knowledge.

Despite these challenges, a deeper analysis of the market reveals key trends across its various segments:

- **Components**: The market is comprised of hardware, software, and service components. As illustrated in Figure 8, hardware (including sensors, gateways, and other data acquisition devices) claimed a foundational 45.7% of the market share in 2024. However, the value is increasingly shifting towards analytics, with the software segment projected to expand at the fastest rate, registering a 36.5% CAGR through 2030.

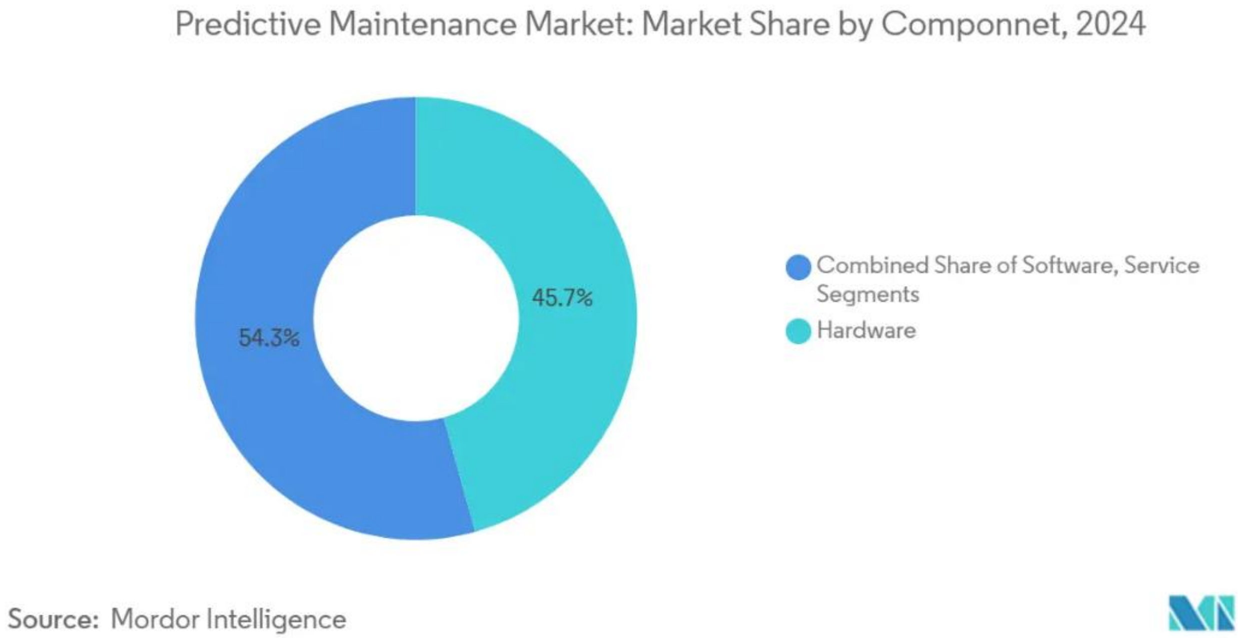


Figure 8 - Predictive Maintenance Market: Market Share by Component, 2024. Source: Mordor Intelligence [21].

- **Deployment**: The method of deployment is a critical factor, with a clear preference for cloud-based solutions. Figure 9 shows that cloud platforms represented a dominant 67.2% of the market in 2024. The flexibility, scalability, and lower upfront capital expenditure of cloud deployment are driving its rapid adoption, with a projected CAGR of 37.9%.

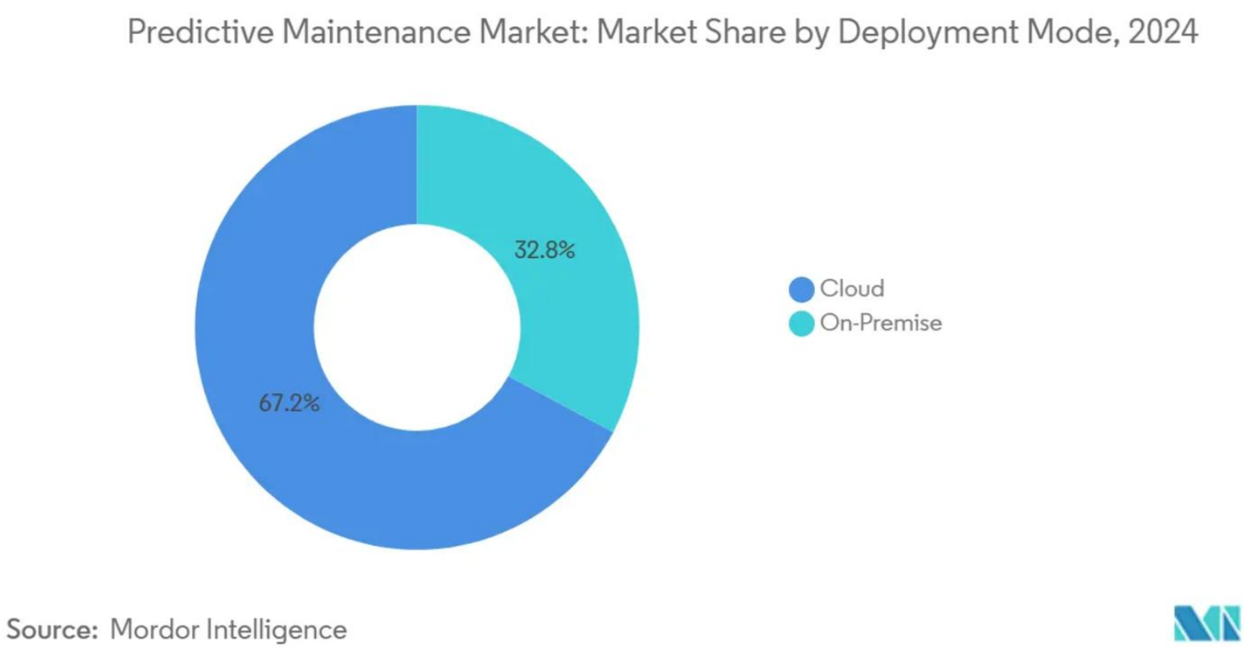


Figure 9 - Predictive Maintenance Market: Market Share by Deployment Mode, 2024. Source: Mordor Intelligence [21].

- **Adopters and Industries**: Large enterprises, with their significant capital and vast fleets of assets, held a majority 64.3% revenue share in 2024. However, the increasing accessibility of cloud-based solutions is democratizing this technology, with small and medium enterprises (SMEs) projected to record the highest CAGR at 37.3% through 2030. In terms of industry, industrial manufacturing led adoption with a 23.4% revenue share in 2024, while the energy and utilities segment is forecast to be one of the fastest-growing, with a projected annual growth of 35.3%.
- **Geography**: While North America commanded the largest revenue share in 2024 with 29.3%, driven by its mature industrial and technological sectors, the most rapid growth is expected elsewhere. As visualized in the global growth map in Figure 10, the Asia-Pacific region is projected to progress at the highest CAGR of 36.1% through 2030, fueled by massive industrialization and widespread "smart factory" initiatives.

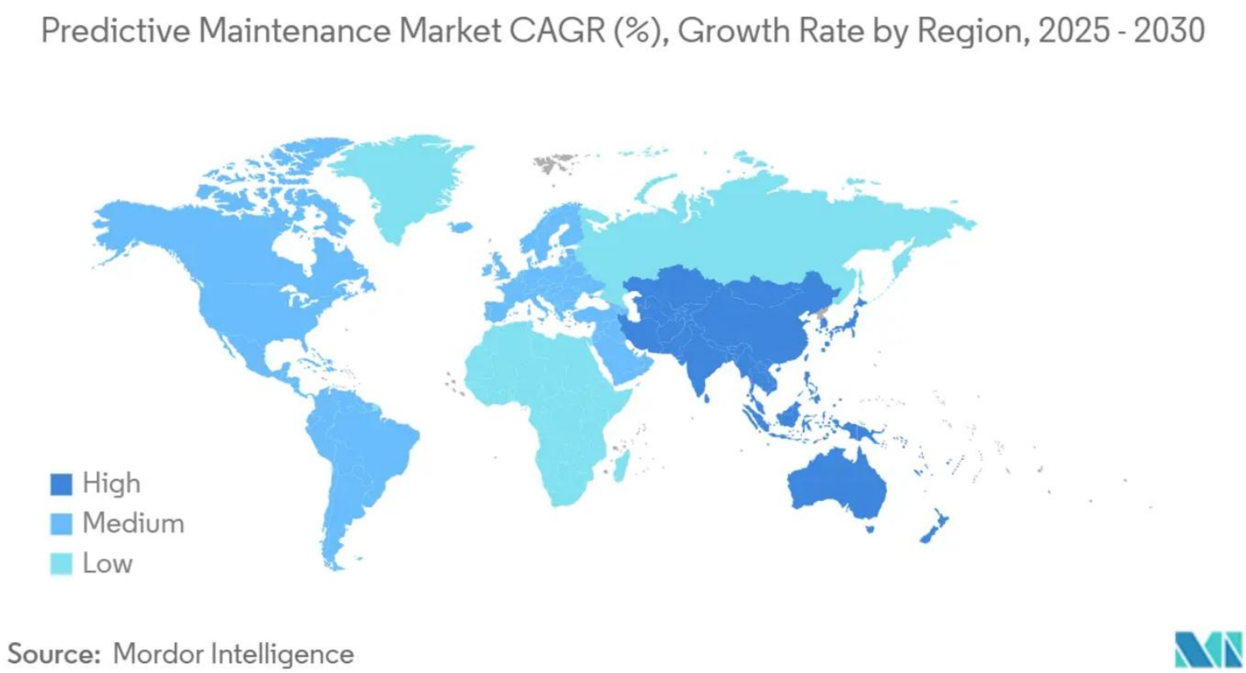


Figure 10 - Predictive Maintenance Market CAGR (%), Growth Rate by Region, 2025 - 2030. Source: Mordor Intelligence [21].

This significant and sustained market growth across all segments and geographies underscores the critical need for the advanced, reliable, and scalable machine learning models that form the central subject of this research.

1.6 The Advent of Artificial Intelligence in Predictive Maintenance

The proliferation of affordable sensors and the rise of the Industrial Internet of Things (IIoT) have created a paradigm shift in data availability. Industrial machinery can now be instrumented to generate vast quantities of operational data, often in the range of terabytes. While this data is immensely valuable, its sheer volume, velocity, and variety (often referred to as the "Big Data" challenge) exceed the capacity for traditional analysis methods and manual interpretation. This is where Artificial Intelligence (AI) and its subfield, Machine Learning (ML), have become transformative enablers for PdM.

AI-powered PdM utilizes sophisticated algorithms to autonomously analyze these large datasets, uncovering subtle patterns, complex correlations, and anomalies that would be imperceptible to human operators. These technologies can learn a machine's normal operating behavior, its unique "digital signature", and therefore detect faint signals of impending failure with greater speed and accuracy than ever before. By integrating AI with the rich data streams from condition monitoring, organizations can transition from a reactive or time-based approach to a truly predictive, and ultimately prescriptive, maintenance strategy. This data-driven intelligence not only predicts that a failure will occur but also provides insights into why it is occurring, thereby moving industrial maintenance into a new era of efficiency and reliability.

2. Foundations and Applications of Artificial Intelligence for Predictive Maintenance

The transition from condition-based to predictive maintenance, as outlined in the previous chapter, is fundamentally enabled by the application of Artificial Intelligence (AI). The vast and complex datasets generated by modern condition monitoring systems require computational tools that can move beyond pre-programmed instructions to autonomously learn from data, identify intricate patterns, and make intelligent predictions. This chapter provides a detailed overview of the primary AI methodologies, and specifically the Machine Learning (ML) models, that form the technical core of the research presented in this thesis.

2.1 Computational Intelligence Overview

Artificial Intelligence is a broad and transformative field of computer science dedicated to creating systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, and language understanding. The conceptual origins of AI can be traced back to the mid-20th century, with the 1956 Dartmouth Workshop being a seminal event that coined the term and established AI as a formal academic discipline. The early years were characterized by great optimism but were followed by periods of reduced funding and interest, known as "AI winters," largely because the computational power and data required to realize the theoretical concepts were not yet available.

In recent years, AI has experienced a renaissance, driven by a confluence of three key factors:

- Availability of Big Data: The proliferation of sensors, the Industrial Internet of Things (IIoT), and vast digital storage capabilities have created an unprecedented volume of data, providing the raw material necessary for training sophisticated algorithms.
- Advancements in Computational Power: The development of powerful hardware, most notably Graphics Processing Units (GPUs), has provided the massive parallel processing capability required for the complex matrix computations inherent in modern AI models.
- Algorithmic Innovations: Significant breakthroughs in the design and training of algorithms, particularly in the sub-fields of Machine Learning and Deep Learning, have made it possible to solve complex problems that were previously intractable.

To navigate this complex landscape, it is essential to clarify the terminology and understand the hierarchical and overlapping relationships between its key domains. This is effectively visualized in Figure 11, which presents a broad overview of computational intelligence [22].

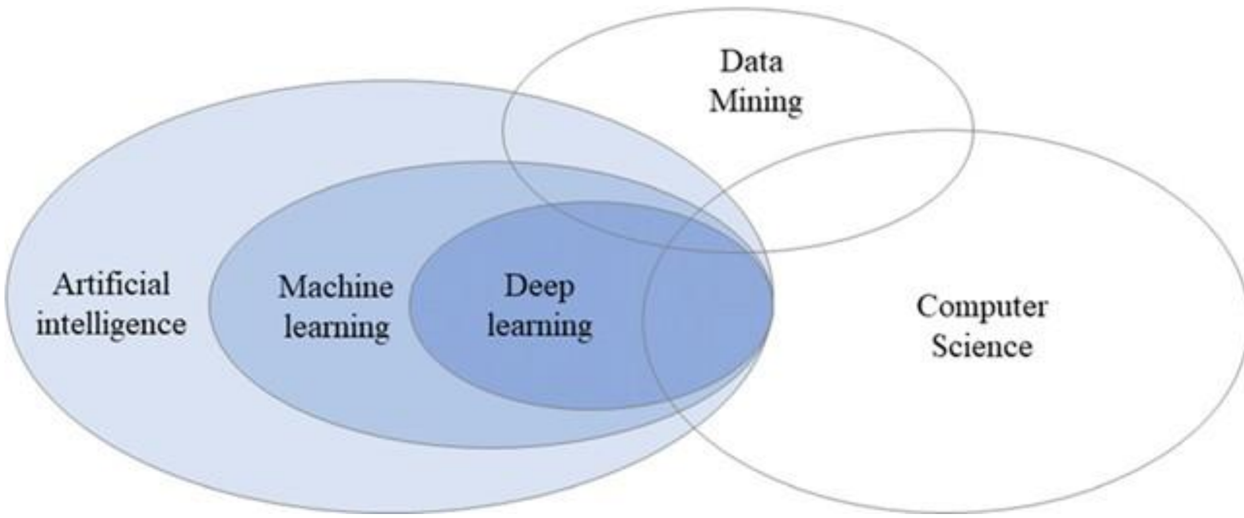


Figure 11 - Venn diagram representing the relationship between Computational Intelligence fields. Adapted from Zhang et al. [22].

At the broadest level is Computer Science, the fundamental study of computation and information processing, which provides the theoretical and practical foundations for all related fields. Within computer science, two major areas of focus are Artificial Intelligence and Data Mining. As previously mentioned, Artificial Intelligence (AI) is the broad endeavor to create intelligent machines that can perform tasks requiring human-like intelligence, such as perception, reasoning, and decision-making. Concurrently, Data Mining is the process of discovering patterns, anomalies, and valuable insights from large datasets, often employing techniques from other domains to achieve its goals.

The intersection of these fields has given rise to Machine Learning (ML), a critical subfield of AI that focuses on training algorithms to learn patterns directly from data without being explicitly programmed. It allows machines to automatically improve their performance on a specific task by learning from experience, providing the practical "learning" capability to AI systems.

A specialized subset of machine learning that has driven many of the most recent breakthroughs is Deep Learning (DL). It utilizes complex, multi-layered artificial neural networks to learn intricate representations of data, making it particularly well-suited for tasks like image and speech recognition. The term 'deep' in Deep Learning refers directly to the high number of layers in these networks (a 'Deep Network'). Historically, training such deep networks was computationally prohibitive. However, the recent advent of powerful parallel processing hardware, particularly Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), have made it feasible to manage the immense computational load, unlocking the full potential of Deep Learning for solving complex numerical problems.

Despite the remarkable power of Deep Learning, it is crucial to approach its application with a strategic mindset. The most powerful tool is not always the most appropriate one. For certain well-defined problems, simpler machine learning models may offer comparable performance with significantly less computational overhead and greater interpretability. In essence, it is not useful to hit a mosquito with a big gun. Therefore, the selection of a model should be driven by the specific requirements of the maintenance task at hand. The following sections will explore the application of various models to the key PdM tasks of anomaly detection, damage classification, and RUL prediction.

The choice of the appropriate model depends on the specific goal, the nature of the available data, and the criticality of the equipment. Generally, these tasks draw from the main paradigms of machine learning: unsupervised learning, which is ideal for anomaly detection when labeled fault data is scarce, and supervised learning, which is used for classification and regression tasks when historical data with known outcomes is available.

2.2 Machine Learning Fundamentals

Following the broad overview of Artificial Intelligence, this section provides a focused examination of its most impactful sub-field for PdM: Machine Learning (ML). At its core, machine learning is a method of data analysis that automates the creation of analytical models. It is founded on the principle that computer systems can learn from data, identify patterns, and make decisions with minimal human intervention, departing from the traditional paradigm of being explicitly programmed for every task. This section deconstructs the fundamental principles that govern how ML models operate, detailing the universal learning process, the primary learning paradigms, and the common tasks these models are designed to solve.

2.2.1 The Machine Learning Process

Regardless of the specific algorithm, the process of "learning" from data in a machine learning model follows a structured, iterative cycle composed of three core components: a model representation, an error function, and an optimization process.

1. Model Representation (The Decision-Making Process): The first step is the selection of a model, which is a mathematical framework or function that takes input data (features) and produces a prediction. This function contains internal parameters that are adjusted during the learning process. The choice of model defines the "hypothesis space," or the range of possible relationships the system can learn. For a simple task, this could be a linear equation $(y = mx + b)$, where the parameters m and b must be learned. For more complex tasks, the model could be a decision tree, a support vector machine, or a multi-layered neural network.
2. Error Function (The Loss Function): To determine how well the model is performing, its predictions are compared to the actual ground-truth outcomes. This comparison is quantified by an error function, also known as a loss or cost function. This function calculates a score that penalizes the model for the discrepancy between its predictions and the true values. For example, in a regression task, a common error function is the Mean Squared Error (MSE), which measures the average squared difference between predicted and actual values. The goal of the learning process is to find the set of model parameters that minimizes this error function.
3. Optimization Process: The core of the "training" phase is the optimization process, where the model's internal parameters are systematically adjusted to minimize the error function. A common and foundational optimization algorithm is Gradient Descent. It works by calculating the gradient (the direction of steepest ascent) of the error function with respect to the model's parameters and then taking a small step in the opposite direction (downhill). This process is repeated iteratively, with the model's parameters being updated at each step, until the error value converges to a minimum. This convergence signifies that the model has learned the best possible mapping from inputs to outputs given the data and its own architecture.

This entire iterative cycle is visually summarized in the workflow presented in Figure 12. The process begins with the Training Data, which serves as the input for the core training phase (Train the ML Algorithm). It is within this step that the previously described components converge: the chosen

Model Representation is systematically adjusted via an Optimization Process with the objective of minimizing the Error Function. The result of this training is a finalized Model with optimized parameters. The model's efficacy is then evaluated by using it to make Predictions on new data and assessing its performance with metrics such as Accuracy. This evaluation is critical: if the performance is unacceptable, the process loops back for further training, often with adjustments to the model's configuration. This feedback loop continues until performance is acceptable, resulting in a Successful Model ready for deployment.

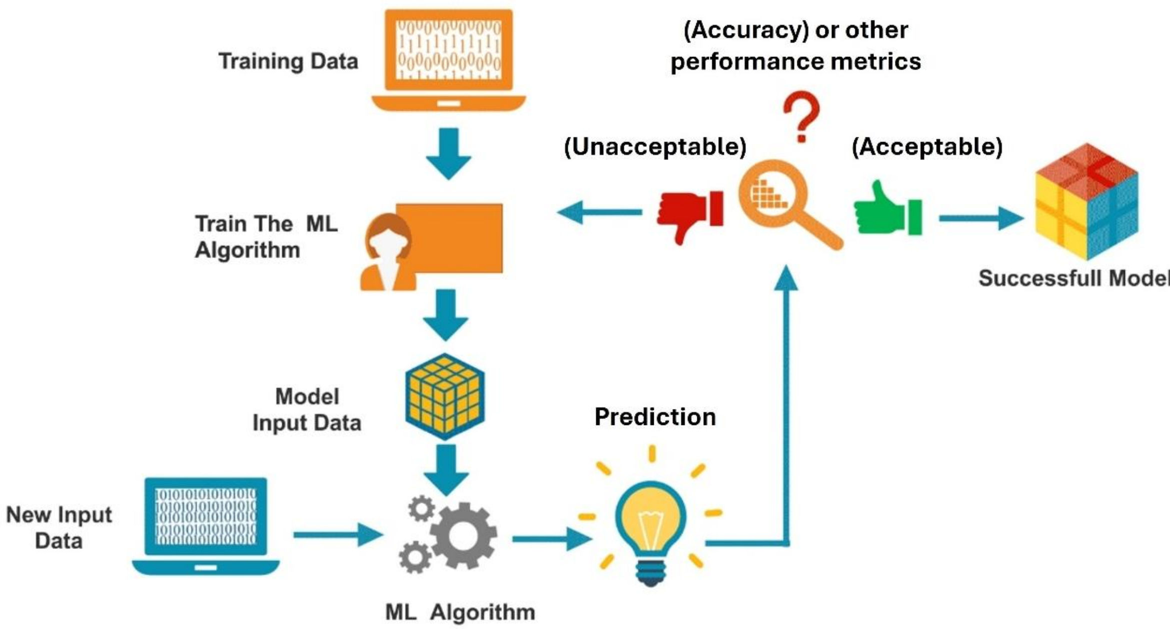


Figure 12 - A visual representation of the iterative Machine Learning process. Based on the standard model described in [23].

2.2.2 Paradigms of Machine Learning

Machine learning algorithms are typically categorized into several learning "paradigms," which are defined by the nature of the data provided and the way learning occurs. These primary paradigms, which are visually summarized in Figure 13, include supervised, unsupervised, semi-supervised, and reinforcement learning.

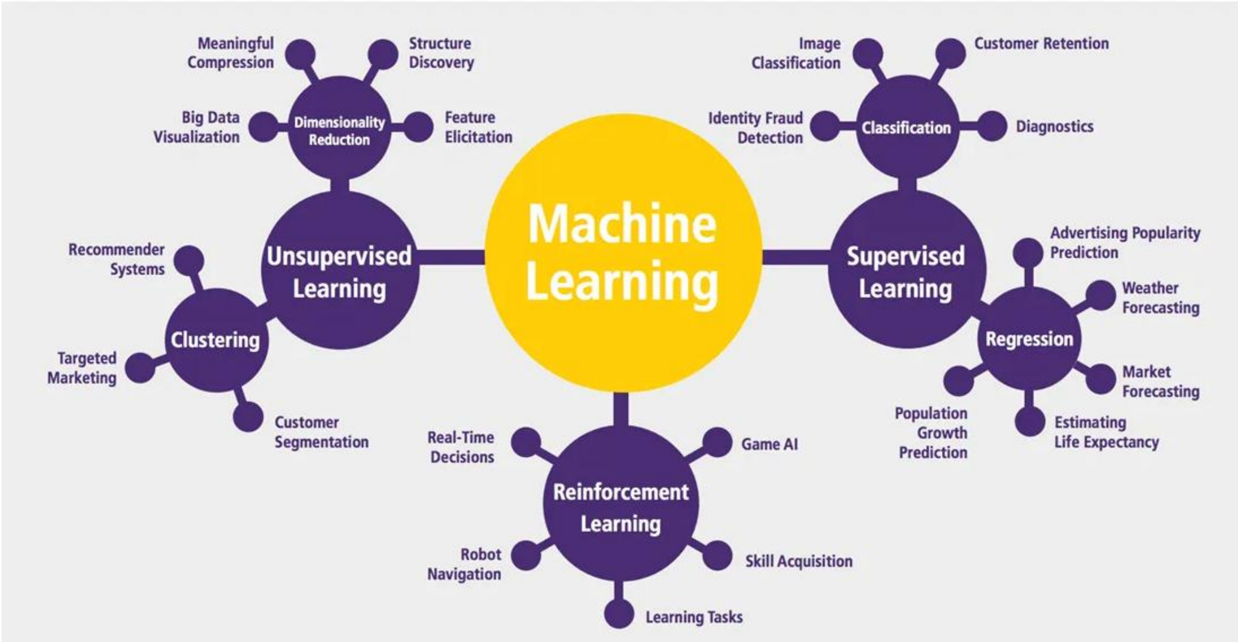


Figure 13 - An overview of the primary paradigms of Machine Learning. Adapted from R. G. [24].

- Supervised Learning: This is the most common paradigm, wherein the algorithm learns from a fully labeled dataset. Each data point consists of a set of input features and a corresponding correct output label. The goal is to learn a mapping function that can accurately predict the output for new, unseen input data. Supervised learning is the foundation for tasks where the desired outcome is known, such as classification and regression.
- Unsupervised Learning: In this paradigm, the algorithm is given a dataset with no predefined labels. The objective is to explore the data and find meaningful structures, patterns, or groupings within it. Instead of predicting a known outcome, unsupervised learning seeks to uncover the underlying distribution of the data. It is the primary approach for tasks like clustering and anomaly detection.
- Semi-Supervised Learning: This approach serves as a bridge between supervised and unsupervised learning, applied in situations with a large amount of unlabeled data but only a small quantity of labeled data. The algorithm leverages the structure found in the large unlabeled dataset to improve the learning process and predictive performance on the small, labeled set. This is particularly useful in industrial settings where data collection is easy but labeling it (e.g., confirming a specific fault) requires costly expert intervention.

- Reinforcement Learning: This paradigm is concerned with learning through interaction. An "agent" (the model) learns to make a sequence of decisions in a dynamic "environment" to maximize a cumulative "reward." Through a process of trial and error, the agent learns which actions yield the best outcomes over time. While highly effective for control systems and autonomous decision-making, it is less commonly applied to the core PdM tasks of classification and RUL prediction that are the focus of this thesis.

2.2.3 Common Machine Learning Tasks

The learning paradigms described above are applied to solve several fundamental types of problems, or "tasks." The three most relevant to predictive maintenance are:

- Classification: A supervised learning task where the goal is to predict a discrete, categorical label. The model learns from labeled data to assign new, unseen data points to one of several predefined classes. In a PdM context, this translates to tasks like diagnosing a specific fault ("bearing fault," "gear mesh fault," "misalignment") or making a binary decision ("healthy" vs. "faulty").
- Regression: A supervised learning task where the objective is to predict a continuous, numerical value. The model learns the relationship between input features and a continuous output variable. The primary application of regression in predictive maintenance is the prediction of Remaining Useful Life (RUL), where the output is a value representing time, cycles, or distance to failure.
- Clustering: An unsupervised learning task that involves automatically grouping a set of data points into clusters. The principle is to place data points that are more similar to each other in the same cluster, while placing dissimilar points in different clusters. In an industrial setting, clustering can be used to automatically identify different operational states or regimes of a machine without needing prior labels.

2.3.4 Dimensionality Reduction

In addition to the primary predictive tasks, a crucial set of techniques in the machine learning workflow involves **Dimensionality Reduction**. Data from Condition Monitoring (CM) systems, originating from multi-sensor arrays and high sampling rates, is often inherently high-dimensional. This "curse of dimensionality" can lead to increased computational complexity, a higher risk of model overfitting, and can make it difficult to visualize and interpret the data.

Dimensionality reduction encompasses a suite of unsupervised learning methods that aim to transform data from a high-dimensional space into a meaningful representation of lower dimension. The primary goal is to reduce the number of input variables (features) while retaining as much of the critical information and underlying structure of the original data as possible. This is typically achieved through two main approaches:

- **Feature Selection:** This involves identifying and selecting a subset of the most relevant original features to use in the model, while discarding redundant or irrelevant ones.
- **Feature Extraction:** This involves creating a smaller set of new, synthetic features by combining and transforming the original features. These new features are designed to summarize the most important information from the original dataset.

In the context of Predictive Maintenance (PdM), dimensionality reduction is a critical pre-processing step that provides several key benefits:

- **Improved Model Performance and Robustness:** By removing noisy and redundant features, it can help subsequent models (e.g., for anomaly detection) to learn the underlying patterns more effectively.
- **Reduced Computational Cost:** Fewer features lead to significantly faster model training and inference times, which is essential for real-time or embedded applications.
- **Enhanced Data Visualization:** Projecting high-dimensional data onto two or three dimensions is the only way to visually inspect its structure, identify natural clusters, and validate the results of anomaly detection models.

The specific dimensionality reduction models employed in this research, such as Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), will be detailed in Chapter 3.

2.3 Machine Learning Applications in Predictive Maintenance

Having established the fundamental principles of machine learning, the focus now shifts to its practical application in solving the core challenges of PdM. The raw data acquired from condition monitoring is not inherently insightful; it requires sophisticated algorithms to transform it into actionable intelligence. The application of machine learning in this domain can be structured as a workflow that addresses a series of progressively more complex questions, moving from simple state detection to detailed diagnosis and future prognosis.

This hierarchical process is clearly illustrated in Figure 14. The workflow begins with Input Data, typically time-series signals from various sensors on the machinery. This data is then fed into a central block of Algorithms, where different machine learning models are applied to answer three critical questions that define the core predictive maintenance tasks:

1. "Is my machine operating normally?" — This question is addressed by Anomaly Detection.
2. "What's wrong with my machine?" — This is the domain of Fault Detection and Diagnosis (also referred to as Damage Classification).
3. "How much longer can I operate my machine?" — This prognostic question is answered by Remaining Useful Life (RUL) Estimation.

The output of this analytical process is a Decision, which provides the clear, data-driven insight needed to schedule maintenance actions efficiently and effectively. The subsequent subchapters will provide a detailed examination of the specific models and methodologies used to address each of these three core tasks in sequence.

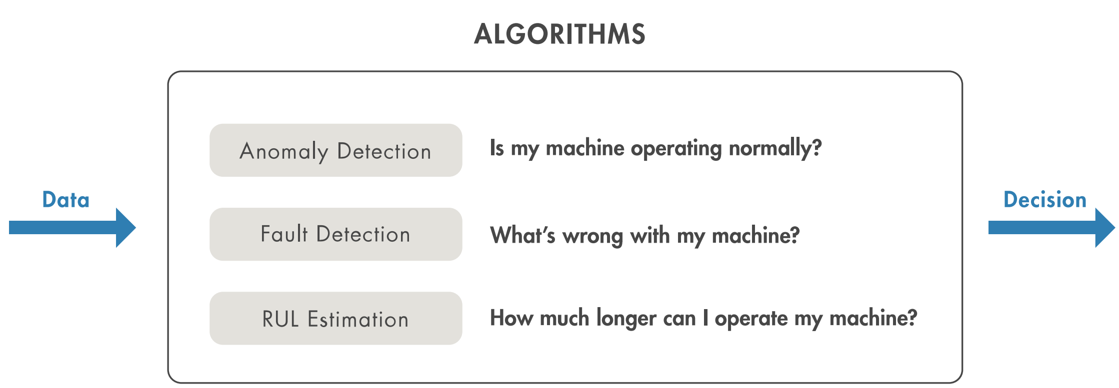


Figure 14 - The core tasks of a predictive maintenance workflow. Adapted from MathWorks [25].

The integration of machine learning into this workflow unlocks a new level of capability, yielding a range of significant benefits that elevate the entire maintenance function, as summarized in Figure 15. The fundamental advantage is the transition to truly Proactive Maintenance; by identifying potential issues long before they become critical, organizations can move decisively away from a costly reactive model. This proactive stance directly translates into substantial Cost Savings, realized by minimizing expensive unplanned downtime, avoiding secondary damage to equipment, and optimizing the allocation of maintenance resources.

Operationally, this leads to Improved Equipment Performance, ensuring that assets operate at peak efficiency, which in turn enhances output and product quality. A direct consequence of preventing severe degradation and catastrophic failures is an Increased Equipment Lifespan, which maximizes the return on capital-intensive assets. Furthermore, by ensuring machinery operates consistently within safe parameters, Enhanced Safety becomes a critical benefit, reducing the risk of workplace accidents associated with equipment malfunctions. Underpinning all these advantages are the Data-Driven Insights generated by the machine learning models. These algorithms can analyze vast and complex datasets to uncover subtle trends and failure patterns that would otherwise go unnoticed, transforming raw data into actionable intelligence for more informed strategic decisions about maintenance, operations, and even future asset procurement.[26]

Figure 15 - Benefits of Machine Learning for Predictive Maintenance, Source: [27]

2.3.1 Anomaly Detection

Anomaly detection is often the first step in identifying a potential problem. It is the process of identifying data points or patterns that deviate significantly from the expected or normal behavior [28], [29]. In an industrial context, an anomaly could be an unusual vibration signature or a sudden temperature spike, signaling a potential fault [16]. Machine learning models, particularly unsupervised learning algorithms, are highly effective at learning the normal operational profile of a machine and flagging any deviations as anomalies, thus providing an early warning of developing issues [28].

Anomaly detection is often the first critical step in a data-driven PdM strategy. It is the process of identifying data points, events, or observations that deviate significantly from the established Normal Operating Condition (NOC) of a system. In the context of industrial machinery, an anomaly could be an unusual vibration signature, a sudden temperature spike, or a subtle change in power consumption that signals the incipient stage of a fault. The primary goal is not to identify the type of fault, but simply to flag that something is abnormal, thereby triggering an alert for further investigation.

This approach is particularly valuable because data from faulty conditions is often rare, expensive, or impossible to collect for all potential failure modes. Therefore, many anomaly detection models are built using unsupervised or semi-supervised learning techniques, which learn the characteristics of normal behavior from a large corpus of healthy-state data.

Within the broader task of anomaly detection, it is useful to distinguish between two subtly different, yet critical, sub-problems: outlier detection and novelty detection [30], [31].

- Outlier Detection: Often referred to as unsupervised anomaly detection, this approach assumes that the training dataset itself is contaminated with outliers—anomalous data points that are already present within the "normal" data. The goal of an outlier detection algorithm is to fit a model to the regions where the training data is most concentrated, effectively ignoring or isolating these pre-existing deviant observations.
- Novelty Detection: Conversely, this is considered a semi-supervised anomaly detection task. This approach operates under the assumption that the training data is "clean," consisting exclusively of normal operating data. The objective is to train a model that learns the precise boundary of this normal behavior, such that it can identify any new, previously unseen data points that fall outside this boundary as novelties or anomalies.

In the context of predictive maintenance, the choice between these two approaches depends on the available data and the specific objective. Outlier detection is most appropriate when analyzing large historical datasets where the data integrity is uncertain, and there is a possibility that unlabeled faults or transient events are already present. The goal here is often to "clean" the dataset or identify these past, undocumented anomalies. Novelty detection is the more common and powerful paradigm for real-time condition monitoring. In this scenario, a model is trained on a trusted baseline dataset, often captured during the machine's initial commissioning phase when it is known to be perfectly healthy. The model is then deployed to monitor the live machine, and any deviation from this learned "healthy" profile is immediately flagged as a novelty, representing the earliest possible sign of degradation or a developing fault.

A broad survey of methodologies is provided in foundational texts on the subject, such as the comprehensive *Outlier Analysis* by Aggarwal [32]. The primary categories of techniques used for anomaly detection include:

- Statistical Methods: These methods build a statistical model of normal data and flag any points that have a low probability of being generated by that model. Techniques like the Z-score, Mahalanobis distance, or Gaussian Mixture Models (GMM) are used to quantify how far a data point lies from the "normal" distribution.

- Density-based Methods: These algorithms identify anomalies by considering the density of data points in a feature space.
    - Clustering (e.g., DBSCAN): Algorithms such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) group similar, high-density data points into clusters. Normal operating data will form dense clusters, while anomalies will be identified as points that do not belong to any cluster (i.e., noise).
    - Local Outlier Factor (LOF): This algorithm computes a score reflecting the local density deviation of a data point with respect to its neighbors. It is particularly effective because it considers varying densities within a dataset. An object is identified as an outlier if its local density is significantly lower than that of its neighbors, indicating it is in a sparser region.
- Classification-based Methods: These methods aim to learn a boundary that separates normal data from anomalous regions.
    - One-Class Support Vector Machine (OC-SVM): As a semi-supervised method, OC-SVM is trained on a dataset containing only normal data. It learns a hypersphere or hyperplane that encloses the majority of the normal data points in a high-dimensional feature space. Any new data point that falls outside this learned boundary is classified as an anomaly.
    - Support Vector Machine (SVM): In its standard, dual form, the SVM is a powerful supervised learning algorithm. Unlike the OC-SVM, this approach requires a training dataset that contains labeled examples of both the normal condition and one or more anomalous/faulty conditions. The SVM algorithm then works by finding an optimal separating hyperplane that maximizes the margin between the two classes. This makes it a highly effective tool for fault classification when historical examples of specific failures are available.[33]
- Isolation-based Methods:
    - Isolation Forest: This is an efficient algorithm based on decision trees. The core idea is that anomalies are "few and different," which makes them easier to "isolate" in a tree structure. The algorithm builds an ensemble of trees, and anomalous points will

require, on average, fewer random partitions to be isolated compared to normal points.

- Reconstruction-based Methods (Deep Learning):
    - Autoencoders: A powerful deep learning technique, autoencoders are a type of neural network trained to reconstruct their input. When an autoencoder is trained exclusively on data from normal operating conditions, it learns to accurately reconstruct healthy data. When it is later presented with anomalous data, which follows a different pattern, the network will struggle to reconstruct it, resulting in a high "reconstruction error." This error can be used as an anomaly score to flag potential faults.

Many of the algorithms discussed, including LOF, Isolation Forest, and GMM, are readily implemented in open-source Python toolkits specifically designed for outlier detection, such as the PyOD library [34].

In Figure 16, a general overview of some of the most important Anomaly Detection Methods are reported divided by affiliation categories.

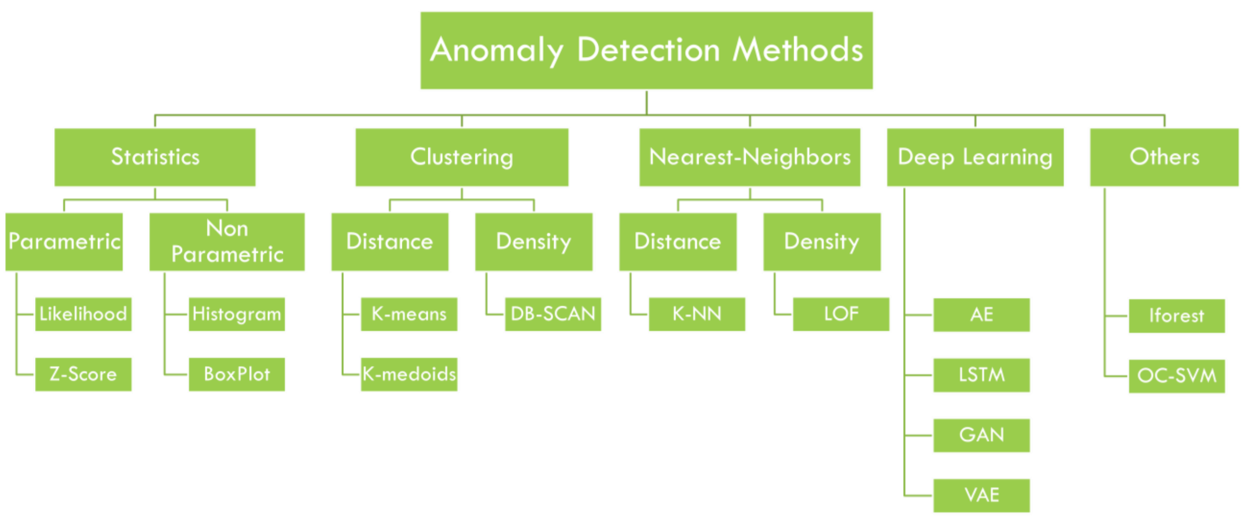


Figure 16 - Anomaly Detection Methods general overview

2.3.2 Fault Detection (Damage Classification)

Once an anomaly has been detected, the next logical and more valuable step is to diagnose the specific type of fault. This is the domain of damage classification. The objective is to automatically assign a label to the detected anomaly, such as "bearing inner race fault," "gear tooth crack," or "shaft misalignment." An accurate diagnosis is crucial for planning maintenance activities effectively, as it allows engineers to prepare the correct tools, spare parts, and personnel, thereby minimizing repair time and costs.

Unlike anomaly detection, damage classification is typically framed as a supervised learning problem. This requires a historical dataset that contains examples of various fault conditions, with each example explicitly labeled with its corresponding fault type. This labeled data is used to train a classification model that learns the relationship between the sensor readings (features) and the specific fault classes.

Key machine learning models for damage classification include:

- Traditional Machine Learning Classifiers: Algorithms like Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests are widely used. These models often require manual feature engineering, where domain experts extract meaningful statistical or frequency-domain features (e.g., RMS, kurtosis, spectral components) from the raw sensor signals to serve as inputs to the model.
- Convolutional Neural Networks (CNNs): CNNs have become state-of-the-art for classification tasks involving raw signal data. They have the significant advantage of being able to automatically learn relevant features directly from the data, eliminating the need for manual feature engineering. For instance, time-series data from vibration sensors can be converted into 2D representations like spectrograms or scalograms, which can then be fed into a CNN as if they were images. The network learns to recognize the unique visual patterns associated with each fault type.

2.3.3 Remaining Useful Life (RUL) Prediction

The estimation of Remaining Useful Life (RUL) represents the pinnacle of PdM, as it directly addresses the question of "how much longer can this component operate safely?" RUL is defined as the expected length of time from the present to the end of the useful life of a component. An accurate RUL prediction enables just-in-time maintenance scheduling, which maximizes the lifespan of components, optimizes spare parts inventory, and allows for downtime to be planned with minimal disruption to production.

RUL prediction is typically formulated as a regression problem in machine learning, where the model is trained to predict a continuous value (time, cycles, distance) based on the evolution of condition monitoring data over time. The sensor data should contain a degradation trend that the model can learn and extrapolate into the future.

The approaches to RUL prediction can be broadly categorized as follows:

- Physics-based Models: These models use mathematical equations based on the physics of the degradation process (e.g., fatigue crack growth laws) to predict failure. They can be highly accurate but require deep domain knowledge and are often difficult to develop for complex systems.
- Data-driven Models: These models rely entirely on historical data to learn the degradation patterns.
    - Survival Models: Statistical methods like Weibull analysis or Cox Proportional-Hazards models are used to model the probability of failure over time based on historical failure data.
    - Machine Learning Regression Models: Algorithms such as Support Vector Regression (SVR) or Gradient Boosting can predict RUL based on extracted features that correlate with system degradation.
    - Deep Learning Models: Recurrent Neural Networks (RNNs), and particularly their advanced variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, are exceptionally well-suited for RUL prediction. These networks are designed to handle sequential data, allowing them to learn the temporal dependencies in the degradation signals as a component wears out over time. A common approach is to feed a sequence of sensor data into an LSTM model, which then outputs the predicted RUL value.

3 Model Used in the Research Project

Building upon the foundational concepts of machine learning discussed in Chapter 2, this chapter details the specific algorithms and methodologies that constitute the analytical framework of this research project. The selection of these models was not arbitrary but was driven by the inherent characteristics of condition monitoring data and the primary research objective: to effectively discriminate between healthy and faulty operational states in mechanical components.

Industrial sensor data is often characterized by high dimensionality, complex temporal dependencies, and subtle, non-linear relationships between variables. To address these challenges, a multi-stage analytical approach was adopted. This chapter presents the specific models employed at each stage of this approach.

The chapter is organized into three main sections, reflecting the data analysis workflow used throughout the case studies presented later in this thesis.

1. The first section, **Dimensionality Reduction Models**, addresses the challenge of data complexity. The techniques described here were used to transform high-dimensional sensor data into lower-dimensional representations, which are crucial for effective visualization, feature extraction, and reducing the computational burden on subsequent models.
2. The second section focuses on **Univariate Anomaly Detection**. These methods are fundamental for analyzing individual sensor signals in isolation, providing a baseline for identifying deviations from normal behavior in key operational parameters.
3. The final section details the **Multivariate Anomaly Detection** models. These more sophisticated techniques were employed to capture the complex interrelationships between multiple sensor streams simultaneously, enabling the detection of subtle anomalies that would be invisible to univariate analysis.

Collectively, the models described in this chapter form the core analytical toolkit that was applied to the various industrial datasets investigated in this PhD project.

3.1 Dimensionality Reduction Model

Data from modern industrial environments, often sourced from a multitude of sensors, is inherently high-dimensional. This characteristic presents a significant challenge known as the **"curse of**

dimensionality." While adding more features might seem intuitively beneficial for providing more information to a model, the practical reality is more complex. Beyond a certain point, increasing the number of features without a corresponding exponential increase in the number of training samples can degrade, rather than improve, the performance of a machine learning model.

This phenomenon is often visualized as shown in Figure 17. The curve illustrates that as dimensionality increases, a classifier's performance initially improves because the added features provide more discriminatory information. However, the performance peaks at an optimal number of features. Further increasing dimensionality leads to a performance decline.

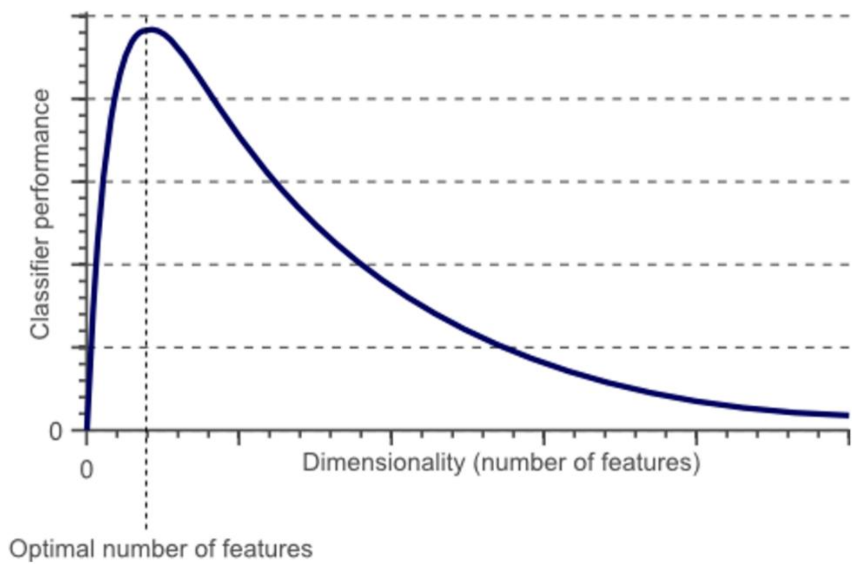


Figure 17 - Model Performance vs. Dimensionality (number of features), Source: [35].

This decline occurs because, as the number of dimensions grows, the volume of the feature space expands exponentially. The available training data becomes increasingly sparse within this vast space, meaning the distance between any two data points becomes very large and less meaningful. From a geometric perspective, as the dimensionality of the feature space approaches infinity, the ratio of the difference between the maximum and minimum Euclidean distance from any sample point to all other points, and the minimum distance itself, tends to zero. In simpler terms, in a high-dimensional space, the concept of a "nearest neighbor" becomes ill-defined because all points appear to be almost equidistant from each other [36].

This sparsity directly contributes to overfitting. With a limited number of training samples spread thinly across a high-dimensional space, a model can easily learn the noise and random fluctuations specific to the training set rather than the true, underlying data distribution. When this overfitted model is exposed to new, unseen data, it fails to generalize and its performance suffers.

To combat the curse of dimensionality and the associated risk of overfitting, two primary strategies are employed:

1 **Feature Selection**: These algorithms aim to select the best subset of the original features. They score and rank the existing features based on their relevance and contribution to the predictive task, discarding those that are redundant or irrelevant.

2 **Feature Extraction**: Instead of selecting a subset, these methods create new, more informative features by combining and transforming the original ones. These techniques project the data from a high-dimensional space into a lower-dimensional one, while seeking to preserve the most important information. A primary example of this approach is Principal Component Analysis (PCA), which will be detailed in the following section.

These dimensionality reduction techniques are often used in conjunction with robust model validation methods, such as **cross-validation**. By systematically splitting the training data into multiple subsets for training and validation, cross-validation provides a more reliable estimate of a model's performance on unseen data. This helps in identifying the optimal number of features or components to use, effectively locating the peak of the performance curve and preventing the model from being either too simple (underfitting) or too complex (overfitting).

These techniques transform the data from a high-dimensional space into a feature space of a lower dimension, while aiming to retain as much of the meaningful variance in the data as possible. The application of these models serves several critical purposes within a predictive maintenance workflow. It enables a form of **Meaningful Compression**, which reduces data storage requirements and the computational load of subsequent analyses. More importantly, it facilitates **Feature Elicitation** or **Extraction**, where new, more informative features (or components) are synthesized from the original, often correlated, sensor signals. This process is instrumental for **Structure Discovery**, revealing intrinsic patterns, clusters, or manifolds within the data that are not apparent in its original high-dimensional form. A primary and indispensable application of this is **Big Data Visualization**; by projecting the data onto two or three dimensions, it becomes possible to visually inspect its structure, a crucial step for both exploratory analysis and for communicating the results of clustering or classification models.

The following sections will detail the specific linear and non-linear techniques used in this research to achieve these objectives.

3.1.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a cornerstone of linear dimensionality reduction and is one of the most widely used unsupervised feature extraction techniques. The fundamental principle behind PCA is to transform a dataset of possibly correlated variables into a smaller, new set of linearly uncorrelated variables called principal components. This transformation is designed in such a way that the first principal component accounts for the largest possible variance in the original data, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal (uncorrelated) to the preceding components.

The PCA algorithm achieves this transformation through a series of well-defined mathematical steps:

1. **Standardization**: The process begins by standardizing the original dataset. Each feature is scaled to have a mean of zero and a standard deviation of one (a z-score, this concept will be described in detail in the chapter 3.2.1). This step is crucial because PCA is sensitive to the scale of the original variables; if one feature has a much larger variance than others, it will dominate the first principal component, leading to biased results.
2. **Covariance Matrix Computation**: Next, the covariance matrix of the standardized data is computed. This square matrix describes the variance of each feature and the covariance between each pair of features, effectively capturing the relationships within the data.
3. **Eigenvalue Decomposition**: The core of PCA is the eigenvalue decomposition (or Singular Value Decomposition) of the covariance matrix. This process yields two key outputs:

3.1 **Eigenvectors**: These are the directions of the new axes where the data has the most variance. Each eigenvector represents a principal component. They are orthogonal to each other, forming a new coordinate system for the data.

3.2 **Eigenvalues**: These are scalar values that indicate the magnitude of the variance captured by each corresponding eigenvector (principal component).

4. **Component Selection and Projection**: The eigenvectors are sorted in descending order based on their corresponding eigenvalues. The first principal component (the eigenvector with the largest eigenvalue) is the direction of maximum variance. The second principal component is the direction of the next largest variance, and so on. The primary application of

PCA for dimensionality reduction is achieved by retaining only the first 'k' principal components that cumulatively explain a desired percentage of the total variance (e.g., 95%). Finally, the original high-dimensional data is projected onto this new, lower-dimensional feature space defined by the selected principal components.

In the context of this research, PCA was primarily used for data visualization by projecting complex, multi-sensor data onto the first two or three principal components. This allows for the visual inspection of clusters and the identification of potential anomalies. Additionally, it serves as a **pre-processing step** to reduce the feature space before feeding the data into more complex anomaly detection models, thereby reducing noise and improving computational efficiency.

A key limitation of PCA, however, is its inherent linearity. It assumes that the underlying structure of the data can be best described by a linear subspace, which may not be the case for complex systems with non-linear relationships. Furthermore, the principal components, being linear combinations of all original features, can sometimes be difficult to interpret in a physical sense.

The output of PCA transformation can be effectively visualized as illustrated in Figure 18. Subplot 17(a) provides a geometric interpretation, showing the original data points in a 3D space with the new principal axes overlaid. These axes represent the directions of maximum variance, and the ellipsoid, typically drawn to encompass a set number of standard deviations, visually summarizes the covariance structure and orientation of the data along these new components. Complementing this visual is the scree plot, shown in Subplot 17(b). This plot quantifies the amount of variance captured by each principal component by displaying their corresponding eigenvalues in descending order. The magnitude of each eigenvalue, often annotated with the percentage of total variance it explains, is used to determine the number of components to retain for an effective and meaningful dimensionality reduction.

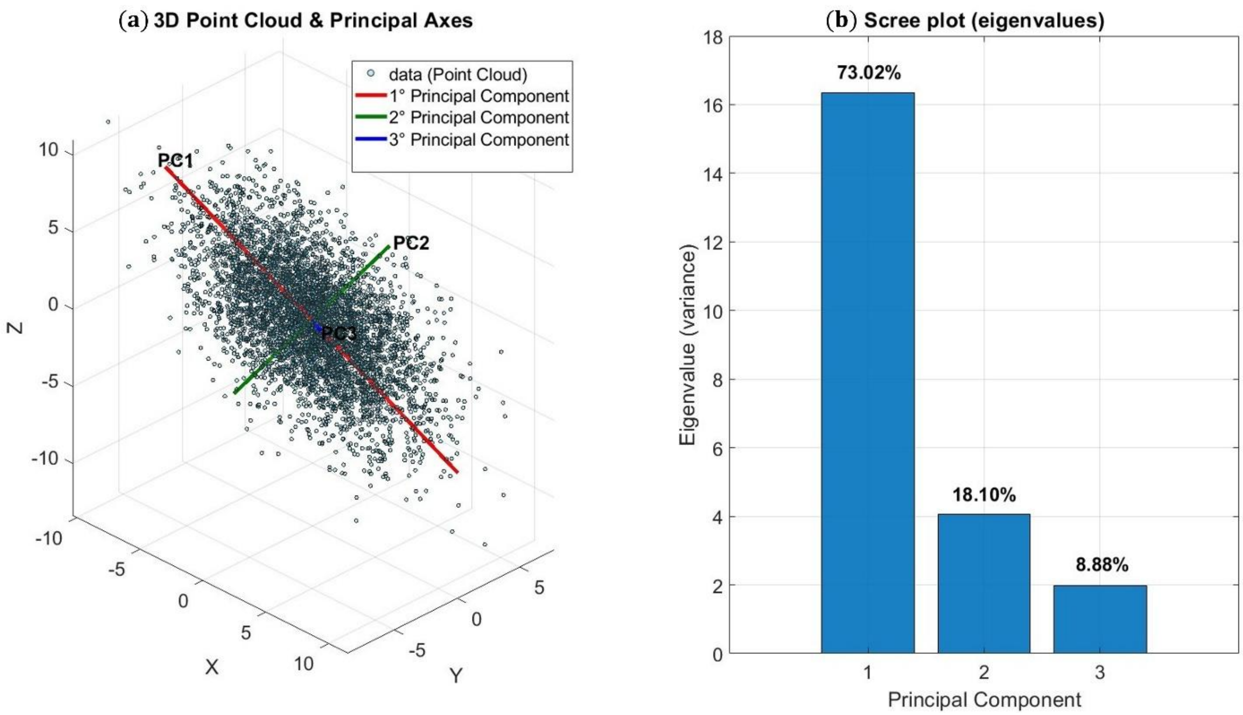


Figure 18 - (a)3D Point Cloud & Principal Axes, (b) Scree plot (eigenvalues).

A practical application of this projection is illustrated in Figure 19, which demonstrates the result of reducing a three-dimensional dataset to a two-dimensional representation. The new 2D space is defined by the first two principal components (PC1 and PC2), which represent the orthogonal axes along which the data exhibits the most variance. The key metric justifying this reduction is the cumulative explained variance. In this example, the first two principal components collectively account for 91.117% of the total variance present in the original 3D dataset. This indicates that the dimensionality has been successfully reduced with only a minimal loss of information, allowing for the effective visualization and analysis of the data's primary structure in a lower-dimensional space.

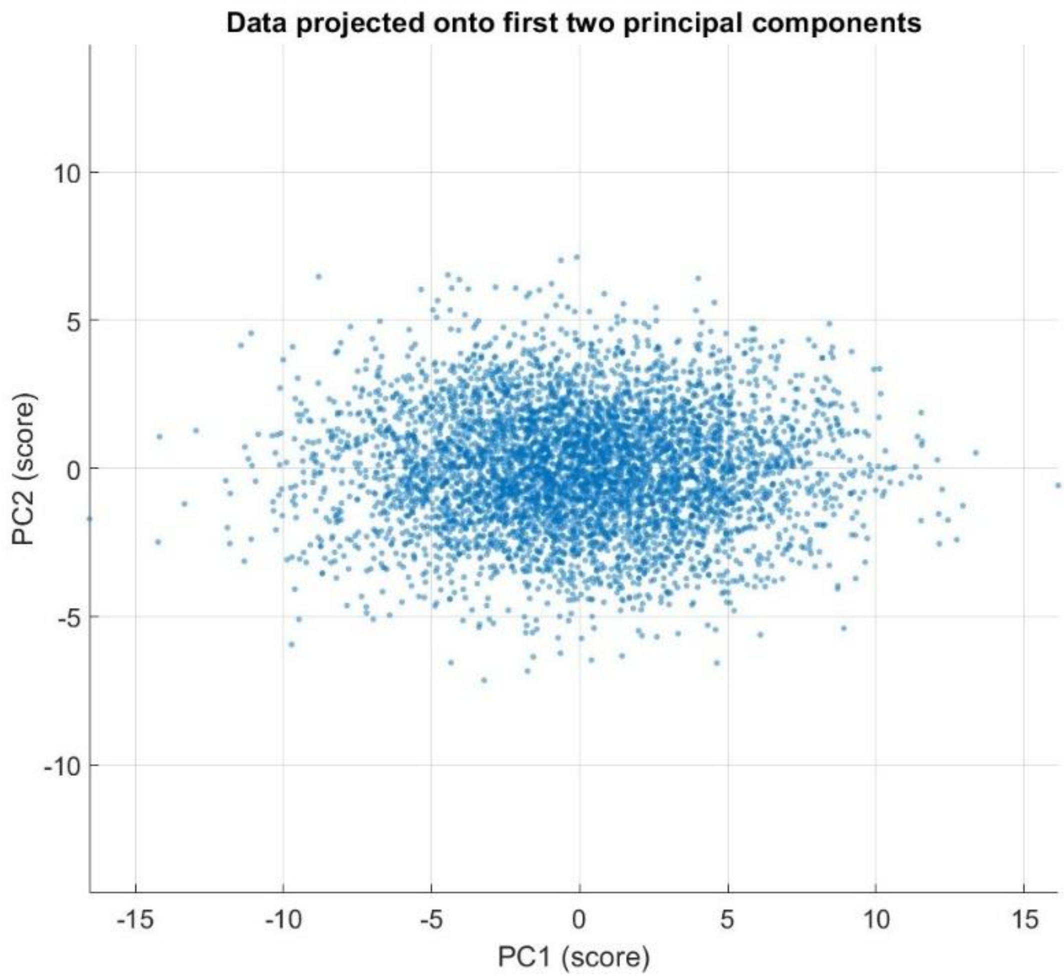


Figure 19 - Representation of data in a lower-dimensional space

3.1.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)

In contrast to the linear approach of PCA, t-Distributed Stochastic Neighbor Embedding (t-SNE) is a powerful non-linear dimensionality reduction technique designed primarily for the visualization of high-dimensional datasets. Its main objective is to create a low-dimensional (typically 2D or 3D) "map" of the data that preserves the local structure of the original high-dimensional space. In essence, it aims to ensure that data points that are close neighbors in the high-dimensional space are also represented as close neighbors in the low-dimensional map.

The t-SNE algorithm, as introduced by van der Maaten and Hinton,[37] operates in a probabilistic manner:

1. **Modeling High-Dimensional Similarities**: First, t-SNE constructs a probability distribution over pairs of high-dimensional data points. The similarity between two points is modeled as

a conditional probability that one point would pick the other as its neighbor, assuming neighbors are picked in proportion to their probability density under a Gaussian distribution centered at that point. Points that are close together have a high probability of being chosen as neighbors, while distant points have an infinitesimally small probability.

2. **Modeling Low-Dimensional Similarities**: Next, t-SNE constructs a similar probability distribution over the points in the low-dimensional map. However, for this second distribution, it uses a heavy-tailed Student's t-distribution with one degree of freedom. The use of the t-distribution is a key feature of the algorithm; its heavy tails allow for dissimilar points to be modeled by larger distances in the map, which helps to alleviate the "crowding problem" (where points tend to clump together in the center of the map) and often results in cleaner, more separated clusters.
3. **Minimizing Divergence**: Finally, the algorithm iteratively adjusts the locations of the points in the low-dimensional map to minimize the divergence between the two probability distributions (the one from the high-dimensional space and the one from the low-dimensional space). This is typically achieved by minimizing the Kullback-Leibler (KL) divergence using gradient descent.

In the context of this research, t-SNE was not used as a pre-processing step for feature extraction but rather as a crucial tool for **exploratory data analysis** and **results validation**. By applying t-SNE to the raw sensor data, it was possible to visually investigate whether natural clusters corresponding to different operational states existed. Furthermore, it was used to visualize the outputs of anomaly detection models, helping to confirm whether the identified anomalies formed distinct clusters separate from the normal operating data.

Despite its power for visualization, it is crucial to interpret t-SNE plots with an understanding of its limitations. The algorithm preserves local structure at the expense of global structure. Therefore, the relative sizes of the clusters and the distances between different clusters in a t-SNE plot are not necessarily meaningful. Its primary strength lies in revealing the presence and separation of groupings within the data, not in representing their global geometry. Additionally, t-SNE is computationally intensive and has key hyperparameters, such as "perplexity" (which relates to the number of nearest neighbors considered for each point), that require careful tuning.

Characteristic	t-SNE	PCA
Type	Non-linear dimensionality reduction	Linear dimensionality reduction
Goal	Preserve local pairwise similarities	Preserve global variance
Best used for	Visualizing complex, high-dimensional data	Data with linear structure
Output	Low-dimensional representation	Principal components
Use cases	Clustering, anomaly detection, NLP	Noise reduction, feature extraction
Computational Intensity	High	Low
Interpretation	Harder to interpret	Easier to interpret

Table 1 - Comparison of t-SNE and PCA side by side, [38].

In summary, as highlighted in Table 1, PCA and t-SNE are complementary techniques with distinct objectives. PCA is a linear method that excels at preserving global variance, making it ideal for feature extraction and noise reduction. In contrast, t-SNE is a non-linear method that prioritizes the preservation of local neighborhood structures, making it the superior choice for visualizing complex datasets and identifying clusters. This powerful visualization capability comes at the cost of higher computational intensity and more challenging interpretation compared to the efficient and more interpretable principal components derived from PCA.

3.2 Univariate Anomaly Detection

Univariate anomaly detection refers to the process of identifying anomalies by analyzing a single variable or a single time-series data stream in isolation. These methods form the foundation of many condition monitoring systems, where they are applied to individual sensor signals (e.g., a single temperature, pressure, or vibration reading) to detect deviations from an expected or normal behavior. The primary goal is to identify data points that are statistically unusual with respect to the historical distribution of that single variable.

The primary strength of univariate methods lies in their simplicity, computational efficiency, and high degree of interpretability. Because they analyze each data stream independently, they are straightforward to implement and can operate in real-time with minimal computational resources. When an anomaly is flagged, its source is unambiguous, it is tied directly to the specific sensor being monitored.

However, this focus on a single variable is also their main limitation. Univariate techniques are incapable of detecting anomalies that only manifest through the complex interplay or correlation between multiple variables. A system might be in a faulty state where several parameters are within their individual normal ranges, but their combined state is anomalous. Such scenarios require the more sophisticated multivariate approaches detailed in a later section.

The following sections will detail several fundamental statistical and time-series-based univariate techniques that were used in this research for baseline anomaly detection.

3.2.1 Z-Score (Standard Score)

The Z-Score, also known as the standard score, is a fundamental statistical measurement that quantifies the distance of a data point from the mean of its distribution, measured in terms of standard deviations. For a given data point x, from a distribution with a mean μ and a standard deviation σ, the Z-Score is calculated using the formula:

$$Z = (x - \mu)/\sigma \qquad 3.1$$

A positive Z-Score indicates the data point is above the mean, while a negative score indicates it is below the mean. The magnitude of the Z-Score represents how many standard deviations are away from the mean the data point is.

In the context of anomaly detection, the Z-Score is used to identify data points that are statistically improbable. This method operates on the core assumption that the "normal" data follows a Gaussian (normal) distribution. For a normal distribution, the empirical rule states that approximately 99.7% of the data points are expected to lie within three standard deviations of the mean. Therefore, a common practice is to set a threshold, typically $|Z| > 3$. Any data point whose Z-Score exceeds this absolute value is considered a statistical outlier and is flagged as an anomaly. Depending on the desired sensitivity of the detection, this threshold can be adjusted (e.g., to 2 or 2.5). Despite its simplicity and efficiency, the Z-Score method has a significant limitation: it is not robust to the presence of extreme outliers in the dataset used for its calculation. Both the mean (μ) and the standard deviation (σ) are sensitive to extreme values. A single large outlier can inflate the standard deviation and shift the mean, potentially "masking" other, less extreme outliers by reducing their Z-Scores. Consequently, this method is most effective for datasets that are approximately normally distributed and are not already heavily contaminated with outliers.

3.2.2 Interquartile Range (IQR) Method

While the Z-Score method is effective for normally distributed data, it is susceptible to the influence of extreme values, which can skew the mean and standard deviation. The Interquartile Range (IQR) method offers a more robust alternative for univariate anomaly detection, as it relies on quartiles rather than the mean and standard deviation. This makes it less sensitive to extreme outliers and applicable to data that may not follow a perfect Gaussian distribution.

The IQR method is best understood through its visual counterpart, the box plot (or box-and-whisker plot). As illustrated in Figure 20, a box plot provides a standardized way of displaying the distribution of data based on a five-number summary: the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum.

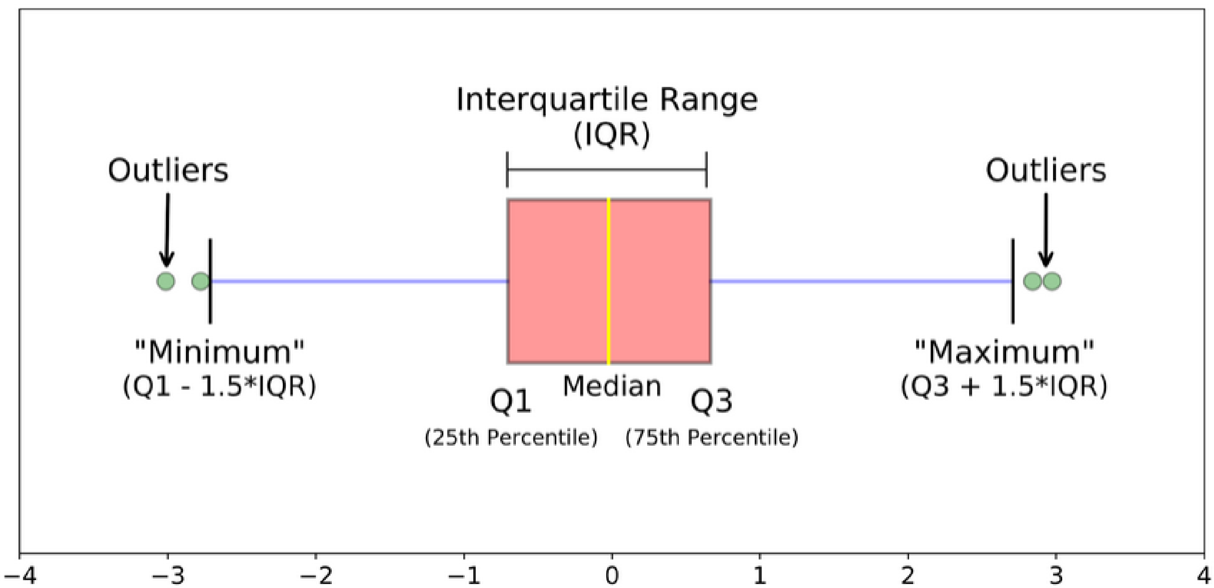


Figure 20 - How to read a boxplot, Source: [39].

The key components of this method, as shown in the figure, are:

- The Interquartile Range (IQR): This is the range between the first quartile (25th percentile) and the third quartile (75th percentile). It is calculated as IQR = Q3 - Q1. The box in the plot represents this range, containing the middle 50% of the data.

- The Median (Q2): The line inside the box represents the median value of the dataset, which separates the higher half from the lower half of the data.
- The "Whiskers": These lines extend from the box to indicate variability outside the upper and lower quartiles. They are typically set to extend to the most extreme data point that is no more than 1.5 times the IQR from the edge of the box.
  - Upper Whisker Limit: Q3 + 1.5 * IQR
  - Lower Whisker Limit: Q1 - 1.5 * IQR
- Outliers: Any data point that falls outside the whiskers is considered an outlier. In Figure 20, these are represented as individual points beyond the whisker limits.

For anomaly detection, these whisker limits serve as robust thresholds. Any sensor reading that exceeds the upper limit or falls below the lower limit is flagged as a potential anomaly. This method is particularly valuable in early exploratory analysis to quickly identify and filter out gross errors or extreme transient events in raw sensor data.

3.2.3 Moving Average and Moving Median Methods

Moving average and moving median methods represent fundamental statistical techniques for anomaly detection in time series data, particularly suitable for condition monitoring applications where sensor measurements exhibit temporal dependencies. These methods belong to the class of sliding window-based approaches that leverage local temporal context to identify deviations from expected behavior patterns [40], [41].

Moving Average (MA): The moving average method computes the arithmetic mean of observations within a sliding temporal window of fixed size *w*. For a time series *x(t)*, the moving average at time *t* is defined as:

$$\mathrm{MA}_w(t) = \frac{1}{w}\sum_{i=0}^{w-1} x(t-i) \qquad 3.2$$

where *w* represents window size, a critical hyperparameter that determines the method's sensitivity to local versus global patterns.

The anomaly score for a given observation can be computed as the absolute deviation from the moving average:

$$s(t) = | x(t) - \mathrm{MA}_w(t) | \quad 3.3$$

Alternatively, a normalized score considering the local standard deviation can be employed:

$$s_{\mathrm{norm}}(t) = \frac{| x(t) - \mathrm{MA}_w(t) |}{\sigma_w(t)} \quad 3.4$$

where $\sigma_w(t)$ is the standard deviation computed over the same window.

An observation is classified as anomalous when its score exceeds a predetermined threshold θ:

$$\mathrm{Anomaly}(t) = \begin{cases} 1 \ if \ s(t) > 0 \\ 0 \ \ otherwise \end{cases} \quad 3.5$$

Moving Median (MM): The moving median method offers increased robustness to outliers compared to the moving average by computing the median rather than the mean:

$$\mathrm{MM}_w(t) = \mathrm{median}\{x(t), x(t-1), \ldots, x(t-w+1)\} \quad 3.6$$

The median, being the 50th percentile of the ordered observations within the window, is less influenced by extreme values, making it particularly valuable when the time series may contain sporadic outliers that should not affect the baseline estimation.

The anomaly score for the moving median approach follows a similar formulation:

$$s(t) = | x(t) - \mathrm{MM}_w(t) | \quad 3.7$$

For robust estimation of scale, the Median Absolute Deviation (MAD) is often preferred over standard deviation:

$$\mathrm{MAD}_w(t) = \mathrm{median}\{| x(t-i) - \mathrm{MM}_w(t) | : i = 0, \ldots, w-1\} \quad 3.8$$

The normalized anomaly score becomes:

$$s_{\mathrm{norm}}(t) = \frac{| x(t) - \mathrm{MM}_w(t) |}{1.4826 \cdot \mathrm{MAD}_w(t)} \quad 3.9$$

where the constant 1.4826 ensures consistency with the standard deviation for normally distributed data.

Window Size Selection: The choice of window size w critically affects the method's performance and represents a trade-off between sensitivity and robustness:

- **Small windows** ($w \approx$ 5-20): High sensitivity to local changes, rapid adaptation to shifts, but increased susceptibility to noise and false positives.
- **Medium windows** ($w \approx$ 20-100): Balanced approach suitable for most applications, provides stability while maintaining reasonable responsiveness.
- **Large windows** (w > 100): Captures long-term trends, robust to noise, but may miss short-duration anomalies or exhibit delayed detection.

For condition monitoring applications, domain knowledge regarding the typical duration of anomalous events and the sampling frequency should guide window size selection. Cross-validation techniques or grid search approaches can systematically optimize this parameter.

3.2.4 Seasonal-Trend Decomposition (STL)

Seasonal-Trend decomposition using LOESS (STL) is a powerful and versatile method for decomposing a time series into three constituent components: a trend component, a seasonal component, and a remainder (or residual) component. This technique is particularly effective for univariate anomaly detection in condition monitoring data, where signals often contain strong periodic patterns (e.g., daily operational cycles, weekly schedules) and underlying long-term trends (e.g., gradual degradation).

The fundamental principle of STL for anomaly detection is that the predictable, structured parts of the time series—the trend and seasonality—can be separated from the unpredictable noise. Any significant anomalies will be captured within this leftover remainder component. The decomposition is typically additive, where the observed data $Y(t)$ at time t is represented as:

$$Y(t) = T(t) + S(t) + R(t) \qquad 3.10$$

where:

- $T(t)$ denotes the **Trend** components, capturing long-term progressive changes in the series

- $S(t)$ represents the **Seasonal** component, modelling periodic fluctuations with fixed period
- $R(t)$ constitute the **Remainder** (residual) component, containing irregular variation and potential anomalies

The STL algorithm itself is an iterative procedure based on a sequence of LOESS (Locally Estimated Scatterplot Smoothing) regressions, which makes it robust to outliers and capable of handling complex seasonal patterns.

For anomaly detection, the process involves two main steps:

1. **Decomposition**: The STL algorithm is applied to the time-series data to isolate the three components, with the most critical output being the remainder series $R(t)$.
2. **Remainder Analysis**: With the trend and seasonal effects removed, the remainder component $R(t)$ ideally represents a stationary, zero-mean signal consisting of random noise. Any significant anomalies will manifest as extreme values within this remainder series. Standard statistical methods, such as the Z-Score (3.2.1) or the Interquartile Range (IQR) Method (3.2.2), can then be applied to the remainder series $R(t)$ to identify these outliers with a much higher degree of confidence than if they were applied to the raw signal.

A key hyperparameter for STL is the **seasonal period**, which must be defined based on domain knowledge of the underlying process generating the data (e.g., the number of samples in a 24-hour cycle). The primary advantage of this method is its ability to "de-noise" a time series by removing its most predictable patterns, thereby making the underlying anomalies much more prominent and easier to detect [42].

In Figure 21, an example of a time series that has an increasing trend, annual seasonality, and noise is reported.

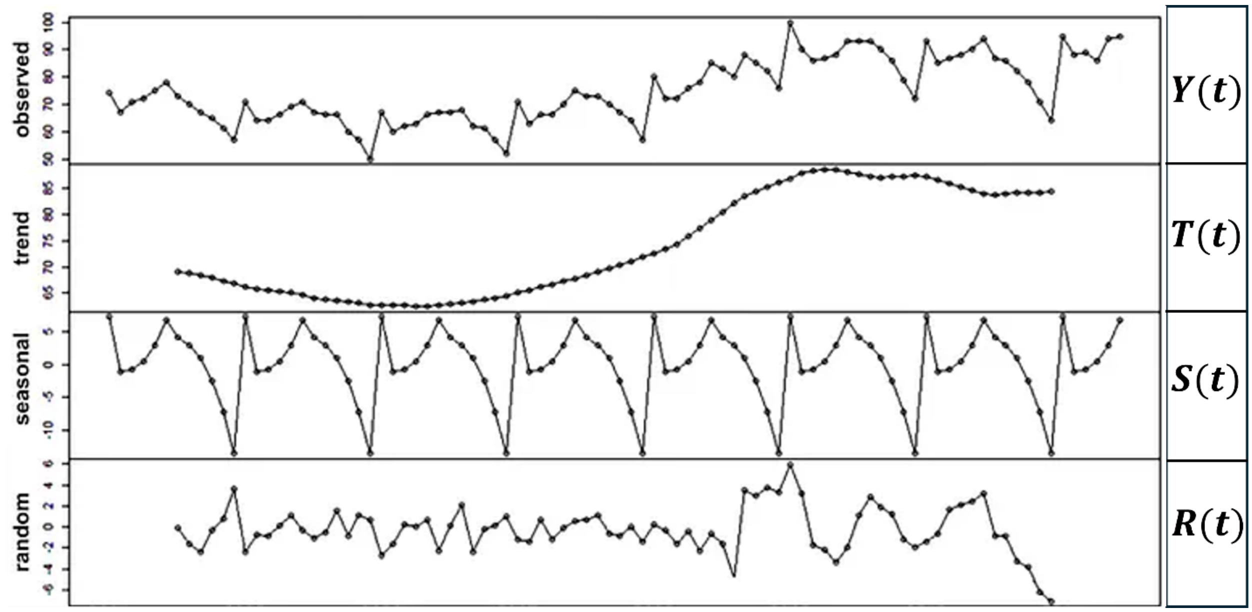


Figure 21 - Seasonal-Trend Decomposition, Source: Medium [43].

3.3 Multivariate Anomaly Detection

While univariate methods are effective for monitoring individual signals, they are fundamentally limited in that they cannot detect anomalies that arise from the complex interplay and correlations between multiple variables. Multivariate anomaly detection addresses this critical gap by simultaneously analyzing an entire set of features (a feature vector) to identify anomalous combinations. For instance, the temperature and pressure of a system might both be within their individual acceptable ranges, but their combined state could signify a developing fault. Such "contextual" anomalies, which are common in complex mechanical systems, are invisible to univariate analysis [41].

To provide a comprehensive and practical understanding of the models discussed in this section, each theoretical description will be accompanied by a practical demonstration on a consistent synthetic dataset. This approach serves a dual purpose: it provides a clear, visual intuition for how each algorithm operates and establishes a common benchmark for comparing their behavior.

For each model, the demonstration will include a visualization of the resulting anomaly score across the 2D domain of the synthetic data. This will illustrate the decision boundary or the regions that each model identifies as anomalous. Furthermore, for selected models where hyperparameter tuning is particularly critical, the results of a parameter sweep will be presented to demonstrate how the model's sensitivity and performance are influenced by its configuration. This combination of theoretical explanation and practical illustration will provide a thorough foundation for understanding the application of these models in the real-world case studies presented later in this thesis.

3.3.1 Feature Scaling Techniques

Before applying most multivariate anomaly detection algorithms, it is a critical pre-processing step to scale the features of the dataset. Multivariate models that are distance-based or assume a certain data distribution (like many classifiers) can be highly sensitive to the scale of the input features. If one feature has a much larger range of values than others, it can dominate the distance calculations and bias the model, leading to poor performance. Feature scaling transforms the data so that all features have a similar scale, ensuring equal contribution to the model's learning process.[44]

The following scaling techniques were considered and used in this research:

- **Standardization (Z-Score Normalization):** This is the most common scaling technique and is particularly effective when the data follows an approximately Gaussian distribution. It rescales each feature i to have a mean (μ_i) of 0 and a standard deviation (σ_i) of 1. The transformation for the entire data matrix x is given by:

$$\hat{x}[:,i] = \frac{x[:,i]-\mu_i}{\sigma_i} \tag{3.11}$$

$$where\ \mu_i = \frac{1}{N}\sum_{k=1}^{N} x[k,i],\ \ and\ \sigma_i = \sqrt{\frac{1}{N-1}\cdot\sum_{K=1}^{M}(x[k,i]-\mu_i)^2} \tag{3.12}$$

This method is advantageous as it does not bind the data to a specific range, which can be beneficial for algorithms that assume zero-centered data, like PCA. The Z-scores themselves can also be used directly for simple threshold-based anomaly detection, as was detailed in section 3.2.1 .

- **Min-Max Normalization**: This technique rescales the data to a fixed range, typically $[0,1]$. For each feature i, it subtracts the minimum value and divides by the range (the difference between the maximum and minimum value):

$$\hat{x}[:,i] = \frac{x[:,i]-\min(x[:,i])}{\max(x[:,i]-\min(x[:,i]))} \tag{3.13}$$

This is useful for algorithms that require data to be on a normalized scale, such as certain neural networks, but it is more sensitive to outliers in the training data than standardization.

- **Unit Vector Normalization (Scaling to Unit Length):** This method scales each sample (a row j in the data matrix) independently so that it has a unit norm (a length of 1). The transformation is given by:

$$\hat{x}[j,:] = \frac{x[j,:]}{\|x[j,:]\|} \quad 3.14$$

When applied to an entire dataset, this can be visualized as projecting all data points onto the surface of a D-dimensional unit sphere. It is particularly useful when the direction of the data vectors is more important than their magnitude.

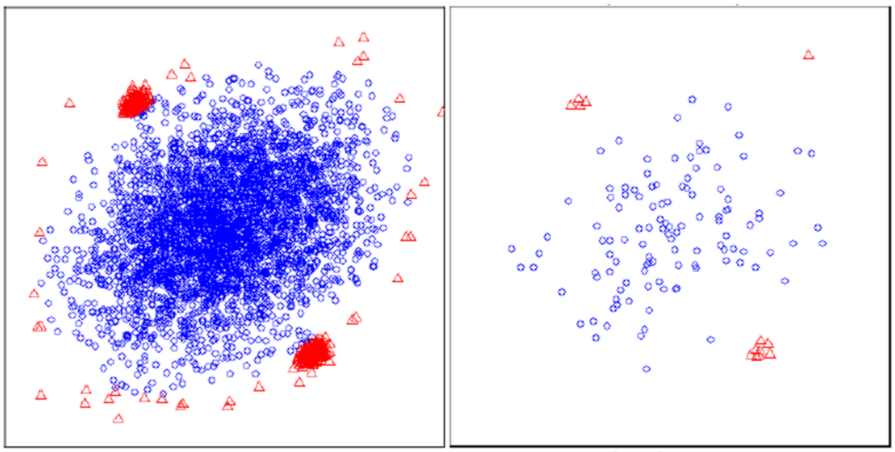
3.3.2 Isolation Forest

Isolation Forest is an efficient and widely-used unsupervised anomaly detection algorithm that operates on the principle of **isolation** rather than profiling normal data points. Unlike distance- or density-based methods, Isolation Forest does not require such metrics to detect anomalies. This unique approach provides several significant advantages: it exhibits a linear time complexity with a low constant and has minimal memory requirements. Consequently, Isolation Forest has the capacity to scale up to handle extremely large datasets and high-dimensional problems, even those with a large number of irrelevant attributes.

The algorithm builds an ensemble of **Isolation Trees (iTrees)**, which are a form of random binary tree, to isolate observations. It then returns a score indicating the likelihood of each point being an anomaly. The process is governed by three main hyperparameters: the sample size φ (the number of observations drawn to build each tree), the number of trees t in the ensemble, and the contamination parameter, which is the expected fraction of anomalies in the training data [45].

The methodology can be broken down into three main stages: sampling, tree building, and scoring.

Given a dataset X with n observations and m dimensions, the process begins by creating random subsamples of the data to ensure that each iTree is diverse, as reported in Figure 22.



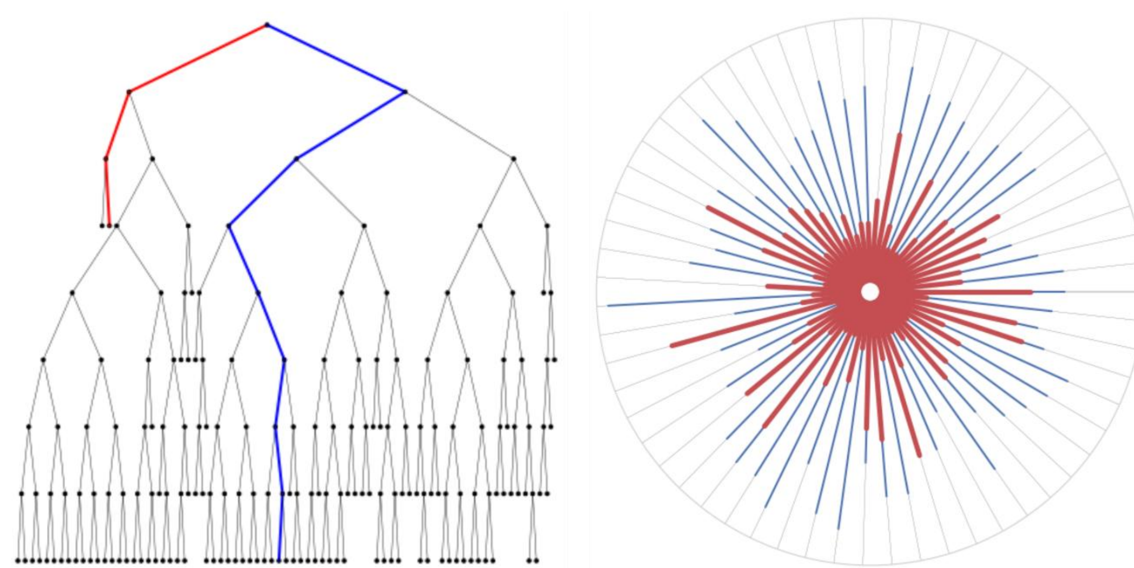
(a) Original sample
(4096 instances)

(b) Sub-sample
(128 instances)

Figure 22 - Random Sampling Process, Source: [45].

A random subset of φ observations is drawn from the original dataset X. This sampling is performed t times, with each unique sample being used to construct one of the t trees in the forest. By default, common implementations often use a sample size φ of 256 and a total of t = 100 trees, as these values have been shown to perform well across a wide range of problems.

Each of the t subsamples is used to build an individual Isolation Tree. The tree-building process is a recursive partitioning of the data, as reported in Figure 23.



(a) Representation of a single tree in a forest.

(b) Representation of a full forest where each radial line corresponds to a tree.

Figure 23- Schematic representation of a single tree (a) and a forest (b) where each tree is a radial line from the center to the outer circle. Red represents an anomaly while blue represents a nominal point. Source: [46].

For each node in the tree, the data is split until one of the following stopping conditions is met: the node contains only a single data point (complete isolation), the tree reaches its maximum depth, or all data points in the node have identical values. The maximum tree depth is typically limited to $\max(depth) = ceiling(\log_2(\varphi))$ to prevent overly complex trees. The splitting process at each internal node involves:

1. Randomly selecting one attribute d from the m available data attributes.
2. Randomly choosing a split value v between the minimum and maximum values of the selected attribute d in the current node.
3. Partitioning the data into two daughter nodes: the left node containing points where the value of attribute d is less than v, and the right node containing points where it is greater than or equal to v.

The core principle of Isolation Forest is that anomalies are "few and different" and are therefore easier to isolate than normal points. A relatively small number of random partitions is needed to isolate an anomalous data point. Conversely, when a forest of random trees globally requires a large number of partitions to isolate a particular point, it is very likely to be a normal inlier, as reported in Figure 24.

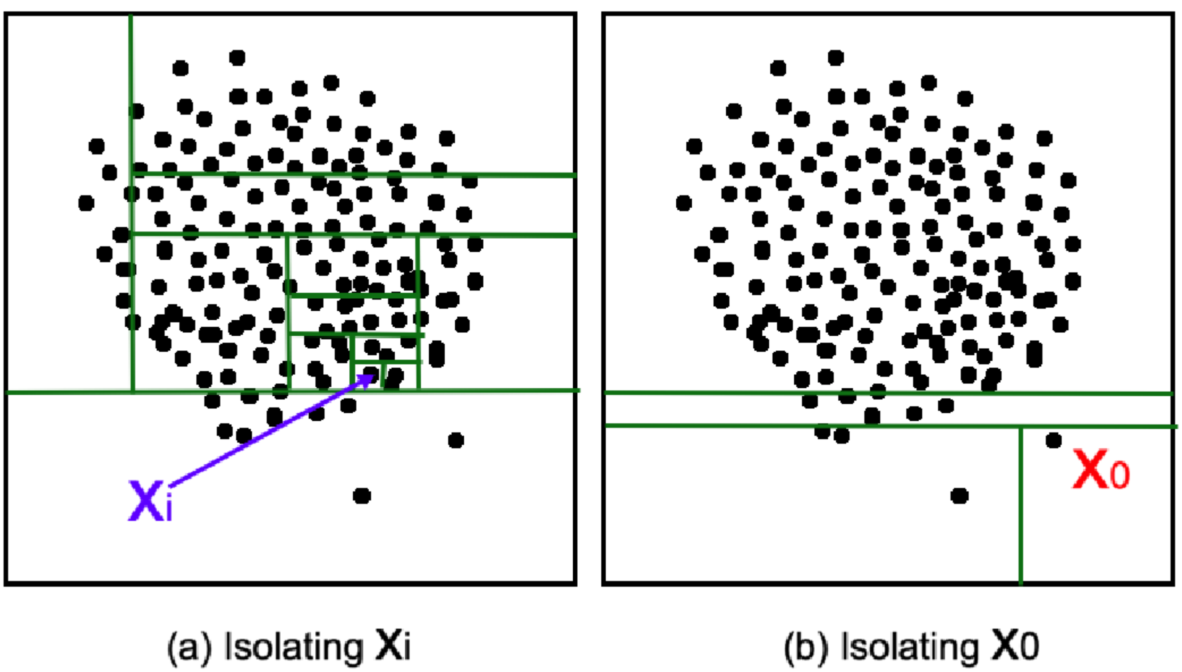


Figure 24 - Different Number of partitions between inlier and outlier, adapted from [45].

The Figure 25 shows a different way of displaying the decision tree forest and the difference in scores between an inlier and an outlier.

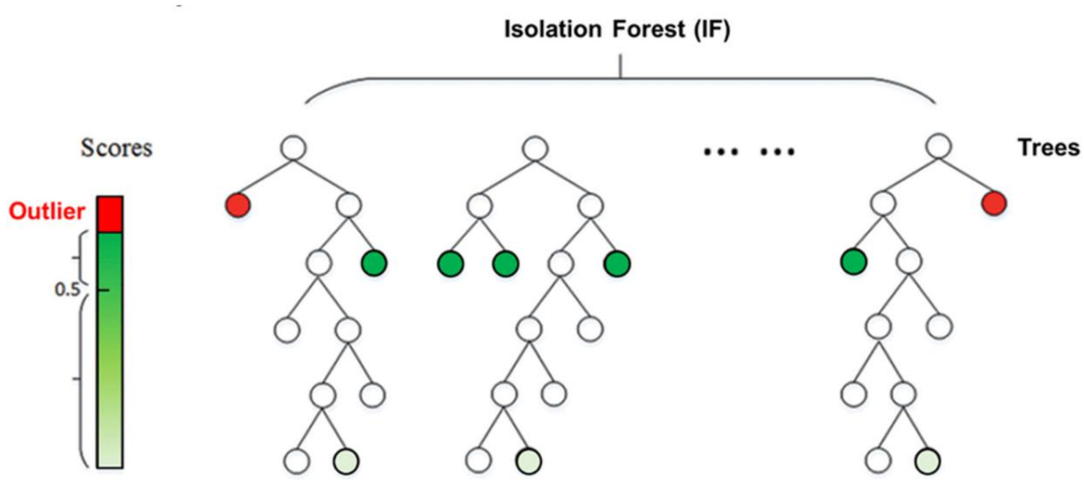


Figure 25 - Decision Tree Forest view, Source Medium [47].

This principle is quantified by calculating an anomaly score for each observation x. This score is based on the average path length $E(h(x))$ of the observation across all t trees in the forest. The path length $h(x)$ is the number of edges an observation traverses from the root of a tree to a terminal leaf node. The final score $S(x,\ \varphi)$ for a sample of size φ is normalized and calculated as:

$$S(x, m) = 2^{\frac{-E(h(x))}{c(n)}} \quad 3.15$$

Where:

- $E\big(h(x)\big) = \frac{\sum_{i=1}^{t} h_i(x)}{t}$ is the average path length for observation x over t trees.
- $c(n) = 2H(n-1) - \frac{2(n-1)}{n}$ is the average path length of an unsuccessful search in a Binary Search Tree, serving as a normalization factor. H is the harmonic number.

The interpretation of the score is as follows:

- If $E\big(h(x)\big)$ is much smaller than $c(n)$, then $S(x,\ \varphi)$ is close to 1, indicating that the point has a very short average path length and is therefore highly likely to be an **anomaly**.

- If $E\left(h(x)\right)$ is approximately equal to $c(n)$, then $S(x,\ \varphi)$ is close to 0.5, suggesting the point is a **regular/normal** instance.
- If $E\left(h(x)\right)$ is much larger than $c(n)$, then $S(x,\ \varphi)$ is much smaller than 0.5. This indicates a very long path length, meaning the point is a deeply embedded **inlier**.

3.3.3 Local Outlier Factor

The Local Outlier Factor (LOF) is a powerful, density-based unsupervised anomaly detection algorithm. Its primary strength lies in its ability to identify anomalies by considering the local density of a data point relative to its neighbors, rather than relying on global data distribution. This makes LOF particularly effective for datasets where different regions may have different densities, a common characteristic of real-world industrial data [48].

The LOF algorithm operates through a sequential process of calculating local densities, reachability, and ultimately, an outlier score for each data point.

1. Calculation of Local Density

The foundation of the LOF algorithm is the concept of local density. For each data point p in the dataset, the algorithm first identifies its local neighborhood. This is typically done by finding the k nearest neighbors using a distance metric, most commonly the Euclidean distance. A point's local density is implicitly defined by its distance to these neighbors. Points with a high local density (i.e., whose neighbors are very close) are considered normal, while those with a low local density (i.e., whose neighbors are far away) are potential anomalies, as reported in Figure 26.

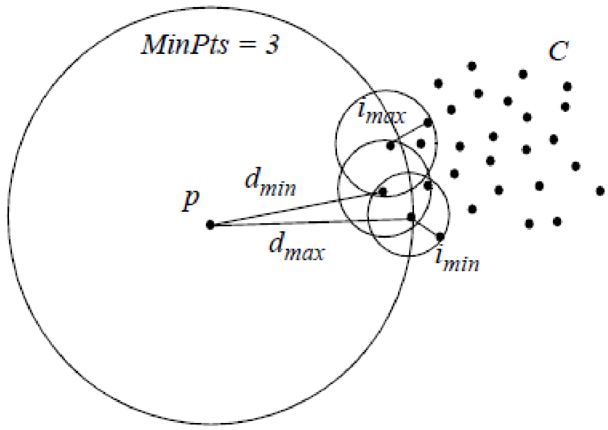


Figure 26 - Calculation of Local Density, Source: [48].

The formulas for the calculation of k-distance and k-distance neighborhood for a given point p and parameter k:

k-distance: The distance to the k-th nearest neighbor of p:

$$k - distance(\mathbf{p}) = d(\mathbf{p}, \mathbf{o}) \tag{3.16}$$

where o is the k-th nearest neighbor of p.

k-distance neighborhood: The set of all points within k-distance from p:

$$N_k(p) = \{q \in D \setminus \{\mathbf{p}\}: d(\mathbf{p}, q) \leq k - distance(\mathbf{p})\} \tag{3.17}$$

Note: $|N_k(p)| \geq k$ due to potential ties at the k-distance threshold.

2. **Calculation of Local Reachability Density (LRD)**

To formalize and stabilize the density measurement, the algorithm first computes the reachability distance of a point p with respect to another point q. This is defined as the maximum of either the true distance between p and q or the k-distance of q. The k-distance (q) is the distance from point q to its k-th nearest neighbor.

The formula for reachability distance from point p to point o is:

$$reach - dist_k(\mathbf{p}, \mathbf{o}) = \max\{k - distance(\mathbf{o}), d(\mathbf{p}, \mathbf{o})\} \tag{3.18}$$

This step effectively smooths the distances, preventing statistical fluctuations for points that are very close to each other. Using this, the Local Reachability Density (LRD) of a point p is then calculated as the inverse of the average reachability distance from p to its neighbors in its k-neighborhood, $N_k p$.

$$LRD_k(p) = \frac{1}{\frac{\sum_{o \in N_k(p)} reach\text{-}dist_k(\mathbf{p},\mathbf{o})}{|N_k(\mathbf{p})|}} \tag{3.19}$$

Or equivalently:

$$LRD_k(p) = \frac{|N_k(\mathbf{p})|}{\sum_{o \in N_k(p)} reach\text{-}dist_k(\mathbf{p},\mathbf{o})} \tag{3.20}$$

A high LRD value indicates that the point is in a dense region (short average reachability distance), while a low LRD value indicates it is in a sparser region.

3. Calculation of Local Outlier Factor (LOF)

The local outlier factor of point **p**:

$$\text{LOF}_k(\mathbf{p}) = \frac{\sum_{\mathbf{o} \in N_k(\mathbf{p})} \frac{\text{LRD}_k(\mathbf{o})}{\text{LRD}_k(\mathbf{p})}}{|\ N_k(\mathbf{p})\ |} \tag{3.21}$$

or equivalently:

$$\text{LOF}_k(\mathbf{p}) = \frac{1}{|\ N_k(\mathbf{p})\ |} \sum_{\mathbf{o} \in N_k(\mathbf{p})} \frac{\text{LRD}_k(\mathbf{o})}{\text{LRD}_k(\mathbf{p})} \tag{3.22}$$

Interpretation:

- LOF ≈ 1: **p** has similar density to its neighbors (normal point)
- LOF ≫ 1: **p** has much lower density than its neighbors (outlier)

- LOF < 1: **p** has higher density than its neighbors (inlier in dense region)

Anomaly criterion: Typically, points with LOF > τ (where $\tau \approx 1.5 - 2.0$) are classified as anomalies.

The LOF algorithm presents several key advantages:

- **Sensitivity to Local Density**: LOF's ability to adapt to varying local densities allows it to detect anomalies in datasets where the definition of an outlier is context-dependent.
- **Detection of Complex-Shaped Anomalies**: As it makes no assumptions about the global data distribution, LOF can effectively identify outliers in non-spherical clusters.
- **Unsupervised Nature**: LOF does not require training labels, making it ideal for real-world scenarios where labeled anomaly data is rare.

However, it also has notable limitations:

- **Difficulty in Score Interpretation**: The raw LOF score is a ratio and is not normalized, making it difficult to set a universal threshold. A score of 1.1 might be a significant outlier in one dataset, while a score of 2.0 could still be an inlier in a dataset with strong local fluctuations.
- **Sensitivity to Parameter Choice**: Performance is highly dependent on the choice of k, which often requires careful tuning.
- **Computational Complexity**: The need to calculate distances to nearest neighbors makes LOF computationally expensive, particularly for large datasets, with a complexity of at least $O(n^2)$.
- **The Curse of Dimensionality**: In high-dimensional spaces, the concept of a meaningful "neighborhood" breaks down as all points tend to become equidistant, which can degrade LOF's effectiveness, see Chapter (3.1).

3.3.4 One-Class Support Vector Machines (OC-SVM)

The One-Class Support Vector Machine (OC-SVM) is a widely-used algorithm for novelty detection. Unlike traditional two-class SVMs that learn a hyperplane to separate two distinct classes of data, the OC-SVM is trained on a dataset containing only "normal" instances. Its objective is to learn a decision boundary, or a high-dimensional frontier, that encloses the majority of the normal data points. Any new data point that falls outside this learned boundary is then classified as an anomaly or a "novelty."

The core principle of OC-SVM is to find a hyperplane in a high-dimensional feature space that separates the data from the origin with maximum margin. To capture complex, non-linear boundaries, the algorithm employs the kernel trick. This technique implicitly maps the original input data into a much higher-dimensional feature space where a linear separation becomes possible. The Radial Basis Function (RBF) kernel is the most frequently used for this purpose, due to its flexibility [49], [50].

The operational mechanism of OC-SVM can be understood through the following stages.

The first step is to implicitly transform the input data from its original space x into a high-dimensional feature space $\varphi(x)$, where a linear separation boundary is more feasible. This non-linear mapping is achieved through a kernel function (e.g., Linear, RBF, Polynomial), which efficiently calculates dot products between data points in this high-dimensional space without ever explicitly computing the coordinates. This is the "kernel trick." Figure 27 illustrates this conceptual transformation.

$$x = (x_1, \ldots, x_m) \mapsto \varphi(x) = (\varphi(x_1), \ldots, \varphi(x_m))) \quad 3.23$$

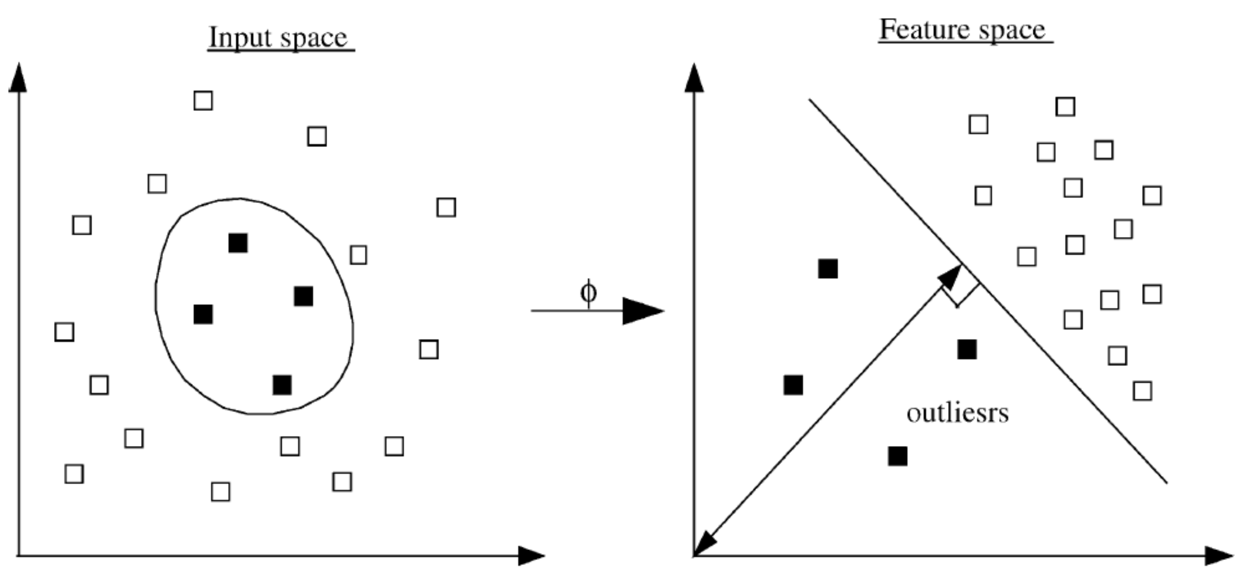


Figure 27 - Non-linear transformation from input space to high-dimensional feature space. Source: [50].

In the new feature space, the OC-SVM algorithm seeks to find an optimal hyperplane, defined by a weight vector w and an offset p, that best separates the mapped data points $\varphi(x)$ from the origin. The concept of margin maximization, which in a two-class SVM refers to the separation between classes (as analogized in Figure 28), is adapted in OC-SVM to mean the separation of the data *from the origin*.

The optimization goal is to push the hyperplane as far as possible from the origin while keeping most of the data points on the other side.

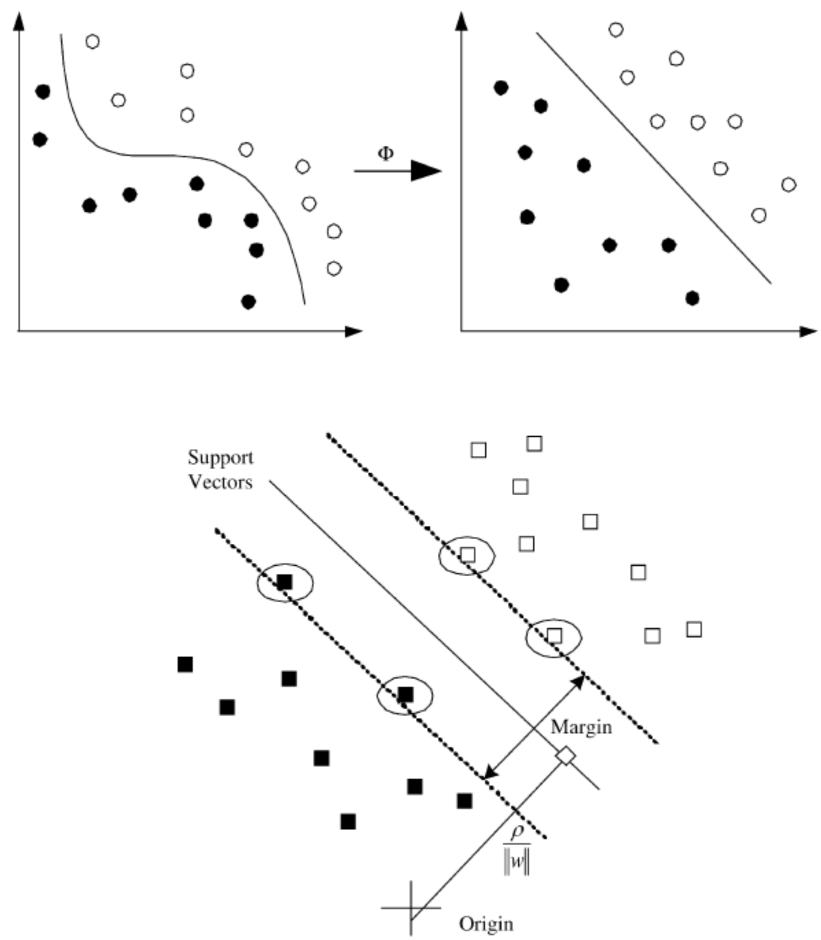


Figure 28 - Feature mapping and two-class separation. Source [50].

This involves solving a quadratic programming problem. To allow for some data points to be on the "wrong" side of the margin (i.e., to handle outliers in the training set), slack variables ξ_i are introduced. The formal optimization problem is:

$$\min_{w_i,\xi_i,p} \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m}(\xi_i - \nu p) \tag{3.24}$$

Subject to the constraints: $\langle w \cdot \varphi(x_i)\rangle \geq p - \xi_i, \xi_i \geq 0$, for all data points x_i.

Here, the term $\frac{1}{2}\ ||w||^2$ maximizes the margin, $\Sigma_i\, \xi_i$ is the penalty for points that do not meet the margin requirement, and v is a hyperparameter that controls this trade-off.

The hyperparameter nu (ν), a value in the range (0, 1], is critical. It controls the trade-off between maximizing the margin and tolerating misclassifications in the training data. The nu parameter has a dual interpretation:

- It is an **upper bound** on the fraction of training data points that can be considered outliers (i.e., those that lie on the origin-side of the hyperplane).
- It is a **lower bound** on the fraction of training data points that will become support vectors (the points that define the boundary).

Choosing an appropriate value for nu is crucial. A small nu will create a "tighter" boundary that attempts to enclose most of the training points, while a larger nu will create a "looser" boundary, allowing more training points to be flagged as anomalies.

After training, the OC-SVM model produces a decision function that classifies new data points. The hyperplane found during optimization serves as this decision boundary. As illustrated in Figure 29, all points lying on one side of the hyperplane are classified as "normal," while points on the other side (the side containing the origin) are classified as "abnormal" or anomalies. This creates the desired frontier that encloses the normal data region.

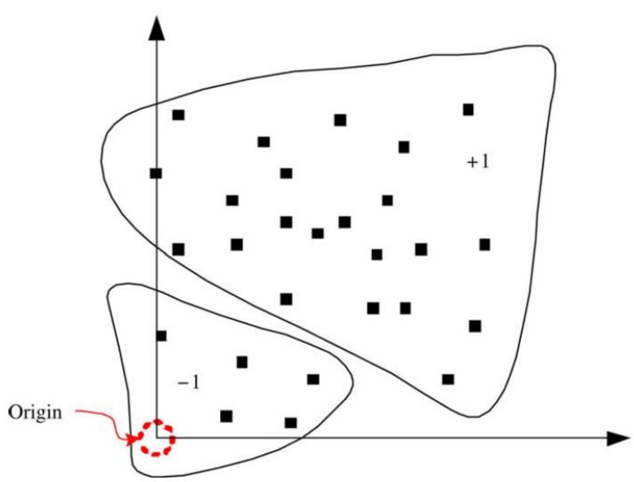


Figure 29 - Classification by one-class SVMs. Source: [50].

3.3.5 Mahalanobis Distance and Robust Covariance

The Mahalanobis Distance, introduced by P.C. Mahalanobis in 1936, is a powerful statistical measure used to calculate the distance of a point from the center of a multivariate distribution. Unlike the Euclidean distance, which treats all dimensions equally, the Mahalanobis distance is scale-invariant and, critically, accounts for the covariance among the variables. This makes it an ideal tool for detecting anomalies in multivariate sensor data where features are often correlated. An anomaly is identified as a point that lies far from the central tendency of the data, considering the shape and orientation of the data cloud.[51]

Unlike Euclidean distance, which measures the straight-line distance between two points in space, Mahalanobis distance takes into account the correlations between the variables in the dataset and is invariant with respect to scale. In other words, Mahalanobis distance evaluates the number of standard deviations of a point from the mean of the distribution, also considering the covariance between variables. This feature makes it very useful in identifying anomalous or unusual points within a multivariate dataset. This property is shown in Figure 30.

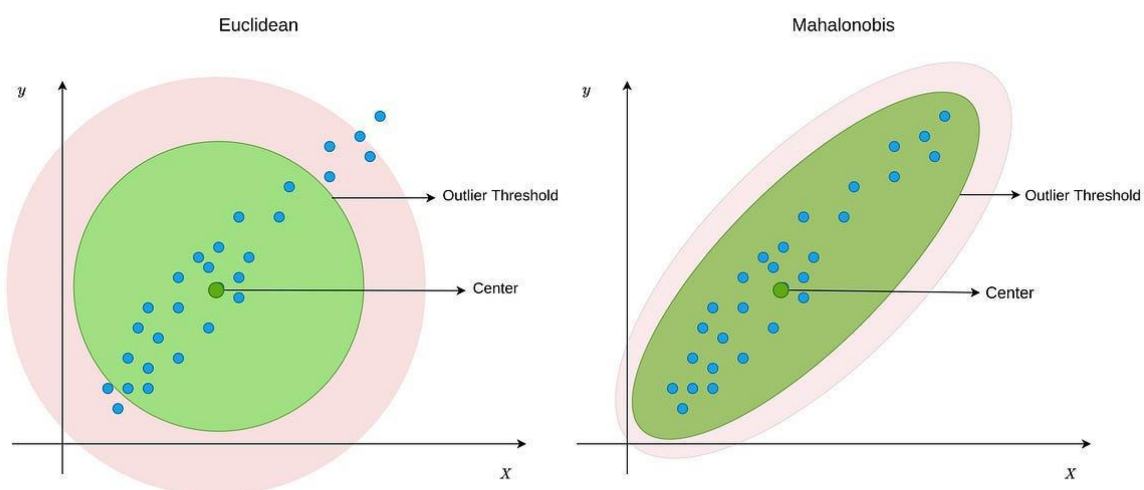


Figure 30 - Visualization of the difference between Euclidean distance and Mahalanobis distance on a plane. Source: [52].

The anomaly detection process using this robust approach involves:

- Applying the MCD algorithm (often the computationally efficient FAST-MCD variant) to the training data to obtain robust estimates of μ and Σ.
- Calculating the robust Mahalanobis distance for all data points (both training and new data) using these robust estimates.

- Classifying a point as an anomaly if its calculated distance exceeds a certain threshold.

1. Calculation of the mean and Covariance Matrix

The mean vector, denoted as μ, represents the central point in the multivariate space around which the data points are distributed. It is a vector composed of the mean values of each variable (or feature) in the dataset. To calculate the mean vector, the mean of each variable is taken across all data points. Mathematically, if you have a data set with n observations and p variables, the mean vector μ is calculated as:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad 3.25$$

Where X_i is the vector representing the *i-th* observation in the dataset.

2. Calculation of the Mahalanobis Distance

The covariance matrix, referred to as Sigma, is a square matrix containing the covariances between all pairs of variables in the data set. This matrix captures how the variables vary together and is fundamental to understanding the shape and orientation of the data distribution in multivariate space. The covariance between two variables X and Y is calculated using the following formula:

$$\mathrm{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y) \qquad 3.26$$

Where μ_X and μ_Y are the means of X and Y, respectively, and *n* is the number of observations in the dataset. The covariance matrix is then constructed by calculating the covariance for each pair of variables in the dataset. For a dataset with p variables, the covariance matrix will be a *p x p* matrix.

3. Identification of Anomalies

After calculating the Mahalanobis distance for each data point in a dataset, anomalies (or outliers) are identified by comparing these distances to a threshold. This threshold is determined based on the *chi-square* $\chi 2$ distribution, given the relationship between the squared Mahalanobis distance and $\chi 2$. It is usually based on the *chi-square* distribution corresponding to a certain confidence level. For example, if a confidence level of 95% is chosen, the value in the *chi-square* distribution table that corresponds to 95% of the area under the curve with the appropriate degrees of freedom will be sought. Any data point whose Mahalanobis squared distance exceeds this threshold value is considered an outlier.

3.3.6 Practical Comparison of the Anomaly Detection Method Described

To synthesize the theoretical descriptions provided in the preceding sections, this subchapter presents a direct, practical comparison of the multivariate anomaly detection models on a controlled, benchmark on a synthetic dataset. This empirical approach serves to visually demonstrate the unique characteristics, decision boundaries, and sensitivities of each algorithm, providing a clear intuition for their behavior before their application to more complex, real-world data.

To create a consistent and interpretable testbed, a two-dimensional synthetic dataset was generated. The "normal" data consists of three distinct clusters, each drawn from a bivariate Gaussian distribution. Each cluster is defined by:

- **A mean vector (μ):** This defines the center point of the cluster in the 2D space.
- **A covariance matrix (Σ):** This matrix controls the shape and orientation of each cluster. The diagonal elements represent the variance in each of the two dimensions, while the off-diagonal elements control the correlation between them, allowing for the creation of both spherical and elliptical clusters.

To ensure a fair and standardized comparison across the different algorithms, a consistent set of assumptions was used for model training. The key hyperparameter for many unsupervised anomaly detection models is the expected fraction of outliers in the data, often referred to as the contamination parameter. For this benchmark, a contamination fraction of 0.01 was used for all models that support this hyperparameter.

To evaluate the models performance, three anomalous points were then manually added to the dataset. As shown in Figure 31, these points were strategically placed in low-density regions, far from the main clusters, to represent clear and unambiguous outliers.

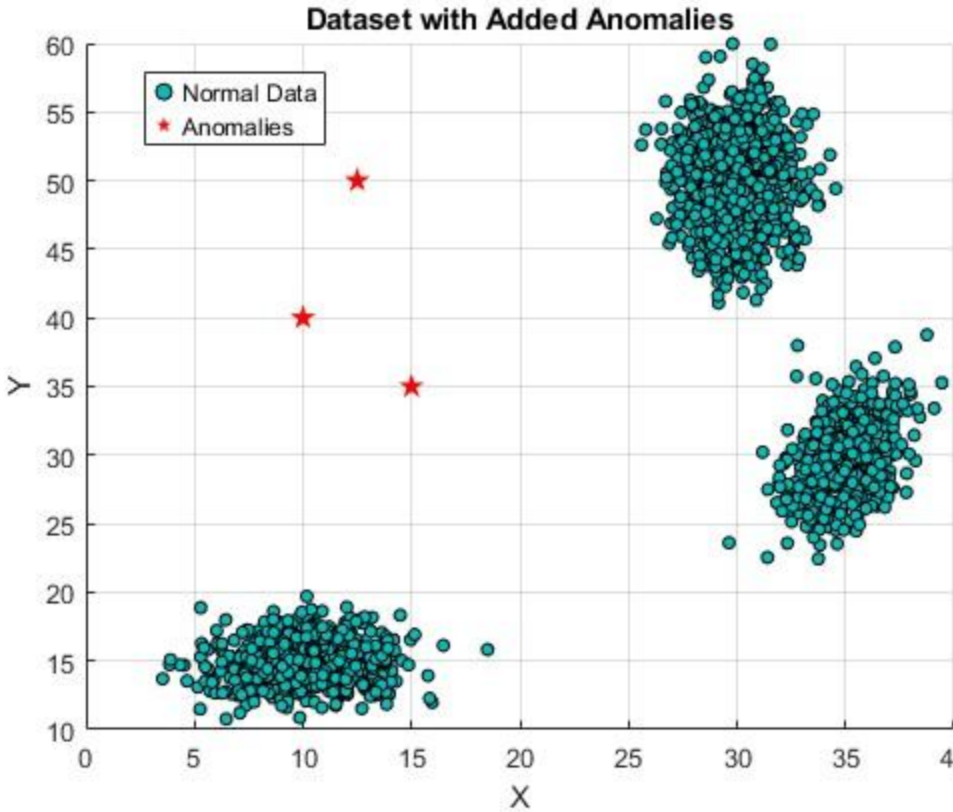


Figure 31 - Dataset with Added Anomalies.

Each of the previously discussed multivariate models was applied to this benchmark dataset to identify the anomalous points. The binary classification results, highlighting which points were flagged as outliers by each model, are presented in Figure 32.

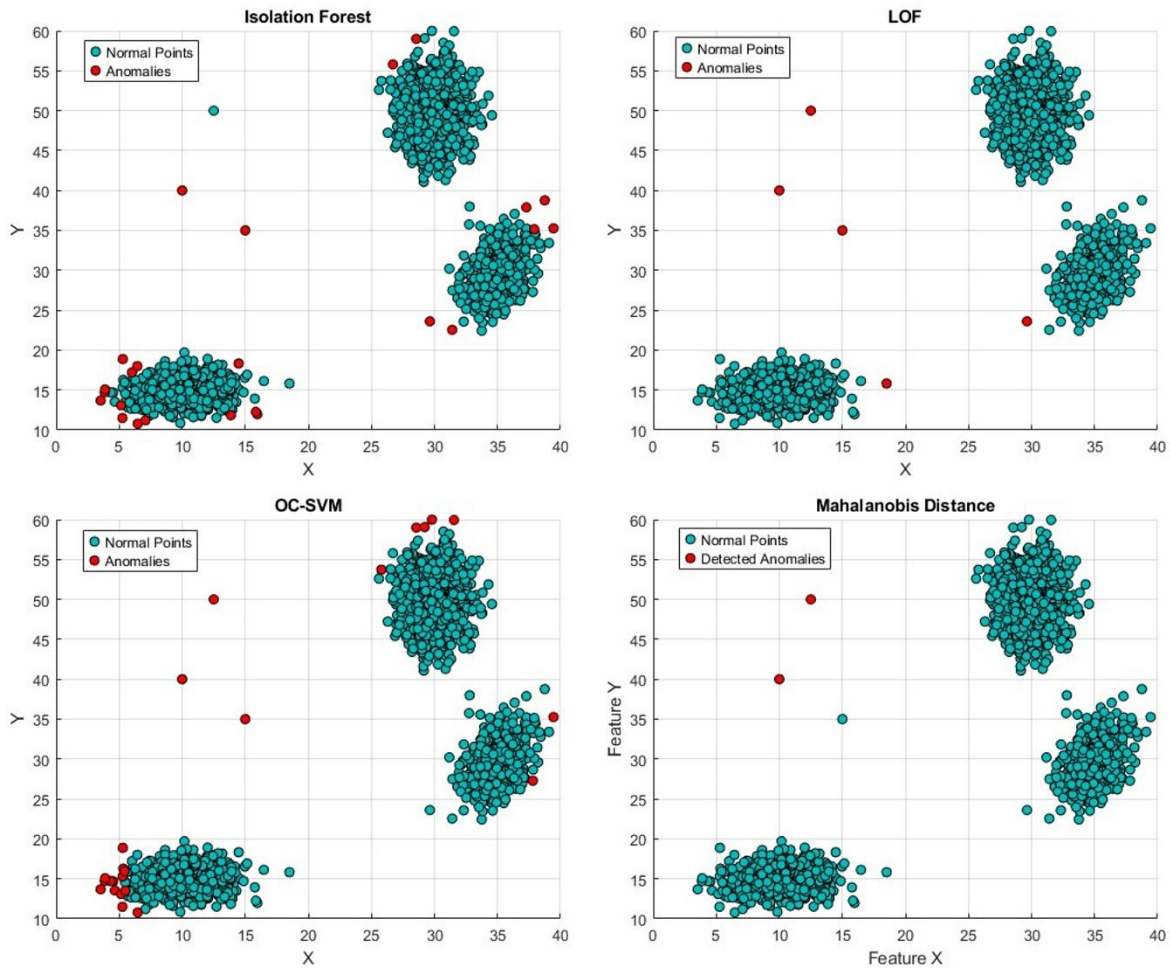


Figure 32 - Outlier Detection Results for Each Model.

Beyond simple binary classification, a more insightful comparison is achieved by visualizing the anomaly score map for each model. This involves calculating the anomaly score for every point in a fine grid spanning the 2D domain, which effectively reveals the model's underlying decision function and the regions it considers to be anomalous. Figure 33 presents a side-by-side comparison of these anomaly score maps for all the tested models. This visual benchmark clearly illustrates the fundamental differences in their approaches—from the elliptical boundaries of Mahalanobis distance to the complex, density-driven contours of LOF and the recursively partitioned regions of Isolation Forest. This comparative analysis provides a critical foundation for selecting the appropriate model for the specific data characteristics encountered in the real-world case studies of this thesis.

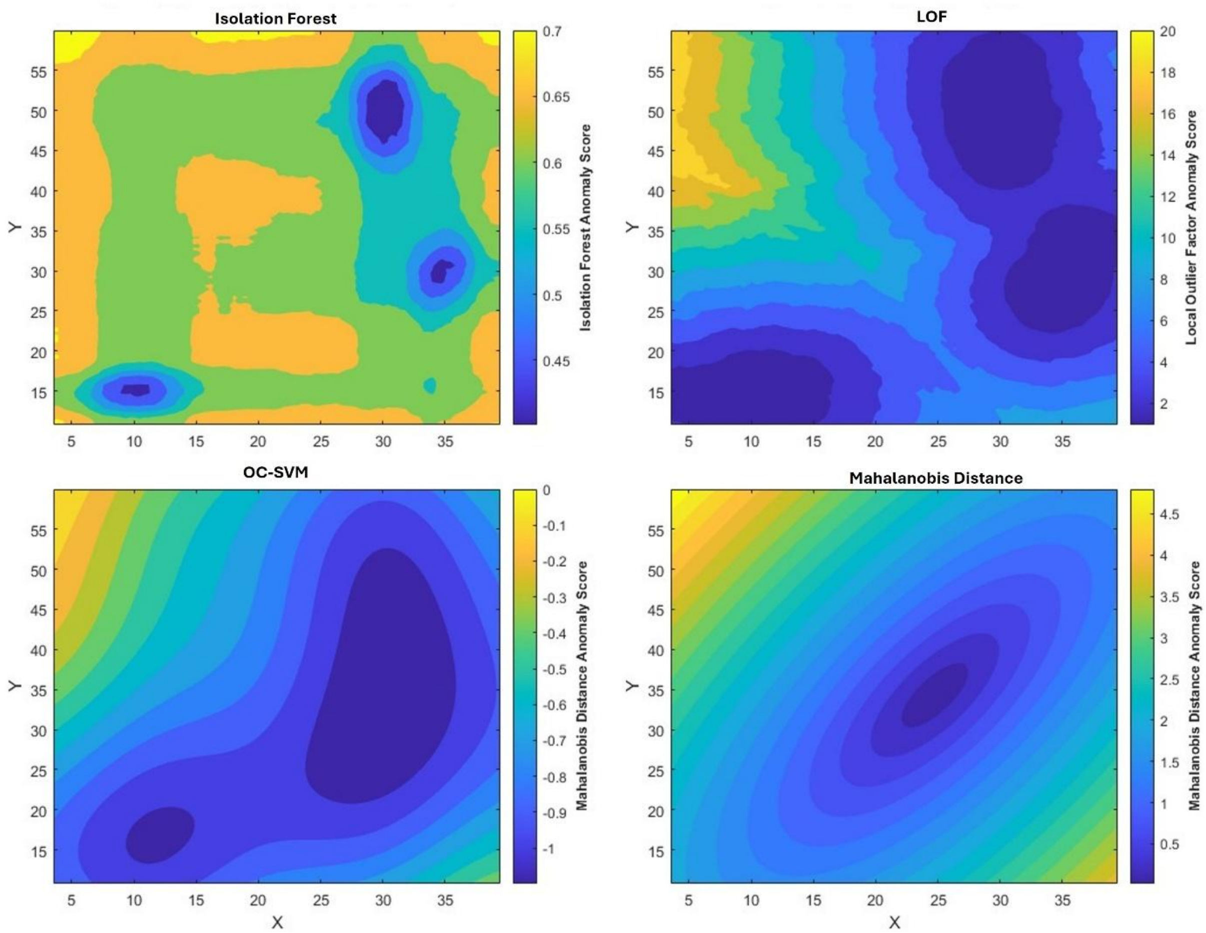


Figure 33 - Anomaly Score Map Comparison.

Upon examining the results in the Anomaly Score Map Comparison Figure 34, a notable artifact is visible in the visualization for the standard Isolation Forest. The model produces distinct rectangular, high-scoring regions in areas of the feature space where no data actually exists. These "ghost clusters" are a direct consequence of the algorithm's reliance on axis-parallel splits, which can artificially assign high anomaly scores to empty regions that fall between clusters.

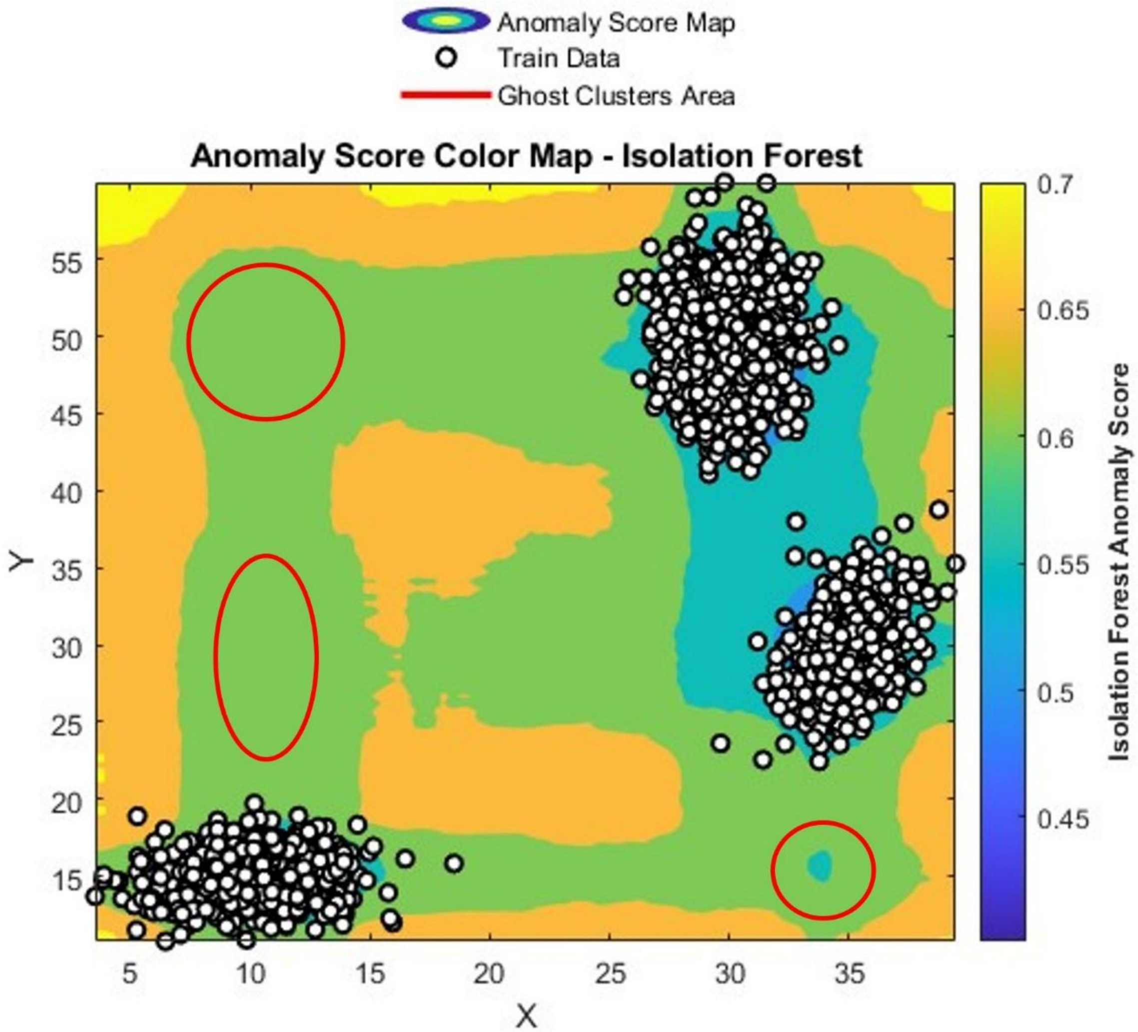


Figure 34 - Ghost Cluster Area, Isolation Forest.

This limitation has been addressed by a proposed extension to the algorithm known as the Extended Isolation Forest (EIF). Instead of using axis-parallel splits, the EIF algorithm creates splits using hyperplanes with random slopes. This provides a more flexible partitioning of the feature space. A single branching process for branch cuts with random slopes in our 2-D example is visualized in Figure 35 for both cases of anomaly and nominal data points, analogous to Figure 24 for the standard Isolation Forest.

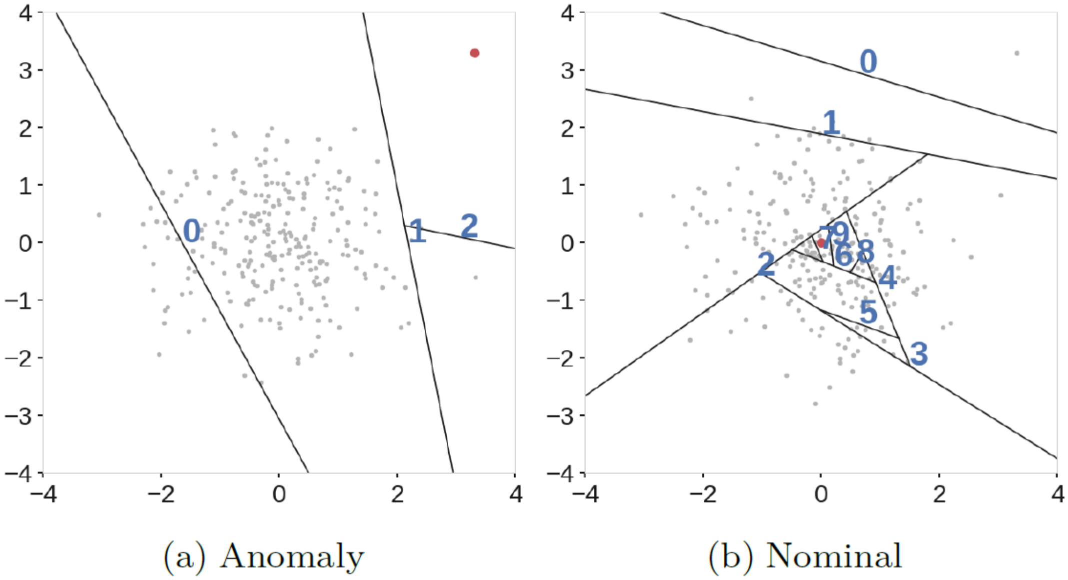


Figure 35 - Branching process for Extended Isolation Forest. Source [46].

Figure 36 directly compares the output of the standard Isolation Forest with the Extended Isolation Forest on the same dataset. The visualization clearly demonstrates that the EIF produces a much smoother and more intuitive anomaly score map, successfully mitigating the creation of the artifactual ghost clusters and improving the model's reliability.

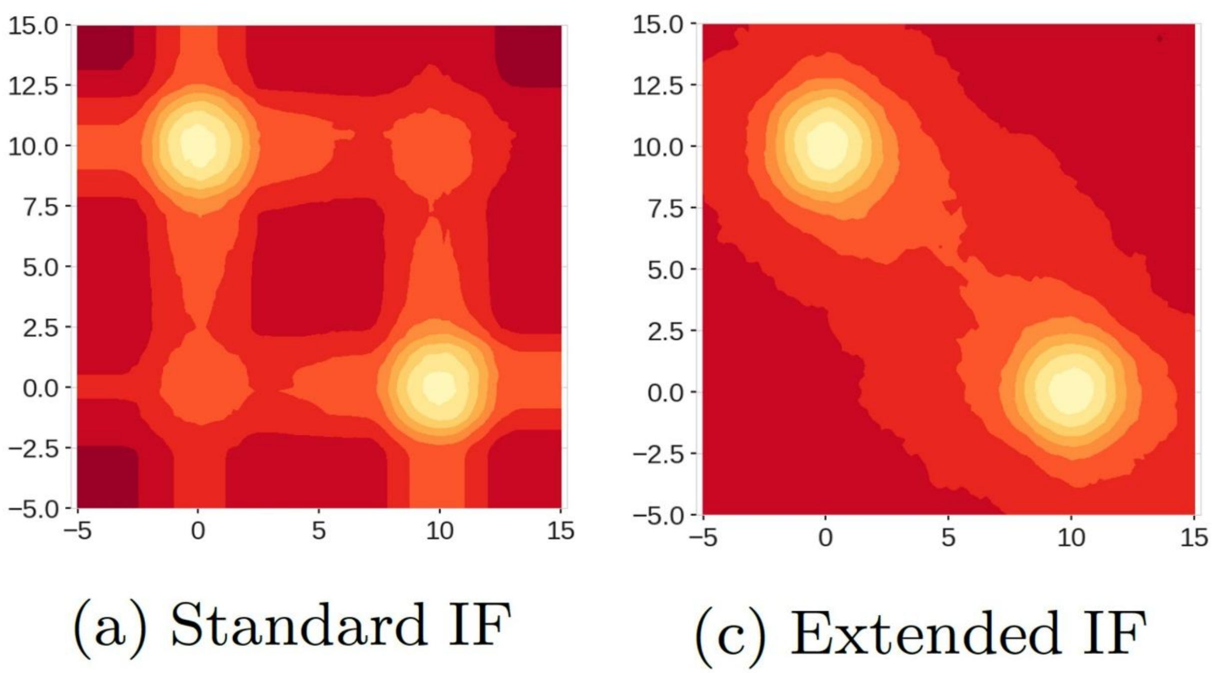


Figure 36 - Comparison of the standard Isolation Forest and Extended Isolation Forest for the case of two blobs. Source [46].

4 Signals, Sensors, and data relevant to Maintenance Practices

The preceding chapters have established the strategic importance of predictive maintenance and detailed the machine learning algorithms that provide the analytical power for its implementation. However, the success of any predictive model is fundamentally predicated on the quality, relevance, and richness of the input data. The most sophisticated algorithm will fail if it is trained on data that do not accurately reflect the physical state of the machinery. This chapter, therefore, shifts the focus from the analytical methods to the raw material of any PdM strategy: the physical signals and data acquired from monitoring conditions.

A thorough understanding of the physical origins of these signals, their acquisition methods, and the information encoded within them is essential for effective feature engineering, model development, and the correct interpretation of results. This chapter provides a detailed overview of the primary data sources and signal processing domains that are central to modern industrial maintenance and form the basis for the experimental case studies presented in this thesis.

The chapter will begin with a comprehensive exploration of vibration analysis, which is arguably the most powerful and widely used technique for monitoring the health of rotating machinery. The discussion will cover the fundamental language of vibrations, the different physical parameters used for analysis, and the key features extracted from the time, frequency, and time-frequency domains. Subsequently, the chapter will provide an overview of other critical condition monitoring technologies that were relevant to this research, including structure-borne ultrasound, oil quality analysis, and thermal monitoring. The objective is to build a foundational understanding of what is being measured and how that measurement relates to the underlying health of mechanical components.

4.1 Vibration

Vibration, defined as the oscillatory motion of a mechanical system about an equilibrium position, is an inherent and unavoidable characteristic of all rotating and reciprocating machinery. While all machines produce a baseline level of vibration, this signal carries a wealth of diagnostic information. A healthy machine exhibits a stable and repeatable vibration "signature"; the development of mechanical faults—such as imbalance, misalignment, bearing degradation, or gear tooth wear—inevitably alters this signature in specific, identifiable ways.[53]

The analysis of these changes is arguably the most powerful and widely-used condition monitoring technique for rotating equipment. This is because vibration is often the earliest detectable indicator of incipient mechanical failure, providing a significant warning period (the P-F interval, see 1.2) long before other symptoms like increased heat or audible noise become apparent. The raw data for this analysis is typically acquired using sensors, most commonly accelerometers, mounted at key locations on the machine.

A comprehensive understanding of vibration analysis requires a specific vocabulary to describe the signal's characteristics (4.1.1), a grasp of the different physical parameters that can be measured (4.1.2), and knowledge of the signal processing techniques used to extract meaningful information from the time and frequency domains (4.1.3). Furthermore, internationally recognized standards exist to provide a framework for assessing vibration severity (4.1.4). The following sections will provide a detailed examination of these foundational aspects.

4.1.1 The language of vibrations

To analyze and interpret vibration signals effectively, it is essential to understand the fundamental terminology used to describe their characteristics. A vibration signal, in its simplest form, can be represented as a waveform that plots the oscillatory motion of a point over time. These core concepts are visually summarized in the waveform diagram presented in Figure 37.

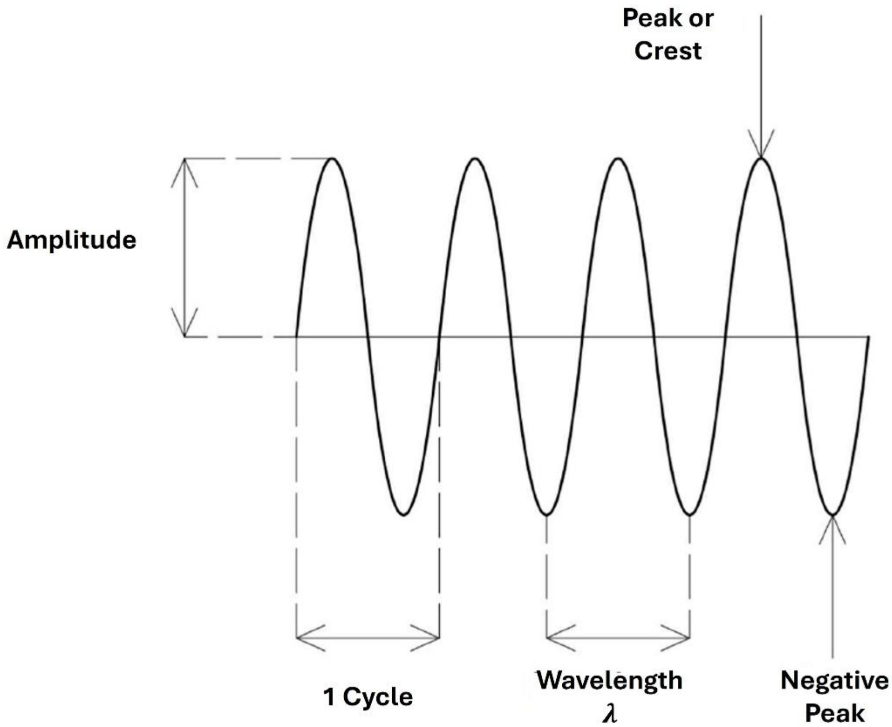


Figure 37 - Terminology used to describe waveforms, adapted from [54].

The key characteristics used to describe this waveform are Amplitude, Frequency, and Period.

Amplitude quantifies the magnitude, or intensity, of the vibration and is the primary indicator of the severity of a fault. There are several ways to measure and express amplitude from a waveform:

- **Peak Amplitude (0-P):** This measures the maximum displacement or intensity from the zero or equilibrium position to the highest point (crest) of the waveform. Peak amplitude is particularly useful for indicating short-duration, high-impact events, such as the shock caused by a cracked gear tooth.
- **Peak-to-Peak Amplitude (P-P):** This represents the total excursion of the vibrating component, measured from the lowest point (trough) to the highest point (crest) of the waveform. It is valuable for assessing the total displacement of a component, which is critical in cases where mechanical clearances are a concern.
- **Root Mean Square (RMS) Amplitude:** This is the most widely used and meaningful measure for overall vibration severity. The RMS value is calculated as the square root of the mean of the squared values of the waveform. For a pure sine wave, this is equal to the peak amplitude multiplied by 0.707. The RMS value is directly related to the energy content and therefore the

destructive potential of the vibration, making it the best overall indicator of a machine's condition. Most vibration standards, including those from ISO, use RMS values to define severity thresholds. It is calculated as:

$$\mathrm{RMS} = \sqrt{\frac{1}{T}\left(\int_0^T x^2(t)dt\right)} \qquad 4.1$$

Frequency is the number of oscillatory cycles that occur in a unit of time. It is arguably the most powerful diagnostic parameter in vibration analysis because it provides information about the source of the vibration. It is typically measured in Hertz (Hz), which corresponds to cycles per second. In industrial contexts, it is also common to express frequency in Cycles Per Minute (CPM), where CPM = Hz × 60.

Different mechanical components and fault types generate vibrations at specific, predictable frequencies. For example, shaft imbalance will typically produce a strong vibration at a frequency equal to the shaft's rotating speed (often referred to as 1x). Similarly, gear mesh problems, bearing defects, and blade pass issues all have unique frequency signatures that allow an analyst to pinpoint the root cause of a problem.

The period is the time required to complete one full cycle of vibration. It is the reciprocal of frequency and is typically measured in seconds (s).

$$T[s] = \frac{1}{F[Hz]} \qquad 4.2$$

While frequency is the primary parameter used for diagnostics, understanding the period is fundamental to the time-domain analysis of the waveform.

4.1.2 Analysis of the Vibration

The vibrations generated by rotating machines can be broadly classified as deterministic. This implies that the motion, being periodic and driven by predictable physical phenomena like shaft rotation or gear meshing, can be described by a mathematical model. This is in contrast to random vibrations, which are non-periodic and arise from stochastic processes such as fluid turbulence, necessitating a statistical approach for their analysis. The scope of this thesis is primarily concerned with the deterministic vibrations characteristic of rotating machinery.

A comprehensive analysis of such signals requires examination from multiple perspectives, or domains. Both the time domain and the frequency domain are considered throughout this research, as each provides essential and often complementary information regarding the health of a machine. The time-domain waveform reveals the signal's evolution over time and its instantaneous magnitude, while the frequency domain breaks the signal down into its constituent frequency components, revealing the sources and energy distribution of the vibration.[53]

The physical manifestation of this vibratory motion can be quantified by three key, kinematically related parameters: **displacement**, **velocity**, and **acceleration**. While these parameters are mathematically linked through integration and differentiation, they can also be measured directly using different types of sensors. The choice of which parameter to measure and analyze is a critical decision in condition monitoring, as it depends on the specific goals of the investigation and, most importantly, the frequency range of the faults being targeted.

4.1.2.1 Displacement

Vibration displacement refers to the total distance a component travels from its position of rest, or equilibrium, during one cycle of oscillatory motion. It provides a direct measure of the physical movement of a machine part and is typically quantified in units of micrometers $[\mu m]$ or, in imperial systems, mils (thousandths of an inch).

In industrial condition monitoring, displacement is most often measured using non-contact eddy current proximity probes. These sensors are ideal for monitoring the relative motion of rotating shafts within their bearings, a critical application for large machinery equipped with fluid-film bearings.

The primary diagnostic value of displacement is as an indicator of low-frequency vibration. At low frequencies, significant vibratory energy can manifest as large physical displacements, even if the associated acceleration is minimal. Consequently, displacement is the preferred parameter for identifying faults that occur at or near the shaft's rotational frequency (often referred to as 1x or 2x), such as mass imbalance, shaft misalignment, or a bent shaft. The magnitude of the displacement in these cases is directly related to the mechanical stress on the component.

Conversely, displacement is a poor indicator of high-frequency vibrations. The minute physical movements associated with high-frequency energy, such as those generated by bearing element defects or gear mesh issues, result in extremely small displacement values that are often masked by the lower-frequency components of the signal. Therefore, relying solely on displacement measurements would mean failing to detect the onset of many critical machinery faults.[55]

4.1.2.2 Velocity

Vibration velocity is the time rate of change of displacement, quantifying how fast a component is moving during its oscillation. It is typically expressed in millimeters per second $\left[\frac{mm}{s}\right]$ or inches per second $\left[\frac{in}{s}\right]$.

In diagnostic terms, velocity is often considered the most valuable and representative parameter for assessing the overall health of a machine across the mid-frequency range (typically 10 Hz to 2,000 Hz). The reason for its effectiveness is that the amplitude of vibration velocity is directly related to the destructive energy of the vibration. High velocity signifies rapid stress cycles, which are a primary contributor to fatigue failure in mechanical components. Unlike displacement, which is biased towards low frequencies, and acceleration, which is biased towards high frequencies, velocity provides a more uniform and "flat" response across the most common fault frequencies found in industrial machinery.

For this reason, vibration velocity—specifically its RMS value—is the most widely used metric in international standards, such as those from the ISO (e.g., ISO 10816 series), for assessing the operational health of general rotating machinery. It is an excellent indicator for a broad range of faults, including general looseness, flow-related problems, and the intermediate stages of bearing and gear wear.

While velocity can be measured directly using electrodynamic velocity transducers, it is more commonly obtained in modern systems by single integration of the signal from a piezoelectric accelerometer. This approach combines the robustness and wide frequency range of an accelerometer with the diagnostic strength of the velocity parameter.[53]

4.1.2.3 Acceleration

Vibration acceleration is defined as the time rate of change of velocity ($a = dv/dt$). It is typically quantified in meters per second squared $\left[\frac{mm}{s^2}\right]$ or, more commonly in industrial practice, in units of gravitational acceleration, 'g' (where 1 g ≈ 9.80665 m/s^2).

The physical significance of acceleration is paramount in vibration analysis. According to Newton's Second Law of Motion (F=ma), acceleration is directly proportional to the dynamic forces acting within the machine. Therefore, measuring acceleration provides a direct indication of the forces that cause component wear, fatigue, and eventual failure.

The primary diagnostic strength of acceleration is its high sensitivity to high-frequency vibrations. Even minute physical displacements, if they occur at a high frequency, will produce a very large acceleration signal. This makes it the ideal parameter for the early detection of faults that generate high-frequency stress waves, such as the initial stages of rolling-element bearing defects (e.g., pitting and spalling), gear mesh problems, and cavitation in pumps.

The most common sensor used for its measurement is the piezoelectric accelerometer. This robust and versatile sensor has a very wide frequency range and is the standard for most modern predictive maintenance programs. While indispensable for high-frequency analysis, acceleration is less effective for diagnosing low-frequency phenomena like imbalance, as the large displacement signals at low frequencies are often associated with very small acceleration amplitudes, which can be lost in the noise floor of the signal.

A specialized technique known as acceleration enveloping (or demodulation) is often applied to accelerometer data. This process filters the high-frequency signal to isolate the repetitive, low-energy impacts characteristic of bearing faults and then demodulates this signal to identify the underlying fault frequency [55].

4.1.2.4 The relationship between parameters

Displacement, velocity, and acceleration are not independent quantities; they are kinematically linked through the mathematical operations of differentiation and integration with respect to time. Starting with displacement d(t), a single differentiation yields velocity v(t), and a second differentiation yields acceleration a(t). Conversely, integrating acceleration once yields velocity, and a second integration yields displacement, as reported in Figure 38.

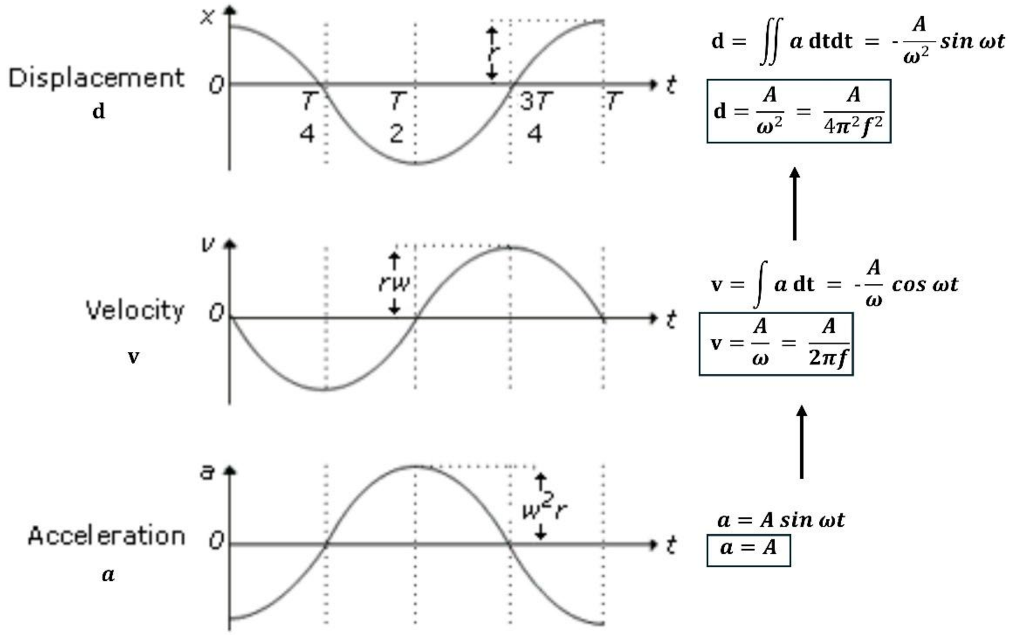


Figure 38 - Integration from acceleration to displacement. Source [56].

This mathematical relationship has a profound, frequency-dependent effect on the amplitude of each parameter. For a simple sinusoidal vibration at a given frequency ω (in radians per second), the peak amplitudes of the three parameters are related as follows:

$$\mathrm{v}_{peak} = \omega \cdot \mathrm{d}_{peak} \tag{4.3}$$

$$a_{peak} = \omega \cdot \mathrm{v}_{peak} = \omega^2 \cdot \mathrm{d}_{peak} \tag{4.4}$$

This means that for the same level of vibratory energy, the measured amplitude of each parameter will be dramatically different depending on the frequency of the vibration.

This frequency-dependent relationship is visually summarized in Figure 39. The plot clearly illustrates why a specific parameter is preferred for a given frequency range to maintain a relatively constant or "flat" response to faults of similar severity. The example shows a constant speed spectrum of 7.6 mm/s.

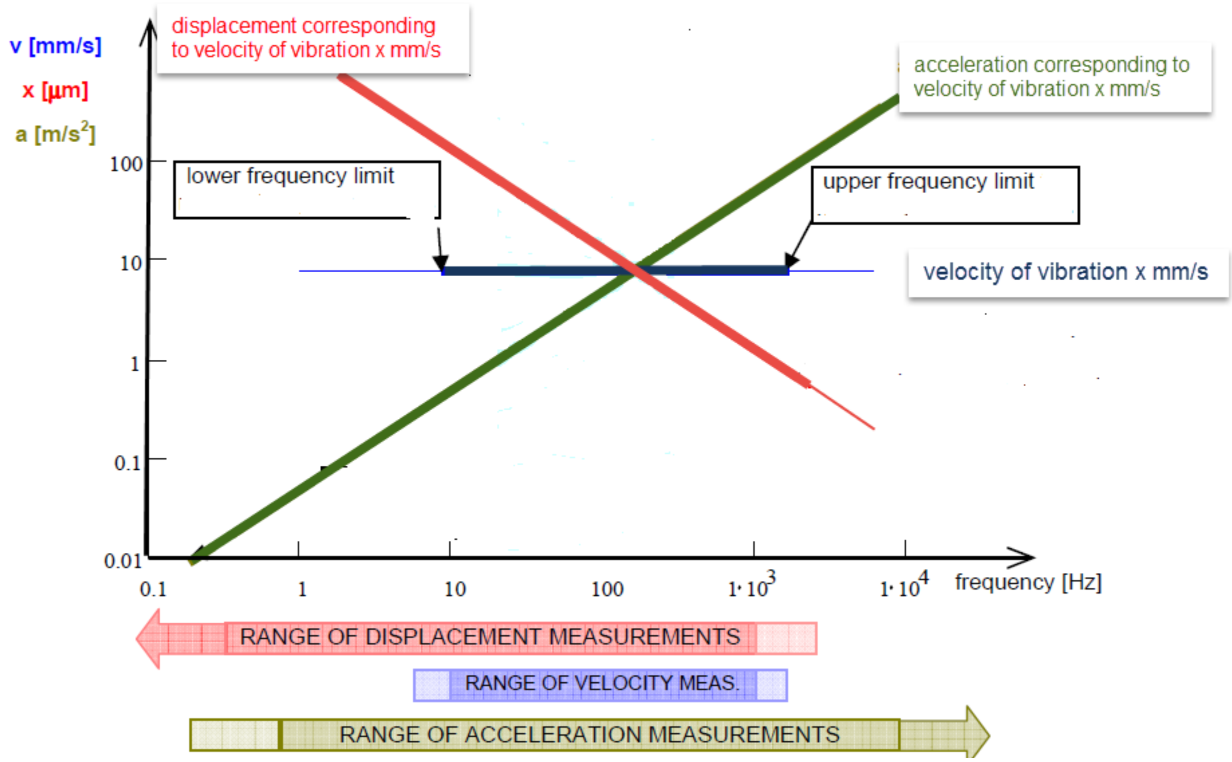


Figure 39 - The relationship between displacement, velocity and acceleration. Source [57].

As the figure demonstrates:

- **At low frequencies**, the acceleration and velocity amplitudes are very small, while the displacement amplitude is dominant. This confirms that displacement is the most sensitive and appropriate parameter for analyzing low-frequency phenomena like imbalance or misalignment.
- **In the mid-frequency range**, the velocity amplitude provides the most uniform and representative measure of vibration severity. This is the range where most common machinery faults appear, which is why velocity is the basis for most international vibration standards.
- **At high frequencies**, the ω^2 factor causes the acceleration amplitude to become very large, while displacement becomes infinitesimally small. This confirms that acceleration is the only suitable parameter for detecting the low-energy, high-frequency impacts characteristic of early-stage bearing and gear faults.

Therefore, the choice of which parameter to analyze is not arbitrary but is a critical diagnostic decision based on the type of fault being investigated and its characteristic frequency signature [40].

4.1.3 Signal based Condition Indicators

While the raw time-series signal contains all the available diagnostic information, its high dimensionality and complexity make it generally unsuitable for direct input into traditional machine learning models or for simple trend analysis. To make this data useful, it is necessary to distill it into a set of quantitative, low-dimensional metrics known as Condition Indicators (CIs) or features.

This process, often referred to as feature engineering or feature extraction, is a critical step that bridges the gap between the physical sensor signal and the mathematical input required by a predictive model. The goal is to compute a set of scalar values that effectively summarize specific characteristics of the vibration signal, reducing its dimensionality while retaining or even enhancing its discriminatory information regarding the machine's health state.

The extraction of these indicators is performed by analyzing the signal from different mathematical perspectives, or domains. The three most important domains for vibration-based condition monitoring are the time domain, the frequency domain, and the time-frequency domain. The time domain analyzes the waveform directly, the frequency domain decomposes the signal to reveal its energy distribution across different frequencies, and the time-frequency domain provides a hybrid view of how the frequency content evolves over time. The following sections will detail the specific condition indicators that can be extracted from each of these domains [58], [59].

4.1.3.1 Time-domain features

Time-domain features, also known as statistical features, are calculated directly from the raw vibration waveform x(t). They provide a quantitative summary of the signal's characteristics without transforming it into another domain. These indicators are computationally inexpensive and highly effective for tracking overall changes in a machine's health, making them a cornerstone of many condition monitoring programs. While they are excellent at detecting if a change has occurred, they generally do not provide diagnostic information about the specific source or frequency of a fault, which requires frequency-domain analysis [58].

The most common and informative time-domain features used in this research are:

- **Root Mean Square (RMS):** As previously discussed, the RMS is the most common and powerful statistical measure of a signal's magnitude. It is directly related to the signal's energy content and therefore its destructive potential, making it the best overall indicator of a machine's health. The calculation of it has already been explained in (4.1.1).

- **Peak and Peak-to-Peak:** The Peak value is the maximum absolute amplitude in the signal, while the Peak-to-Peak value is the difference between the maximum and minimum amplitude. These are useful for detecting short-duration, high-energy events like impacts.
- **Crest Factor:** This is the ratio of the Peak amplitude to the RMS value of the signal. It is an excellent indicator of the "impulsiveness" of a signal. A healthy machine with smooth, continuous vibration will have a low crest factor. The emergence of sharp impacts from a developing fault (e.g., a crack in a bearing race) will cause the peak value to rise much faster than the overall RMS, leading to a significant increase in the crest factor.

$$Crest\ Factor\ =\ Peak\ /\ RMS \tag{4.5}$$

- **Skewness:** This is the third statistical moment, which measures the asymmetry of the signal's probability distribution. A perfectly symmetric distribution (like a pure sine wave) has a skewness of zero. A positive skew indicates a tail on the right side of the distribution, while a negative skew indicates a tail on the left. In diagnostics, a non-zero skewness can be indicative of one-sided faults, such as a single cracked gear tooth that produces impacts in only one direction of rotation. For a signal with N number of samples, mean μ and standard deviation σ, it is calculated as:

$$\boldsymbol{Skewness} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^3 \tag{4.6}$$

- **Kurtosis:** This is the fourth statistical moment, which measures the "tailedness" or "peakedness" of the distribution compared to a Gaussian (normal) distribution. A normal distribution has a kurtosis of 3. Like the crest factor, it is highly sensitive to the impulsiveness of a signal. A signal with sharp spikes from incipient bearing faults will have a high kurtosis value because these spikes create long heavy tails in the probability distribution. It is one of the most classic indicators for early-stage bearing fault detection.

$$\boldsymbol{Kurtosis} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^4 \tag{4.7}$$

- **Signal-to-Noise Ratio (SNR):** This is a measure that quantifies the level of a desired signal relative to the level of background noise. It is defined as the ratio of the power of the signal (P_{signal}) to the power of the noise (P_{noise}). In a time-domain context, this can be estimated using the ratio of the variance of the signal to the variance of the noise. A change in SNR over time is a key indicator of a changing health state. For example, a developing fault might introduce broadband noise into the system (e.g., from severe rubbing or cavitation), which would decrease the SNR. Conversely, if a fault creates a strong, clear periodic signal, it may increase the signal's power relative to the background noise, thereby increasing the SNR.

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right) = 20\log_{10}\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right) \quad 4.8$$

Where:

- A_{signal} is the amplitude (RMS) of the signal
- A_{noise} is the amplitude (RMS) of the noise

4.1.3.2 Frequency-domain features

While time-domain features are excellent for detecting changes in the overall energy or impulsiveness of a signal, they provide limited diagnostic information about the root cause of a fault. To understand the source of the vibration, it is necessary to analyze the signal in the frequency domain [53]. Figure 40 provides a direct visual comparison of these two domains, illustrating how a complex time-domain waveform is transformed into an interpretable frequency spectrum where distinct periodic components are clearly visible.

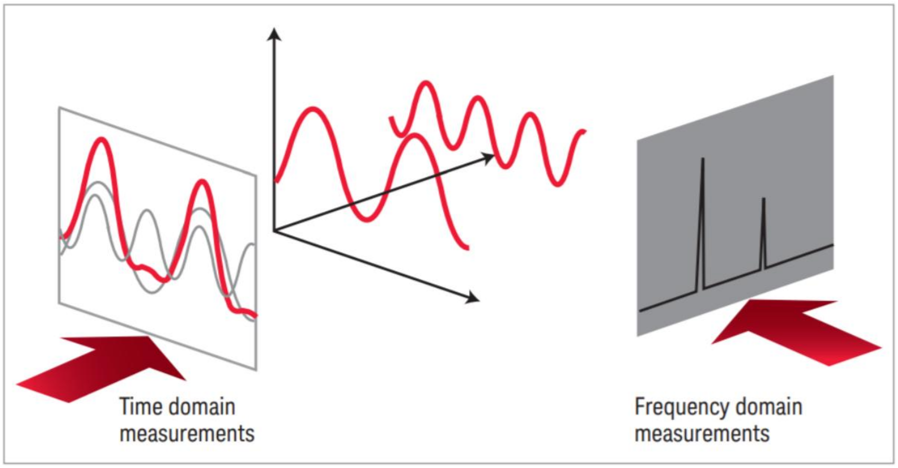


Figure 40 - Time-domain and Frequency-domain a visualization comparison. Source [60].

The primary tool for this transformation is the Fast Fourier Transform (FFT), an efficient algorithm that computes the Discrete Fourier Transform (DFT). The FFT decomposes a finite time-series signal x(t) into its constituent frequency components, revealing the amplitude (or power) of the vibration at each specific frequency. The result of this transformation is the frequency spectrum, a plot of amplitude versus frequency that is the most fundamental and powerful tool in vibration diagnostics. Peaks in this spectrum correspond to dominant periodic events within the machine, allowing for the direct identification of the components responsible for the vibration.

From this spectrum, several key features can be extracted to serve as condition indicators:

- **Peak Frequency and Amplitude:** The most basic and informative features are the frequency and amplitude of the highest peaks in the spectrum. The frequency of a peak is used for diagnosis (e.g., a peak at 1x the shaft speed indicates imbalance), while its amplitude quantifies the severity of the fault at that specific frequency.
- **Spectral Centroid:** This is the "center of mass" of the spectrum, representing the weighted mean of the frequencies present in the signal, with the magnitudes of the frequencies serving as the weights. A shift in the spectral centroid over time is a powerful indicator of a change in the machine's state of health. For example, an increase in the centroid value indicates that energy vibration is shifting towards higher frequencies, which can be a sign of worsening wear or friction. For a spectrum $S(k)$ with N frequency bins, it is calculated as:

$$\textbf{Spectral Centroid} = \frac{\sum_{k=1}^{N} S(k) \cdot f(k)}{\sum_{k=1}^{N} S(k)} \; where\ f(k)\ is\ frequency\ of\ bin\ k \quad 4.9$$

- **Spectral Kurtosis (SK):** This is an advanced statistical tool that indicates how the impulsiveness of a signal varies with frequency. Unlike time-domain kurtosis, which provides a single value for the entire signal, SK calculates the kurtosis for each individual frequency band. It is particularly powerful for detecting incipient, non-stationary faults, such as the repetitive impacts from an early-stage bearing defect, which often appear as high kurtosis values in a specific high-frequency band long before they are visible in the overall spectrum [61].
- **Harmonics and Sidebands:** These are not single scalar features but are critical diagnostic patterns within the spectrum.
  - **Harmonics** are integer multiples of a fundamental frequency (e.g., 2x, 3x, 4x the shaft speed). The presence and amplitude of harmonics can indicate non-linearity or distortion in the system, often caused by mechanical looseness, rubbing, or severe misalignment.
  - **Sidebands** are peaks that appear at frequencies equally spaced around a central carrier frequency. They are caused by modulation effects and are classic indicators of gear tooth faults (where the sidebands are spaced at the gear's rotational speed around the gear mesh frequency) or specific bearing defects.

4.1.3.3 Time-frequency domain features

The primary limitation of the traditional Fast Fourier Transform (FFT) is that it assumes the signal is **stationary—**meaning its statistical properties (like mean and variance) and frequency content do not change over time. While this is a reasonable approximation for machines operating under constant load and speed, many real-world signals are **non-stationary**. This includes signals from machines with varying speeds, transient events like impacts, or faults that only appear intermittently. The FFT provides information on what frequencies are present in the signal, but it loses all information about when they occur [62].

To analyze these non-stationary signals, time-frequency analysis is required. These techniques produce a two-dimensional representation of the signal, showing how its frequency content evolves over time. The two most common methods for this are the Short-Time Fourier Transform (STFT) and the Wavelet Transform (WT).

- **Short-Time Fourier Transform (STFT):** The STFT is the most intuitive approach to time-frequency analysis. It works by taking a long time-series signal, dividing it into smaller, overlapping segments of equal length (a process called "windowing"), and then applying the FFT to each segment individually. The result is a series of frequency spectra, one for each point in time, which are combined to create a spectrogram. The spectrogram is a 2D plot with time on one axis, frequency on the other, and the amplitude (energy) represented by color or intensity. While effective, the STFT is limited by the Heisenberg-Gabor uncertainty principle: the choice of window size creates a fixed trade-off between time resolution and frequency resolution. A short window provides good time resolution but poor frequency resolution, while a long window provides good frequency resolution but poor time resolution.
- **Wavelet Transform (WT):** Wavelet Transform overcomes the fixed-resolution limitation of the STFT. Instead of using a fixed-size window of sinusoids, the WT decomposes the signal by using small, wave-like functions called wavelets, which are localized in both time and frequency. A key advantage of the WT is its multi-resolution analysis capability. It uses short, high-frequency wavelets to analyze high-frequency components of the signal, providing excellent time resolution for transient events like sharp impacts. Simultaneously, it uses long, low-frequency wavelets to analyze low-frequency components, providing excellent frequency resolution for slower-varying phenomena. This makes the WT particularly well-suited for analyzing machinery faults that often manifest as low-energy, high-frequency transients. The output of a Continuous Wavelet Transform (CWT) is often visualized as a scalogram [63].

From these 2D time-frequency representations, features can be extracted. The entire spectrogram or scalogram can be treated as an "image" and fed directly into deep learning models like Convolutional Neural Networks (CNNs). Alternatively, statistical features (e.g., energy, entropy) can be calculated from the coefficients produced by a Discrete Wavelet Transform (DWT) or a Wavelet Packet Decomposition (WPD), which provide a more compact set of condition indicators.

4.1.4 ISO standards

To move from a qualitative assessment to a quantitative, objective evaluation of machinery health, the industry relies on internationally recognized standards. The International Organization for

Standardization (ISO) has developed a series of standards that provide a framework for measuring and evaluating the severity of mechanical vibration. These standards offer a common language and a set of established alarm limits for different classes of machinery, providing a crucial baseline for condition monitoring programs.

The most widely referenced of these was the **ISO 10816 series**. However, these standards have been progressively updated and integrated into a more comprehensive framework. A key example relevant to industrial machinery is the **ISO 10816-3:2009**, which has now been revised by **ISO 20816-3:2022**. [64] These standards provide general guidelines for the evaluation of machine vibration by measurements on non-rotating parts (e.g., bearing housing).

The core methodology of these standards involves:

1. **Machine Classification:** Machines are categorized into different groups based on their type, size (power rating), and the rigidity of their foundation (e.g., rigid or flexible).
2. **Measurement Parameter:** The primary parameter for vibration severity evaluation is the Root Mean Square (RMS) value of vibration velocity, typically measured in a broad frequency band (e.g., 10 Hz to 1,000 Hz). As discussed previously, velocity is chosen because of its direct relationship to the destructive energy of the vibration.
3. **Severity Zones:** The standard defines a set of four severity zones, often visualized in a chart, which are used to assess the machine's operational condition. A generalized example of this severity chart is shown in Table 2. These zones provide a simple, "traffic-light" style assessment:
    - **Zone A (Green):** Good. The vibration values are typical of new or newly refurbished machines.
    - **Zone B (Yellow):** Satisfactory. Machines in this zone are considered acceptable for unrestricted long-term operation.
    - **Zone C (Orange):** Unsatisfactory. Machines operating in this zone are not considered safe for long-term operation. Corrective action should be taken to identify and address the root cause of the increased vibration.
    - **Zone D (Red):** Unacceptable. The vibration values in this zone are considered critical and are likely to cause damage to the machine.

While these ISO standards are an indispensable tool for setting alarms and making general assessments of machine health, they are fundamentally **threshold-based**. They can effectively indicate that a problem exists once the vibration level has crossed a predefined limit, but they do not provide in-depth diagnostic information or predictive capabilities on their own. They serve as a crucial first line of defense, which can then be complemented by the more advanced diagnostic and prognostic models detailed in this thesis to determine the specific fault type and its future progression.

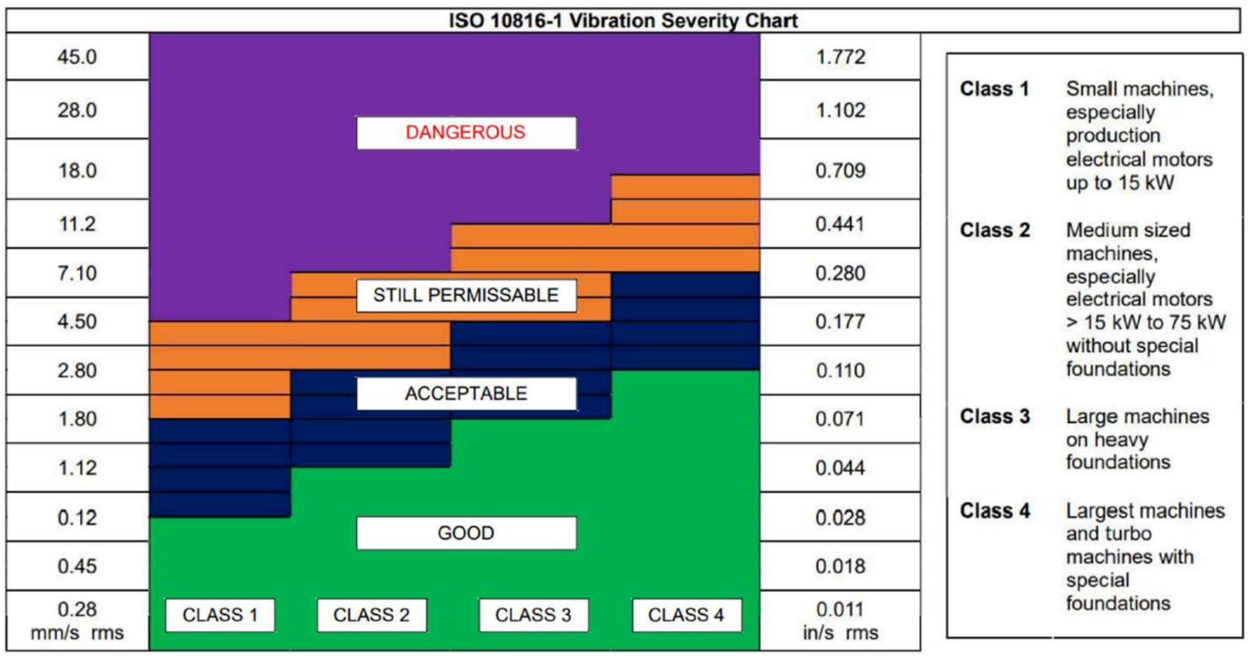


Table 2 - ISO 10816-1 Vibration Severity Chart

4.2 Ultrasound

Ultrasonic monitoring is a condition monitoring technique that focuses on the detection of high-frequency, low-energy stress waves, typically in the range of 20 kHz to 100 kHz, which are well beyond the limit of human hearing. In an industrial machinery context, these ultrasonic signals are generated by three primary sources, often referred to by the acronym **FIT**: **F**riction, **I**mpacting, and **T**urbulence [65].

- **Friction**: This is a common source in rotating machinery, particularly from inadequate lubrication in bearings, which causes microscopic rubbing and asperity contact.

- **Impacting**: This refers to the repetitive, low-energy impacts produced by phenomena such as the rolling elements of a bearing passing over a microscopic pit or spall on the raceway, or a cracked gear tooth making contact.
- **Turbulence**: This is generated by the chaotic flow of fluids in pipes or by pressure leaks and is also associated with cavitation in pumps.

The primary advantage of ultrasonic monitoring is its remarkable sensitivity to these incipient failure modes. The low-energy stress waves generated by early-stage faults are often detectable with ultrasound long before they manifest as a significant change in the lower-frequency vibration spectrum or as a noticeable increase in temperature. This makes ultrasound a powerful tool for providing the earliest possible warning of developing problems, particularly in rolling-element bearings [66].

This research focuses specifically on **structure-borne ultrasound**, where the high-frequency waves are transmitted through the solid structure of the machine and are captured using a contact sensor, similar to an accelerometer. The raw, high-frequency ultrasonic signal, however, is difficult to analyze and store directly. To make this data useful and audible to a human analyst, it must be processed and transformed into a lower-frequency range. The most common method for achieving this is heterodyning.

4.2.1 Heterodyned transformation

To make the high-frequency, inaudible ultrasonic signals captured by sensors useful for analysis, they must be translated into the human auditory range. This is achieved through a signal processing technique known as **heterodyning**. The principle was first demonstrated in 1901 by Reginald Fessenden, an inventor who also pioneered amplitude modulation (AM) radio, and it remains the foundation of all modern, high-quality ultrasound measurement instruments. The heterodyning process effectively gives an inspector a "sixth sense" by shifting an inaudible frequency range into an audible one, allowing for a direct, intuitive interpretation of high-frequency phenomena.

The human auditory range is limited, typically from 20 Hz to 20 kHz, while the ultrasonic signals of interest for condition monitoring exist at much higher frequencies. As illustrated in Figure 41, heterodyning provides the bridge between these two domains.

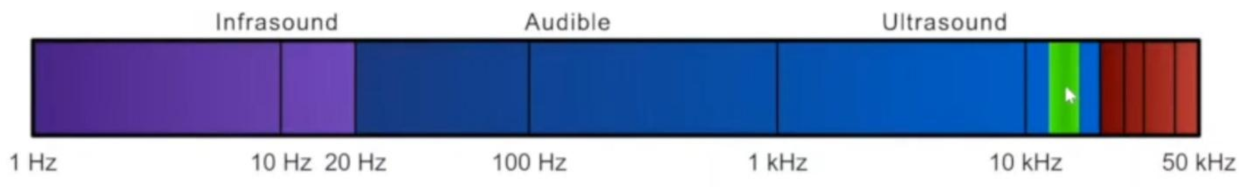


Figure 41 - Sound Spectrum. Source: Mobius Institute [67].

The process involves a series of well-defined electronic steps:

1. **Filtering**: The raw signal captured by the sensor contains a broad spectrum of frequencies. The first step is to apply a bandpass filter to remove all audible and other unwanted frequencies, isolating only the specific ultrasonic frequency range of interest.
2. **Mixing**: Next, a pure-tone sinusoidal signal of a fixed, high frequency, known as the local oscillator or mixer signal, is generated internally by the instrument. The precise frequency of this mixer varies by manufacturer; for instance, some instruments operate with a mixer frequency of approximately 38.5 kHz.
3. **The Heterodyne Process**: The isolated ultrasonic signal is then electronically multiplied (or modulated) with this mixer signal. This modulation is the core of the heterodyning process. As a result of this multiplication, two new signals are produced: a sum frequency component $(\boldsymbol{\theta} + \boldsymbol{\varphi})$ and a difference frequency component $(\boldsymbol{\theta} - \boldsymbol{\varphi})$. Crucially, the difference frequency component now falls within the human audible range, while the sum component remains at a very high, inaudible frequency and is filtered out. Figure 42 provides a schematic representation of this process.

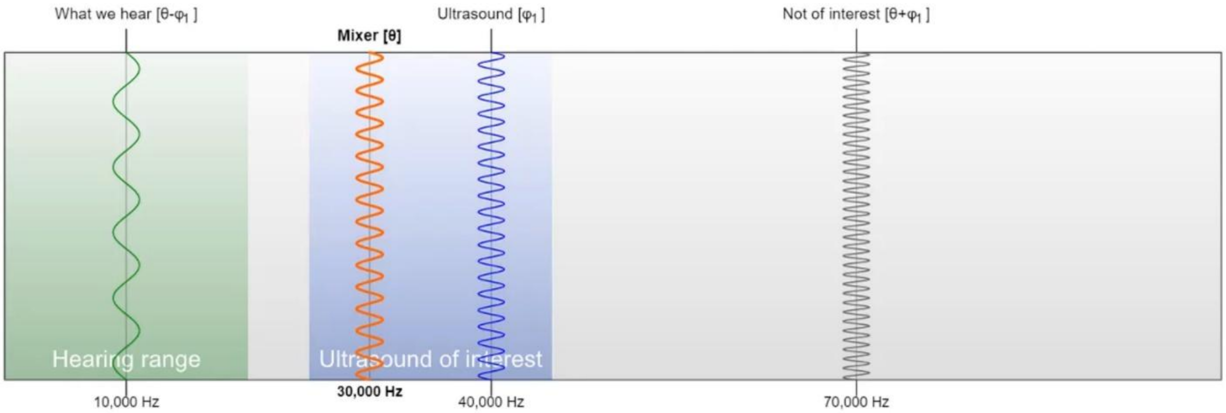


Figure 42 - Heterodyned Transformation. Source: Mobius Institute [67].

What is remarkable about this process is that the temporal characteristics of the original ultrasonic signal are preserved. A transient "click" that occurred with a one-second period in the ultrasonic range will be perceived as an audible "click" with the same one-second period. This means that the rhythm, pattern, and character of the original signal are maintained, which is critical for diagnostics.

The mathematical foundation for this phenomenon is the product-to-sum trigonometric identity, shown below. This identity demonstrates that the multiplication of two sine waves (the mixer signal $\boldsymbol{\theta}$ and the original ultrasound signal $\boldsymbol{\varphi}$) results in two new cosine waves, one at the difference frequency and one at the sum frequency.

$$\boldsymbol{sin\theta sin\varphi} = \frac{\boldsymbol{cos(\theta-\varphi)-cos(\theta+\varphi)}}{\boldsymbol{2}} \quad 4.10$$

In practice, an ultrasonic signal from machinery is not a single pure tone but a complex signal containing multiple frequencies. The heterodyning process acts on all these frequencies simultaneously, translating the entire ultrasonic "soundscape" down into the audible range. As illustrated in Figure 43, this means that the rich diagnostic information contained in the complete high-frequency signal is made directly available for analysis.

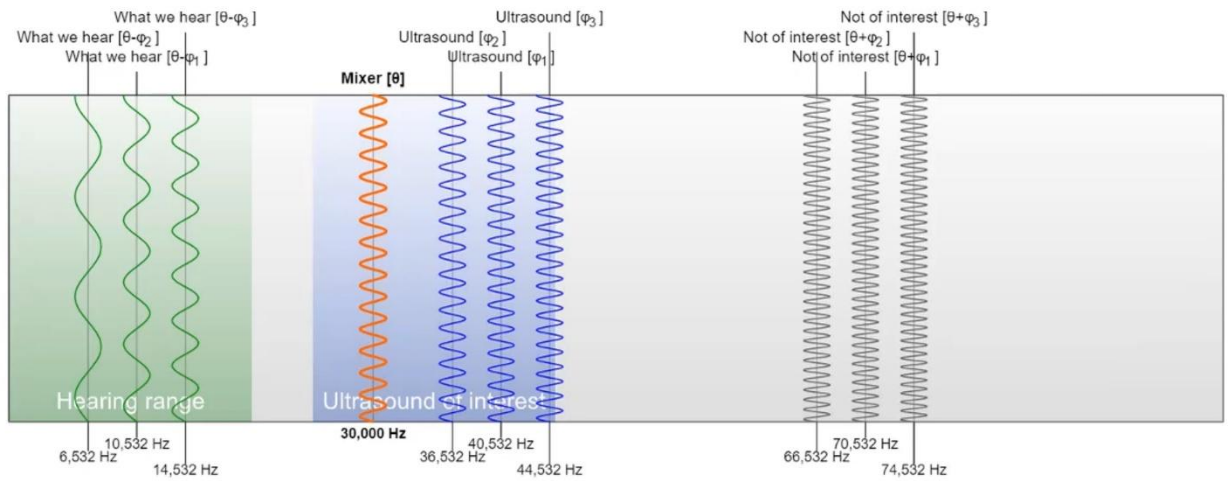


Figure 43 - Heterodyned Transformation for multiple signals. Source: Mobius Institute [67].

4.2.2 Instrumentation: SDT COMMONSense Sensor

The theoretical principles of structure-borne ultrasound and heterodyning were practically applied in this research through the use of a specific sensor: the “SDT COMMONSense 4-20 mA Contact Dynamic”. This sensor is designed for permanent mounting on critical assets to provide continuous condition monitoring data. The sensor offers multiple output modes, with the **dynamic mode** being the one used for acquiring the raw, time-domain waveform necessary for advanced signal analysis. The key technical specifications of this mode are detailed below.

The relationship between the measured ultrasonic signal and the sensor's output voltage in dynamic mode is defined by the sensor's output equation. This equation is fundamental for correctly scaling the acquired voltage signal back to its physical units.

$$\boldsymbol{Sensor\ signal}\ [\boldsymbol{mV}] = \frac{(\boldsymbol{current}[\boldsymbol{A_{DC}\ in\ mA}]-\mathbf{12}\ [\boldsymbol{A_{bias}}])\boldsymbol{x}\mathbf{208{,}25}[\boldsymbol{\Omega}]}{\boldsymbol{linear\ Gain}} \tag{4.11}$$

To accommodate a wide range of signal amplitudes, from subtle friction to high-impact events, the sensor features selectable gain settings. The sensitivity of the dynamic output is directly dependent on the chosen gain, as detailed in the manufacturer's specifications in Table 3.

Gain $[dB]$	Linear Gain	Sensitivity $[V]/[mA]$	Offset/$A_{bias}[mA]$	Optimal range $[mV]$
0	1	208.25	12	[-1666 ; +1666]
12	4	52.06	12	[-418 ; +418]
24	16	13.02	12	[-105 ; +105]
36	63	3.31	12	[-26 ; +26]
48	251	0.83	12	[-6.6 ; +6.6]
60	1000	0.208	12	[-1.7 ; +1.7]

Table 3 - Sensitivities according to the gain, in dynamic mode for COMMONSense sensor. Source: [68].

A critical characteristic of any sensor is its frequency response curve, which describes its sensitivity to different frequencies within its operational range. Figure 44 presents the normalized heterodyned response curve for the COMMONSense sensor. This curve illustrates the sensor's peak sensitivity and its effective bandwidth. Understanding this response is crucial for accurately interpreting the resulting frequency spectra, as it indicates the frequency ranges where the sensor is most and least sensitive to ultrasonic activity. These technical specifications form the basis for the correct acquisition and interpretation of the ultrasonic data presented in the case studies.

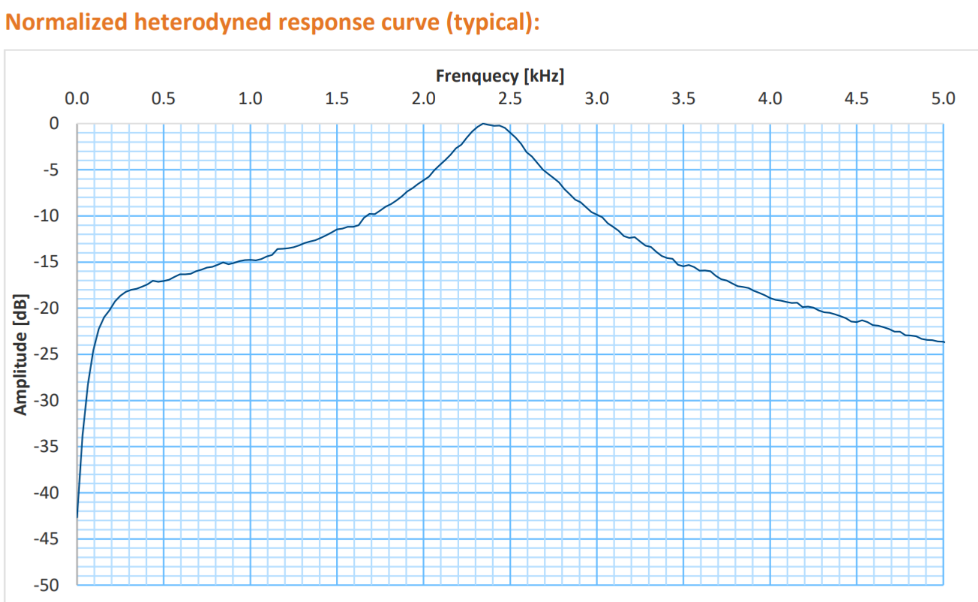


Figure 44 – Normalized heterodyned response curve for COMMONSense sensor. Source: [68].

In subsequent chapters, when reference is made to the sensor, this acronym will be used: Contanct Ultrasound Sensor (CUS).

4.3 Oil Quality Analysis

Lubricating oil is the lifeblood of most industrial machinery. Beyond its primary function of reducing friction between moving parts, it also serves to dissipate heat, clean internal components by carrying away contaminants, and protect surfaces from corrosion. The condition of this lubricating oil is, therefore, a direct and comprehensive indicator of the health of both the oil itself and the machine it lubricates. The analysis of lubricating oil, a discipline also known as tribology, is a cornerstone of predictive maintenance for many types of equipment, particularly gearboxes, engines, and hydraulic systems.

Oil analysis provides diagnostic information across three primary categories:

1. **Fluid Properties:** This involves tracking the degradation of the oil's physical and chemical properties over time. Key parameters include:
   - **Viscosity:** This is the oil's most critical property. A significant change in viscosity (either an increase or decrease) indicates that the oil has degraded, overheated, or become contaminated, compromising its ability to lubricate effectively.
   - **Chemical Composition:** Metrics such as the Total Acid Number (TAN) and Total Base Number (TBN) are used to track the chemical breakdown of the oil and the depletion of its additive packages. An increase in TAN, for example, can indicate oxidation, which leads to sludge and varnish formation.
2. **Contamination:** This category focuses on identifying the presence of harmful foreign substances in the oil. Contaminants can be external, such as dirt/silica (indicating a faulty air filter) or water (indicating a seal failure or condensation), or internal, such as soot or fuel in an engine (indicating poor combustion). The detection and quantification of these contaminants are crucial for identifying operational issues and preventing accelerated wear.
3. **Wear Debris Analysis:** This is the most powerful diagnostic aspect of oil analysis for predictive maintenance. As mechanical components wear, they release microscopic metallic particles into the oil. By analyzing the concentration and elemental composition of these particles, it is possible to identify precisely which components are degrading. For example, a high concentration of iron might point to gear or bearing wear, while the presence of copper could indicate wear in bushings or thrust washers.

Traditionally, oil analysis has been performed **offline**, where a sample is periodically drawn from the machine and sent to a laboratory for detailed analysis using techniques like spectrometry and ferrography. While highly accurate, this approach provides only intermittent snapshots of the machine's condition. In line with the data-driven focus of this research, **online oil quality sensors** are increasingly being deployed. These sensors are installed directly on the machinery and provide a continuous, real-time data stream on key oil properties, such as viscosity, dielectric constant (which is sensitive to oxidation and contamination), moisture content, and even particle count. This continuous data is ideal for integration into a machine learning framework for real-time anomaly detection and trend analysis. This specific methodology, leveraging online sensors for continuous oil condition monitoring, is the approach adopted for the experimental case studies detailed in subsequent chapters [69].

4.4 Thermal Monitoring

Thermal monitoring is a condition monitoring technique based on the fundamental principle that nearly all forms of energy loss in a mechanical or electrical system ultimately manifest as heat. An abnormal or unexpected change in the operating temperature of a component is therefore a reliable, non-invasive indicator of a developing fault or an inefficient process.

In any machine, heat is generated by intended processes (e.g., combustion) and unintended losses. From a maintenance perspective, it is these unintended heat sources that provide diagnostic information. The most common sources include:

- **Friction:** From inadequately lubricated bearings, misaligned components, or mechanical rubbing.
- **Electrical Resistance:** Generated by loose connections, corroded terminals, or failing electrical components in motors and circuits.
- **Fluid Dynamics:** Resulting from issues like pump cavitation or internal leaks in hydraulic systems.

Temperature data can be acquired through two main categories of sensors: non-contact methods, such as infrared thermography, and direct contact sensors. This research is concerned with the latter, where sensors are physically mounted on the equipment to provide a continuous data stream. Common contact sensors include:

- **Thermocouples:** These are robust, inexpensive, and versatile sensors that generate a voltage proportional to a temperature difference. They are widely used for monitoring bearing and fluid temperatures.
- **Resistance Temperature Detectors (RTDs) and Thermistors:** These devices measure temperature based on the predictable change in their electrical resistance. They are often used when higher accuracy is required compared to thermocouples.

An unexpected and sustained increase in temperature is a clear sign of an abnormal condition. It can be one of the most effective ways to detect issues such as bearing lubrication failure, overheating in motor windings, cooling system inefficiencies, and high-resistance electrical faults.

While highly effective, it is important to note that a significant temperature increase is often a **lagging indicator** of failure, especially for mechanical faults. In the context of the P-F curve, a thermal anomaly may only become apparent in the later stages of degradation. Techniques like vibration and ultrasound analysis can often detect the root cause of a fault (e.g., the microscopic impacts of a bearing spall) long before it generates enough frictional energy to produce a measurable increase in heat. Therefore, thermal monitoring is an excellent and often essential complementary technology, but it is most powerful when used in conjunction with other condition monitoring techniques [65].

Beyond its direct application in fault detection, thermal monitoring also served a fundamental role in defining the boundary conditions for the experimental testing within this research. By establishing and maintaining controlled thermal parameters (such as various component temperatures, ambient temperature, and humidity), it was possible to generate consistent datasets. This experimental control is crucial, as it minimizes the influence of environmental or operational variations, thereby ensuring that any observed changes in other signals can be more confidently attributed to the evolving health state of the machine under observation.

4.5 Contextual and operational Parameters

In addition to the primary condition monitoring signals detailed previously, a comprehensive data acquisition strategy also involves recording a suite of contextual and operational parameters. While not always direct indicators of a specific fault, these variables are essential for understanding the machine's operating state, for data segmentation, and for correlating changes in condition indicators with operational demands.

These parameters, often acquired directly from the test bench control system or supplementary sensors, typically include:

- **Rotational Speed and Torque:** These are the most critical operational parameters, defining the load and speed conditions under which the primary CM data (e.g., vibration) is acquired. As demonstrated in the methodologies, they are fundamental for data segmentation and working point identification.
- **Electrical Parameters (Current, Voltage):** For electrically driven systems, motor current and voltage are crucial. As established, motor current often serves as a reliable proxy for mechanical torque and can also be used for Motor Current Signature Analysis (MCSA) to detect electrical or electro-mechanical faults.
- **Environmental Parameters (Humidity, Ambient Temperature):** These variables provide context for the operational environment. Significant changes in ambient conditions can influence component temperatures or even the properties of lubricants and recording them is essential for ruling out external factors when analyzing other signals.
- **System Pressures:** In hydraulic or lubricated systems, parameters like oil pressure are critical for ensuring that the system is operating within its specified limits.

By recording these parameters alongside the high-frequency vibration and ultrasound data, a complete and holistic view of the machine's behavior is captured. This rich dataset enables a more robust analysis, allowing for the clear separation of changes caused by operational variability from those that are true indicators of a developing fault.

5 Research Framework and Methodological Approach

The preceding chapters have established the necessary theoretical foundations, covering the strategic context of predictive maintenance, the mathematical principles of the machine learning models employed, and the physical nature of the signals used for condition monitoring. This chapter now serves to bridge this theory with the practical application by outlining the overarching framework, objectives, and standardized workflow that guided this research project. The aim is to clarify the project's primary goals, detail the general methodology applied across the various case studies, and provide a brief overview of the data and code management strategies employed.

1.1 Research Aims and Objectives

The primary aim of this research is to develop and validate a systematic framework for the application of Artificial Intelligence Techniques to industrial Condition Monitoring and Predictive Maintenance (CM-PdM). While numerous AI-based approaches are available to address CM-PdM challenges, there is no single "best" solution for all scenarios. The central focus of this PhD project is therefore the definition of a framework that allows for the selection and adaptation of the optimal technique for a given industrial application.

The development of this framework is guided by a set of critical, application-specific considerations, including:

- **Hardware Implementation Constraints:** Evaluating the suitability of a model for a given hardware target, whether it is a resource-constrained embedded system for real-time, on-device inference or a powerful cloud-based platform for complex, offline analysis.
- **Software and Development Platform:** Assessing the implementation within different software ecosystems, such as MATLAB or Python, and leveraging their respective strengths in signal processing, algorithm development, and deployment.
- **Fault Dynamics:** Considering the time-to-failure characteristics of the targeted faults. Rapidly developing faults require low-latency models, while slow degradation allows for more complex, post-processing-based approaches.

- **Nature of the Data:** Tailoring the approach to specific data characteristics, such as the signal's frequency content (high-frequency vs. low-frequency), and determining the necessity and complexity of feature engineering.
- **Expected Analytical Output:** Aligning the choice of model with the desired outcome, whether it is simple anomaly detection, multi-class fault classification, or the regression-based prediction of Remaining Useful Life (RUL).
- **Operational Context:** Differentiating between the requirements for a real-time, continuous monitoring application versus a periodic, post-processing analysis.

1.2 General Methodological Workflow

To address these objectives in a structured and repeatable manner, a standardized methodological workflow was adopted for the development and evaluation of all predictive models. The process begins with the acquisition of data that describes the system in both **healthy** and, where available, **faulty** conditions. This raw data is then subjected to a series of processing steps to transform it into a format from which informative features can be extracted, which are then used to train and validate a machine learning model.

This standardized process is visually summarized in Figure 45, which outlines the general workflow applied for anomaly detection tasks throughout this research.

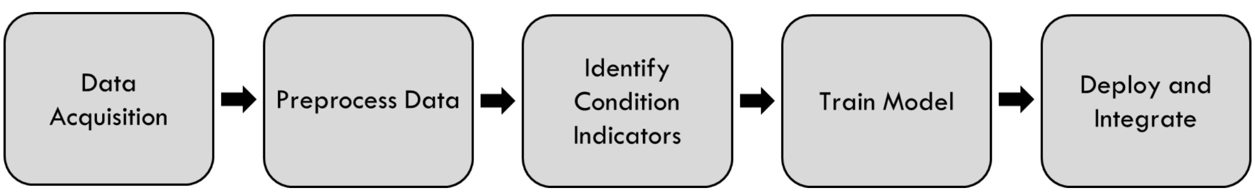


Figure 45 - Workflow methodology for Anomaly Detection, general approach.

One of the most critical stages in this workflow is the identification of **Condition Indicators**. This phase requires a deep understanding of the underlying physics of the system and the data to engineer or extract features that can effectively distinguish between a "**healthy**" and a "**deteriorating**" operational state. The success of this feature definition stage is heavily reliant on the quality of the acquired data. It is therefore essential that the data is collected under well-defined and standardized boundary conditions, a principle that was strictly adhered to in the experimental setups of this project.

The goal of applying this workflow can be categorized into a hierarchy of analytical objectives, as was first introduced in Chapter 2.3. Listed in order of increasing complexity and diagnostic value, these are:

- **Detect Anomalies:** To identify when the machine deviates from its normal operating condition.
- **Classify Failures:** To diagnose the specific type of fault that is occurring.
- **Estimate Remaining Useful Life (RUL):** To predict the future time-to-failure.

The primary focus of the results obtained throughout this research project is centered on the first and most fundamental objective: anomaly detection. As will be detailed in the case studies, the developed methods successfully identify the onset of degradation. In certain applications, by leveraging a specific set of engineered features, it was also possible to achieve the second objective and classify different types of failures using a similar modeling approach.

Crucially, the output of the models developed in this research is typically not a single scalar value or a binary "yes/no" flag. Instead, the output is framed as a Health Index (or a Warning Index), which is a continuous time-series trend that represents the evolving health state of the component under observation. This continuous health trend is a powerful diagnostic tool in its own right. Furthermore, it serves as a critical prerequisite for the final objective of RUL estimation. By extrapolating the trajectory of this Health Index to a predefined failure threshold, it becomes conceptually possible to estimate the Remaining Useful Life, representing a clear pathway for future development beyond the current scope of this work.

The specific application of this workflow, including the data pre-processing, feature selection, and model tuning for each distinct industrial application, will be described in detail in the subsequent chapters dedicated to the case studies.

1.3 Research Data and Code Management

Effective management of data, source code, and documentation is critical for ensuring the integrity, reproducibility, and long-term viability of a complex, multi-year research project. To address these requirements, a structured management framework was established, centralizing the various assets of the project into distinct, interconnected repositories. The architecture of this framework is visually summarized in Figure 46 and is comprised of four main components.

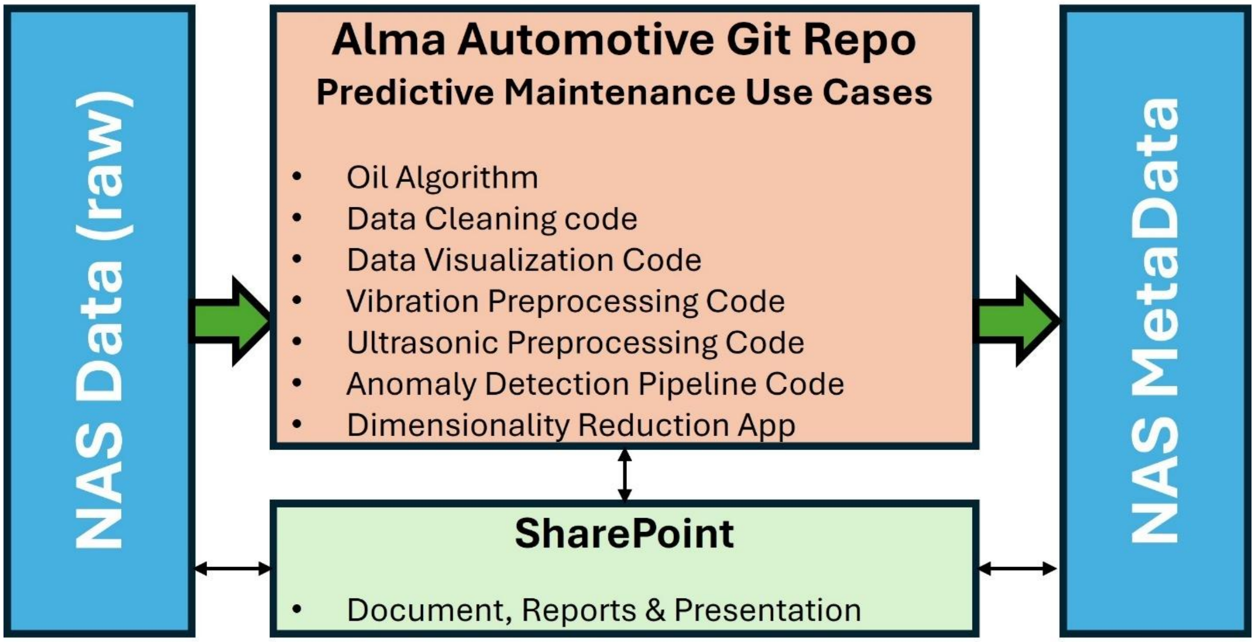


Figure 46 - Data, Code & Reports management.

1. **Raw Data Repository (NAS):** This component serves as the primary, centralized archive for all foundational data generated from the experimental test campaigns. This includes the high-resolution raw sensor signals as well as initial metadata that has been partially post-processed and transmitted by the edge devices installed on-site. This repository acts as the immutable "single source of truth" for the project's foundational data, ensuring that both the original measurements and the preliminary on-site calculations are preserved and can be revisited at any stage of the research.

2. **Source Code Version Control (Git Repository):** All source code developed throughout the research project is managed in a dedicated Git repository. This includes scripts and applications for data pre-processing (vibration, ultrasonic), data cleaning and visualization, the implementation of anomaly detection pipelines, and custom applications for tasks such as dimensionality reduction. The use of a version control system like Git is essential for tracking changes, enabling collaboration, and ensuring the full reproducibility of all analytical results.

3. **Documentation and Reporting Archive (SharePoint):** This platform is used for the management and dissemination of all human-readable documentation. This includes

scientific reports, progress updates, formal presentations, and technical documents related to the research. It serves as the central hub for knowledge sharing and project reporting.

4. **Processed Data and Metadata Repository (NAS):** This repository designates storage for all derived data products. After the raw data is processed by the source code, the outputs—such as cleaned datasets, extracted features, model results, and generated Health Indices—are stored here. This clear separation of raw and processed data ensures data integrity and provides a clear lineage from the original measurements to the final analytical results.

The workflow established by this framework is cyclical and ensures clear data provenance. Raw data is ingested from the **Raw Data Repository** and processed using the version-controlled code from the **Git Repository**. The outputs of this processing are twofold: the resulting processed data and metadata are archived in the **Processed Data Repository**, while the insights and conclusions are formalized into documents and reports stored in the **Documentation Archive**. This systematic approach guarantees that all aspects of the research are organized, traceable, and securely managed. It is important to note that the **Git Repository** also contains the source code for the theoretical and comparative analyses presented in this thesis, including the scripts used for generating the synthetic dataset and for the practical comparison of the anomaly detection models detailed in Subchapter 3.3.6.

6 Test Campaign on Mechanical gearbox – case studies

The experimental data analyzed in this chapter originates from a series of durability tests conducted in collaboration with S.T.M. Riduttori, and Alma Automotive s.r.l. The test campaign comprised four distinct durability tests, all performed on the same Device Under Test (DUT) to ensure a consistent experimental baseline.

The Device Under Test (DUT) for all experiments was an industrial, four-stage helical gearbox manufactured by S.T.M. Riduttori. It is designed for high-torque applications and features a high reduction ratio of approximately 168:1. The gearbox architecture consists of four parallel-shaft helical gear stages and one orthogonal input stage, which utilizes a helical bevel gear pair. This configuration makes it representative of a wide range of heavy-duty industrial power transmission systems.

While all tests utilized the same gearbox, they were designed to explore different operational conditions, varying primarily in three aspects: the type of lubricating oil used, the duration and torque profile of the work cycle, and consequently, the specific faults that developed during operation. A summary of the entire experimental campaign is provided in Table 4. Although the predictive maintenance methodology presented was developed and refined using data from all four tests, this chapter will focus specifically on the analysis of the test that featured the longest duration and resulted in the most significant and diagnostically interesting failure.

Name of the Test	Working Cycle	Oil Used	Faults	Working Hours
Test7A	Running-in + Nominal	Castrol Optigear Synthetic X 320	Breaking Z1-P-AX3	480
Test7B	Running-in + Nominal	Castrol Optigear Synthetic X 320	Pitting Z1-P-AX3	2304
Test7C	Running-in + Nominal + Lubrication Test	Fuchs Renolin Unisyn XT 320	Bearing Fault AX3 Pinion fault Z1-P-AX2	3850
Test7D	Running-in + Nominal + Overload	PETRONAS GEAR SYN 220	Pitting Z1-P-AX3	1827

Table 4 - Summary table of the entire experimental campaign

A detailed schematic of the gearbox's key mechanical components is provided in Figure 47, which illustrates the shafts, bearings, and gears that are the focus of the condition monitoring analysis. The diagnostic approach relies on analyzing the characteristic frequency orders associated with each component, defined as the signal frequency normalized by a reference shaft's rotational frequency. Phenomena such as gear tooth wear, shaft imbalance, or bearing race defects generate distinct vibration signatures at predictable orders that are dictated by the component's geometry and kinematics [70].

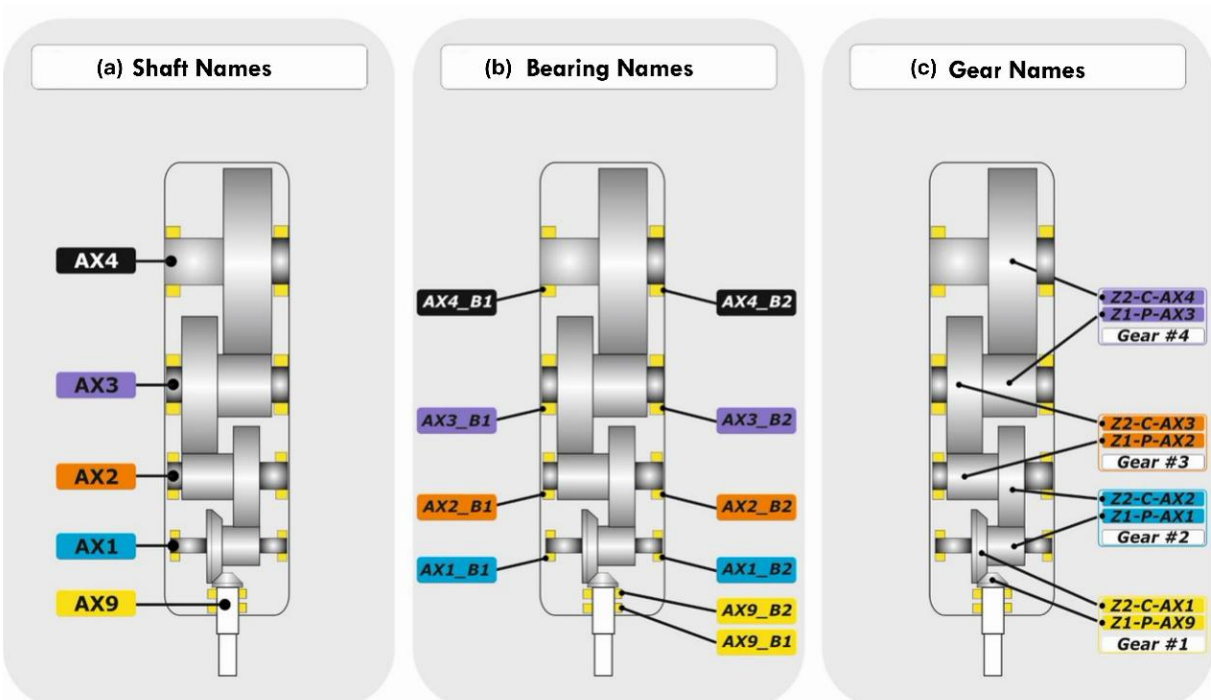


Figure 47 - Gearbox components, schemes and names. (a) Shafts Names, (b) bearing Names, (c) gear Names.

To systematically reference these components in the subsequent analysis, the following naming convention, corresponding to the labels in Figure 47, was established:

- **Shafts**: Labeled AX9 for the input shaft, and AX1 through AX4 for the subsequent stages.
- **Bearings**: Labeled AX(N)_B(S), where 'N' is the number of the shaft on which the bearing is mounted, and 'S' is an index (1 or 2) indicating its position on that shaft.
- **Gears**: Labeled based on their type and shaft, such as Z1_P_AX(N) for a pinion ('P') or Z2_C_AX(N) for a crown gear ('C'), where 'N' is the corresponding shaft number.

Table 5 compares the main characteristics of accelerometers, ultrasonic and other sensors installed in the DUT. In the next chapter another scheme will explain the position of the sensor as Figure 55 for Test7B, Figure 68 for Test7C and Figure 75 for Test7D.

Sensor type	Sensor Name	Physical quantity	Physical scale	Frequency Band
IEPE Piezoelectric	HS-100SF Hansford Sensors	Acceleration (m/s^2)	$100\ mV/g$	0.2 Hz – 15 kHz
CUS Ultrasonic	SDT Ultrasound	Amplitude Ultrasonic Vibration	$dB\mu V \rightarrow V$	0.25 – 4 kHz (Heterodyned)
Current Transducer	CT132TRAN	AC Motor Current	0-5V / 0-10V / 4-20mA DC	N/A
Hall Effect Sensor	AA080005-4 (M18x1.5)	Rotational Speed	Pulse Train (TTL/Digital)	N/A
Oil Debris Sensor	WearDetect Oil Debris Sensor	Ferrous Debris Volume (Fine & Coarse)	% FSC (via CAN bus)	N/A

Table 5 - Specifications of the primary sensors used in the experimental setup.

Among the installed sensors, a key instrument for defining the ground-truth health state of the gearbox throughout this research is the online oil debris sensor, referred to hereafter as OIL_SENS The specific model used is the WearDetect Oil Debris Sensor, and its position will be indicated in the sensor layout diagrams for each of the test campaigns discussed. This sensor is specifically designed to detect ferrous metal particles in the lubricating oil, providing a direct and early indication of wear or developing failure in gears and bearings. The sensor integrates:

- Ferrous Metal Detection: permanent magnet attracts debris, differentiating fine particles (<0.5 g) and coarse chips, reporting their volume as % of Full-Scale Deflection (FSD).
- Temperature and Oil Presence Detection: dielectric sensor monitors oil condition; absence of oil or >10% water contamination are flagged.

Debris accumulation is measured only on the sensor face. Overload occurs when collected debris saturates the sensing surface. Calibration of fine and coarse channels is performed at 100% FSD using reference debris samples.

6.1 Test Procedure

To ensure the repeatability and comparability of results across the different durability tests, a standardized Nominal Working Cycle was applied to the gearbox throughout the entire experimental campaign. The profile of this work cycle is illustrated in Figure 48, which displays the target input speed and the corresponding motor current setpoints for two complete cycles. Each discrete step in the cycle represents a specific working point.

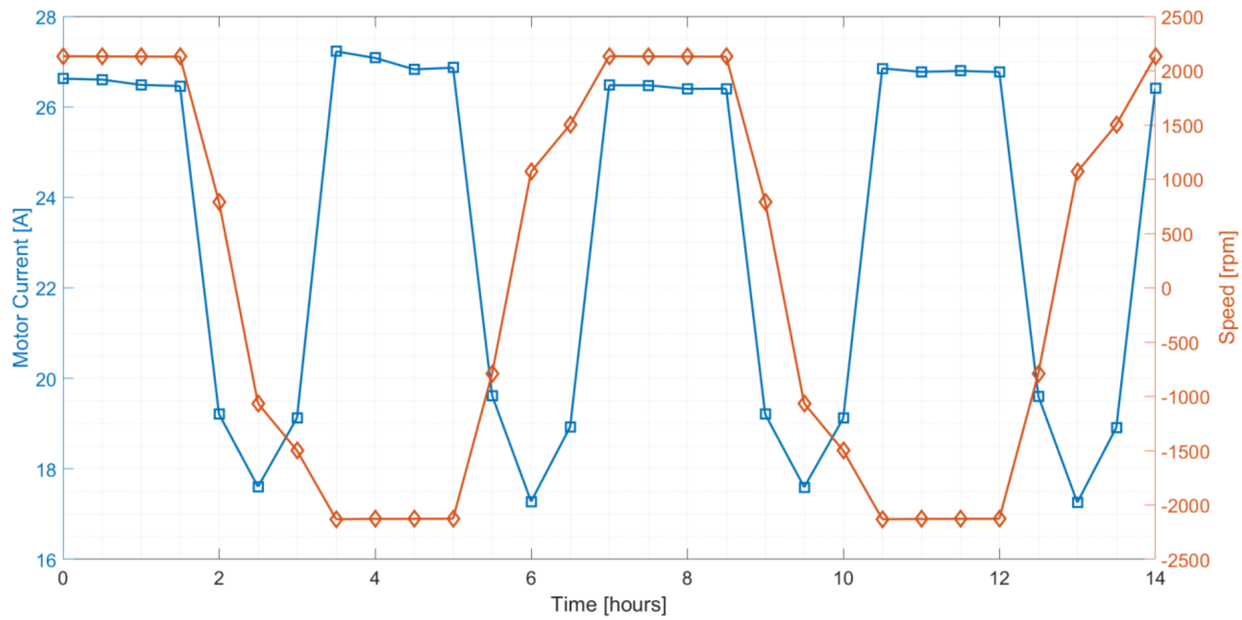


Figure 48 - Target Speed and Torque for two entire Nominal Working Cycle.

The nominal work cycle consists of eight operating points (four clockwise and four counterclockwise) repeated continuously. High-frequency acquisitions (50 kHz, ~150 s duration) were performed every 30 minutes using the condition monitoring system supplied by Alma Automotive (Warnn/AlmaLink). The 150s window was chosen to guarantee at least one full rotation of the output (slower) shaft even at low speeds (790 rpm input speed, corresponding to 4.71 rpm output speed), ensuring complete capture of cyclical phenomena.

As a direct measurement of the input torque was unavailable, the motor current was adopted as a reliable and proportional proxy, a common and effective practice in industrial motor applications [71]. The work cycle was designed to subject the gearbox to a range of representative operational loads, corresponding to nominal output torque values of 3500 Nm, 4500 Nm, and 5500 Nm.

While the figure shows the target setpoints, the actual measured data from the full durability campaign provides insight into the operational stability of these working points. Table 6 presents the statistical summary of the measured working points throughout the longest durability test Test7C, after the removal of outliers resulting from test restarts or transitional acquisition phases.

Working Point	Mean Current	Mean RPM	Std Current	Std RPM
WP 0	19.02	791.85	0.074	0.445
WP 1	17.35	-1065.49	0.105	0.226
WP 2	18.88	-1498.86	0.066	0.293
WP 3	26.59	-2128.47	0.189	1.690
WP 4	19.33	-790.81	0.088	0.287
WP 5	17.05	1069.09	0.096	0.376
WP 6	18.69	1502.87	0.069	0.390
WP 7	26.19	2130.73	0.205	1.632

Table 6 - Statistics of Working Point about Nominal Working Cycles, Test7C

A key observation from this statistical analysis is that the clusters corresponding to the higher speed and higher current (i.e., higher load) operating points exhibit a greater standard deviation. This increased variability is expected, as higher torque transmission can induce more dynamic fluctuations within the mechanical system. Understanding this baseline variability is crucial for setting appropriate thresholds for anomaly detection models.

6.2 Specific Methods

This section details the specific, multi-stage methodology developed for anomaly detection in the gearbox durability tests. It is important to note that the workflow described herein represents the most refined version of the framework, developed and finalized based on the comprehensive data from the Test7C experimental campaign. Consequently, the analysis of earlier tests, such as Test7B, may refer to a previous iteration of this methodology, with any differences being explicitly noted in the relevant results section.

The overall workflow of the proposed anomaly detection procedure is depicted in Figure 49. The methodology employs a hybrid approach, integrating physics-informed feature extraction with unsupervised machine learning models to identify incipient faults from sensor data.

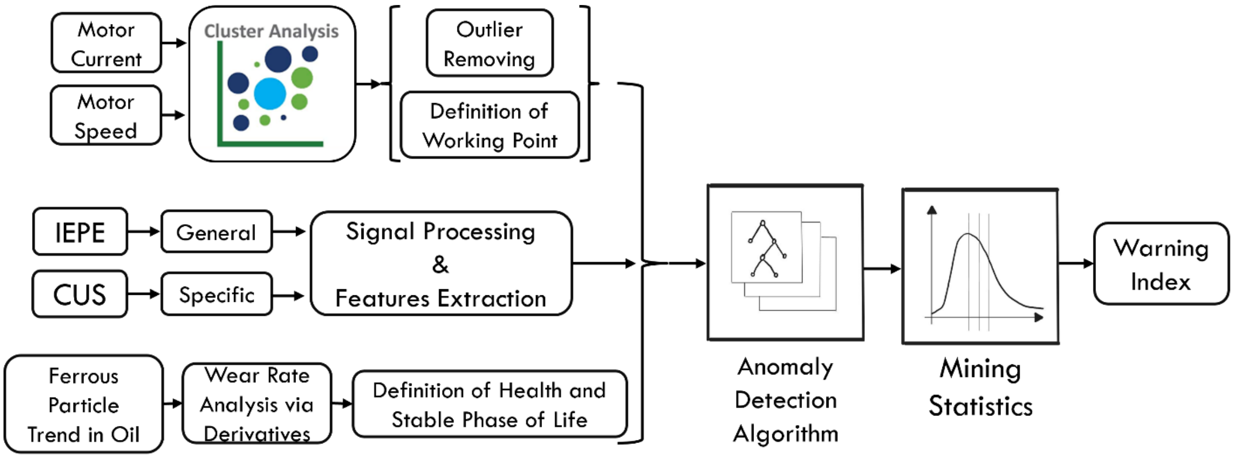


Figure 49 - Workflow of the Anomaly Detection Procedure Specific for this Campaigns Test

The central part of the block diagram illustrates how raw data is processed to determine significant features, which are then strategically grouped according to the mechanical component they are most sensitive to. A critical aspect of the methodology is that the anomaly detection models are only applied when the device under test is operating under stable, pre-defined working points. The definition of the "healthy" training baseline for these models is not arbitrary; it is determined by analyzing the ferrous particle trend from the online oil quality sensor. Once trained, the models evaluate an anomaly detection index for each feature group. To enhance robustness and avoid false positives, the instantaneous values of this index are subjected to further statistical analysis.

The following subsections will provide a detailed, step-by-step description of each major stage in this process, from the initial definition of working points to the final generation of a consolidated health index.

6.2.1 Definition of Working Point and Outlier Removal

In the first stage of the data processing pipeline (Figure 49, top left), motor current and rotational speed are analyzed to distinguish significantly from non-significant operating conditions. Current and speed have been selected due to their availability on the field: although this paper presents data generated during laboratory tests, the anomaly recognition methodology has been developed with the aim of recognizing faults even in standard working conditions, without the need to install additional sensors. This step is essential for rejecting acquisition data associated with transients or other states which are not representative of the gearbox’s nominal load cycle. Data points outside the standard working cycle are systematically removed, ensuring the integrity of the anomaly detection process. Outliers are identified as points that do not belong to the dense clusters corresponding to stable operating conditions. The procedure is fully automated and independent of the specific usage pattern. The largest deviations typically occur during start-up and shut-down phases, often related to scheduled inspections, and are more evident in motor current than in motor speed, which remains comparatively stable throughout the test.

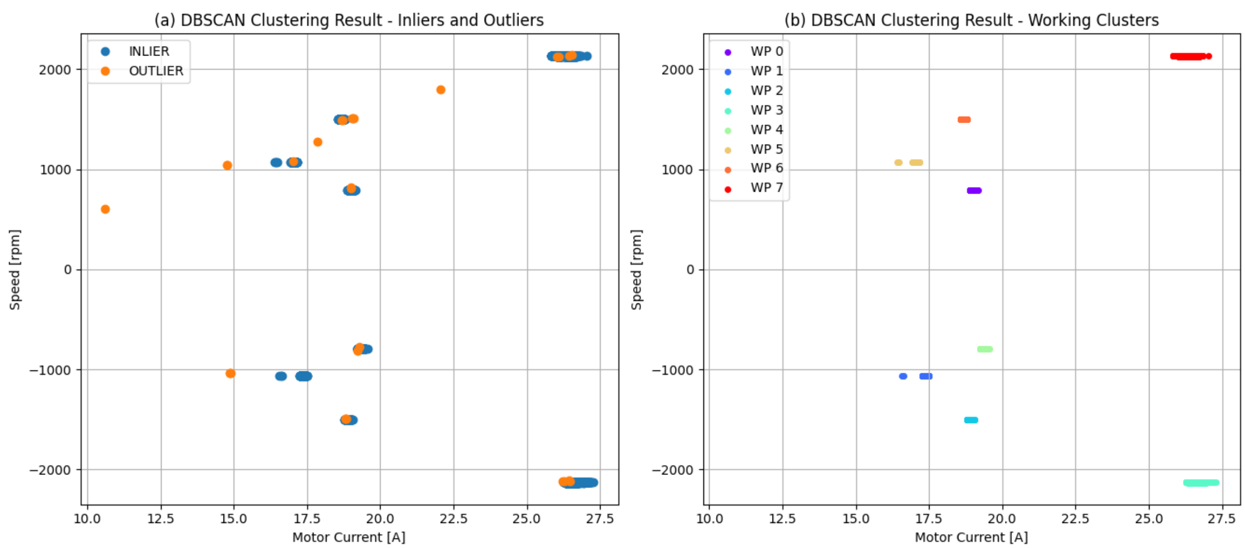


Figure 50 - DBSCAN Clustering Results. (a) Inliers and Outliers. (b)Clusters of Working Point

Efficient visualization of operating conditions is provided by the motor current–motor speed diagram (Figure 50). Here, all data points representing the summary of each acquisition are plotted, and the DBSCAN clustering algorithm is applied to classify them into two groups: inliers (valid working points) and outliers (non-representative conditions). During pre-processing, outliers are eliminated since they are sparse compared to the main clusters and cannot support meaningful statistical analysis. When data refers to an operating condition that has been labeled as an outlier, the anomaly detection algorithm is not applied.

In this context, motor current (measured via clamp meter) is reported as the primary variable, as it directly reflects the torque level at which the gearbox operates. Consequently, all transients are discarded, and subsequent vibration analysis is restricted to the stable working points defined in the diagram.

6.2.2 Definition of Health and Stable Phase of Life via Oil Analysis

Wear debris analysis is widely regarded as a cornerstone of gearbox diagnostics and prognostics [72]. Particles suspended in lubricating oil carry valuable information about the condition of internal components, and their quantity, size, shape, and composition can be linked to specific wear mechanisms. Since elemental spectroscopy is effective for fine particles but less sensitive to coarse debris, it is often complemented by ferrous density analysis to capture a wider range of wear indicators. Recent studies also emphasize the benefits of integrating wear debris data with vibration monitoring for more robust assessments of machine health. Advanced methods, such as multi-variable grey prediction models, aim to interpret complex debris patterns and correlate them with specific component degradation. Ultimately, monitoring wear particle generation over time is considered essential for predictive maintenance and reliability enhancement [73]. However, it is crucial to recognize that this data is inherently linked to the oil change intervals; trend analysis is therefore more significant than absolute values, as oil changes will reset the measurement baseline.

Building on this background, the developed approach exploits oil quality data—particularly ferrous particulate concentration—to define a supervised training set for anomaly detection. Figure 7(a) shows the raw oil quality signal inside the gearbox housing, together with its zero-phase filtered version. The lower subplot Figure 7(b) reports the first and second derivatives of the signal, which were used to segment the gearbox lifecycle into distinct phases. The region highlighted green delineates the period identified as the most stable, 'healthy' state of operation, which occurs immediately following the running-in phase. This segment was selected to serve as the training set for the anomaly detection models. The end of the running-in period is identified by analyzing the first derivative: the interval where the derivative assumes low values, corresponding to slow particulate growth, is selected. During the running-in phase, deposition is accelerated by the fact that tribologically stable contact between components is achieved, as confirmed by the bathtub curve. In the proposed strategy, the 25th percentile of the derivative is used as a threshold. The period that satisfies this condition—representing the healthy operational baseline—is highlighted by the transparent green band in Figure 7(b). This statistical threshold was chosen to provide a robust and objective method for isolating the period of lowest, most stable wear that represents the healthy operational baseline.

The derivative trend resembles the classical "bathtub curve," comprising three stages: (i) early failures or infant mortality (decreasing failure rate), (ii) useful life (approximately constant rate), and

(iii) wear-out (rapidly increasing failure rate due to wear). This correlation between failure rate and the evolution of ferrous debris from gear contacts has been highlighted in prior studies [74], [75], [76].

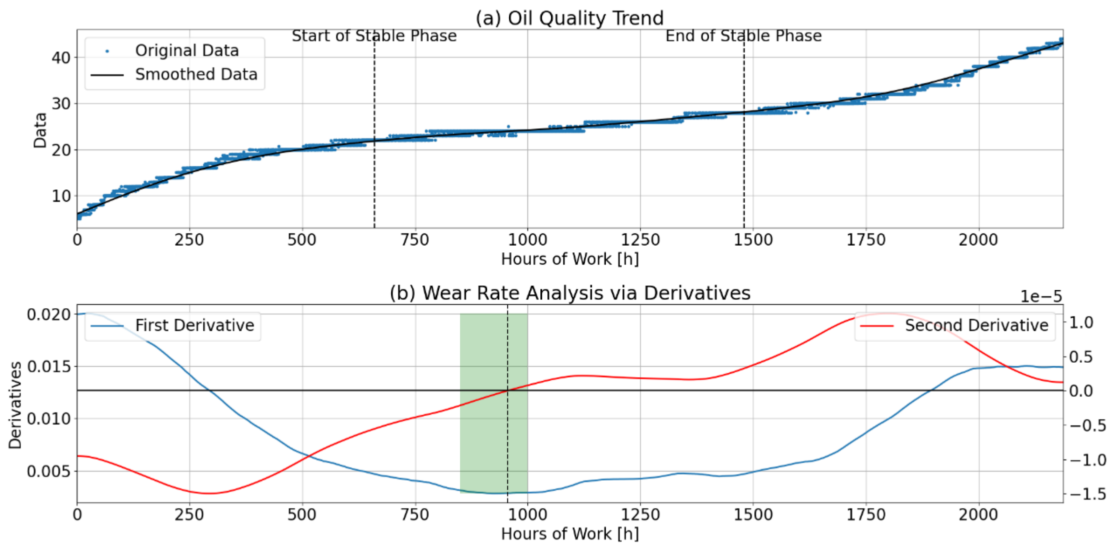


Figure 51 - (a) Oil Quality Trend. (b) Wear Rate Analysis via Derivatives

Accordingly, the training dataset for the anomaly detection model is drawn from the stable phase of gearbox life, thereby avoiding contamination by outliers. For Example, for Test7C case studies results, the training set consists of 600 working hours of data captured immediately after the running-in phase was declared complete. The selection of this duration is guided by two key requirements. First, enough acquisitions, a minimum of approximately one hundred, is needed to ensure that the clusters for all eight nominal working points are statistically well-defined and stable. Second, a longer observation period is necessary to compensate for the non-homogeneous acquisition sequence of the work cycle (see Figure 48), which samples certain high-speed working points in consecutive blocks. The 600-hour duration ensures that all operational states are adequately represented in the training data, preventing any potential bias in the model. This resulting dataset enables the implementation of a novelty detection strategy, in which the model is trained only on healthy data and tasked with detecting new anomalous observations.

6.2.3 Signal Processing and Features Extraction for IEPE and CUS

To translate the diagnostic principles into practice, the first feature extraction stage focused on the **IEPE vibration signals**. Standard time-domain statistical indicators were computed, including RMS, Crest Factor, Kurtosis, Skewness, and Peak value, to capture the overall energy and impulsiveness of vibration. In parallel, the frequency spectrum (FFT) was partitioned into contiguous 1 kHz bands from

500 Hz to 24.5 kHz, and the RMS of each band was calculated to describe how energy is distributed across the spectrum. This broadband vibrational analysis provides a general assessment of machine health, complementing the targeted, order-specific analysis of ultrasonic signals.

For the CUS ultrasonic sensors, a more selective strategy was required. The analysis focused on specific spectral orders known to be sensitive to gearbox components health. To extract a robust feature for each order, a proportional, percentage-based windowing method was adopted to compute RMS values from the FFT spectrum.

This technique addresses a key challenge in analyzing signals with a wide range of frequencies. Instead of using a fixed-width frequency band (e.g., ±10 Hz) for all orders, the analysis window for each order is defined as a percentage of that order's own central frequency. For a given order with a target frequency f_target, the window is defined over the interval [f_target × (1 - p), f_target × (1 + p)], where p is the specified window percentage. This proportional bandwidth ensures that low-frequency orders (e.g., a 0.5x shaft order of the last stage of reduction) are analyzed with a narrow, high-resolution window, while higher-frequency orders (e.g., a 10x GMF harmonic) are analyzed with a correspondingly wider window. This strategy is essential for preventing spectral overlap between adjacent low-frequency orders and ensures a consistent analysis relative to the order's frequency. The feature for each order is then calculated as the Root Mean Square (RMS) value of all the FFT magnitude points that fall within its defined percentage-based window. These features target the most common gearbox failure modes [53]:

- Low-order shaft harmonics (imbalance, misalignment, looseness).
- The fundamental Gear Mesh Frequency (GMF) and its first nine integer harmonics (i.e., orders 1xGMF through 10xGMF). The fundamental GMF is the primary indicator of the overall health of the gear meshing process, while its harmonics are highly sensitive to distributed faults such as generalized tooth wear or lubrication issues.
- Sidebands around the GMF, which are essential for detecting localized faults such as a single cracked or pitted tooth. Specifically, three pairs of sidebands modulated by the shaft rotational frequency (f) were monitored around each GMF: [GMF ± 1f, GMF ± 2f, GMF ± 3f].
- Characteristic bearing frequencies (defects in cage, rolling elements, outer race, inner race).

In total, 148 orders were extracted per sensor, distributed as follows:

- Shaft harmonics: orders 0.5, 1, 2, 3 for each of the 4 shafts → 16 orders.
- Gear Mesh Frequencies (GMF) and Harmonics: The fundamental GMF (order 1) and its first nine harmonics (orders 2–10) for each of the 4 gear stages → 4 stages × 10 orders/stage = 40 orders.
- Sidebands: Six key sidebands (three pairs: ±1f, ±2f, ±3f) around the GMF for both the pinion and crown gear shafts of each of the four stages → 6 sidebands/shaft × 2 shafts/stage × 4 stages = 48 orders.
- Characteristic bearing frequencies: which correspond to defects in the cage (Fundamental Train Frequency, FTF), rolling elements (Ball Spin Frequency, BSF), outer race (Ball Pass Frequency of the Outer race, BPFO), and inner race (Ball Pass Frequency of the inner race, BPFI). For 10 bearings → 40 orders.

By monitoring these orders, the framework produces a physically meaningful representation of gearbox health, once the levels corresponding to healthy conditions have been assessed.

In addition, envelope detection was applied to the raw ultrasonic signals, demodulating the high-frequency carrier and exposing low-frequency fault impulses [77]. This process is enhanced by the heterodyne principle, which shifts ultrasonic waves (typically 35–45 kHz for bearing defects) into the audible range. The CUS sensor circuitry mixes the ultrasonic signal with a local oscillator, producing new frequencies, including their difference, which can be monitored directly (see Chapter 4.2.1) [78]. This translation not only enables frequency-domain analysis but also allows maintenance professionals to listen to machine condition or store the signals as audio files [79], [80]. The ability to detect subtle, early-stage signatures of component wear makes heterodyned ultrasonic monitoring a cornerstone of modern PdM.

Figure 52 summarizes the feature extraction workflow.

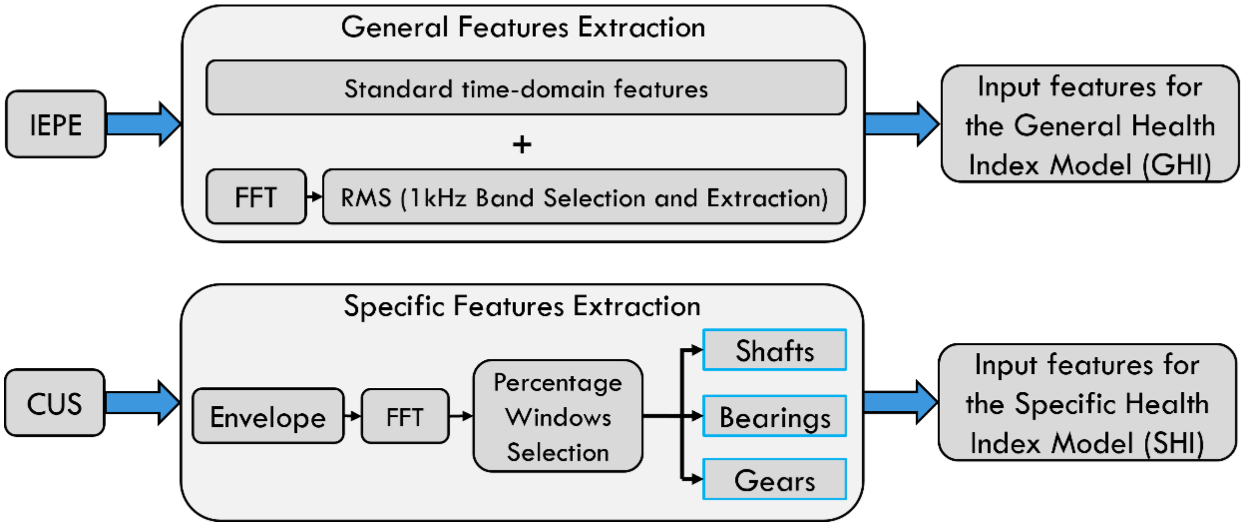


Figure 52 - Signal processing & features extraction workflow

6.2.4 Anomaly Detection and Health Index Generation

The second stage of the workflow leverages unsupervised models to fuse multi-sensor features into a unified health index. This approach is widely supported in the literature: autoencoders are often trained on normal operational data, and their reconstruction error is used as an anomaly score to represent machine health [81], [82]. Similar strategies have been applied in robotics, where unsupervised models, including autoencoders, are used to monitor the condition of robot arms and mobile systems [83], [84]. Multi-sensor data fusion is now considered a pillar of condition monitoring, with deep learning models combining features from different sensors and domains (time and frequency) to improve diagnostic accuracy [85], [86]. More advanced architectures, such as Generative Adversarial Networks (GANs) and autoencoder variants, have also been proposed to capture non-linear correlations in multivariate sensor data, enabling robust early fault detection [87].

In the present study, a parallel analysis strategy was adopted to generate distinct health indices for different sensor types and component groups, as illustrated in the workflow in Figure 53. This approach, which allows for both a general assessment of machine health and a more targeted diagnostic insight, based on the Local Outlier Factor (LOF) unsupervised anomaly detection model [48].

The LOF algorithm was implemented using the *lof* function from the MATLAB Statistics and Machine Learning Toolbox. For this study, the default hyperparameter settings of the toolbox were used. These include a value of NumNeighbors set to min (20, n-1) (where *n* is the number of unique observation), the Euclidean distance metric, and a ContaminationFraction of 0 to perform novelty detection. The models were trained exclusively on data from the stable operational phase, as identified via oil analysis (see 6.2.2). The resulting LOF score for each new acquisition, which represents its deviation from the learned 'healthy' data distribution, serves as the health index.

This trained model was then applied in two parallel streams:

1. A **General Health Index (GHI)** for the entire gearbox was developed by applying the LOF model to the set of general statistical features from the IEPE accelerometers. This provides a single, overarching measure of the system's vibrational health.

2. A highly granular, component-level analysis was conducted using the high-frequency data from the CUS ultrasonic sensors. The 148 order-based features were first segregated based on the specific component they describe. A separate LOF model was then trained for each individual component, yielding a dedicated **Specific Health Index (SHI)** for every monitored element. This results in five distinct SHIs for shafts, five for bearings, and four for gears. This highly targeted methodology moves beyond localizing a fault to a component group and enables the identification of the specific component that is beginning to degrade.

To ensure that the diagnostic output is sensitive to persistent degradation trends rather than spurious single-point anomalies, a final processing step involves applying a moving average filter to the raw SHI and GHI time series.

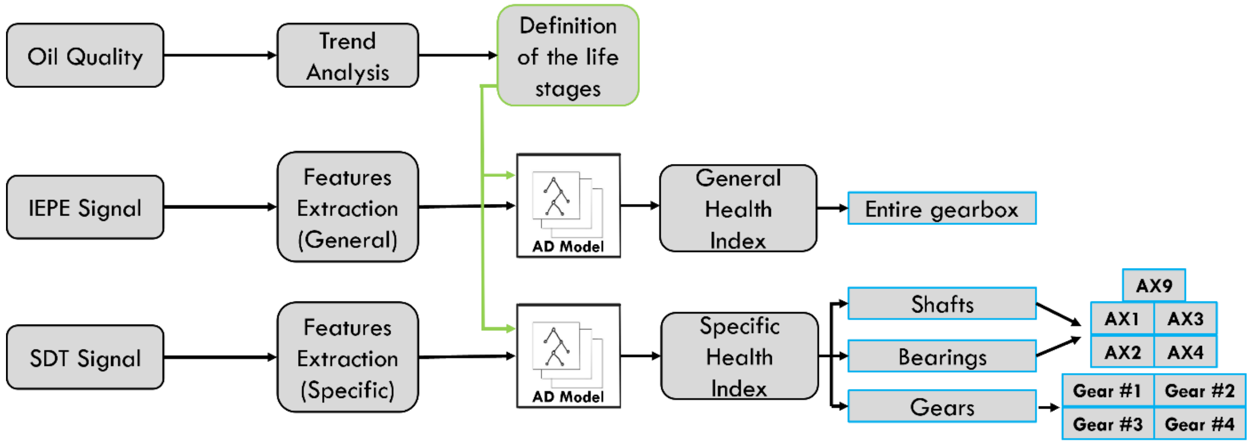


Figure 53 - Anomaly Detection and health Index Generation Pipeline

The final step in the methodology is the definition of alarm thresholds to make the Health Indices actionable for maintenance decisions. Two levels were defined: a 'Pre-Alarm' (Warning) threshold to indicate a significant deviation, and an 'Alarm' threshold to signal a more advanced state of degradation.

Theoretically, these thresholds can be set a priori using a statistical approach based on the distribution of the Health Index during the 'healthy' training period. For instance, a threshold could be defined as the mean plus a multiple of the standard deviation of the training HI values. For what concern Test7C, the mean GHI during training was 1.19, while the mean SHI values for the key axes were 1.75 (AX1), 1.75 (AX2), and 2.58 (AX3).

Given that the average values of these diverse health indices were observed to be of a comparable order of magnitude during the training phase, a fixed-threshold strategy was implemented to provide a simple yet generalizable approach. The fixed thresholds introduced—**a Pre-Alarm of 7.5 and an Alarm of 10.0**—can be considered an approximation of applying a common multiplicative factor to the various baseline health indices. As will be demonstrated in the Results section, these values proved to be highly effective for the LOF-based Health Indices in this test. They provide an excellent balance between sensitivity and robustness, successfully delivering an early and reliable warning of the developing fault while remaining insensitive to normal operational variability. This confirms that for this application, a well-chosen fixed threshold can serve as a simple yet powerful detection method.

A key advantage of this methodology lies in its low computational complexity, which supports real-time implementation on local hardware and ensures operational independence from external servers or cloud platforms.

6.3 Evaluate Campaign Results

This section presents the results from the experimental test campaign, with the primary objective of validating the specific anomaly detection framework detailed in Section 6.2. The following subchapters will discuss the findings from each of the durability tests.

It is important, however, to first clarify the role of the initial test, Test7A. The primary focus of this test was the initial configuration and setup of the experimental environment. This included the validation of the data acquisition system, the final definition of the nominal work cycle, and an initial assessment of sensor sensitivity to various features. While the oil data from this test was used to verify the consistency of ferrous particulate trends, the test was not run to failure and did not produce the extensive dataset required for a full diagnostic analysis. Consequently, a detailed analysis of Test7A is not presented in this thesis; the focus will be placed on the subsequent tests that completed longer durability cycles and developed significant, diagnostically relevant faults.

6.3.1 Test7B

The first durability test subjected to a detailed diagnostic analysis in this research is Test7B. The physical test environment is shown in Figure 54, which depicts the initial installation of the sensor suite within the S.T.M. R&D test room.

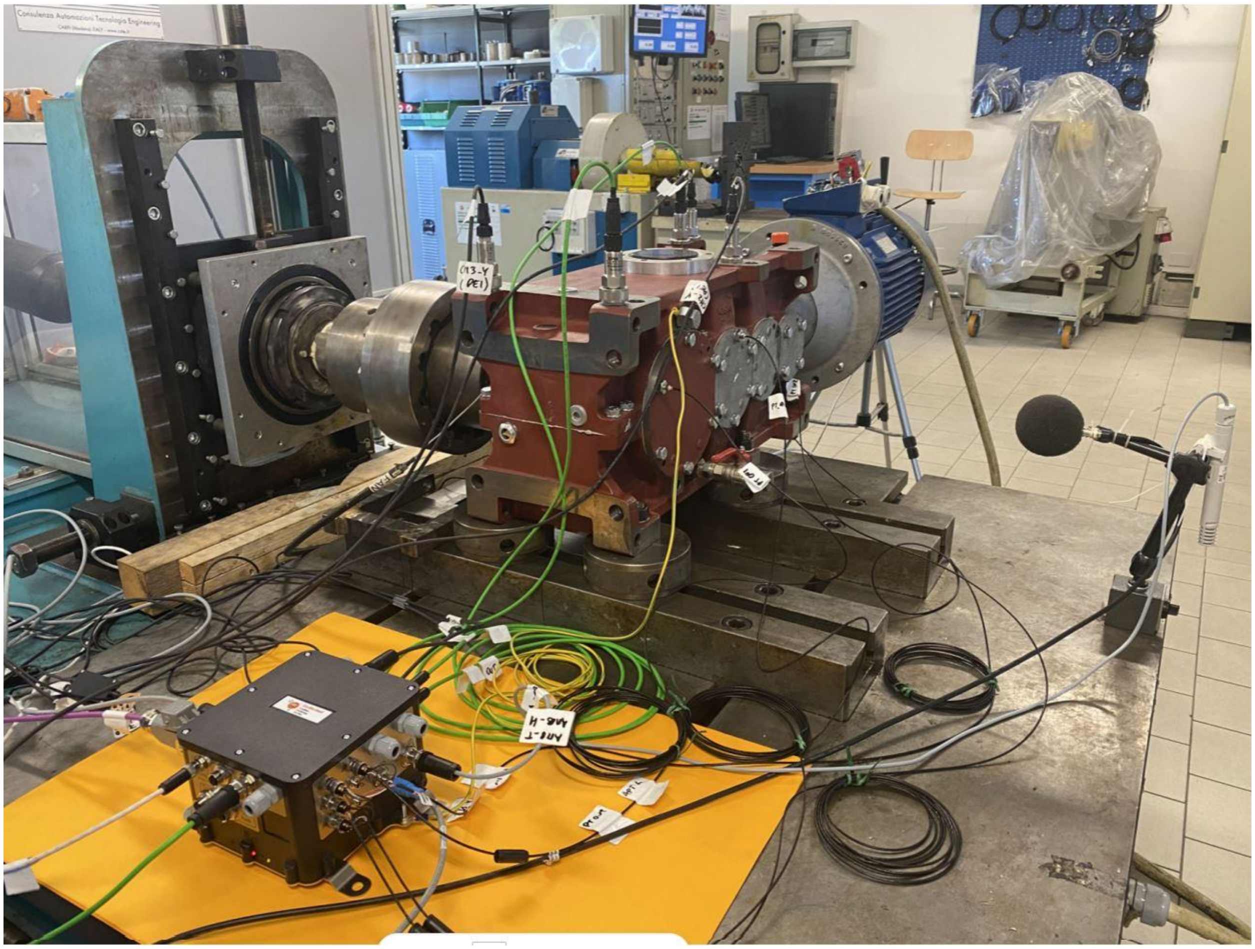


Figure 54 - STM Research and Development test room, Installation of sensors, Test7B.

This test was conducted for a total duration of 2187.5 working hours and was completed without any oil changes, providing a continuous dataset for tracking the natural degradation of both the lubricant and the mechanical components. Throughout this duration, the gearbox was subjected to the standard Nominal Working Cycle, as detailed in Section 6.1. This cycle consists of eight distinct working points (four clockwise and four counterclockwise), with a 150-second high-frequency acquisition recorded every 30 minutes of continuous operation.

The specific configuration and placement of the condition monitoring sensors on the Device Under Test for this test are illustrated in Figure 55.

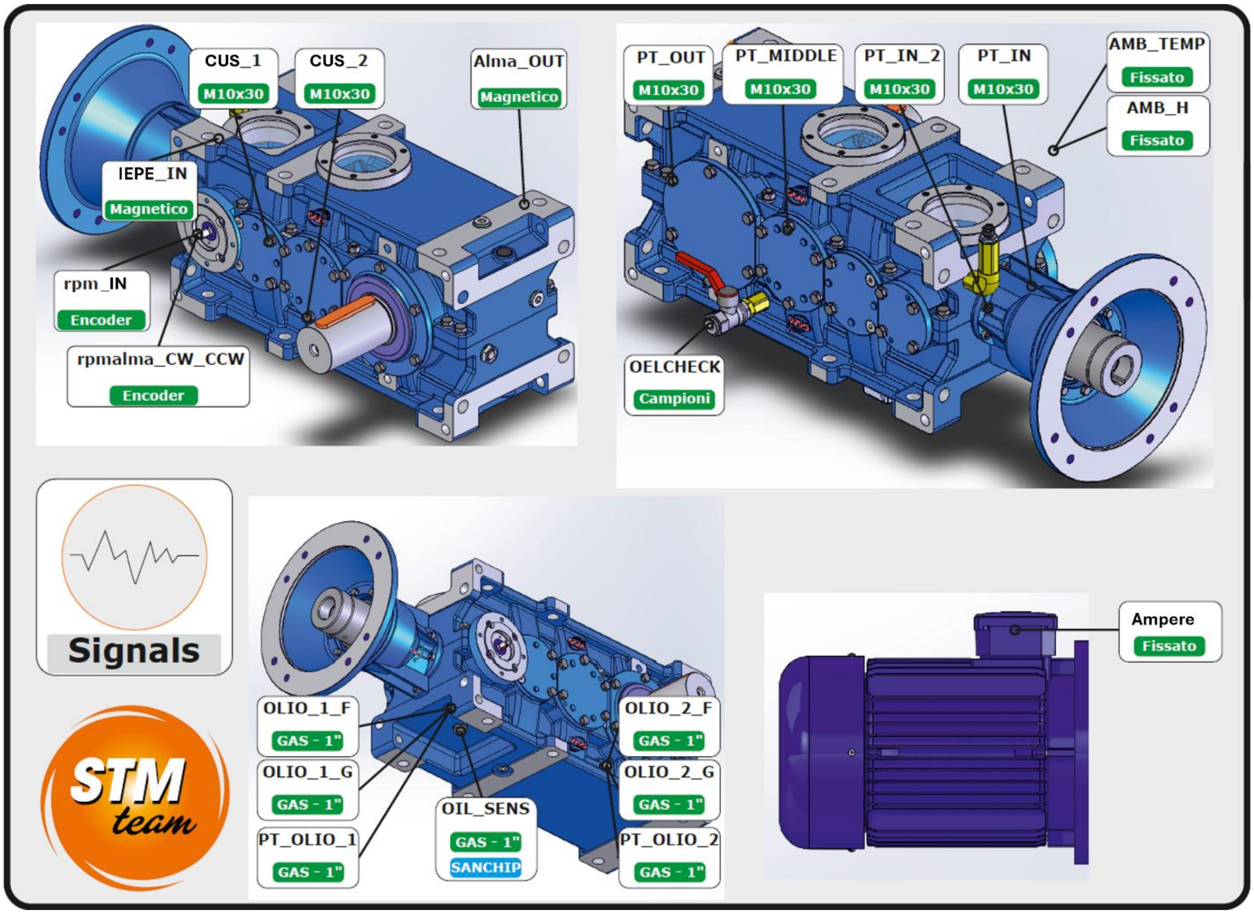


Figure 55 - Position of the sensor in the Device Under Test, Test7B.

The analysis of the Test7B dataset was conducted using an earlier iteration of the signal processing and feature extraction framework, the workflow of which is depicted in Figure 56, is that the data was processed using an earlier iteration of the signal processing and feature extraction framework. For this specific durability test, a key characteristic was that both data acquisition and the subsequent feature extraction were performed in real-time by an onboard Electronic Control Unit (ECU), specifically the Miracle 2 hardware developed by Alma Automotive.

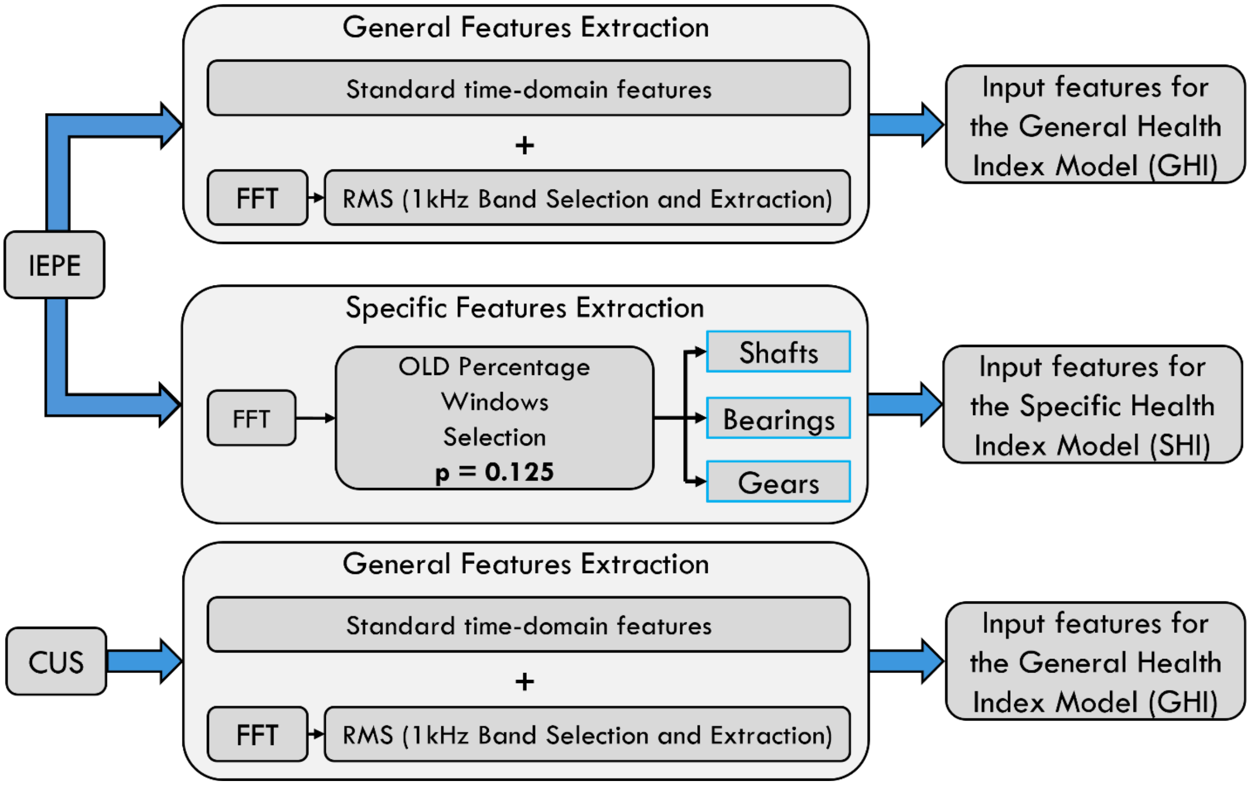


Figure 56 – OLD Signal processing & features extraction workflow, Test7B.

As the workflow diagram illustrates, this earlier methodology differs from the refined framework (previously shown in Figure 52) in several critical aspects. Advanced processing for the ultrasonic sensors, such as envelope filtering and the specific extraction of component-related features, was not yet implemented. Instead, the component-specific features were primarily derived from the IEPE accelerometer signal. Furthermore, the post-processing of this IEPE data also varied slightly, with the RMS calculation for spectral orders employing a wider window (0.125% of the center frequency) compared to the more selective window (0.0125%) used in the current strategy.

An example of a raw feature trend generated by this process is presented in Figure 57. The plot illustrates the unfiltered trend of the second-order sideband of the AX3 shaft frequency around the fourth-stage gear mesh frequency (GMF4), extracted from the IEPE_IN sensor.

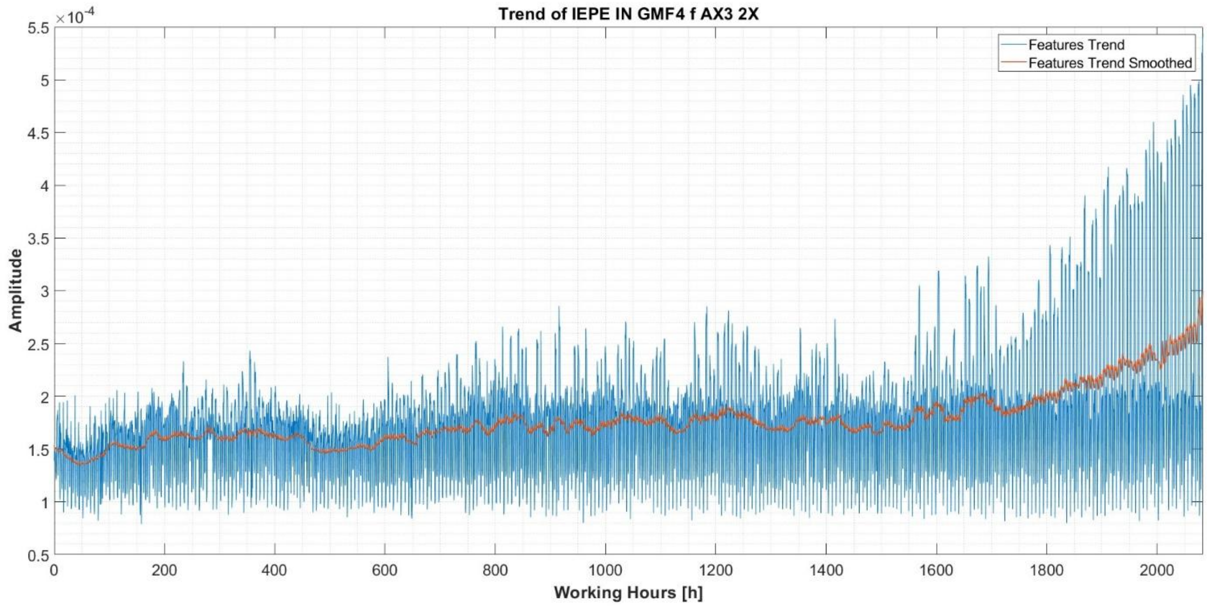


Figure 57 - Trend of the Features GMF4 f AX3 2x of IEPE IN signal, Test7B.

To enhance the interpretability and reveal underlying degradation patterns, a moving average filter was applied to the features, with the data grouped by individual working points. This visualization technique, shown in Figure 58, allows for direct assessment of a feature's behavior under specific load, speed, and rotational direction conditions. This approach makes it possible to understand under which specific operating conditions a potential vibration anomaly is most prominent. The figure also highlights three distinct lifecycle phases, identified for subsequent analysis and model training: the running-in phase (yellow), the stable operational phase (green), and the final wear-out phase (red). It is possible to note that Working Point number 5 described by (19.33 [A] and -790.8 [rpm]) is the one that stands out from all the others in terms of growth trend at the end of the test.

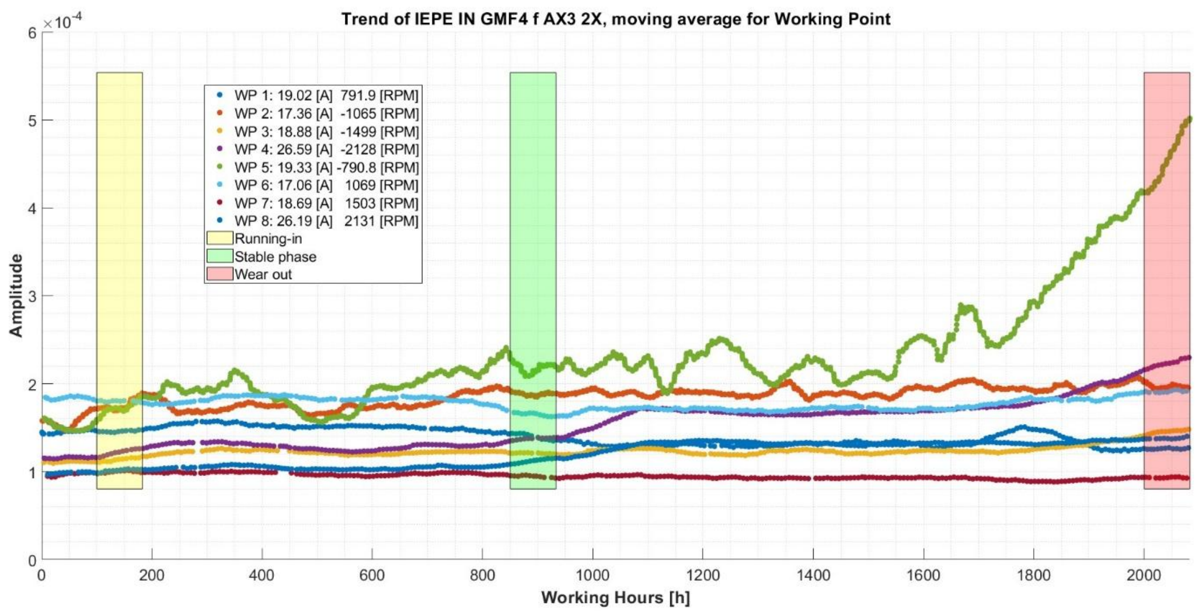


Figure 58 - Trend of the Features GMF4 f AX3 2x of IEPE IN signal, moving average for Working Point, Test7B.

The final diagnostic output of this earlier framework is a component-level Specific Health Index (SHI), which was generated using a One-Class Support Vector Machine (OC-SVM) as the anomaly detection model. Figure 59 presents the resulting SHI values for 27 distinct component features derived from the IEPE_IN signal. The results are visualized as a bar plot, comparing the anomaly scores across the three identified lifecycle phases (running-in, stable, and wear-out). It is important to note another key difference from the current methodology: in this earlier framework, the SHI for gear-related faults was disaggregated, with separate indices for the gear mesh frequency (GMF) and its sidebands, unlike the consolidated gear index used in the refined strategy. The anomaly score bins in which the wear-out phase in relation to the stable phase marks a higher value are those of the fourth stage and partially also of the third stage and meshing.

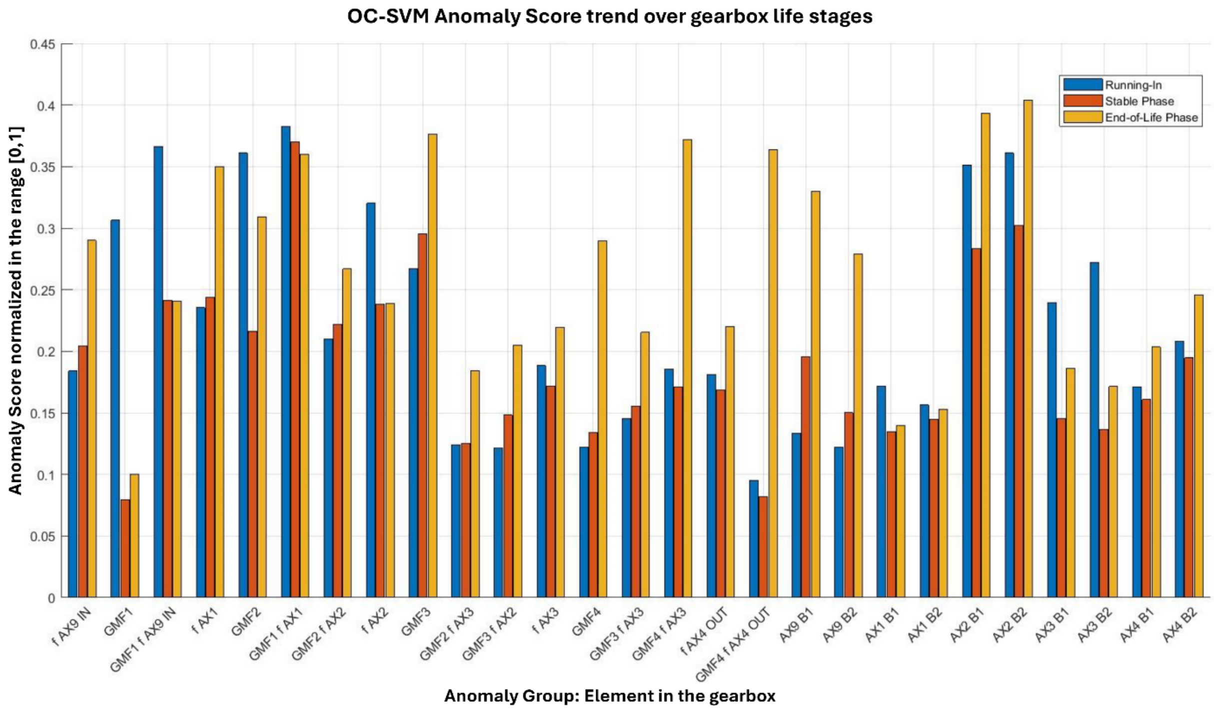


Figure 59 – SHI: OC-SVM Anomaly Score trend over gearbox life stages, Test7B.

To provide a more granular, frequency-domain perspective on the detected anomalies, the analysis also examined the trend of the OC-SVM anomaly score when applied to multiple, distinct frequency bands extracted from the IEPE signal. Figure 60 illustrates the evolution of these scores across the three identified life stages of the gearbox. This type of visualization helps to identify which specific parts of the frequency spectrum are contributing most to the overall degradation trend, offering further insight into the nature of the wear mechanisms present during the test.

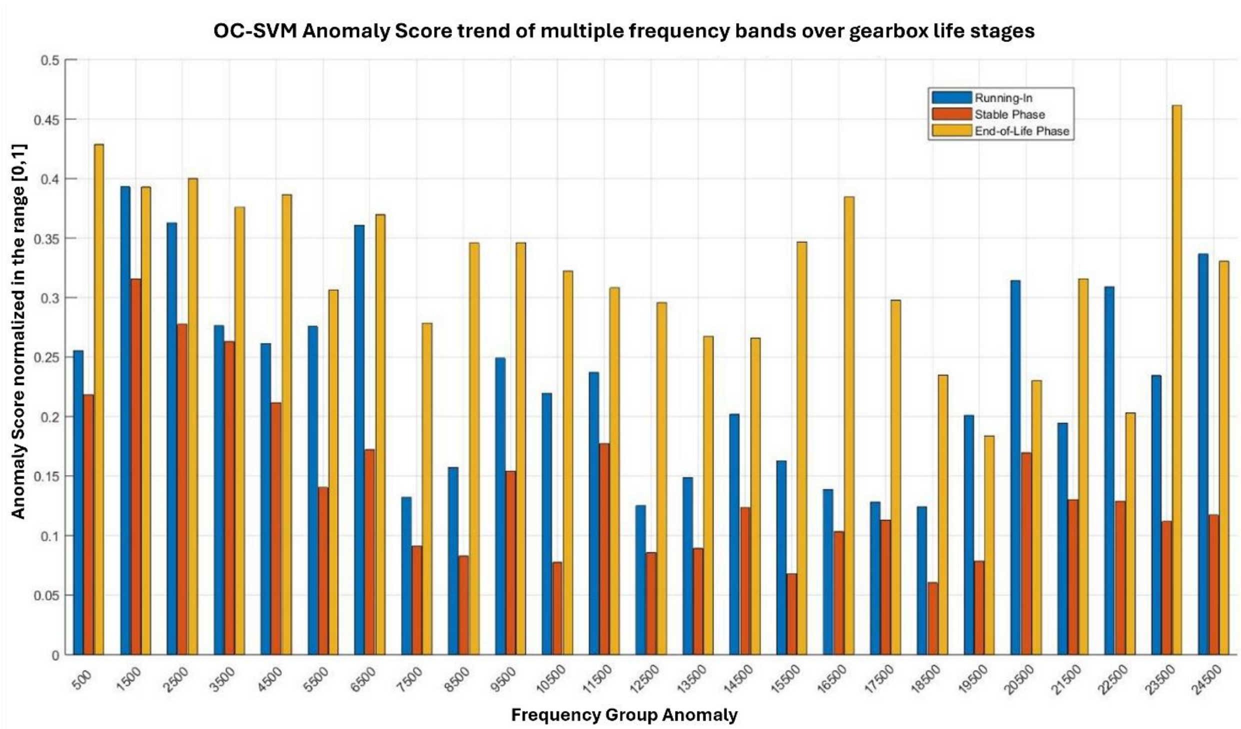


Figure 60 - OC-SVM Anomaly Score trend of multiple frequency bands over gearbox life stages, Test7B.

To quantitatively validate the distinct characteristics of the identified lifecycle phases, a supervised learning approach was also employed. The time windows previously identified as 'Running-in', 'Stable Phase', and 'Wear out' were used to create a labeled subset, representing approximately 20% of the total dataset. A decision tree classifier was then trained on this supervised data. To ensure a robust and generalizable result, the model's training was performed using a Stratified K-Fold cross-validation strategy.

Following the training phase, the validated classifier was applied to the entire dataset to predict the corresponding lifecycle stage for every acquisition. The results of this classification are summarized in Figure 61. The figure utilizes horizontal boxplots to visualize the distribution of the model's predictions over the course of the test. Each boxplot quantifies the statistical spread (median, quartiles, and range) of the predicted classes as a function of working hours, providing a clear visual representation of how the machine's state, as interpreted by the model, transitions through the different phases of its life.

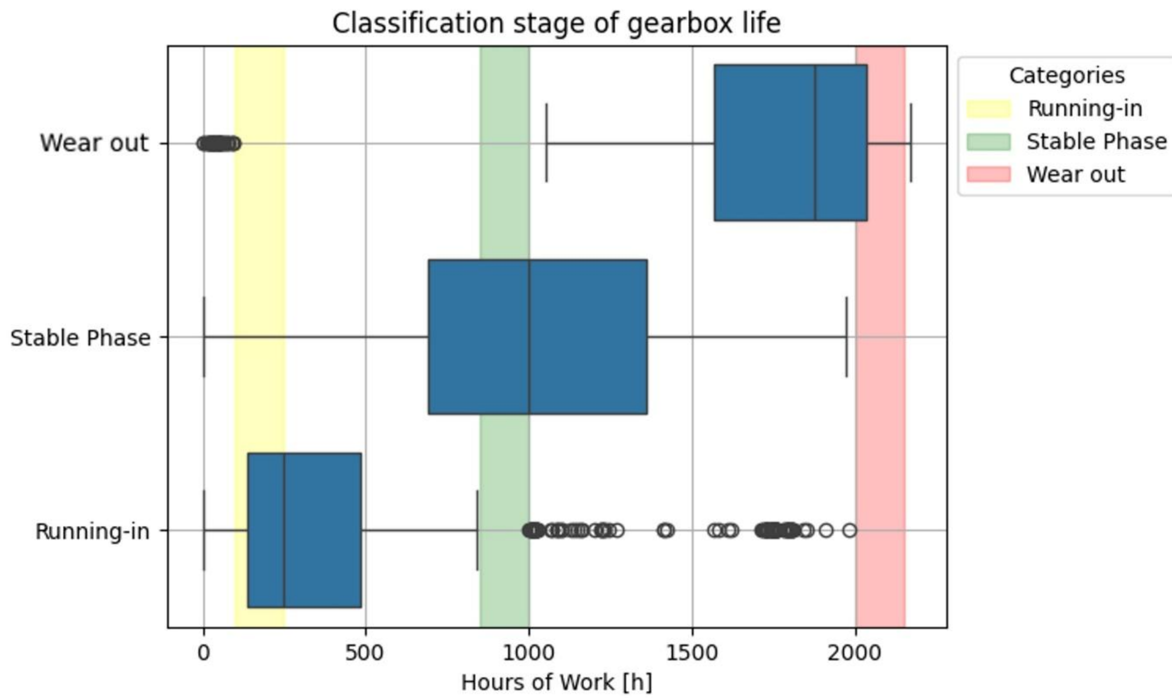


Figure 61 - Classification of the life stage of the gearbox, Test7B.

The quantitative output from the model provides a clear breakdown of the total duration assigned to each phase:

- **Running-in:** 435.5 [h]
- **Stable Phase**: 1191.0 [h]
- **Wear out**: 546.5 [h]

This classification is highly coherent with the expected lifecycle behavior, mirroring the structure of the theoretical bathtub curve. The 'Stable Phase' correctly constitutes the largest portion of the gearbox's operational life, analogous to the "useful life" period of the bathtub curve. This result not only validates the distinctness of the identified operational phases but also reinforces the connection between the vibration-based features and the underlying tribological wear patterns of the machine.

To further explore the distribution of the classification results, Figure 62 presents the same data using a violin plot. This visualization combines a traditional boxplot with a kernel density plot (the 'violin' shape), which illustrates the probability density of the data at different values. The inner boxplot provides the standard quartile information, while the outer shape reveals the full distribution. This dual representation offers a more nuanced view than a simple boxplot, clearly showing the concentration of predictions for each life stage at different points in the gearbox's operational history.

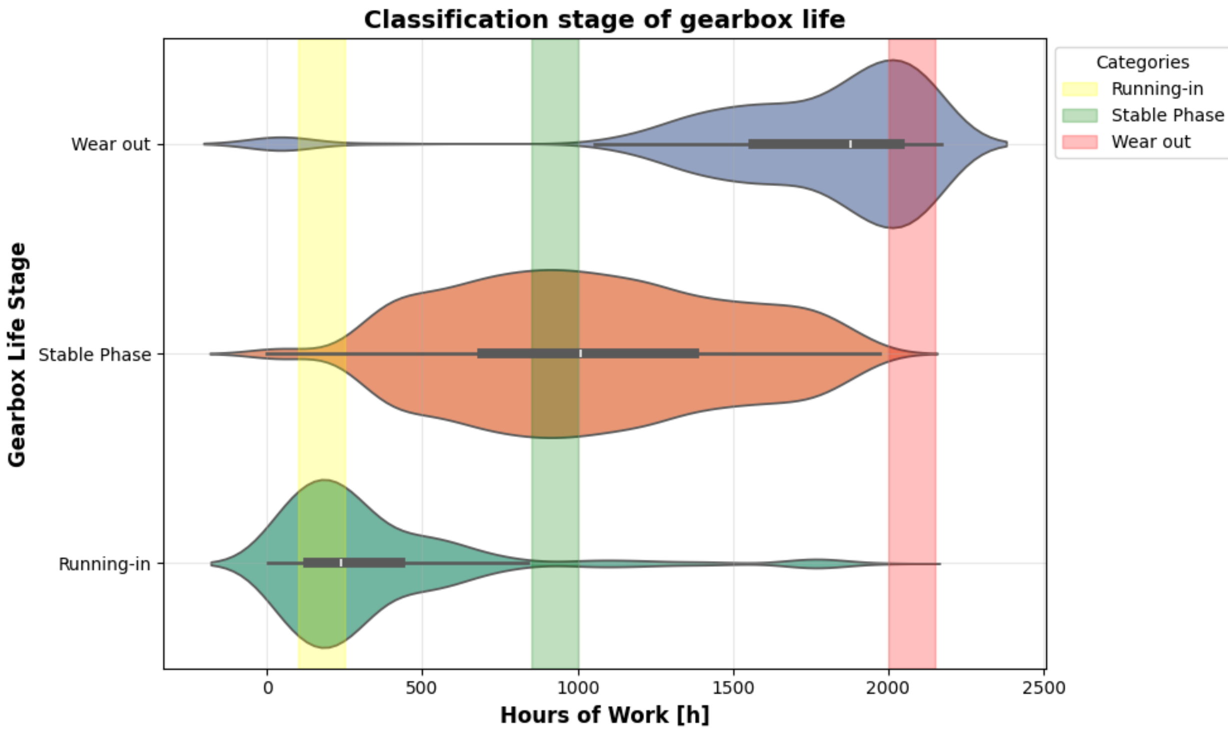


Figure 62 - Classification stage of gearbox life using violin plot, Test7B.

An alternative, highly granular visualization is provided in Figure 63, which displays the results as a swarm plot. This type of plot positions each individual data point along the categorical axis, with points adjusted horizontally to avoid overlap. The primary advantage of this visualization is that it shows the distribution of every single prediction without losing the granularity of the individual data points, offering a clear view of the density and spread of the classifications over time.

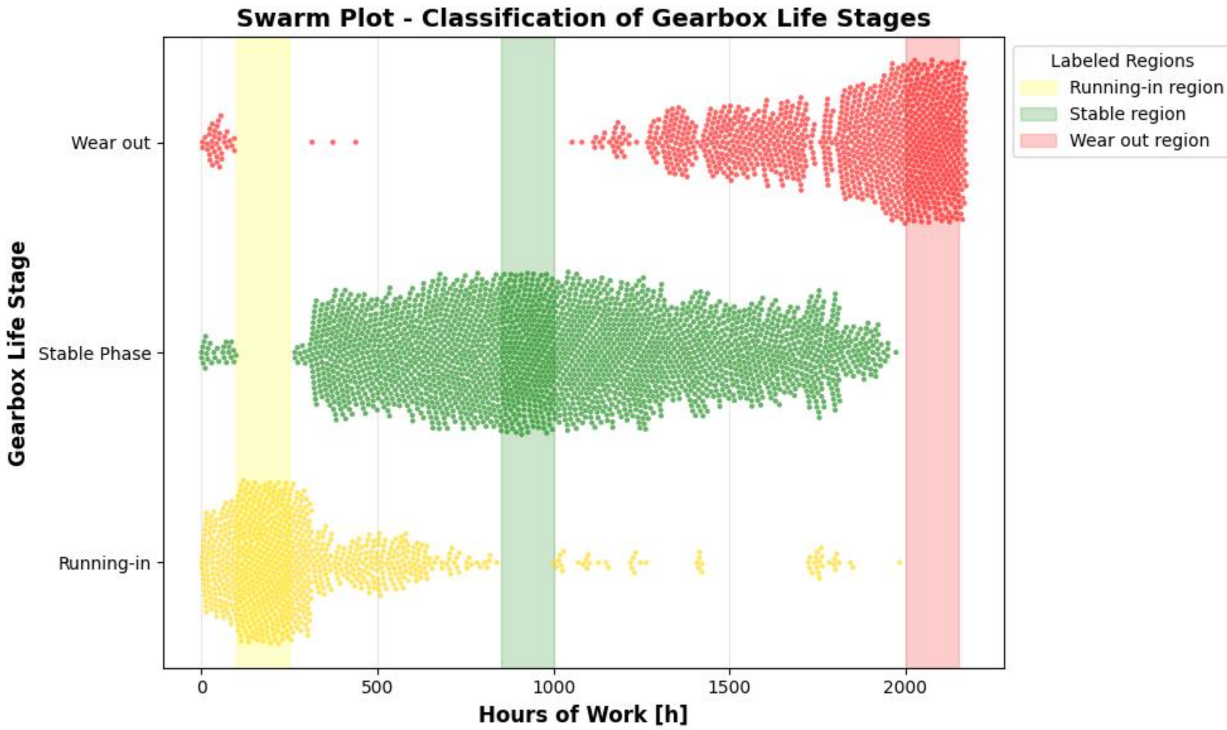


Figure 63 - Classification stage of gearbox life using swarm plot, Test7B.

Finally, to analyze the temporal evolution and total duration of each life stage, Figure 64 presents the cumulative duration of the classifications over the course of the gearbox's operation. This plot tracks the total accumulated time that the model has assigned to each of the three stages ('Running-in', 'Stable Phase', 'Wear out') as the test progresses. By visualizing the cumulative time, this representation provides a clear overview of the total life expectancy spent in each stage, reinforcing the finding that the 'Stable Phase' constitutes the majority of the gearbox's operational life.

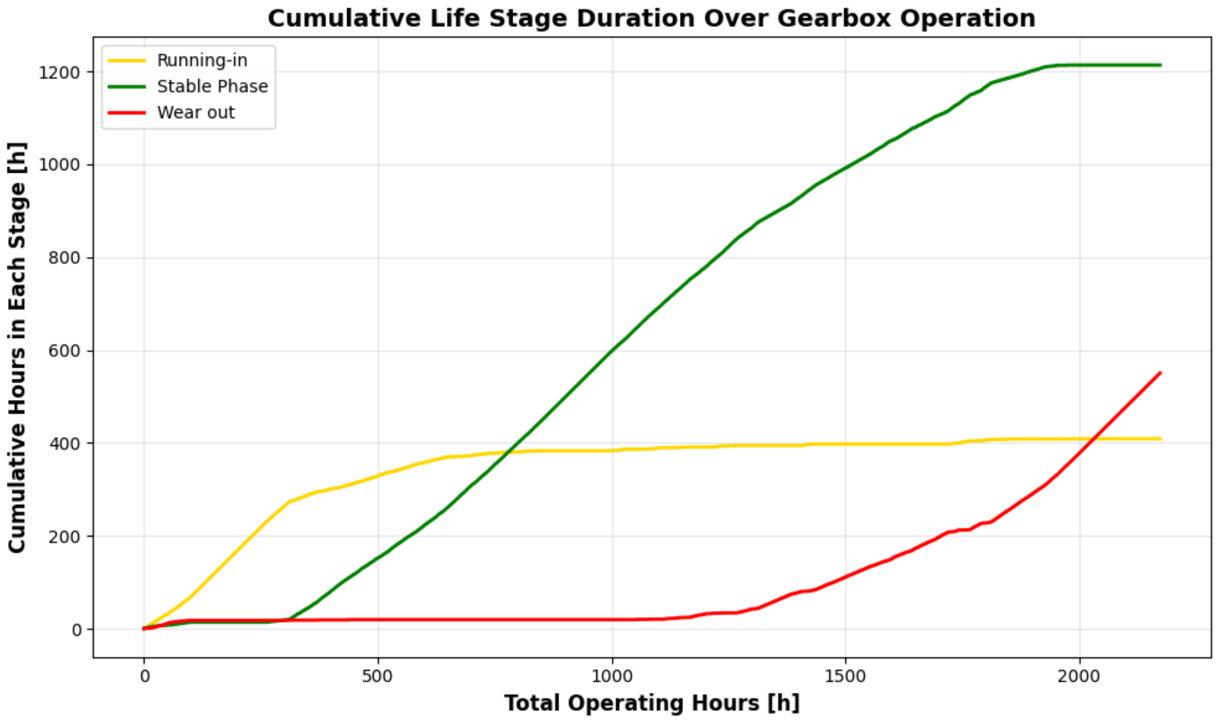


Figure 64 - Cumulative Life Stage Classification Duration Over Gearbox Operation, Test7B.

While this supervised classification successfully validates the distinctness of the identified lifecycle phases, the approach has notable practical limitations that must be considered for broader application. The primary limitation is its reliance on *a priori* labeled data. The model requires a pre-defined, supervised dataset where segments of the operational history are already manually classified into the three different life stages. This presupposes that the health evolution of a unit is already known, making the approach more suitable for retrospective analysis than for a truly predictive, forward-looking application on a new unit.

Furthermore, the generalizability of a model trained on a single durability test is a significant concern. For the classifier to be effective across a fleet of similar gearboxes, the statistical distribution of the input features must remain highly consistent from one unit to another. Any significant variation in manufacturing tolerances, assembly, or operational conditions could alter the feature space, likely leading to a high rate of misclassification when the pre-trained model is applied to a new gearbox. Therefore, without a robust process for model recalibration or domain adaptation for each new unit, this supervised classification method is not readily scalable as a universal diagnostic tool.

6.3.1.1 Discussion

Following the completion of the 2187.5-hour durability test, the gearbox was disassembled for a detailed inspection to identify the root cause of the degradation trends observed in the condition monitoring data. A visual overview of the gearbox's internal state post-disassembly is provided in Figure 65, with a closer examination of the individual meshing stages shown in Figure 66.

Figure 65 - Photo of the gearbox after disassembly at the end of the test, Test7B.

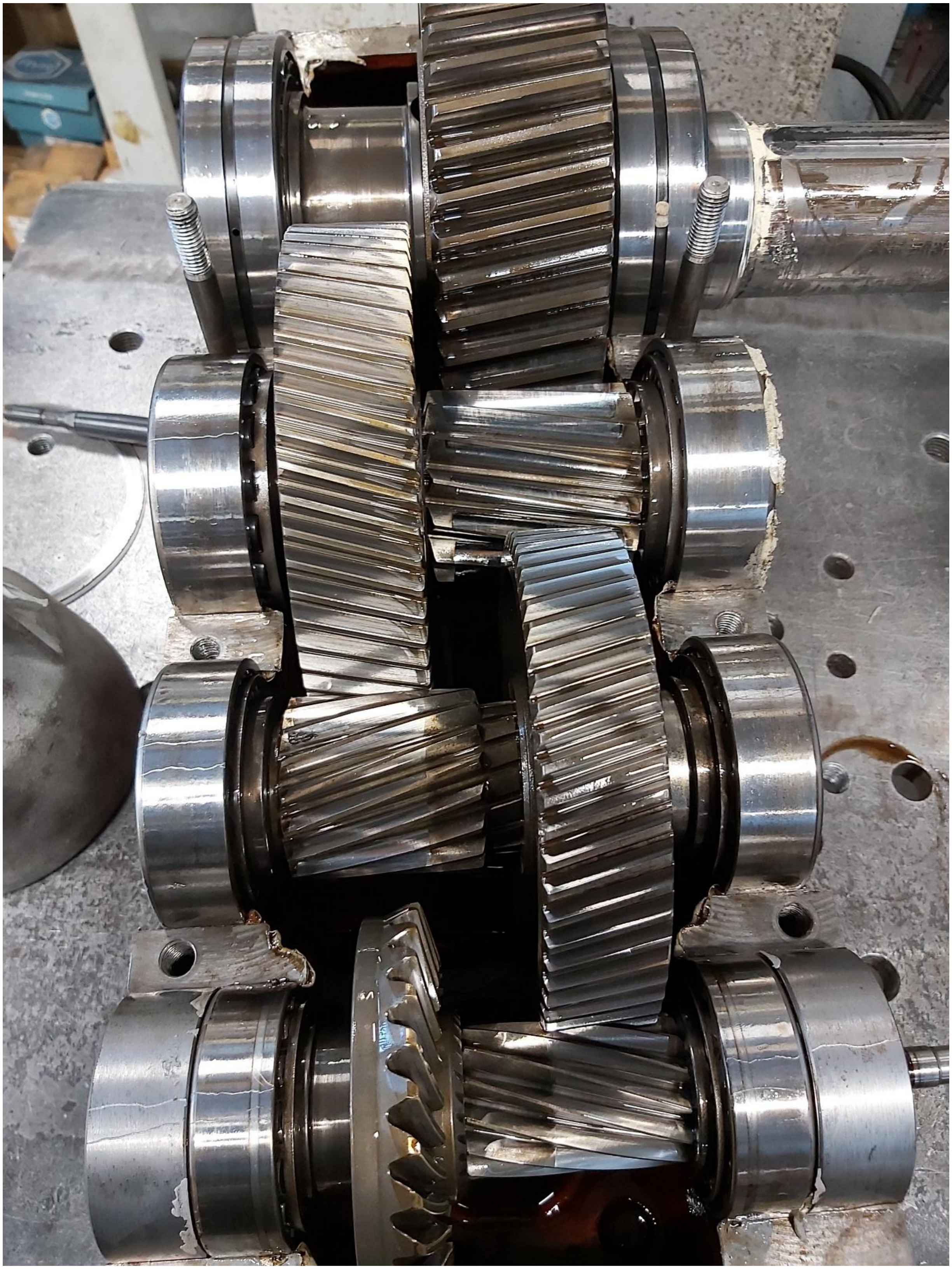


Figure 66 - Close-up view of meshing stages, Test7B.

The inspection revealed the primary source of the wear-out signature: the development of significant pitting on the pinion of the third axis (AX3), which is part of the final, fourth stage of reduction. A detailed photograph of this fault is presented in Figure 67.

Figure 67 - Detail showing pitting on the AX3 pinion, last stage, Test7B.

This physical evidence provides a direct and definitive confirmation for the trends observed in the feature analysis. The upward trend previously noted in the sideband features of the AX3 shaft frequency, specifically around the fourth-stage gear mesh frequency (GMF4) and extracted from the IEPE_IN sensor, was a direct result of the repetitive impacts generated by this pitting.

Furthermore, the inspection revealed that the pitting was concentrated on only one side of the tooth flank. This observation perfectly explains the behavior seen previously in Figure 58, where the anomalous trend was only evident when the gearbox was operating in the counterclockwise direction. In this specific rotational direction, the load was applied to the damaged flank of the tooth, causing the impacts to manifest in the vibration signal. Conversely, in the clockwise direction, the load was applied to the healthy flank, and the fault signature was not present. This confirms the diagnostic power of analyzing features on a per-working-point basis to isolate the specific operational conditions under which a fault is active.

6.3.2 Test7C

Figure 68 shows the sensor configuration on the DUT, including two IEPE accelerometers (IEPE_IN and IEPE_OUT), three ultrasonic sensors (CUS_1–3), several thermocouples (PT_IN, PT_IN_2, PT_MIDDLE, PT_OUT), and the encoder signal (rpm_IN), all acquired as analog inputs. Oil quality and motor current were also recorded via Controller Area Network (CAN) bus. As reported in the figure, while CUS sensors have been placed on bolts holding the carters covering the three intermediate shafts housings, IEPE accelerometers have been positioned on the gearbox case, to monitor the input and output shafts.

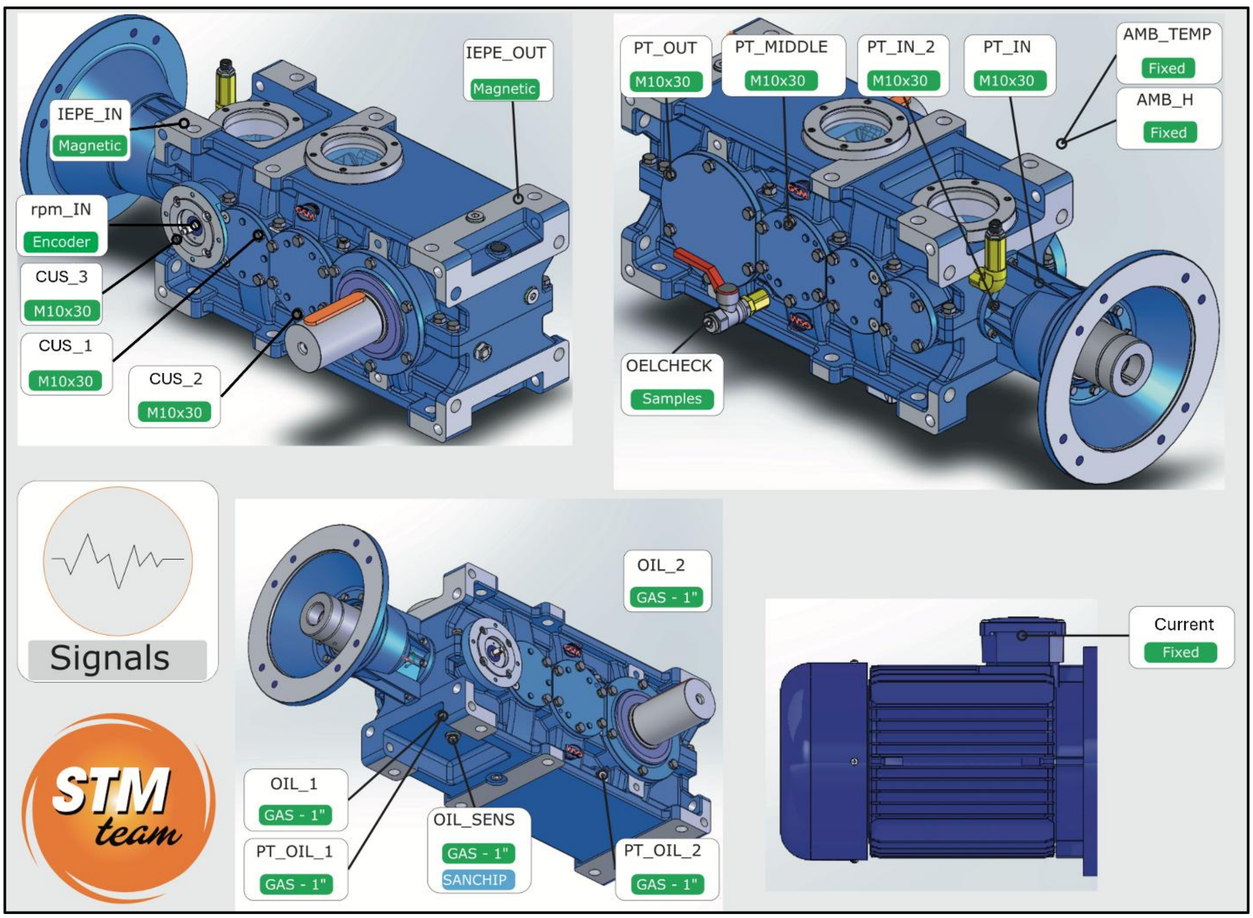


Figure 68 - Position of the sensor in the Device Under Test, Test7C

Figure 69 62 shows the trend of fine ferrous particulate matter over the entire test, highlighting its correlation with key chronological events:

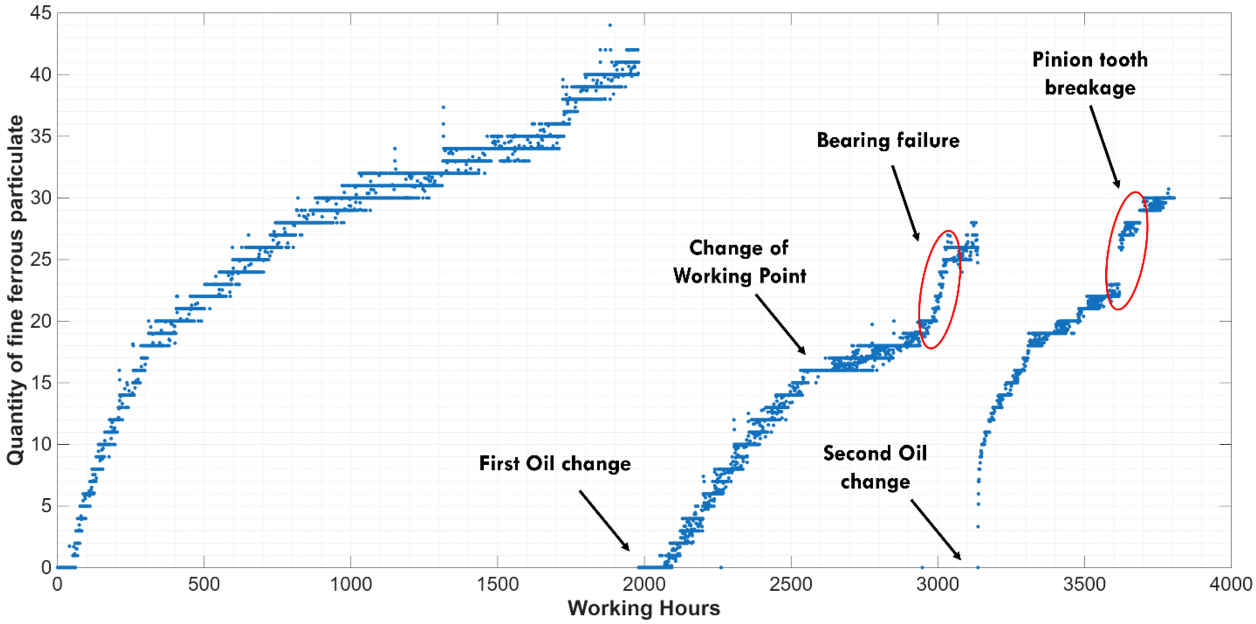


Figure 69 - Ferrous Particle Trend during Test7C

- ≈ 2000 h – First oil change; ferrous particulate level reset to zero.
- 2541 h – Change of working point; torque reduced by 1000 Nm from nominal values; change in slope observed in particulate growth curve.
- ≈ 3000 h – Suspect bearing failure; significant increase in growth rate of ferrous particulate matter.
- 3135.5 h – First test stop due to excessive vibration and acoustic emissions from the DUT. Change of oil and replacement of the faulty bearing.
- 3800 h – Second and last stop of the test.

A first interesting result of the proposed approach consists in assessing the efficacy of the signal post-processing and feature extraction workflow in distinguishing between healthy operational wear and a distinct component fault. Figure 70 provides a direct visual comparison of the processed ultrasonic signal's frequency spectrum from the ultrasonic sensor CUS_2 (located on AX3) under these different conditions, proving the sensitivity to the considered fault.

To establish a robust "healthy" baseline, three spectral logs are presented from Test7B, on the other separate durability test that was completed without any bearing faults occurring on the third axis (AX3) and with the same nominal working condition of Test7C. These logs represent three distinct lifecycle phases:

- The initial running-in period, characterized by elevated broadband noise.
- The stable operational phase, which exhibits the lowest overall noise floor.
- The final wear-out phase, where noise levels increase generally due to natural degradation, but without forming discrete, fault-related harmonics.

In stark contrast, the fourth log presented in the figure is taken from the primary case study, Test7C, at a point where the AX3 bearing fault was actively developing. This spectrum is fundamentally different from the healthy baselines. It is dominated by a series of high-amplitude, distinct peaks at the fundamental rotational frequency of the third axis (AX3) and its subsequent integer harmonics (e.g., 2x, 3x, etc.). This strong harmonic family is a classic signature of a shaft-synchronous, repetitive impact phenomenon, directly pointing to a fault on a component rotating at that speed, such as the AX3 bearing. This comparison demonstrates that the feature extraction methodology is highly effective at isolating the physically meaningful signatures of a component fault from the general noise of a healthy, aging machine.

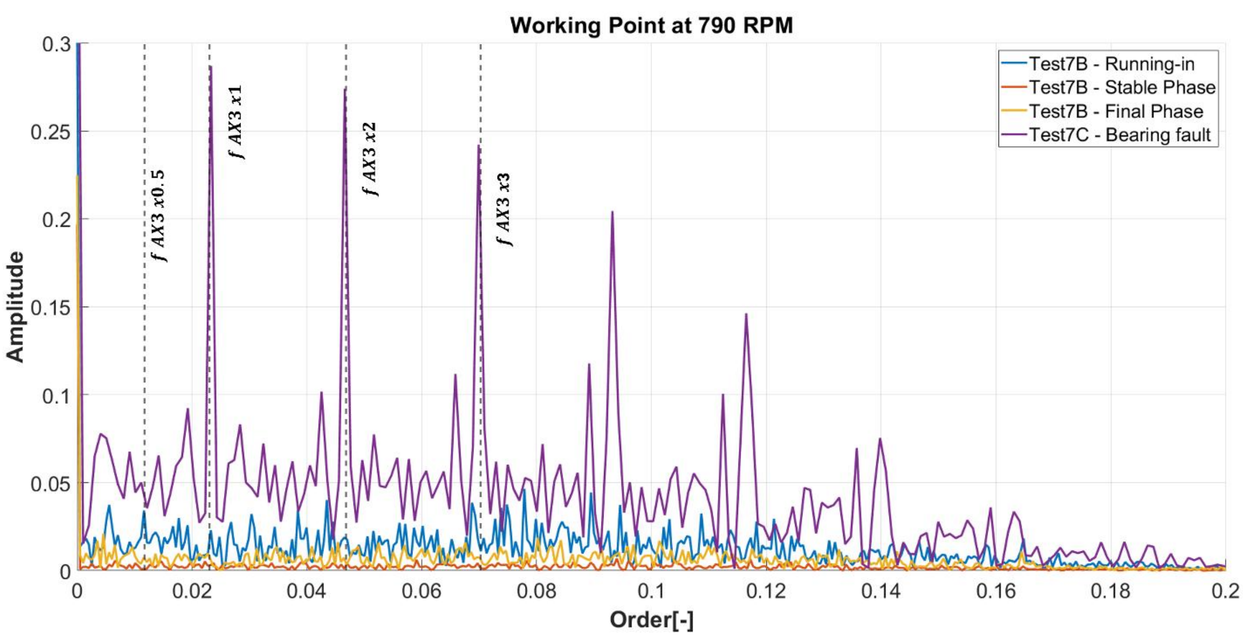


Figure 70 - Different logs that explain different phases of the life of the gearbox. Test7B - Running-in (blue), Test7B - Stable Phase (red), Test7B - Final Phase (yellow) and Test7C - Bearing Fault (purple).

While the same fault signatures were eventually visible in the IEPE accelerometer signals, they appeared with a significant time delay and a lower signal-to-noise ratio compared to the ultrasonic data, underscoring the importance of both sensor selection and appropriate signal processing for effective early fault detection.

The influence of sensor placement was evaluated by comparing signals under fault conditions. As shown in Figure 7164, the most sensitive sensor was CUS 2, installed on the slowest and quietest stage. In contrast, CUS 3, mounted on the first stage, proved very noisy and was unable to highlight the fault on axis 3, even after post-processing. This confirms that sensor positioning plays a critical role in the detectability of localized faults (see Figure 68 for sensor layout).

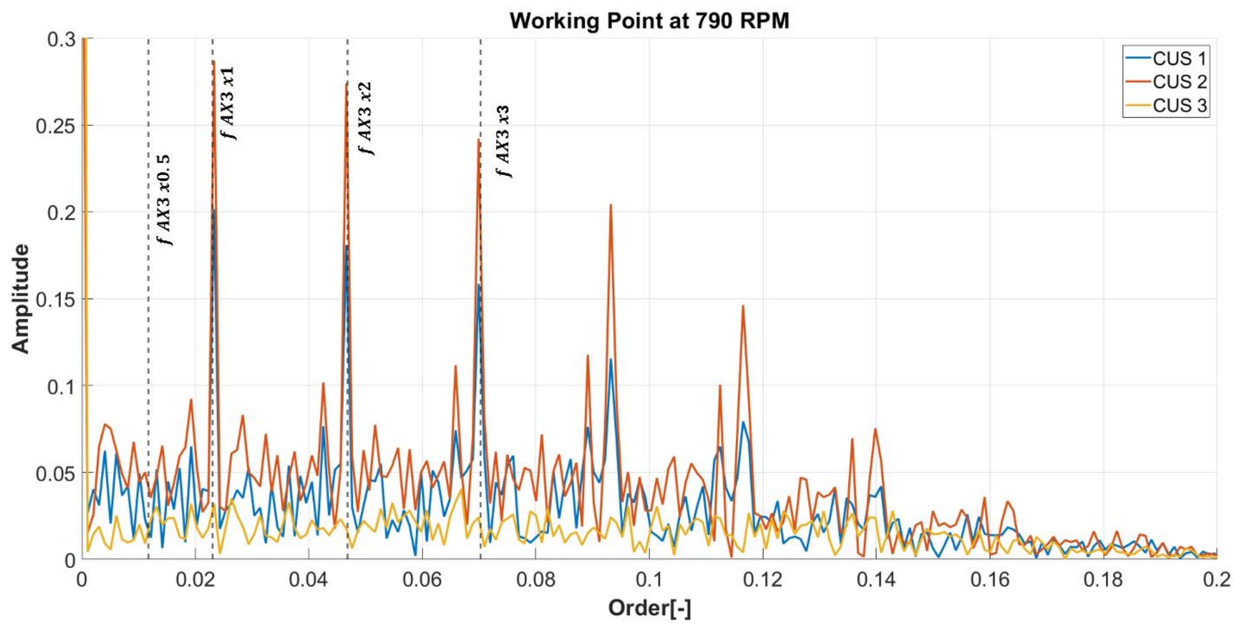


Figure 71 - Comparison between the sensitivity of ultrasonic sensor to mounting position. CUS 1 (blue), CUS 2 (red), CUS 3 (yellow).

Figure 72 illustrates the trend of the Warning Index generated by the Local Outlier Factor (LOF) anomaly detection model. The model was trained on general IEPE features, excluding Crest Factor and Kurtosis, using the shaded green region as the training set. The index crossed **the pre-alarm threshold at ~2430 hours and the alarm threshold at ~2820 hours**, providing substantial lead time before the test was stopped at 3135.5 hours due to excessive vibration and acoustic emissions.

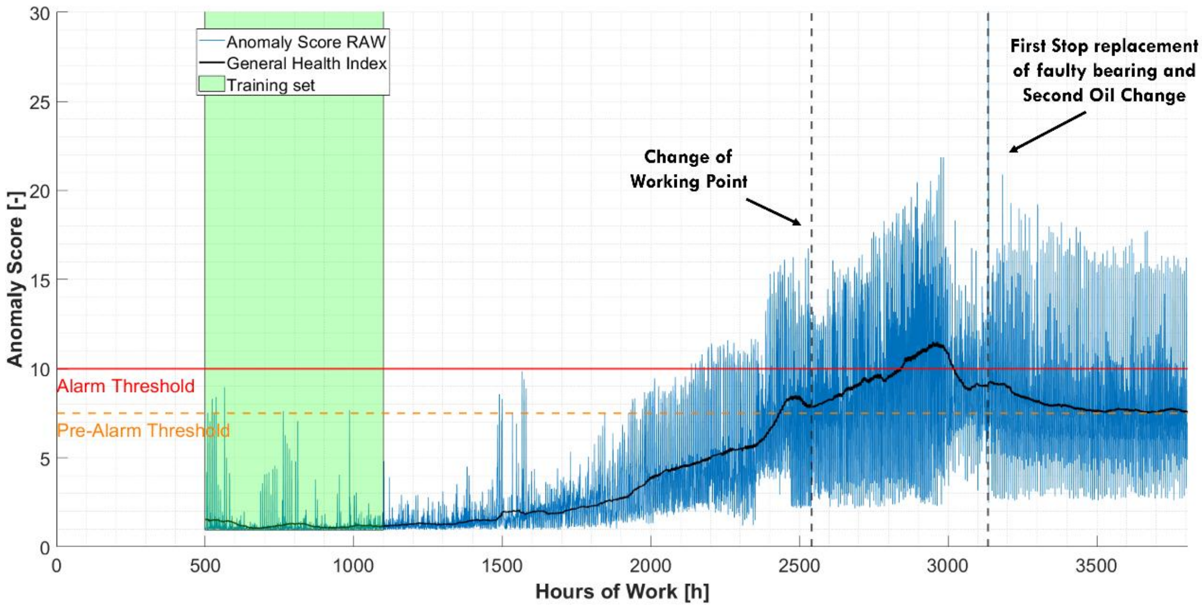


Figure 72 - Trend of the General Health Index (GHI) and definition of Threshold for IEPE Sensor

Figure 7366 reports the comparative trend of the Specific Health Index derived from the CUS ultrasonic sensors for bearings on three different axes (AX1, AX2, and AX3). While a general upward trend is observable across all three indices as the machine ages, indicating a system-wide increase in wear, the index for AX3 provides the earliest and most decisive fault indication. The AX3 index is the first to surpass the **Alarm threshold, doing so at approximately 2380 working hours**, and subsequently deviates from the healthy baseline far more significantly than the others. The index's sensitivity to operational changes is also evident, with a notable drop corresponding to the 1000 Nm torque reduction, followed by a final, sharp rise immediately preceding the confirmed bearing failure at 3135.5 hours when the test was stopped. A similar, though significantly less pronounced, trend was observable in the General Health Index derived from the IEPE sensor, highlighting the superior sensitivity of the ultrasonic-based features for this specific failure mode.

CUS sensors demonstrated markedly higher sensitivity to the transient impacts typical of incipient faults compared to traditional IEPE accelerometers. While this improves early detection capability, it also introduces volatility in the raw anomaly score. To mitigate false alarms from isolated events, the detection criterion was therefore defined not by instantaneous threshold crossings, but by the point at which a moving average of the Specific Health Index consistently exceeds the thresholds as described before.

It is crucial to note that the monitoring strategy and the alarm thresholds shown in the figure are only active after the anomaly detection model has been trained. The training process itself is performed on a set of 'healthy' data collected immediately following the machine's initial running-in period. Therefore, the timeline presented in the figure begins at the start of the active monitoring phase, subsequent to both the running-in and the model training stages.

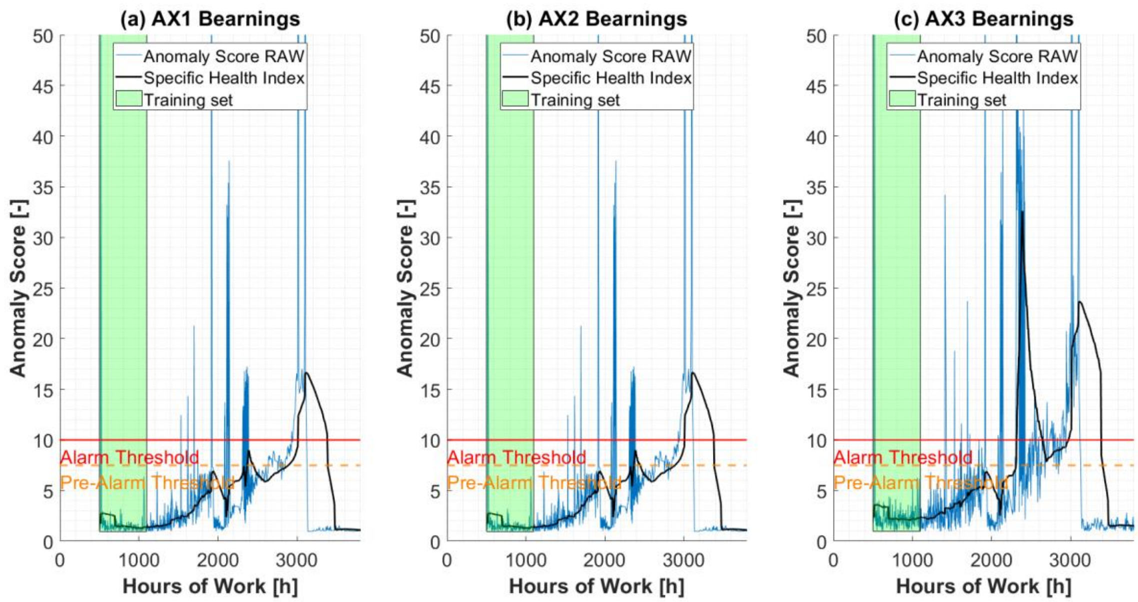


Figure 73 - Specific Health Index for the CUS Ultrasound sensor, focus on bearing on different axis. (a) AX1 bearings, (b) AX2 bearings, (c) AX3 bearings.

6.3.2.1 Discussion

This section assesses the practical effectiveness of the proposed Early Fault Diagnosis (EFD) framework by correlating anomaly detection results with ground-truth events from the Test7C durability test, which ended with a catastrophic bearing failure on the AX3 shaft. The test was stopped at 3135.5 working hours due to excessive vibration and noise. Teardown and inspection confirmed advanced deterioration, with complete cage destruction and displacement of the rolling elements (Figure 14). This catastrophic failure represents the definitive end-of-life point (F-point) for evaluating the lead time provided by the monitoring system.

Figure 74 - AX3 bearing after the first teardown at 3135.5 Working Hours

Analysis of the test timeline alongside model outputs highlights the following events:

- 2315 h – The CUS ultrasound-based Specific Health Index for AX3 first exceeded the second Alarm threshold, providing the earliest definitive indication of fault development.
- 2430 h – The IEPE sensors-based General Health Index exceeded the first threshold
- 2541 h – A sharp increase in noise prompted manual intervention, reducing operational torque by 1000 Nm.
- 2750 h - The CUS ultrasound-based Warning Index for AX3 exceeded the second Alarm threshold for the second time.
- 2840 h - The IEPE ultrasound-based General Health Index exceeded the second Alarm threshold.
- 3135.5 h – The test was halted due to sustained severe vibration and acoustic emissions. The DUT failure is finally diagnosed.

From this timeline, the EFD framework demonstrated clear predictive capability, issuing actionable warnings well before both operational disruption and final failure. In particular, using ultrasound contact accelerometers the system provided:

- ~226 hours of lead time (2541 h – 2315 h) before manual intervention became necessary, reducing the load torque by 1000 Nm
- ~820.5 hours of total lead time (3135.5 h – 2315 h) before catastrophic failure required test termination.

The use of IEPE sensors on the input/output shaft provided fault detection performance aligned with careful manual inspection

- The first alarm threshold was exceeded after 2430 hours, anticipating by 111 hours the change in working point decided by the bench operator due to noise.
- The second threshold was exceeded after 2840 hours, anticipating the actual failure by almost 295 hours

6.3.3 Test7D

The experimental campaign for Test7D was designed with a distinct objective compared to the previous durability-to-failure tests. Its primary purpose was to evaluate and study the behavior of oil aging and to support a specific oil homologation process in following Petronas instruction. To achieve this, the gearbox was subjected to a particular life cycle involving multiple, distinct stages of running-in and loading, as detailed in the overall test summary (Table 4).

Reflecting its specific focus and validating the configuration of the condition monitoring package offered to commercial customers, a more streamlined sensor suite was installed on the DUT. As shown in Figure 7568, this included one IEPE accelerometer and one CUS ultrasonic sensor, strategically placed on the gearbox housing in a location determined to be most sensitive for a general health assessment. This core setup was supplemented by the essential WearDetect oil quality sensor, an encoder for rotational speed, a thermocouple for gearbox temperature, and additional sensors to monitor ambient temperature and humidity.

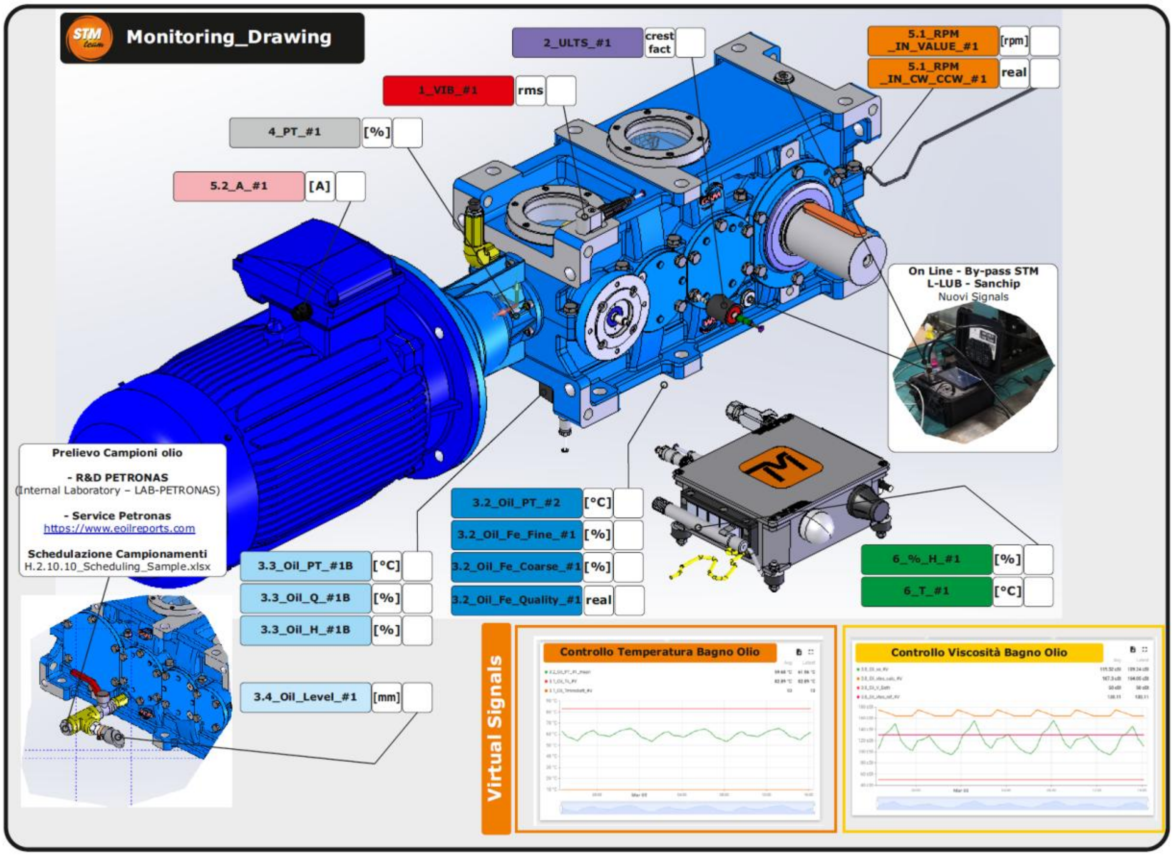


Figure 75 - Position of the sensor in the Device Under Test, Test7D

To achieve the objectives of the oil homologation study, Test7D was subjected to a unique, multi-stage loading profile designed to stress the lubricant under various conditions. As illustrated by the motor current trend in Figure 76, the test cycle included four distinct phases: an initial running-in period without load, a second running-in period at -20% of the nominal cycle load, a period of operation under the standard nominal cycle, and a final overload phase.

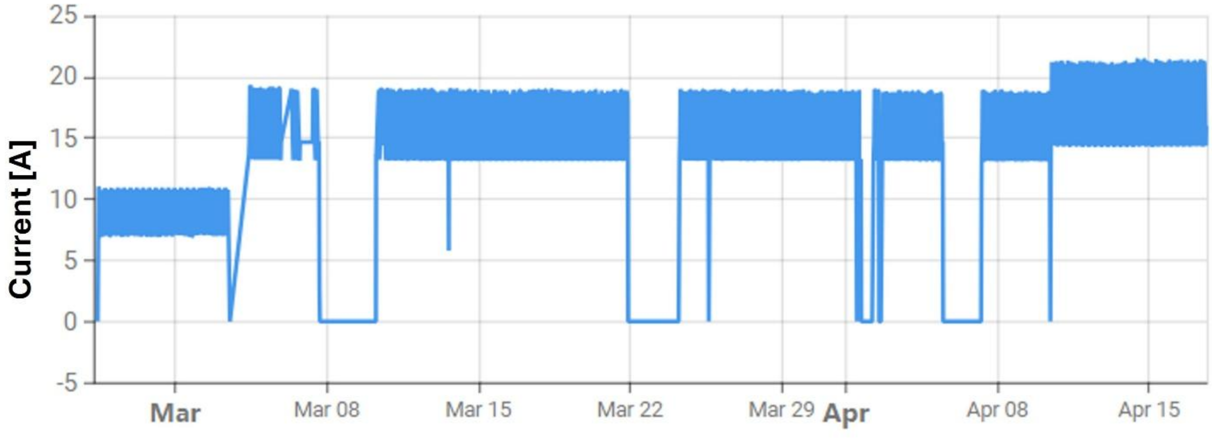


Figure 76 - Level of Current during Working Cycle, Test7D

A key finding from this test was the consistency of the wear debris generation when compared to the previous durability tests, which serves as a validation of the measurement system and the underlying wear characteristics of the gearbox. Figure 77 provides a direct comparison of the ferromagnetic particle trends from tests Test7B, Test7C, and Test7D. As the plot clearly shows, the growth curve of the ferrous particulate matter in Test7D is very similar to the trends observed in the other tests during their stable operational phases. This consistency across different tests and operational cycles reinforces the reliability of the oil debris sensor as a primary indicator of the machine's tribological condition. The figure also highlights some brief period where the test was halted due to a problem with the bench control PLC, an event reflected in the temporary plateau of the wear curve.

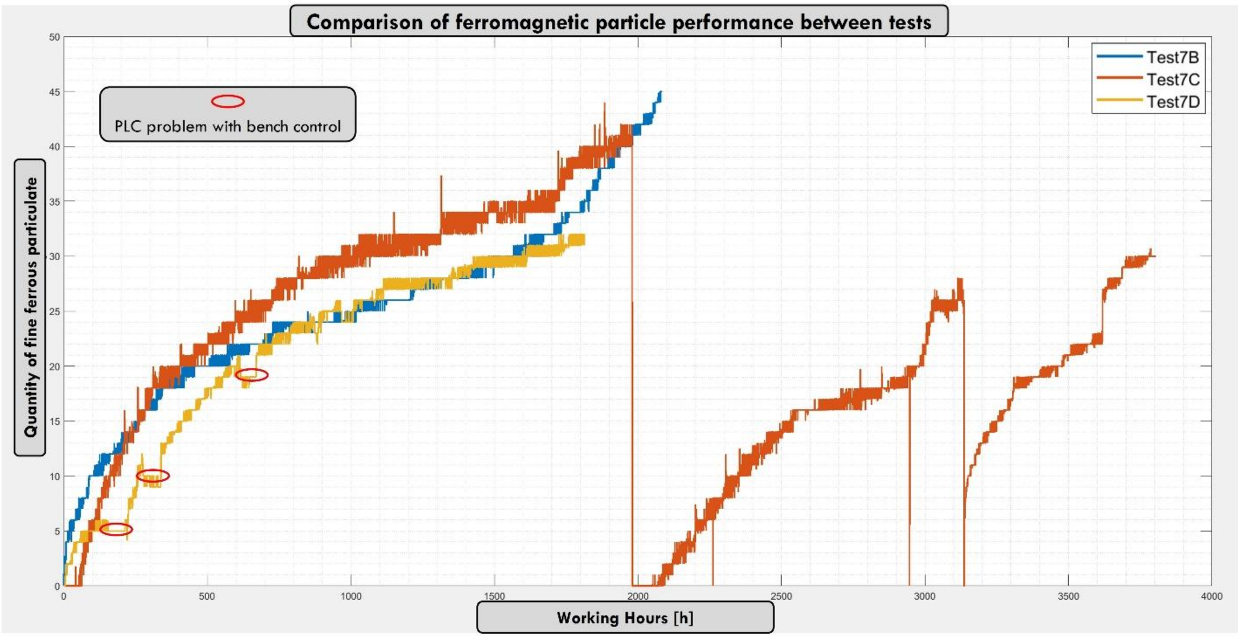


Figure 77 - Comparison of ferromagnetic particle performance between tests.

Beyond the oil homologation study, the continuous monitoring during Test7D revealed a distinct and diagnostically interesting mechanical phenomenon. At approximately 960 working hours, a clear, repetitive impact signature was detected in the ultrasonic sensor data. Figure 78 details this finding from a single log of 150 [s] of acquisition. The upper subplot confirms that the gearbox was operating at a stable working point (approximately -2135 rpm and 26.3 A) during this time. The lower subplot shows the time-domain waveform from the CUS sensor, along with its upper and lower peak envelopes, where the repetitive impact is clearly visible. The period of this impact is approximately 4.7 seconds, which corresponds to a frequency of 0.2128 Hz. When normalized by the input shaft speed, this frequency corresponds to an order of 0.0060, which precisely matches the rotational order of the final output shaft (AX4_OUT), strongly suggesting a fault or an external object touching on this slow-moving component.

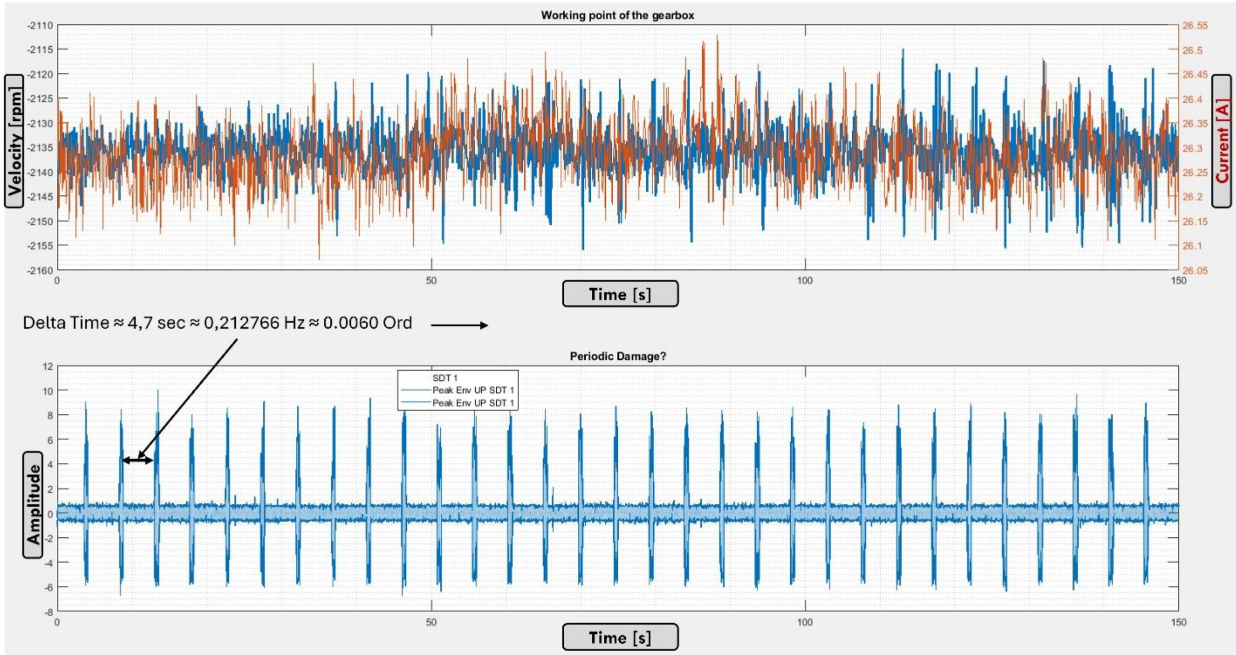


Figure 78 - Impact Phenomenon Visualization, Test7D

The evolution of this impact phenomenon over time is shown in Figure 79. The signature is strongest and most repeatable at 960 working hours. However, in a subsequent log at 1200 working hours, the impact becomes less pronounced and less regular. By 1300 working hours, the distinct impact signature has disappeared entirely, although the overall RMS value and background noise of the signal have increased. As the gearbox from this test has not yet been dismantled for a detailed inspection, the root cause of this phenomenon remains unconfirmed. A leading hypothesis is that a component on the output shaft developed a localized defect that generated the initial impacts. This defect may have then progressively worn away or fractured, causing the distinct impacts to cease while contributing to a general increase in system-wide noise and wear, or as previously mentioned, something external interfered with the output shaft.

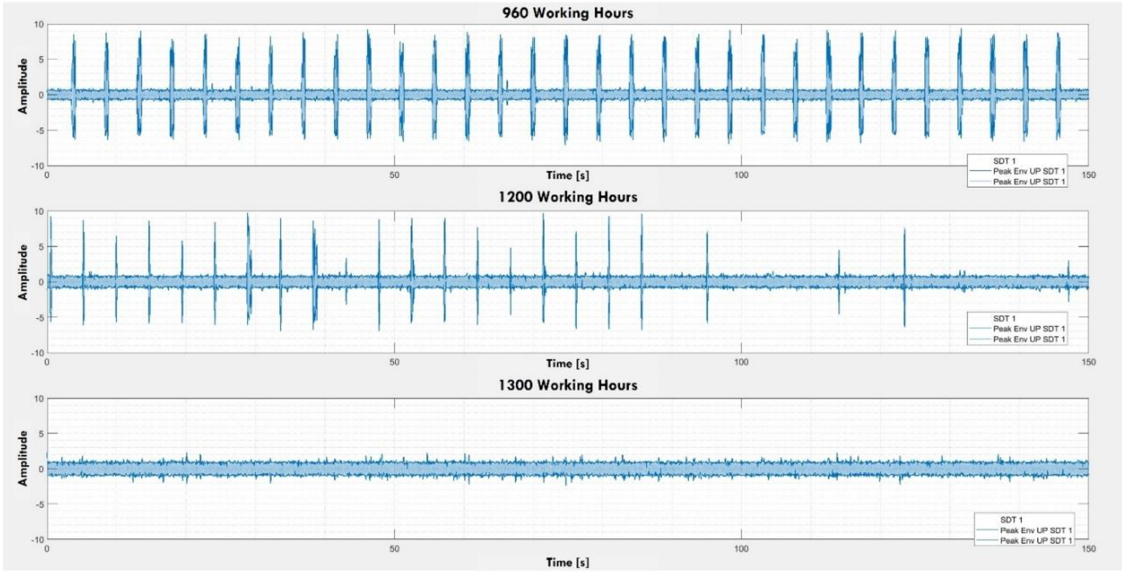


Figure 79 – Evolution of Impact Phenomenon During Time, Test7D

A more detailed analysis of the impact signature reveals that the events are not random but exhibit a highly consistent and repeatable structure. Figure 80 provides a close-up view of three separate impacts captured within the same 150-second acquisition. In addition to the raw ultrasonic signal, the figure also displays the upper and lower Hilbert envelopes, which clearly trace the amplitude modulation of the signal. This analysis shows that each impact event has a consistent duration of approximately 0.5 seconds and a very similar amplitude decay profile.

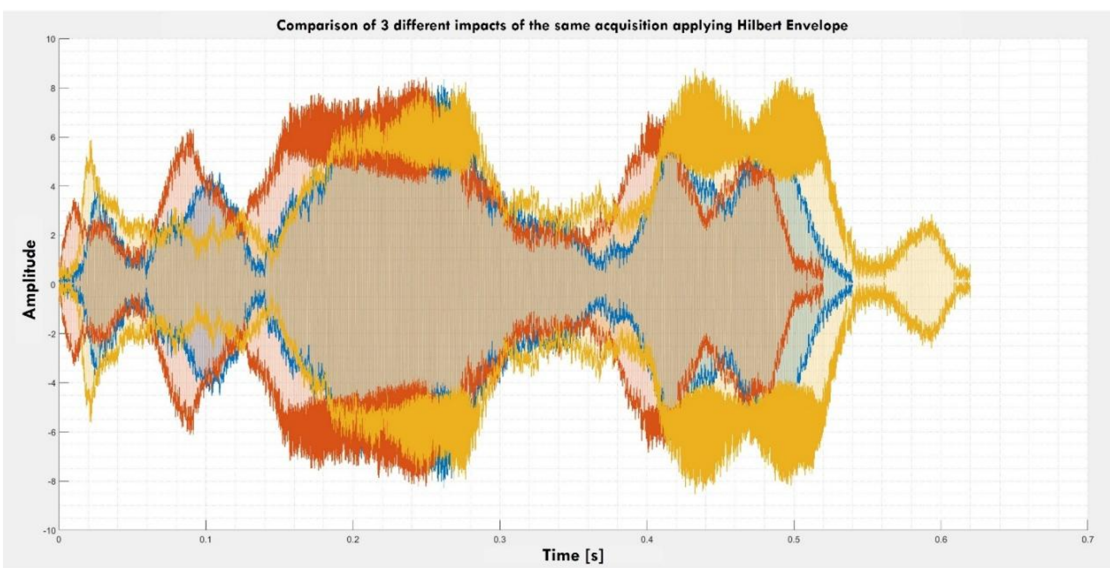


Figure 80 - Comparison of three different impacts of the same acquisition applying Hilbert Envelope, Test7D.

To determine if this phenomenon was detectable by other sensor types, a time-frequency analysis was performed on the signal from the IEPE accelerometer. The spectrogram of the entire 150-second acquisition, shown in Figure 81, reveals that the repetitive impacts are indeed visible in the accelerometer data, concentrated in a high-frequency band.

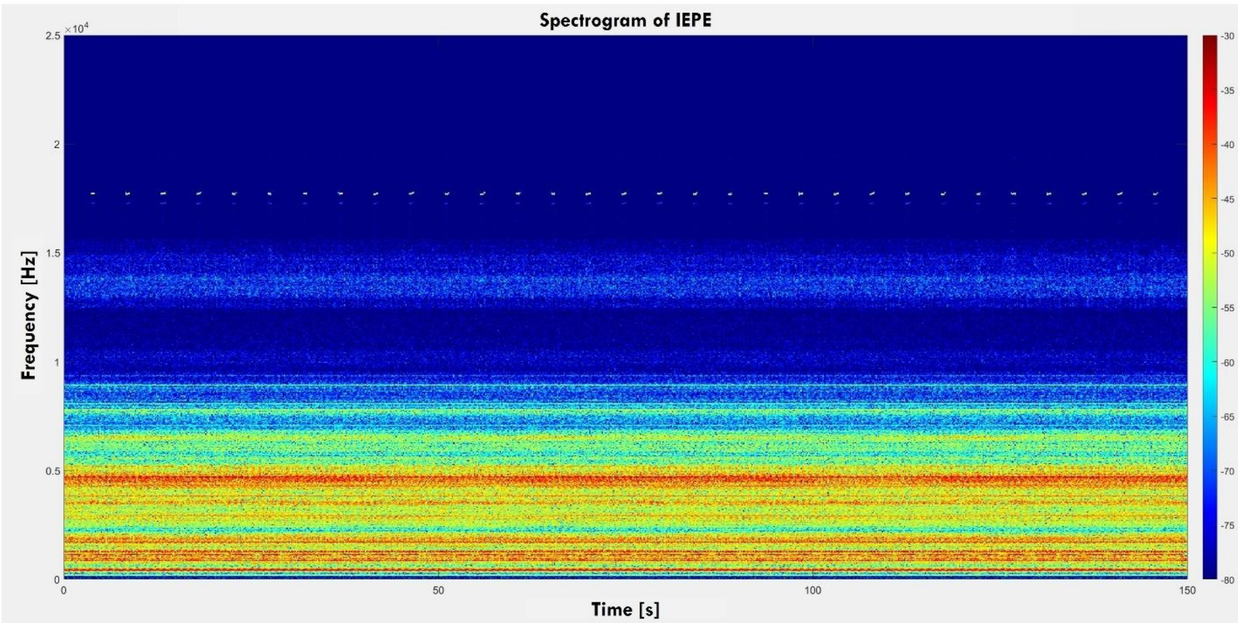


Figure 81 - Impact Phenomenon Test7D Spectrogram of IEPE, specific log at 960 [h].

A zoomed-in view of this spectrogram, presented in Figure 82, focuses on the frequency range between 16 kHz and 18 kHz where the impact energy is most prominent. This detailed view not only confirms the primary carrier frequency of the impacts but also reveals the presence of slight sidebands spaced approximately 400 Hz apart, suggesting a potential modulation effect associated with the fault.

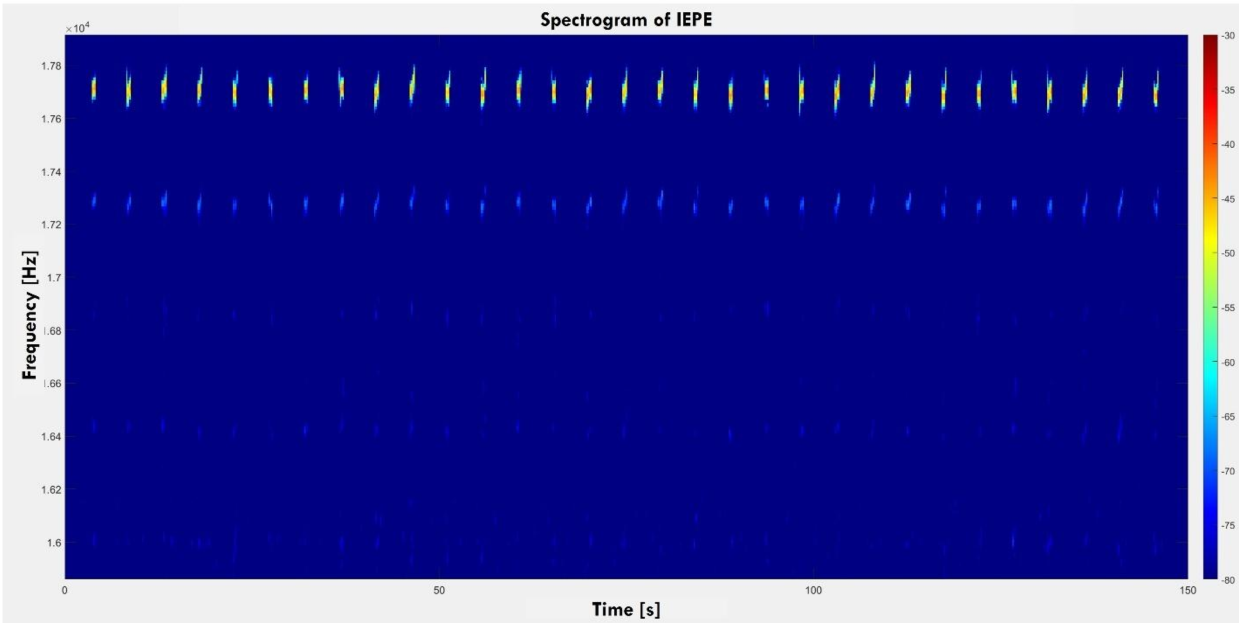


Figure 82 - Impact Phenomenon Test7D Spectrogram of IEPE in the specific windows frequency, specific log at 960 [h].

However, a direct comparison of the time-domain signals highlights the critical difference in sensitivity between the two sensor types for this specific phenomenon. As shown in Figure 83, the amplitude of the impact signature in the raw IEPE signal is approximately one order of magnitude lower than the overall RMS of the signal, making it almost imperceptible without advanced filtering. In stark contrast, the analysis of the ultrasonic signal (previously shown in Figure 78) demonstrates the opposite behavior: the RMS value during an impact is approximately one order of magnitude greater than the signal's baseline RMS. This comparison provides definitive evidence of the superior sensitivity of the ultrasonic sensor for detecting this specific type of low-energy, high-frequency impact event.

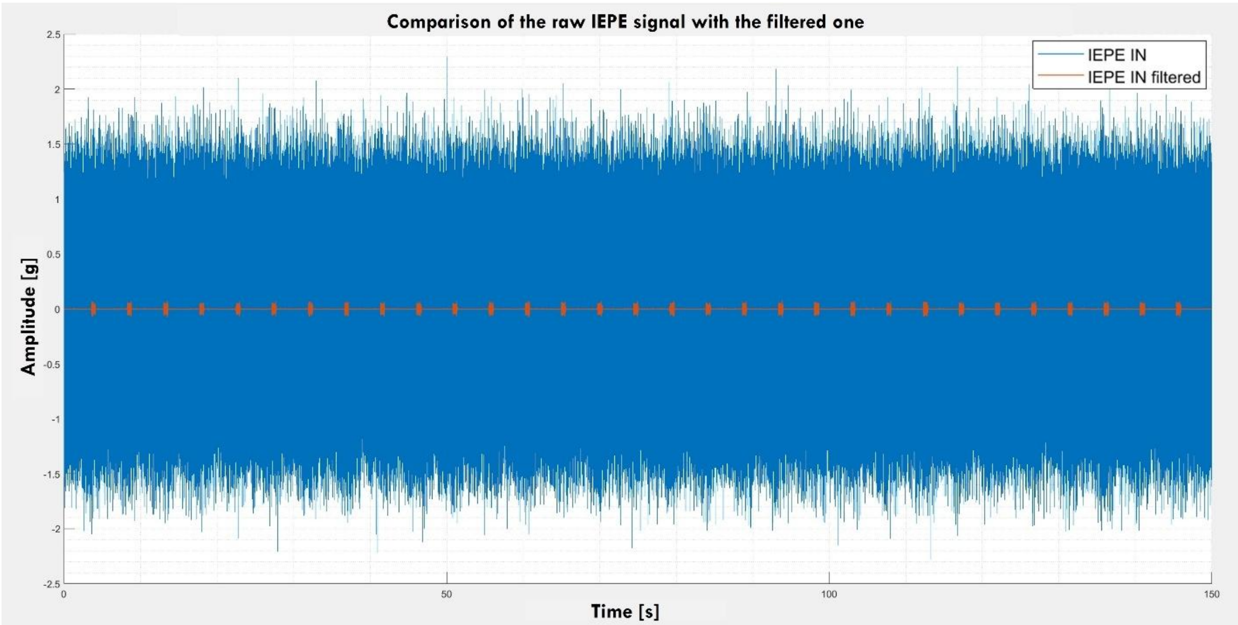


Figure 83 - Impact Phenomenon Test7D, Comparison of the raw IEPE signal with the filtered one, specific log at 960 [h].

6.3.4 Conclusion

All these tests campaign introduces a multi-sensor condition monitoring framework for the early fault diagnosis of industrial gearboxes, validated on a dataset featuring fatal failure (Test7C principally). The methodology leveraged a parallel analysis strategy, employing unsupervised anomaly detection models to generate both a General Health Index from vibration data and component-specific Health Indices from order-based ultrasonic features. The primary findings of this case study are threefold:

1. The supervised definition of a 'healthy' baseline using oil debris analysis was a critical step for ensuring the robustness of the unsupervised anomaly detection models.
2. For the incipient bearing fault investigated, the targeted analysis of high-frequency ultrasonic data provided a significantly earlier and more sensitive detection than the general statistical features from traditional accelerometers.
3. The proposed EFD framework successfully identified the developing bearing fault with a total lead time of over 750 working hours before catastrophic failure, demonstrating its practical value for predictive maintenance.

The main result is the development and validation of a hybrid, multi-stage diagnostic framework. This framework demonstrates a novel methodology where different sensing modalities are assigned distinct, strategic roles: oil analysis is used to supervise the creation of a robust 'healthy' baseline, while high-frequency ultrasonic and vibration sensors are used for real-time, high-sensitivity anomaly detection. The value of this approach is quantified on a full lifecycle durability test, showcasing a practical and effective path to early fault diagnosis.

The findings are, however, subject to certain limitations. This study focused on a single bearing failure mode within a controlled laboratory environment. The robustness and generalizability of the framework should be further validated on a wider range of fault types (e.g., gear tooth pitting, shaft misalignment) and on machinery operating under more variable, real-world field conditions.

Future work could focus on three key areas: (1) extending the methodology to other fault modes and gearbox types; (2) deploying the system on in-field assets to assess its performance against industrial noise and operational variability; and (3) investigating the potential of deep learning models, such as autoencoders, within this multi-index diagnostic framework.

7 Electric Powertrain Case Studies

The transition from internal combustion engines to electric powertrains introduces a new set of components, operational characteristics, and potential failure modes. While the fundamental principles of condition monitoring and predictive maintenance remain the same, their application must be adapted to the specific context of electric motors, power electronics, and high-speed gearboxes. This chapter details the application of the methodologies developed in this thesis to several industrial case studies involving electric vehicle (EV) powertrains.

The development and validation of these systems operate within a stringent regulatory and standardization framework. At the highest level, functional safety for all automotive electrical and electronic systems is governed by ISO 26262, which is an adaptation of the general functional safety standard IEC 61508. While ISO 26262 is primarily concerned with preventing failures that could lead to hazardous events, its principles underscore the importance of reliability and system integrity, which are the core goals of any maintenance strategy [88], [89].

Complementing these safety standards are guidelines focused specifically on the practice of machine monitoring. ISO 17359:2018 provides the general guidelines for establishing a condition monitoring process for machinery, covering aspects from sensor selection and data acquisition to diagnostics and prognostics. The case studies presented in this chapter represent the practical implementation of these principles, tailored to the unique diagnostic challenges of modern electric powertrains [90].

7.1 Brief Survey of CM & PdM Techniques in Electric Vehicles (EVs)

Condition Monitoring (CM) and Predictive Maintenance (PdM) have gained traction in electric vehicles (EVs) due to the high availability of sensor data and the critical need for operational reliability. Commonly used CM techniques in EVs include vibration analysis, current signature analysis (CSA), thermal imaging, and model-based residual analysis [91], [92]. These methods enable early detection of faults in bearings, gearboxes, inverters, and electric motors. Data-driven approaches are often supported by machine learning models that classify system health or estimate Remaining Useful Life (RUL) [93].

For instance, vibration-based monitoring effectively detects mechanical imbalances or bearing degradation, while CSA is valuable for diagnosing stator winding faults or rotor asymmetries. Emerging research emphasizes combining multi-modal sensor data (acoustic, vibration, current, and temperature) to enhance diagnosis accuracy in e-powertrain components [94].

7.2 Test Campaigns & Data Sources

The data and case studies presented in this chapter originate from a six-month industrial internship conducted within the **Department of Powertrain and Power Electronics at Rimac Technology in Zagreb, Croatia**. This collaboration provided access to extensive data from multiple advanced test campaigns on next-generation electric vehicle powertrains. The primary work was conducted at the company's testing facility in **Jankomir**, a principal location for powertrain development and validation.

The findings presented in this chapter are derived from a series of extensive test campaigns conducted on automotive electric powertrains. This section provides a detailed overview of the experimental environment and the nature of the data acquired. The following subchapters will describe the physical test bench configurations, the specific objectives and profiles of the different test types—including accelerated durability, end-of-line, and overload tests—and the key characteristics of the resulting datasets that form the basis for the subsequent analysis.

7.2.1 Test Setup

In this section, the test setup (type of sensor and sampling frequency), the cycles profiles, and the test's scope are described.

Figure 84 - System Testbench 1 (STB1) Jankomir.

Figure 84 shows the configuration of the primary test environment, the System Test Bench 1 (STB1), located at the Jankomir facility. This versatile test bench was used for both the Accelerated Durability Tests and the End-of-Line Tests. It should be noted that while the majority of the data was generated on this bench, some units also underwent partial durability cycles at other company sites, such as the ASAP facility.

The standard layout is shown in Figure 85. The system is configured exactly as it was on BR-1 and Cedar, using only one conditioning unit and no AC measurements. To ensure consistency, each UUT is mounted so that its drive direction faces the wall.

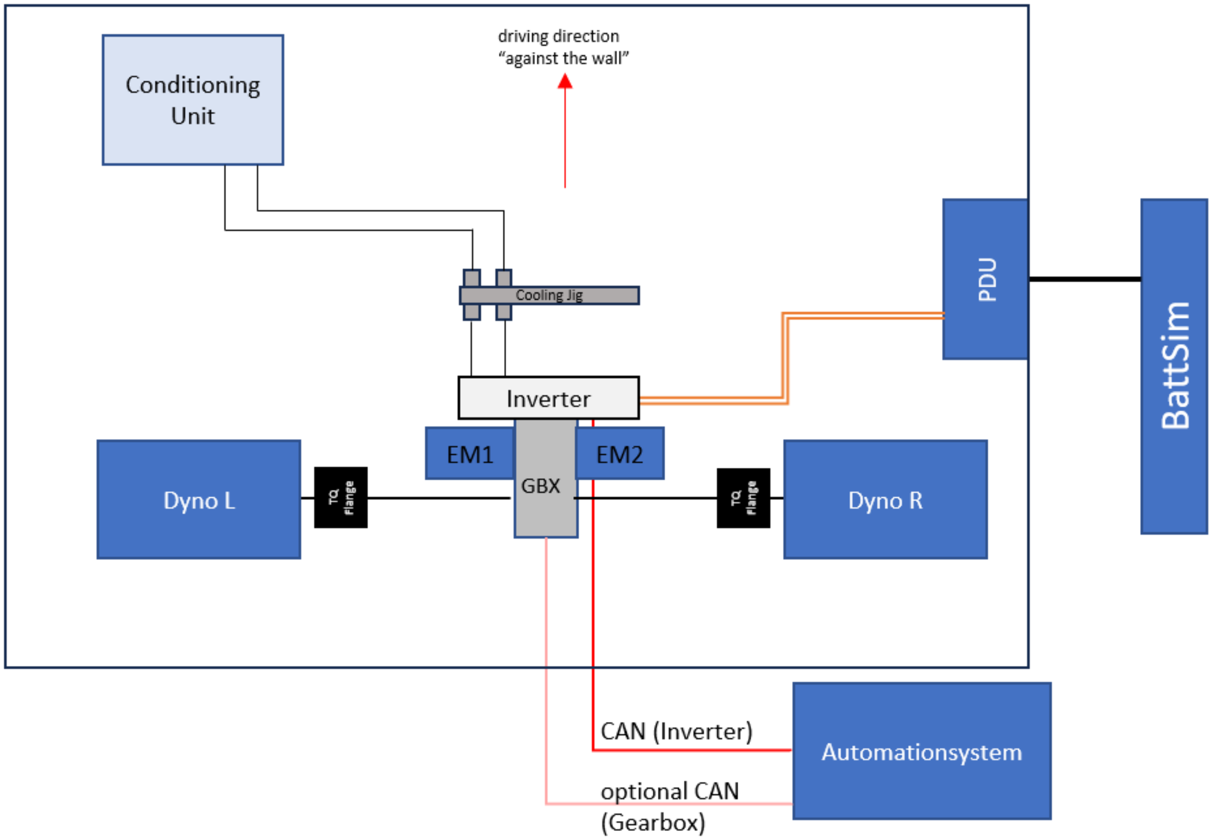


Figure 85 - Standard Layout (STB1).

For the Gearbox Overload test, the setup is different because the device under test is only the gearbox, and the Electric Motor is replaced by another Unit that can provide more than 250% of the Nominal Torque. This different configuration is shown in Figure 86.

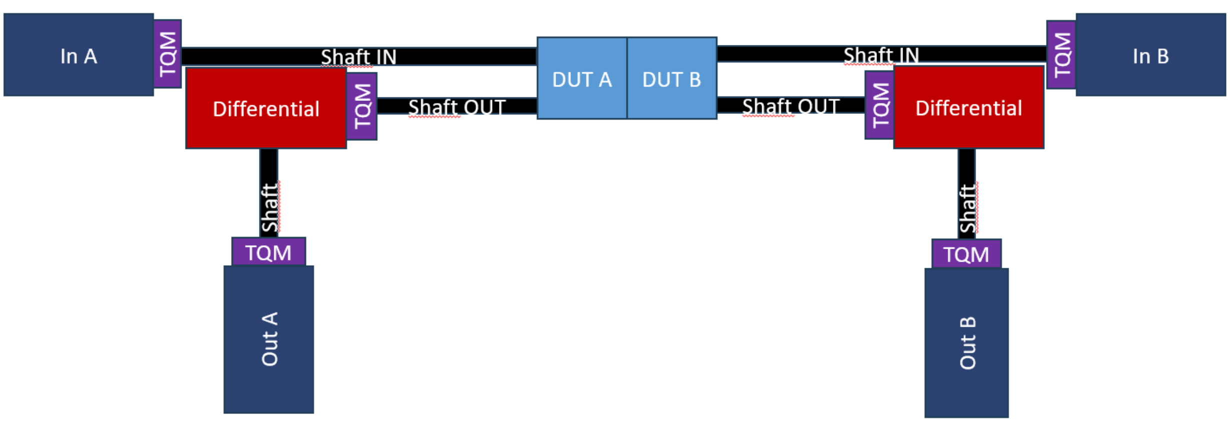


Figure 86 - Test bench layout for contact pattern and overload test.

7.2.2 Accelerated Durability Test

Accelerated Durability Test is designed to validate the lifespan and reliability of EV powertrain components, including electric motors, gearboxes, and power electronics, by subjecting them to stress conditions exceeding normal operational levels, thereby accelerating wear and failure modes in shorter time spans [95] . These tests typically run continuously (24/7) under mechanical loads, thermal cycling, vibration, and electrical stress, often with automated fault injection capability, to identify weak points and confirm expected durability targets (e.g., hours of operation or cycles).

Standardization and State-of-the-Art Practices

There is currently no single ISO standard devoted solely to EV powertrain accelerated durability testing. (Some ISO work does address EV drive systems: e.g., 19453-6 specifies environmental stress tests to observe battery and drive reliability). Instead, industry practice follows general accelerated testing guidelines. Approaches like HALT/HASS (highly accelerated life/stress testing) and environmental stress screening are adopted in the EV context. In these methods, test protocols use established life-stress models (Arrhenius, Eyring, log-linear, etc.) to relate elevated conditions back to normal usage. For example, a mild-hybrid powertrain life test was accelerated by using temperature as a stress factor according to the Arrhenius equation [96]. In advanced rigs, multiple stressors are combined: accelerated multiple environmental testing (AMET) subjects components to simultaneously thermal, electrical, and mechanical loads. Notably, complex mechanical tests (vibration and dynamometer brake testing) are the core of automotive ADT equipment. These "vibro-dyno" setups (often combined with thermal chambers) are designed for full systems and components, echoing environmental screening methods used in other industries [97].

Typical Test Elements for Electric Powertrain

- **Mechanical load cycling:** test push torque and speed beyond normal duty to stress bearings, rotors and gear sets. Overload pulses or extended high-RPM runs simulate aggressive usage.
- **Thermal cycling:** Components undergo repeated hot-cold shifts. EV powertrains are tested across a wide temperature range to induce thermal fatigue in windings, magnets, and electronics. Such cycling reveals insulation breakdown or material fatigue that might occur in harsh environments.

- **Vibration stress:** Multi-axis sine and random vibrations are applied to uncover resonances and structural weaknesses. Vibration rigs subject the entire assembly to road-like shake and high-G shocks. For example, industry practice treats vibration testing (random/sine) as a key part of accelerated durability, often using shock absorbers and electrodynamic shakers to reveal fatigue paths.

- **Electrical and fault injection:** Tests deliberately introduce electrical anomalies. This can include injecting voltage spikes, current surges, or simulated short/open circuits to trigger protection circuits and check fault handling. Fault-injection is a standard reliability technique: controlled faults are induced to observe the system's response. For instance, one might override sensor signals or momentarily disconnect supply lines to validate that thermal monitoring or current-signature detection works as intended.

In this case, the Accelerated Durability test is made of 459 repetitions of the cycles described in Figure 87, each one is 976.8 s in duration, for a total time of about 447374.4 s, or 124.27 h.

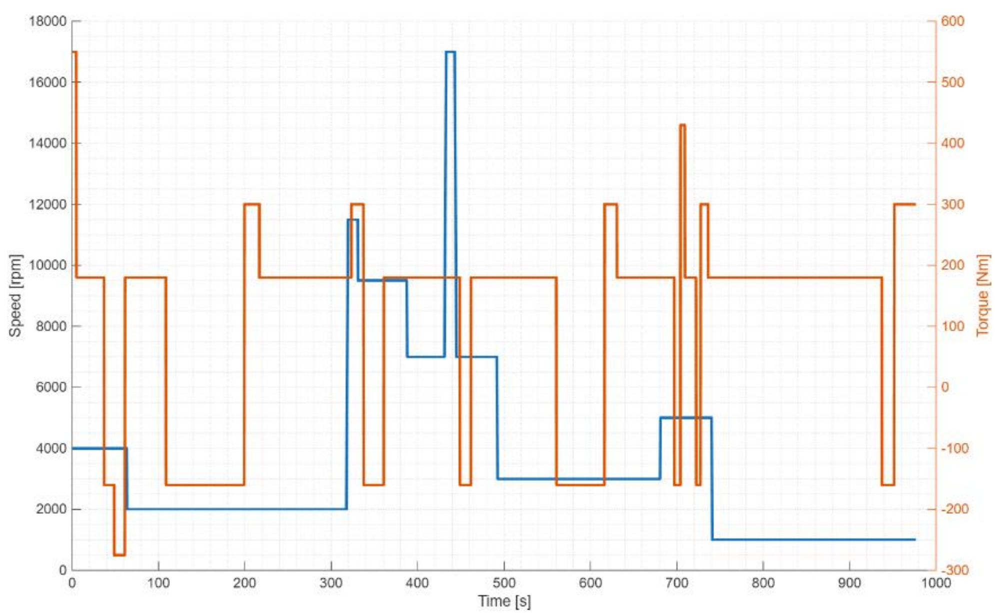


Figure 87 - Working Point of the Accelerated Durability Test.

In Table 7 the main parameters used to perform accelerated durability tests are listed.

Ambient T	25	°C
Voltage	800	V
Sw_freq	10\16	kHz
Coolant IN temp	52	°C
Coolant flow rate	25	L/min
Oil IN temp	52	°C
Oil flow rate	20	L/min
Speed grad limit	5000	rpm/s
Torque grad limit	2000	Nm/s

Table 7 - Working Point of the Accelerated Durability Test, nominal value.

7.2.3 End-of-Line Test

The EOL test is the final, factory-floor verification step for assembled electric powertrain units before vehicle integration. It executes a predefined sequence of functional checks under simulated operating conditions, electrical load profiles, speed-torque curves, thermal cycles, and NVH measurement, while communicating via the vehicle's CAN bus to validate control-systems integrity and sensor functionality. Performance metrics such as torque accuracy, efficiency, vibration, and noise are recorded and compared against tolerance bands determined by manufacturing specifications. Any unit failing to meet these criteria is flagged for rework or rejection, ensuring only fully compliant assemblies proceed down the line [98]. An entire EOL test is reported in Figure 88 for unit EDU008.

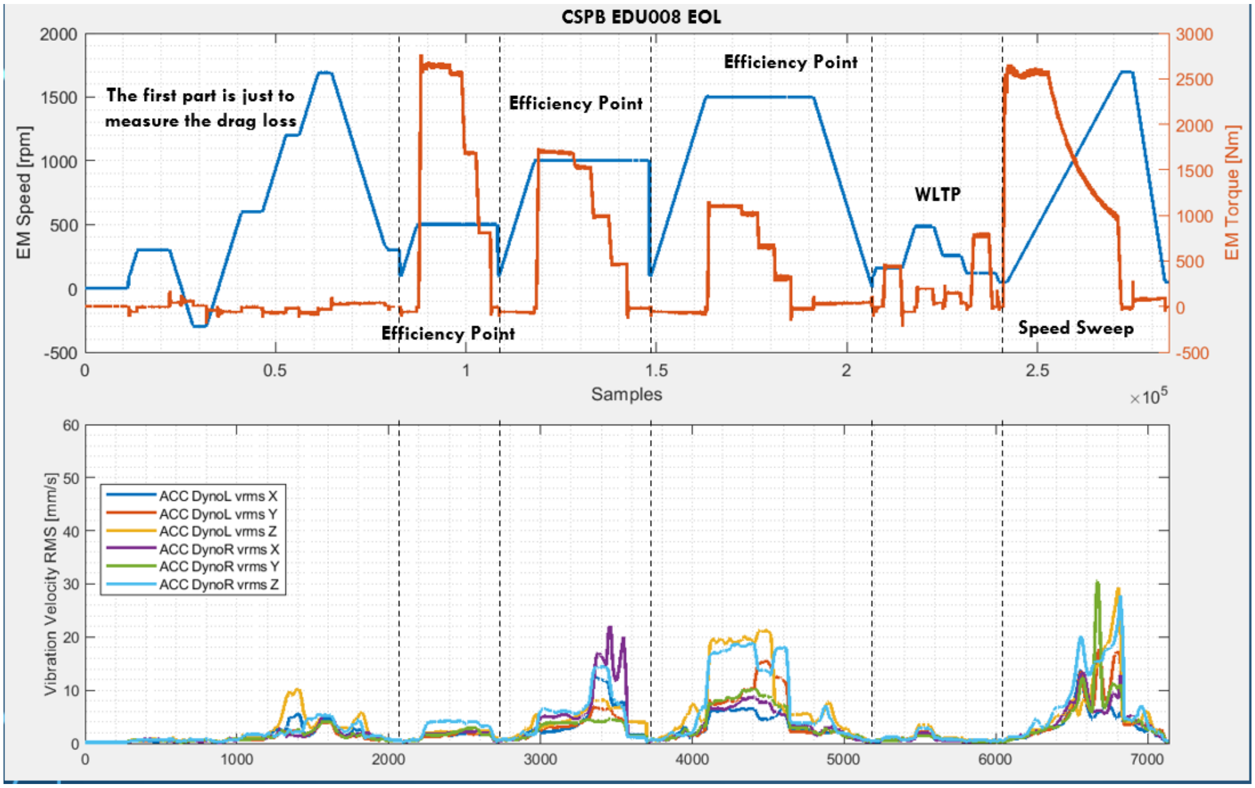


Figure 88 - Working Point & Vibration Velocity RMS for a unit under EOL Test.

7.2.4 Gearbox Overload Test

The gearbox Overload Test verifies the mechanical robustness and thermal endurance of the transmission under torques exceeding its nominal rating. On the specific test bench described before in this section, the gearbox is driven through a sequence of step or ramp torque pulses over thousands of cycles while monitoring torque ripple, temperature rise, vibration, and lubricant conditions [99]. This accelerated stress exposes weak points in gear teeth, bearings, and seals, ensuring that the design meets or exceeds the target durability before vehicle integration [100]. In Figure 89, the number of rotational cycles that every unit is supposed to undergo is reported.

<table>
<tr><th></th><th colspan="4">Overload</th></tr>
<tr><th>Unit</th><th>160%</th><th>200%</th><th colspan="2">250%</th></tr>
<tr><td>U2</td><td>10000</td><td>1000</td><td colspan="2">1</td></tr>
<tr><td>U3</td><td>10001</td><td>1001</td><td colspan="2">1</td></tr>
<tr><td>U4</td><td>10001</td><td>1001</td><td colspan="2"></td></tr>
<tr><td rowspan="7">U5</td><td rowspan="7">10001</td><td rowspan="7">1001</td><td>1</td><td rowspan="7">damage after 2058</td></tr>
<tr><td>100</td></tr>
<tr><td>100</td></tr>
<tr><td>100</td></tr>
<tr><td>200</td></tr>
<tr><td>1000</td></tr>
<tr><td>557</td></tr>
</table>

Figure 89 - Sequence of Working Point for the Overload Test for different units.

7.2.5 Data Characteristics

For both the Durability test and the End-on-Line test, the sampling rates of all the signals recorded in the test bench are the same. Brief summary by physical quantity categories:

- Pressure, Temperature, and Coolant Flow → Sampling Frequency 10 Hz.
- Voltage, Current, Speed, Torque, and Vibration → Sampling Frequency 100 Hz.

During the Speed Sweep stage of the EOL cycle, raw data is recorded at the high sample frequency of real-time acquisition. Usually, this high-frequency acquisition is at a 50 kHz sample rate.

The goal of an NVH bench test for an electric drive unit is to identify and quantify any undesirable noise or vibration sources under controlled conditions. By running the powertrain over its speed-torque envelope on a dyno, it is possible to detect all noise/vibration problems likely to annoy customers and verify that assemblies meet NVH targets. In EVs, the focus is on new sources (e.g., electromagnetic and gear noises) that were once masked and substituted by the engine, which was the only visible and most impactful one. For example, [101] notes that managing EV NVH requires understanding noise from the inverter, the electric motor, and the gears. Typical objectives include detecting assembly faults (e.g., misaligned gears), measuring tonal noises (whine, rattle, etc.), and ensuring repeatable, production-ready behavior.

A common method is to perform a speed sweep at fixed torque or load: the controller ramps motor speed from idle to top speed at a controlled rate, while the inverter supplies the desired torque. Sweep rates are typically a few hundred rpm per second, fast enough to cover the range but slow enough to capture narrow-band phenomena. Sweeps are often done separately in "drive" and “coast” (regen) modes to reveal how loading direction affects noise. After each sweep, order-tracked spectra (torque- and speed-synchronous FFT) are analyzed to pinpoint frequency orders of concern. Steady-state runs at specific speeds/loads may follow to isolate problem speeds found during the sweep [102].

7.3 Specific Methods

The workflow in Figure 90 summarizes and outlines the Real-time for onboard usage strategy at a high level:

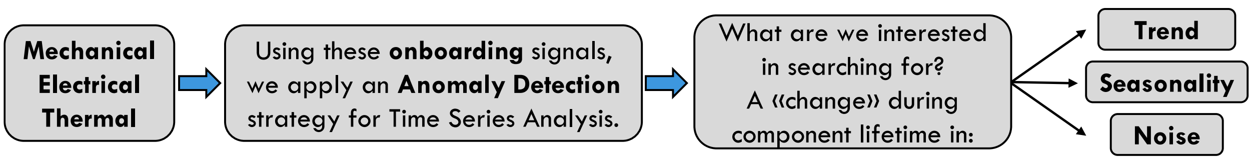


Figure 90 - Aim and scope of the onboard data analysis.

Prior to examining the methods of detecting anomalies in time series, it is imperative to first understand the concept of anomalies and their manifestation in data. An anomaly is defined as a value or event that deviates significantly from the normal trend of the data. These anomalies can be classified into two distinct categories: **punctual** or **collective**. Point anomalies are defined as isolated values that significantly deviate from other values within the time series. Collective anomalies are defined as groups of values that deviate from the remainder of the time series.

- The trend, i.e. the direction and speed of data change over the long run. For example, an increasing trend indicates that the data is increasing over time, while a decreasing trend indicates that the data is decreasing over time.
- seasonality, i.e. the periodic variation of data as a function of time. For example, annual seasonality indicates that the data has a cyclical pattern that repeats itself every year, such as toy store sales increasing in December and decreasing in January.
- Noise, i.e. the random variation of data as a function of time. For example, noise can be caused by measuring, transmitting, or processing errors.

Another main component in the analysis of a time series is cyclicality, i.e. the irregular variation of data as a function of time. For example, economic cyclicality indicates that data has a fluctuating trend that depends on external factors, such as GDP, inflation, unemployment, etc. In Figure 91, the detailed description of the anomaly detection workflow for EOL and Durability test is reported.

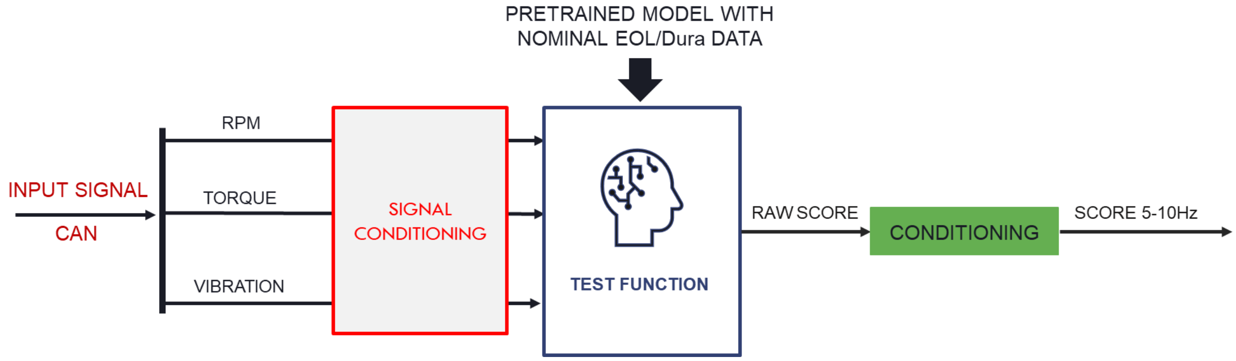


Figure 91 - Detailed description of the anomaly detection workflow.

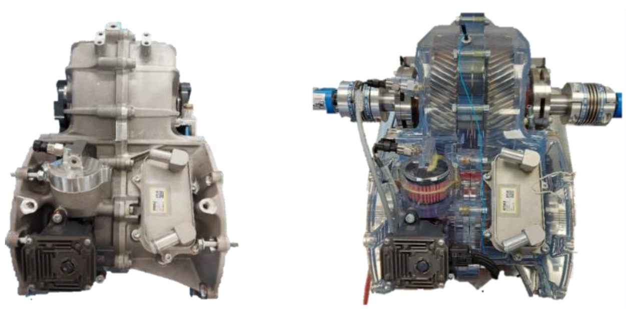
7.4 Results

This section presents the key findings and results from the application of condition monitoring and anomaly detection methodologies to the various electric powertrain test campaigns. The primary objective is to demonstrate the practical effectiveness of the developed algorithms in identifying faults and tracking degradation in a real-world automotive context.

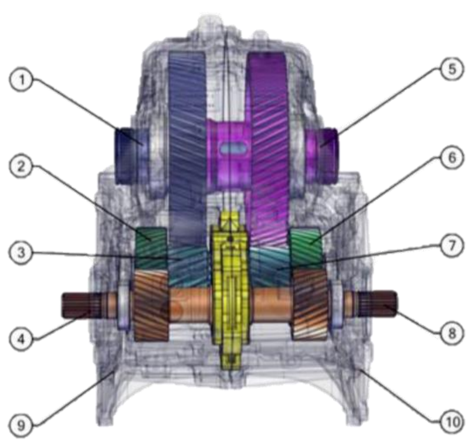
The results are organized by project, with each of the following subchapters detailing the specific analysis and outcomes for a distinct powertrain development program: the **C2 Powertrain**, the **BR1 Project**, and the **CEDAR Project**. For the **CEDAR Project**, a more in-depth analysis will be provided, focusing on the specific sensitivities of the monitoring system to temperature, a direct comparison of acceleration versus velocity-based condition indices, and the identification of abnormal vibrational trends that developed during the accelerated durability test.

7.4.1 C2 Powertrain Development

This case study focuses on the C2 Durability Test, a project centered on a specific electric powertrain, the specifications of which are detailed in Figure 92.



Max. RPM	17700
Max. Torque	2 x 900 Nm
Continuous Torque	2 x 350 Nm
Peak Power	2 x 450 kW
Continuous Power	2 x 250 kW
Center Distance	219,25 mm
Lifetime Durability	100'000 km



Legend	
1	Gear output RS
2	Gear intermediate RS
3	Shaft intermediate RS
4	Gear input RS
5	Gear output LS
6	Gear intermediate LS
7	Shaft intermediate LS
8	Gear input RS
9	Housing RS
10	Housing LS

Figure 92 - Specs and Description of the Device Under Test, C2 Project.

The objective of this analysis was to apply a data-driven model to describe the health state and potentially estimate the Remaining Useful Life (RUL) of the powertrain. The test concluded with a registered mechanical fault in the rear gearbox's left side, making it a valuable dataset for retrospective analysis. The intended approach was to select and compare data logs from three distinct lifecycle stages—"Start Life," "Middle Life," and "End Life"—as defined by a formal durability and homologation testing report, a section of which is shown in Figure 93.

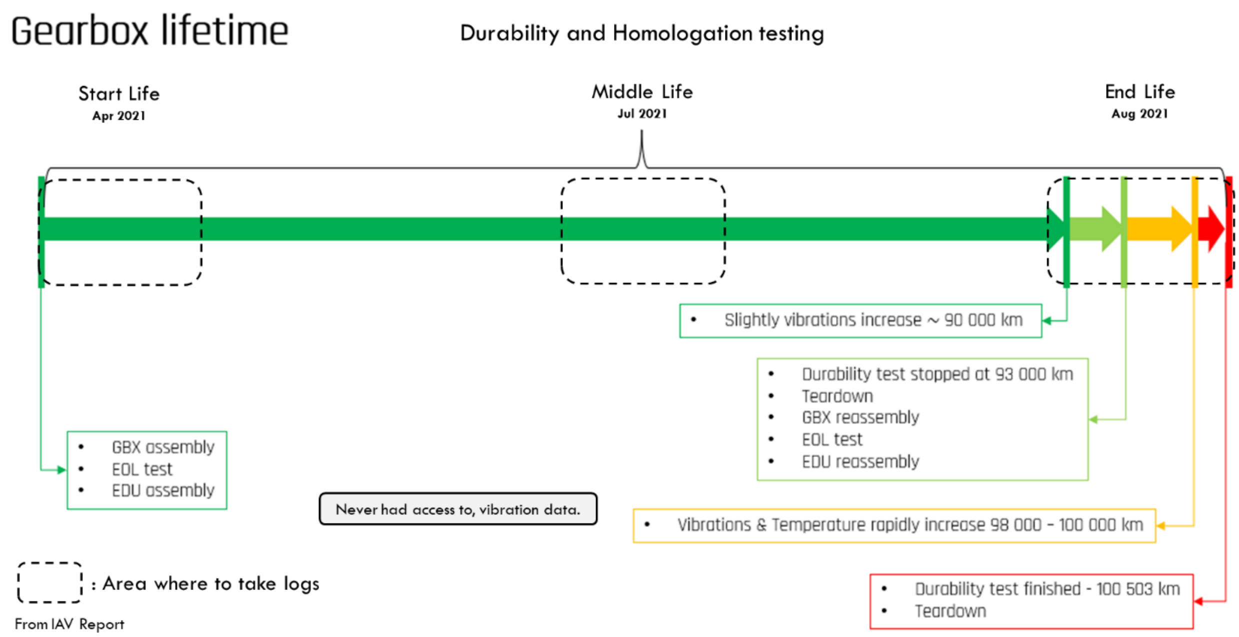
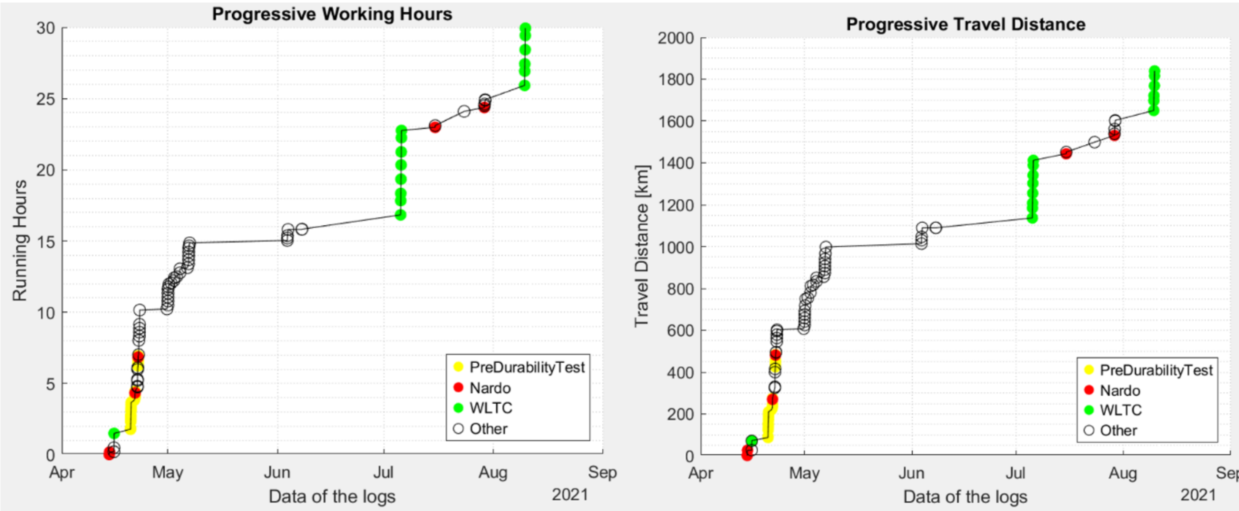


Figure 93 - Durability and Homologation testing, from IAV Report.

The initial step was to identify and categorize comparable data logs from the available dataset that corresponded to these different life stages, as summarized in Figure 94.



Logs Category	Number of Logs	Total hours	Total Distance
Nardò	6	1,0730	141,8021
WLTC	15	12,938	560,5777
PreDurability	26	3,7865	270,2506
«Other»	51	12,5411	917,6773

Figure 94 - Reports and categorization of all the logs found in SharePoint.

However, a preliminary analysis of the available data revealed significant limitations. The cataloged logs represented only approximately 2.3% of the total recorded bench testing hours. This raises critical concerns regarding the statistical representativeness of this small sample. It is uncertain whether the distribution of operating conditions within the available logs accurately reflects the full duty cycle of the complete 100,000+ km durability test, a prerequisite for any valid lifecycle analysis.

Further investigation into the data's consistency revealed a more fundamental problem: a lack of controlled boundary conditions. Figure 95 compares the trends of several key temperature signals across different logs that were nominally recorded during the same "Nardò Cycle" but at different points in the powertrain's life. The thermal profiles are markedly different, indicating that the thermal management of the test bench was not consistent across the acquisitions.

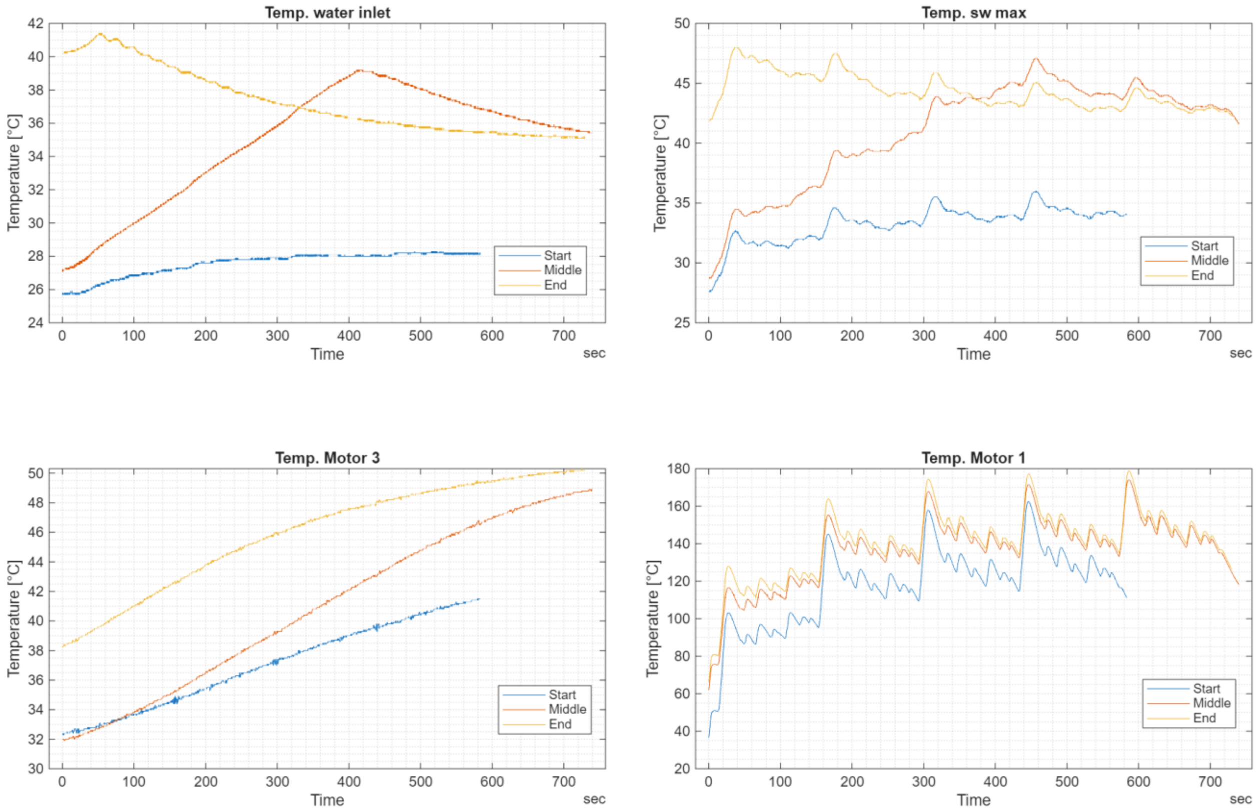


Figure 95 - Comparison of the logs in different periods of the life, Nardò Cycle.

This lack of controlled thermal boundary conditions fundamentally compromises the ability to make a direct and reliable comparison of other condition monitoring signals, such as vibration, as thermal effects can significantly influence their characteristics.

In conclusion, the C2 Durability Test dataset was deemed unsuitable for the intended CM-PdM analysis due to a confluence of critical data quality issues. These included:

- **Unrepresentative Data Sampling:** The available logs constituted a very small and potentially biased fraction of the complete durability test.
- **Inconsistent Boundary Conditions:** The lack of controlled thermal conditions, as demonstrated, prevents a valid comparison of data across different life stages.
- **Data Integrity Problems:** Further practical issues were discovered, such as missing data channels in some logs and the need for complex, undocumented decoding procedures in others.

These limitations highlight a crucial principle: the data was not originally acquired with the specific, stringent requirements of a data-driven CM-PdM analysis in mind. Consequently, this case study serves as a valuable lesson, underscoring the necessity of a structured, "data-first" approach to test design when predictive modeling is a primary objective.

7.4.2 BR1 Project

This case study focuses on an Overload Test conducted on the BR1 powertrain. The configuration of the Device Under Test (DUT) and the corresponding sensor locations are shown in Figure 96.

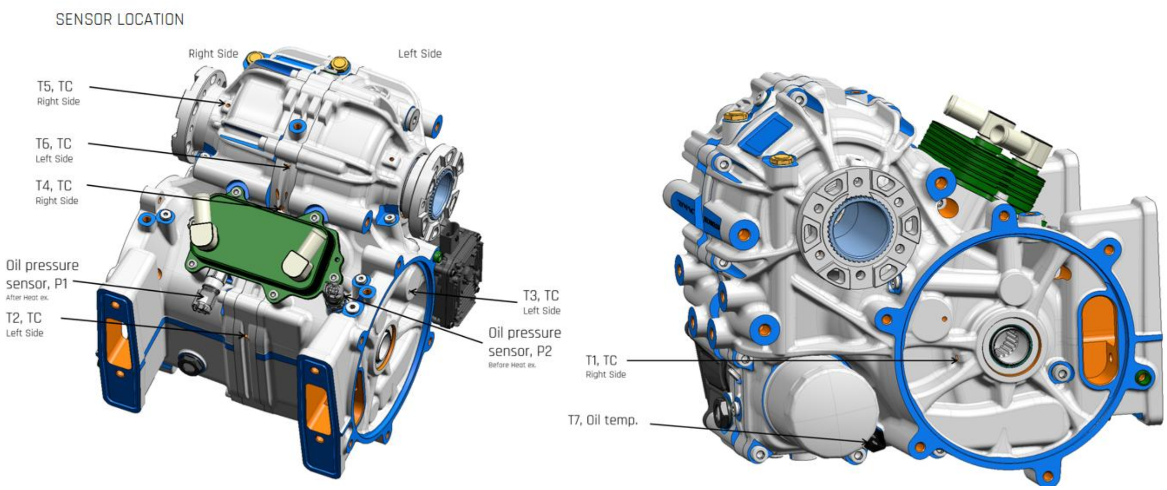


Figure 96 - Sensor Location for the BR1 Project, Overload Test.

The test concluded with a catastrophic mechanical failure on the left side of the unit's gearbox. The moment of failure is clearly captured in the speed sensor data, as presented in Figure 97. The plot shows the input speed of the left-side motor rapidly increasing, while the corresponding output speed abruptly drops to zero, indicating a complete loss of power transmission.

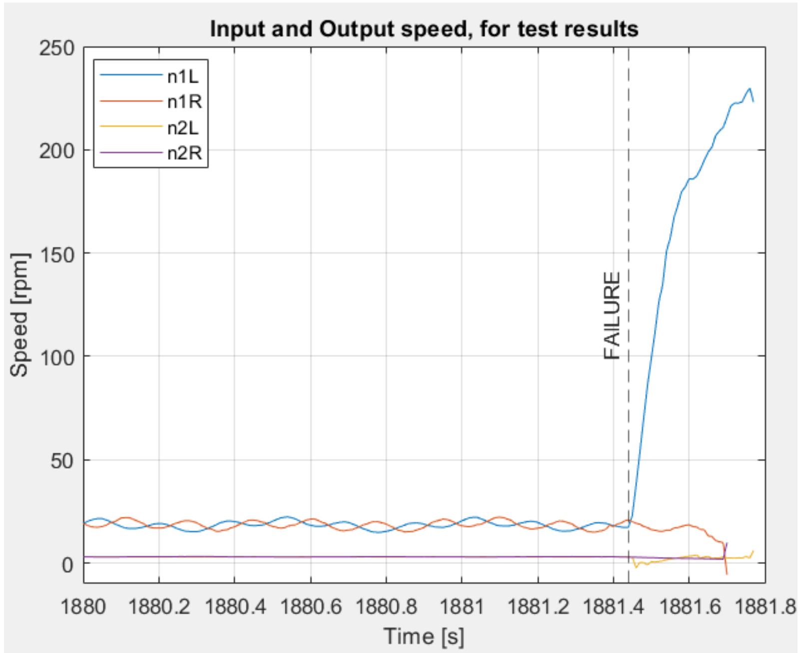


Figure 97 - Input and Output Speed for both sizes of unit 5.

The primary analytical goal was to retrospectively analyze the high-frequency data to determine if this failure could have been anticipated. However, an initial examination of the available signals revealed a fundamental limitation imposed by the data acquisition parameters. An attempt to perform a time-frequency analysis, such as generating a spectrogram, was compromised by significant aliasing. As shown in Figure 98, which presents the frequency analysis of the input speed signal, as the rotational speed increases, the dominant frequency order exceeds the Nyquist frequency limit ($f_{sample}/ 2$). This folding of frequencies makes it impossible to accurately interpret the spectral content of the signal. The same limitation was found to be present in the torque and electrical measurements (Id, Iq, Vd, Vq), precluding their use for high-frequency diagnostic techniques.

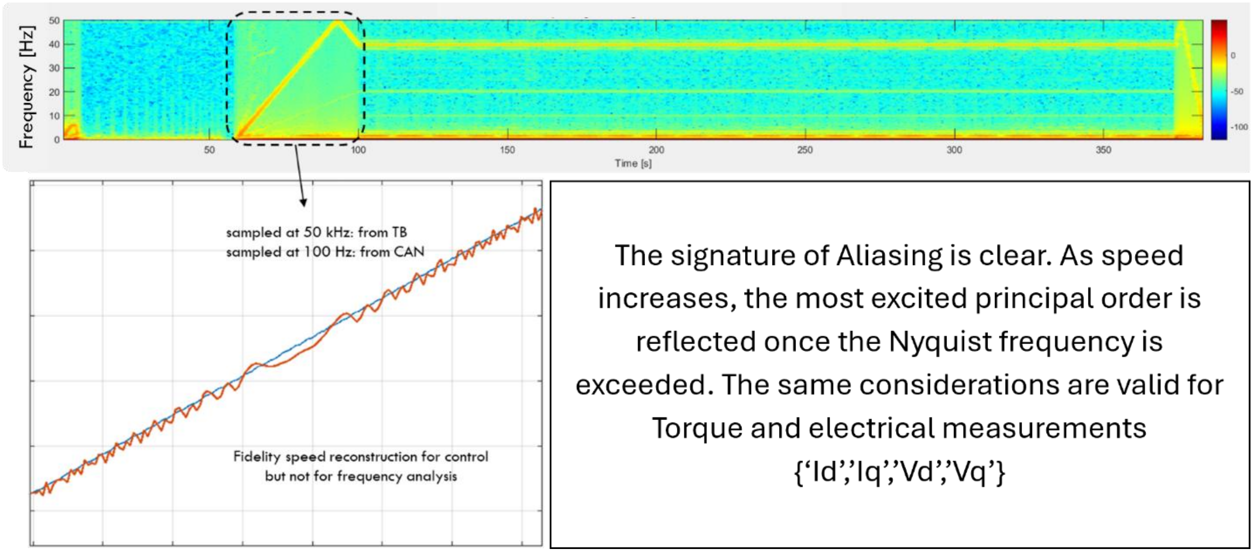


Figure 98 - Frequency Analysis of Speed Input.

This aliasing issue also prevents the application of more advanced techniques, such as inter-order phase analysis, which could otherwise be used to estimate mechanical slack or backlash by comparing the phase shift between identical orders on the motor and the dyno.

Given these data limitations, a different analytical strategy is recommended for this type of test. Since the overload test operates far beyond rated conditions (with torques up to 250% of nominal) and the failure mode is likely driven by mechanical overload, the problem is more appropriately framed as one of oligocyclic fatigue. The goal should be to estimate Remaining Useful Life (RUL) based on the accumulated stress cycles rather than detecting a gradually developing fault signature.

Using only the available onboard sensors, it is not possible to retrospectively determine if the failure of unit "U5" could have been anticipated within the few rotation cycles preceding the event by examining short-term trends. A potential alternative approach, constrained by the low sampling rate, would be to inspect the harmonic content of the instantaneous velocity signals and compare the input and output velocities to reveal torsional irregularities. However, it must be emphasized that any frequency-domain analysis on this dataset is severely limited by the aliasing problem. The usefulness of spectral methods for this application is contingent upon a significant improvement in the data acquisition sampling rate.

7.4.3 CEDAR Project

This case study investigates an Accelerated Durability Test performed on a "CEDAR" project powertrain unit. The physical configuration of the Device Under Test (DUT) and the placement of key sensors are illustrated in Figure 99.

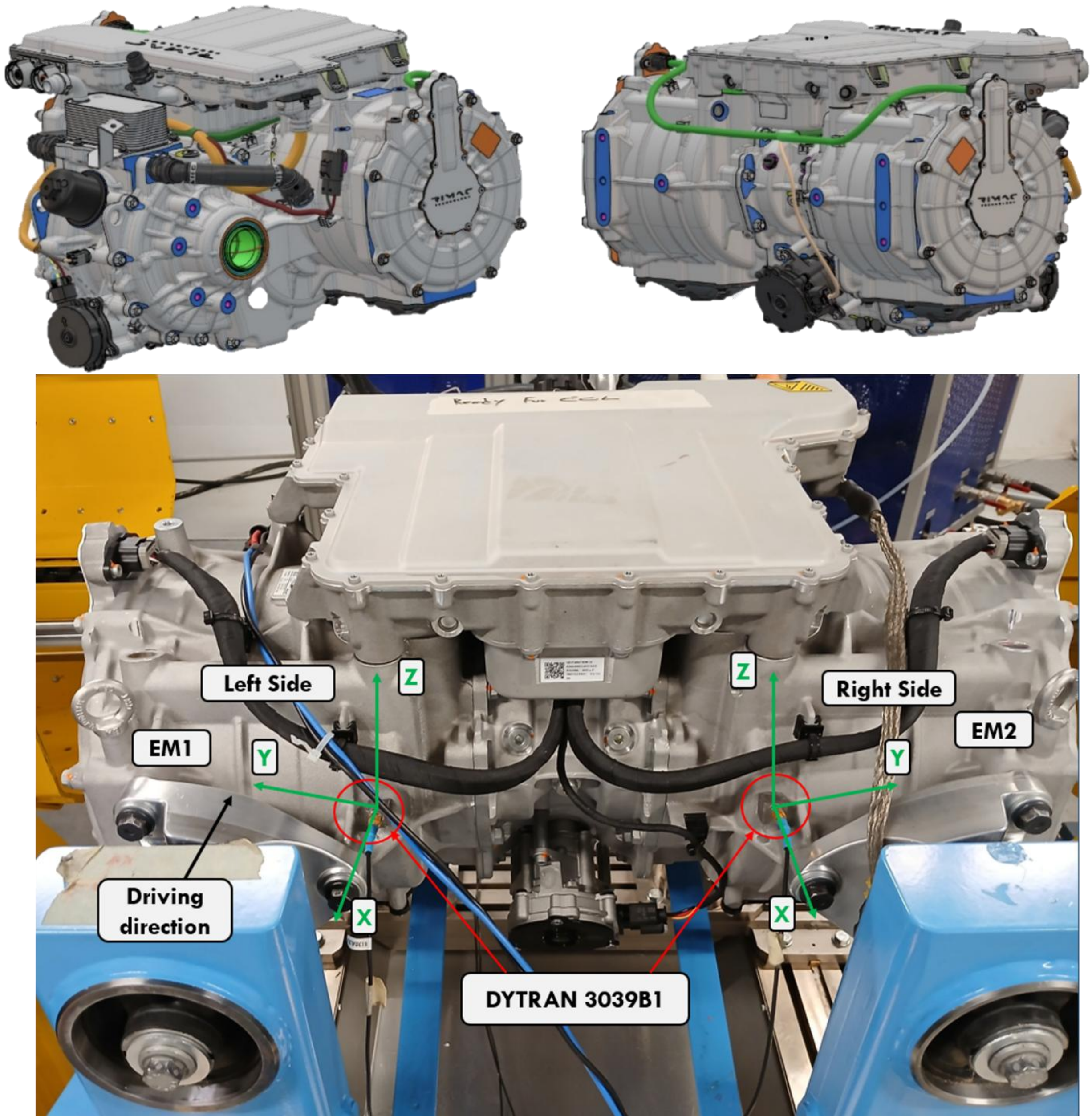


Figure 99 - CEDAR Project, Rear PWT sensor location.

The first unit subjected to this durability test, designated Unit 034, experienced a premature failure after approximately 37 hours of operation. Subsequent teardown and inspection identified the root cause as a catastrophic bearing failure, the details of which are shown in Figure 100.

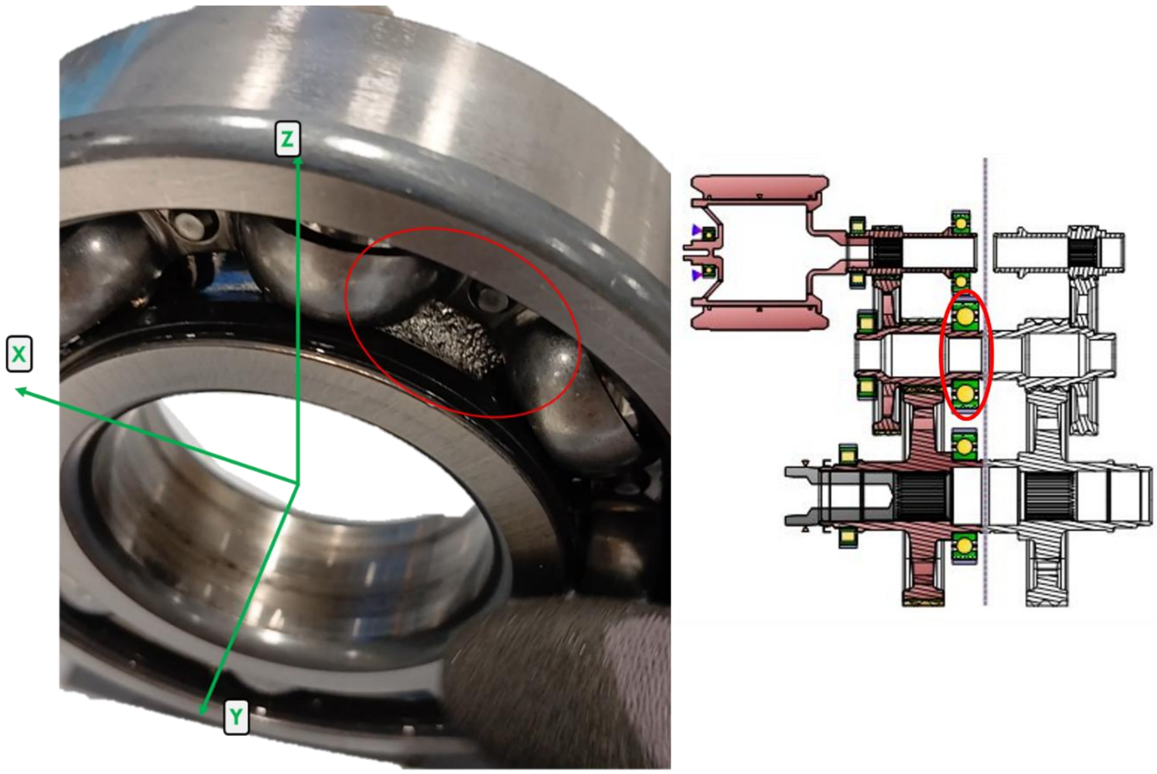


Figure 100 - Element (bearing) and type of fault, CEDAR Project.

Initial analysis focused on the vibration data from the test. Figure 101, shows the Vibration Velocity RMS trends for three different conditioning cycles. For this type of short, variable cycle, the relative change in the vibration signature is more diagnostically significant than its absolute value. The fault signature was found to be most clearly visible in the Z-direction of the accelerometer.

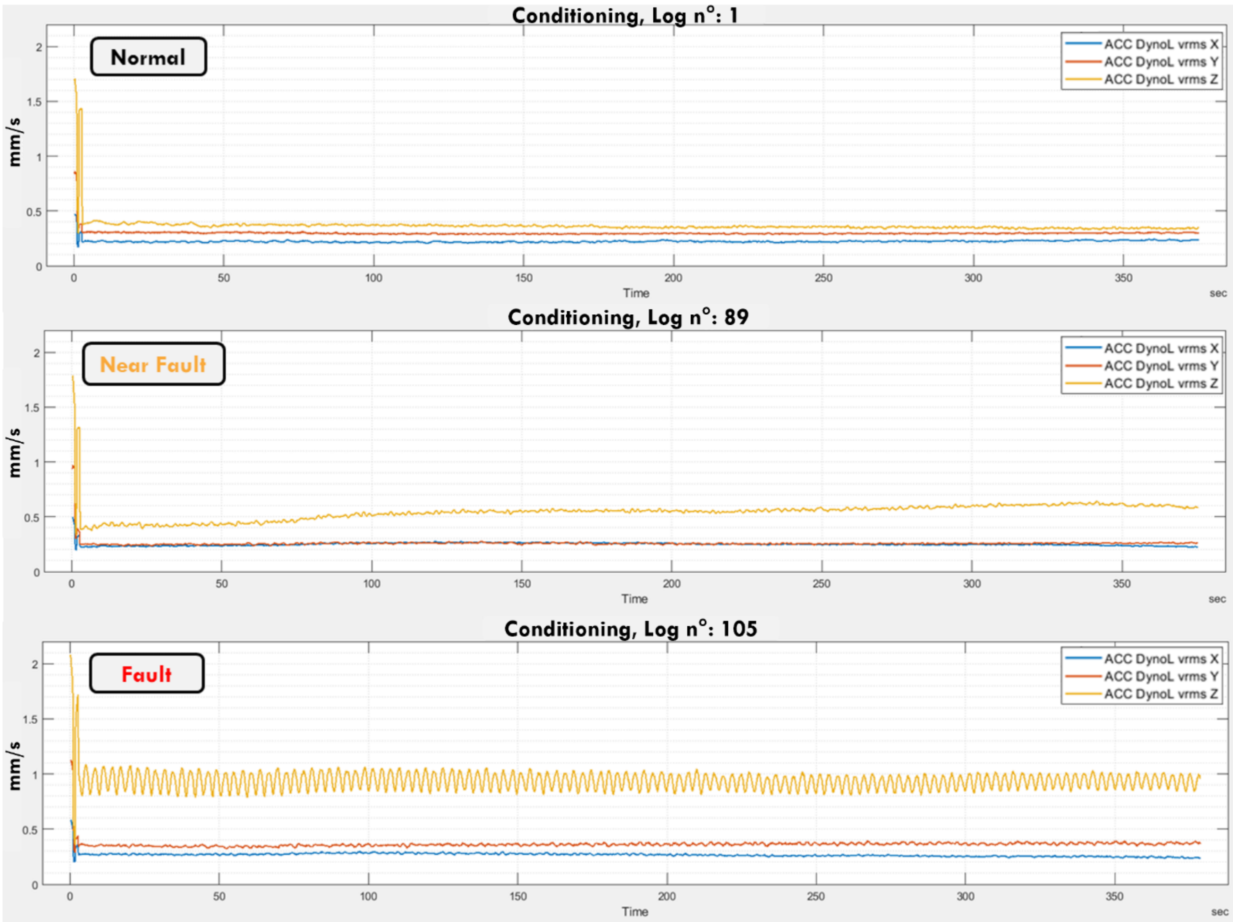


Figure 101 - Vibration behavior for the conditioning cycles.

For the longer, 15-minute working cycles, the analysis was narrowed to a specific 20-second window (from 430 to 450 seconds into the cycle) where the fault signature was most prominent, as shown in Figure 102. The feature used for this analysis was the RMS of the vibration velocity, derived from the raw acceleration signal. The condition monitoring strategy implemented on the test bench utilized fixed thresholds for this feature: a 'Warning' limit at 40 mm/s (rms) and a 'Stop' limit at 60 mm/s (rms), with a 1-second debounce time. While several spikes were observed in the final cycles, only the very last cycle exceeded the stop threshold, which terminated the test.

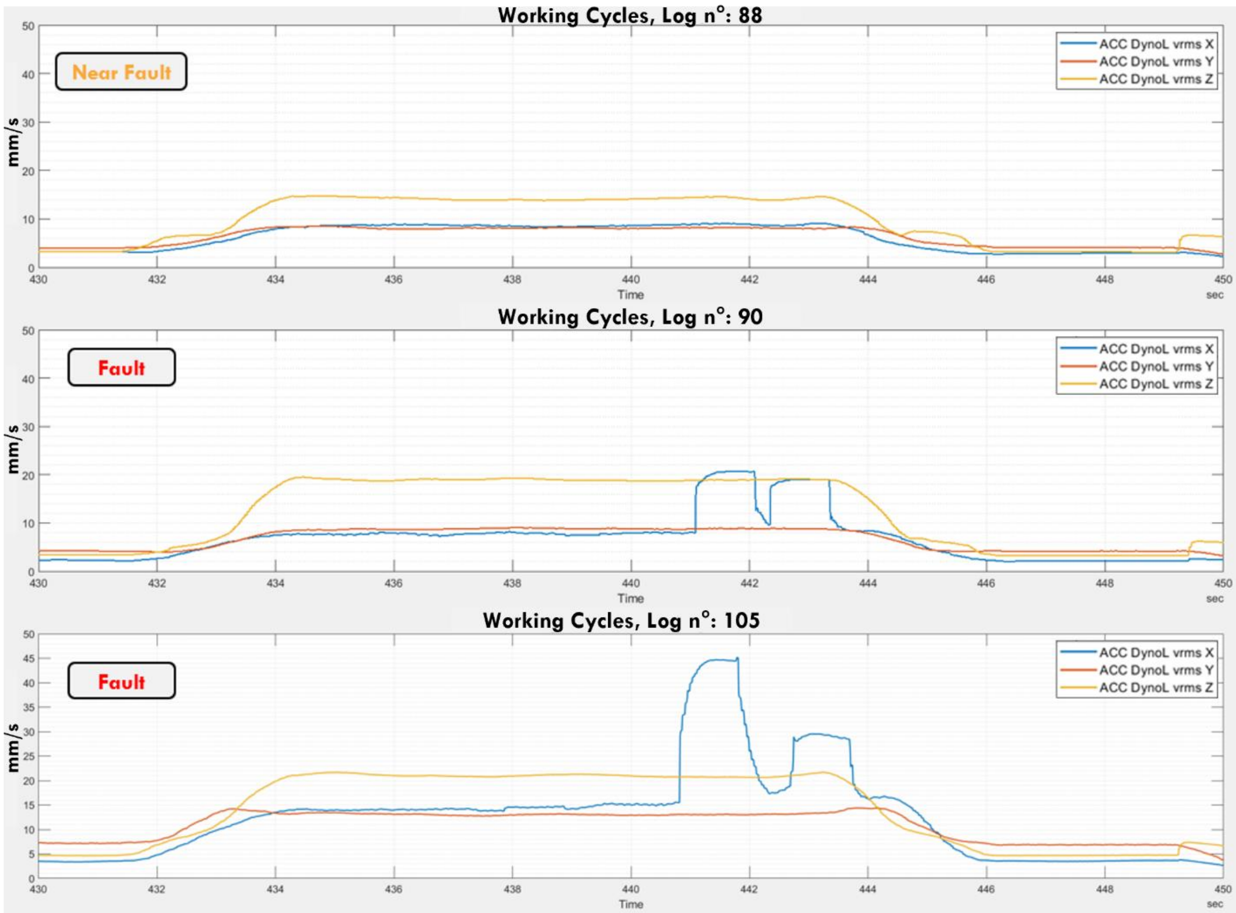


Figure 102 - Vibration behavior for Working Cycles.

In addition to the vibrational anomalies, a significant and progressive increase was observed in the oil pressure trend throughout the test. This prompted a detailed investigation of the lubrication and cooling circuit, a CAD view of which is shown in Figure 103.

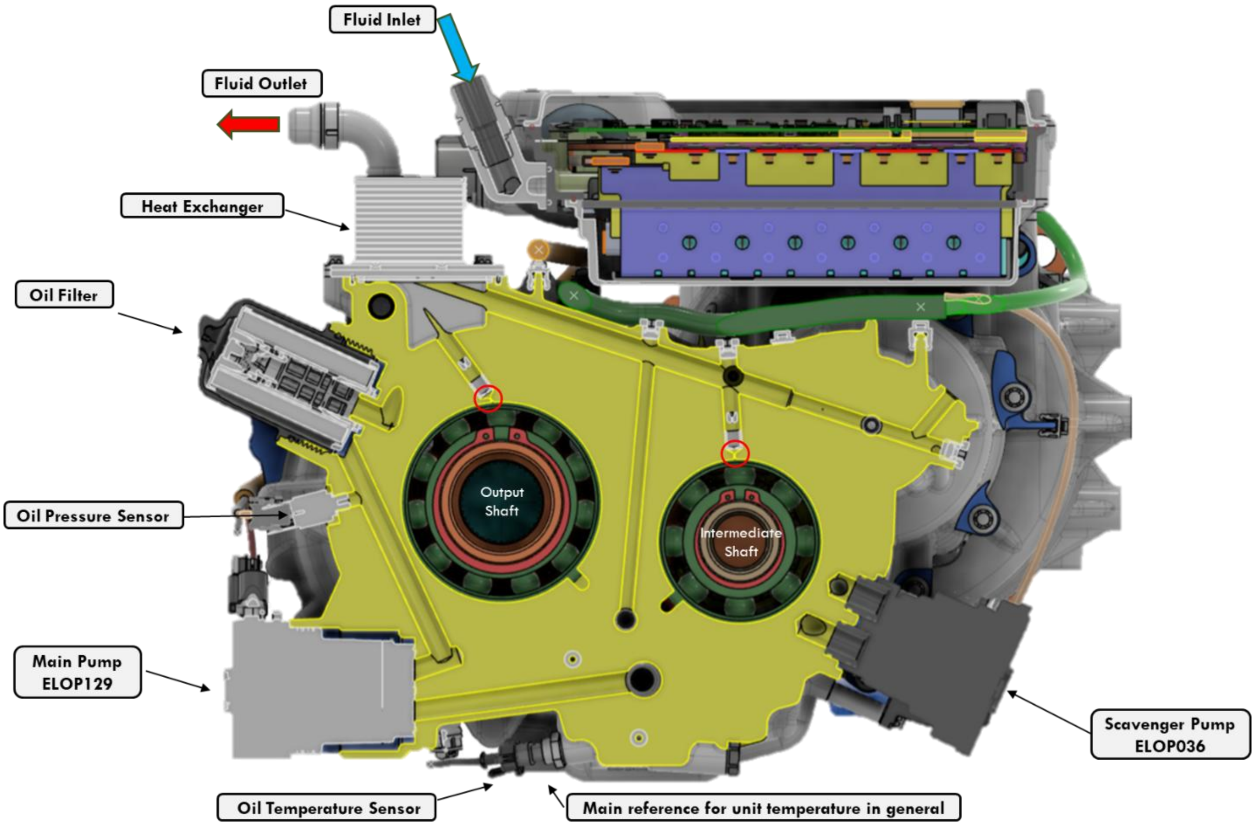


Figure 103 - Oil Circuit view (Forced Lubrication System).

Given that the pressure sensor is located between the pump discharge and the filter inlet, the consistently rising pressure trend strongly suggests a progressive clogging of the oil filter. This phenomenon is clearly visualized in Figure 104, which overlays the oil pressure trends from all 106 working cycles of the durability test. A clear progression is visible, from a mean pressure of approximately 1.8 bar in the initial (blue) cycles to over 3 bar in the final (red) cycle. A similar, though more variable, upward trend is also visible in the conditioning cycles, as shown in Figure 105, where the pressure increase is also influenced by temperature-driven changes in oil viscosity.

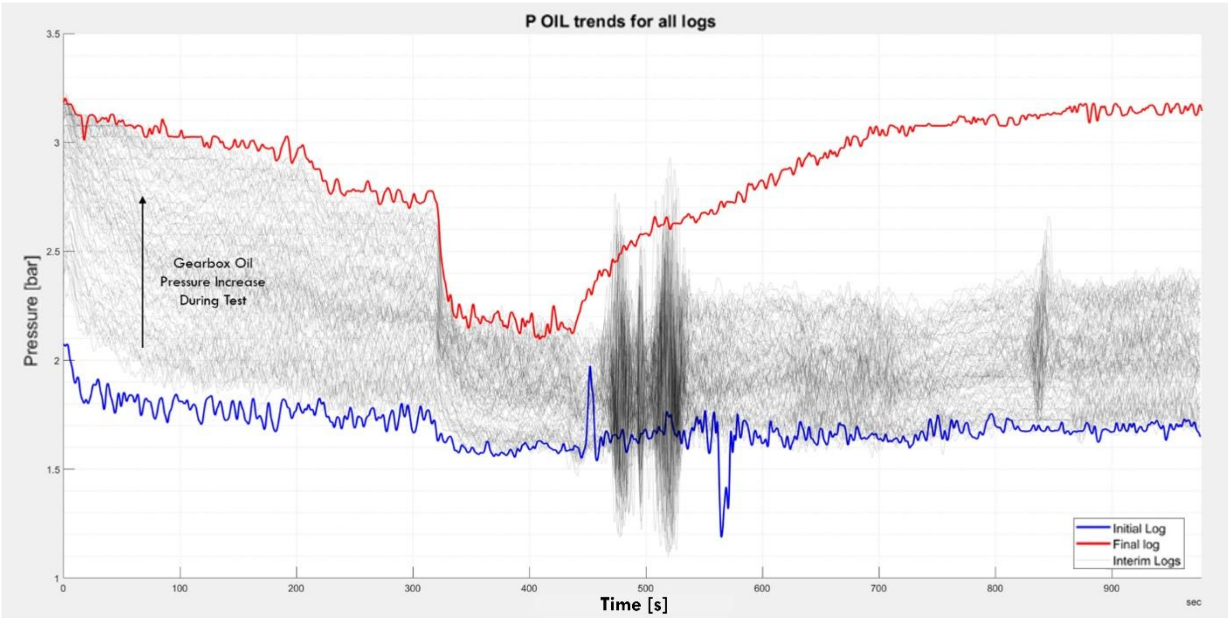


Figure 104 - Oil pressure trends for all the Working Cycles, Unit 34.

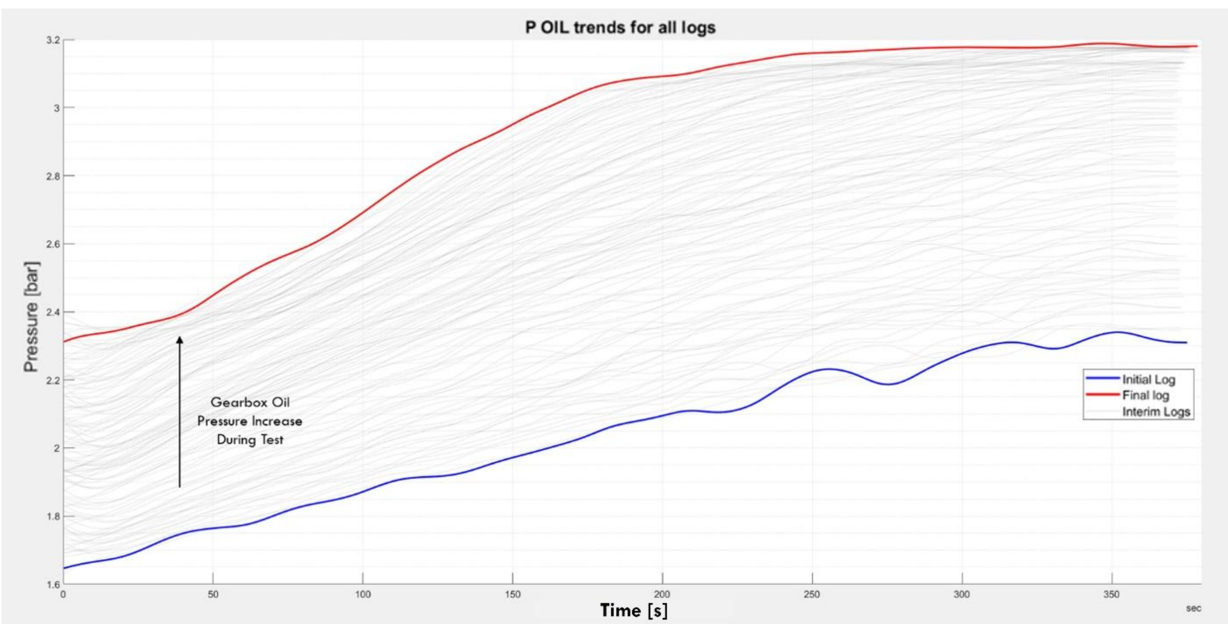


Figure 105 - Oil pressure trends for all the conditioning cycles.

This clear degradation signature in the lubrication system provided an opportunity to develop and validate a multivariate, model-based anomaly detection strategy. The workflow for this onboard strategy, outlined in Figure 106, aims to produce a Warning Index or RUL estimation based on the signals from the main oil pump.

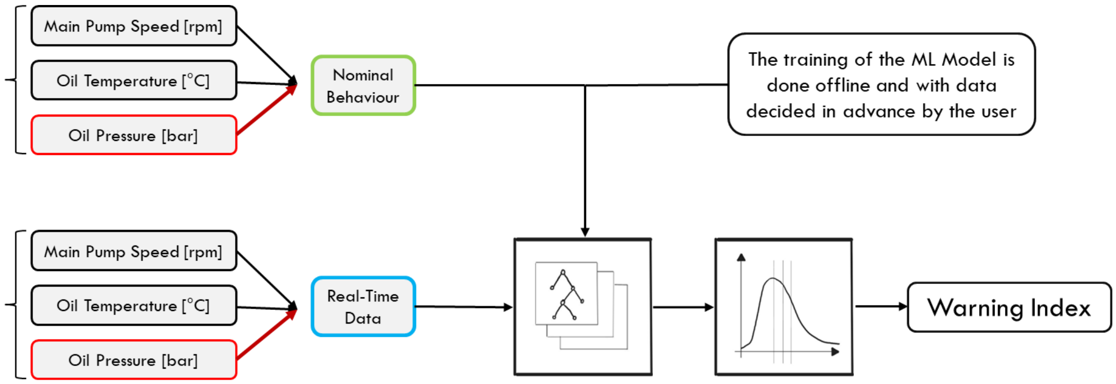


Figure 106 - Warning Index/RUL Estimation model based on the Main Oil Pump signals

Using this workflow and applying it to the dataset of the durability done on unit 34, it is possible to obtain the results reported in Figure 109. An SVM Model is applied for classification, trained with one class that represents the healthy conditions of the system behavior. The healthy conditions are defined simultaneously by the values of the Main Pump Speed, Oil Temperature, and Oil Pressure. By setting up the hyperparameters of the models, it is possible to define the boundary of the healthy area (the grey surface), and with the normalized anomaly scores, it is possible to define how much a point is in a good or bad condition state. An anomaly score of 0 corresponds to a normal point in the distribution of the training and so corresponds to a very healthy state (the blue ones), an anomaly scores higher than 0.5 exceeds the threshold defined by the boundary (the yellow ones). Finally, if we move towards an anomaly score close to 1(the red ones), the condition maximally deviates from nominal behavior, which is seen as the most severe scenario.

The training dataset only grounds on the first cycles of the test since they are the ones that most closely reflect both the **design specification** of the oil circuit and the **deterioration signature** analyzed using the accelerometer signal.

The training dataset remains within the confidence boundary.

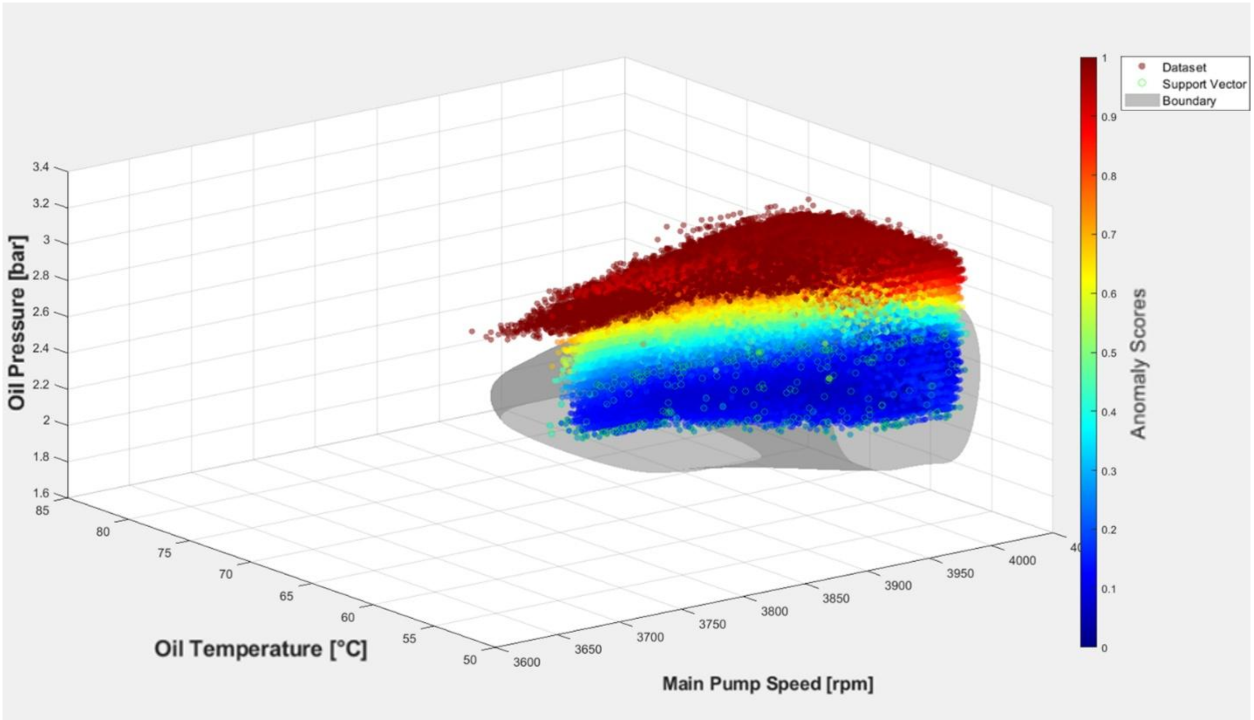


Figure 107 - Results of the Anomaly Detection Model, data from Unit 34.

Figure 108 reports different views of the oil circuit Working Points throughout the test for the conditioning cycles and the corresponding raw anomaly score values calculated by the developed Machine Learning model.

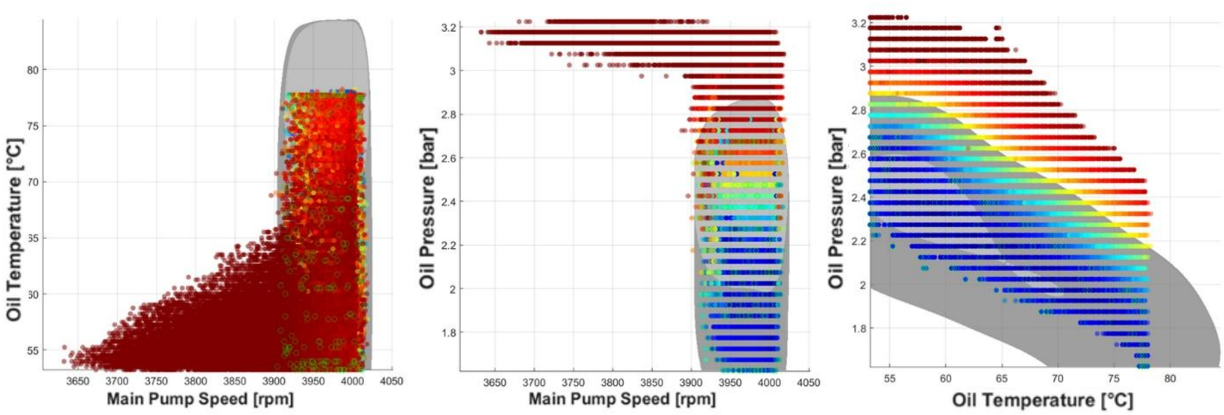


Figure 108 - Section view of the results of the Anomaly Detection Model, data from Unit 34.

The correlation between the health of the lubrication circuit and the mechanical health of the powertrain is demonstrated in Figure 109. This plot shows the trend of the mean and median anomaly scores for each cycle alongside the vibration level. A clear, strong correlation is visible, indicating that the degradation of the lubrication system (i.e., filter clogging) is directly linked to the increase in mechanical stress and vibration that ultimately led to the bearing failure.

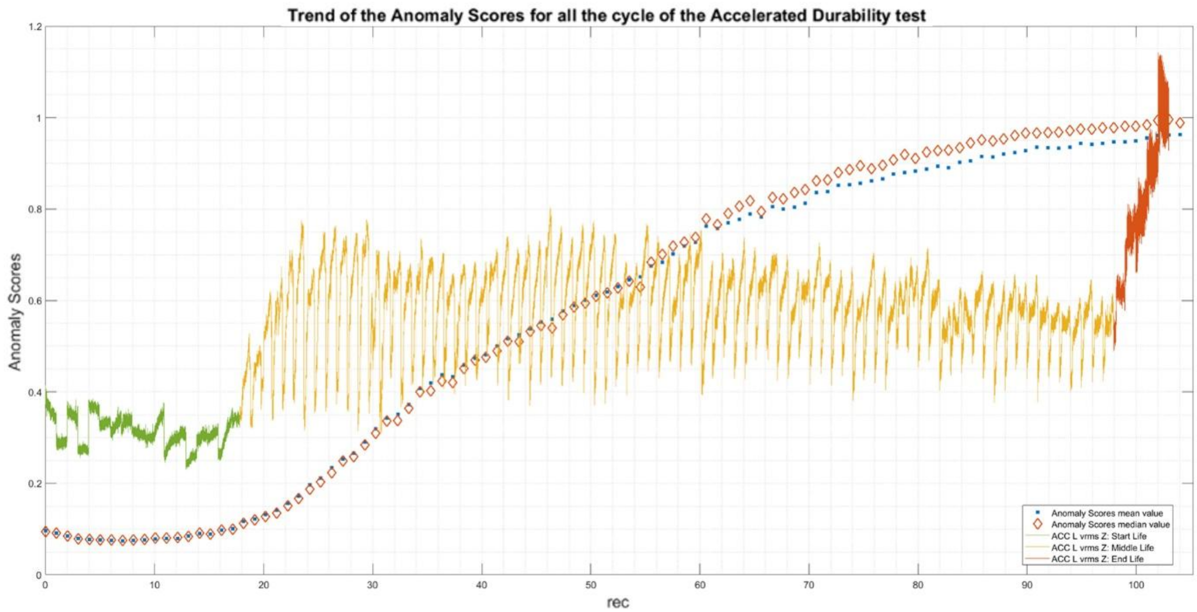


Figure 109 - Trend of the Anomaly Scores for all the cycles of the Accelerated Durability test Unit 034.

The final phase of this work focused on validating a real-time implementation of this monitoring strategy using an ARISBE, an embedded condition monitoring system by Alma Automotive. The hardware architecture for this real-time application is shown in Figure 110.

With the last test in System Test Bench (STB1) during the internship period, the compatibility of this device with a test bench working scenario was tested: I/O communication with the bench and signal processing capability were verified. During this application of the DEMO version Anomaly Score was estimated only for the left accelerometer Z direction and the right accelerometer Y direction.

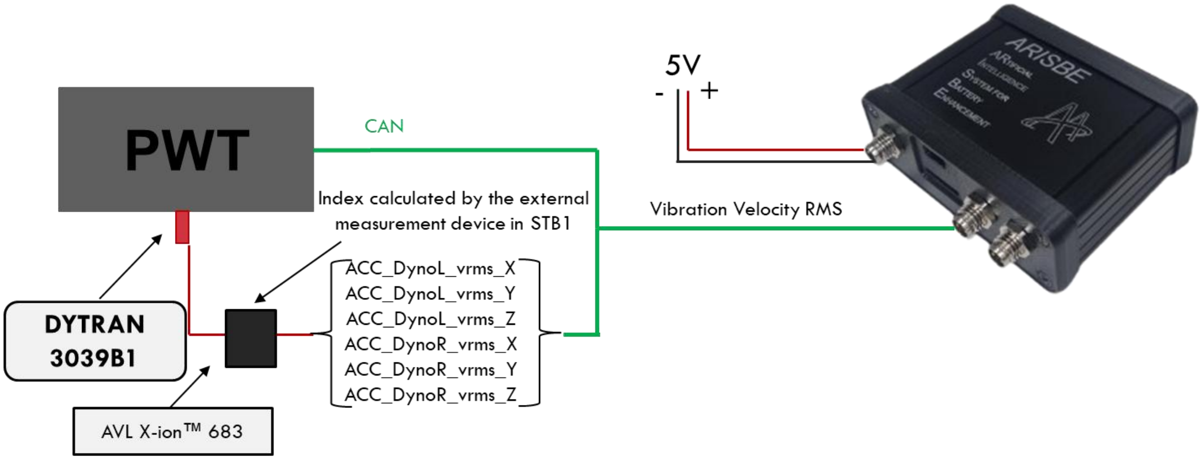


Figure 110 - Hardware Architecture of the real-time Condition Monitoring strategy

For this demonstration, a model was trained on a dataset from an EOL test and then applied in real-time to a durability test. The output of the anomaly score for every cycle of the durability test for the left-side Z-direction accelerometer is shown in Figure 111.

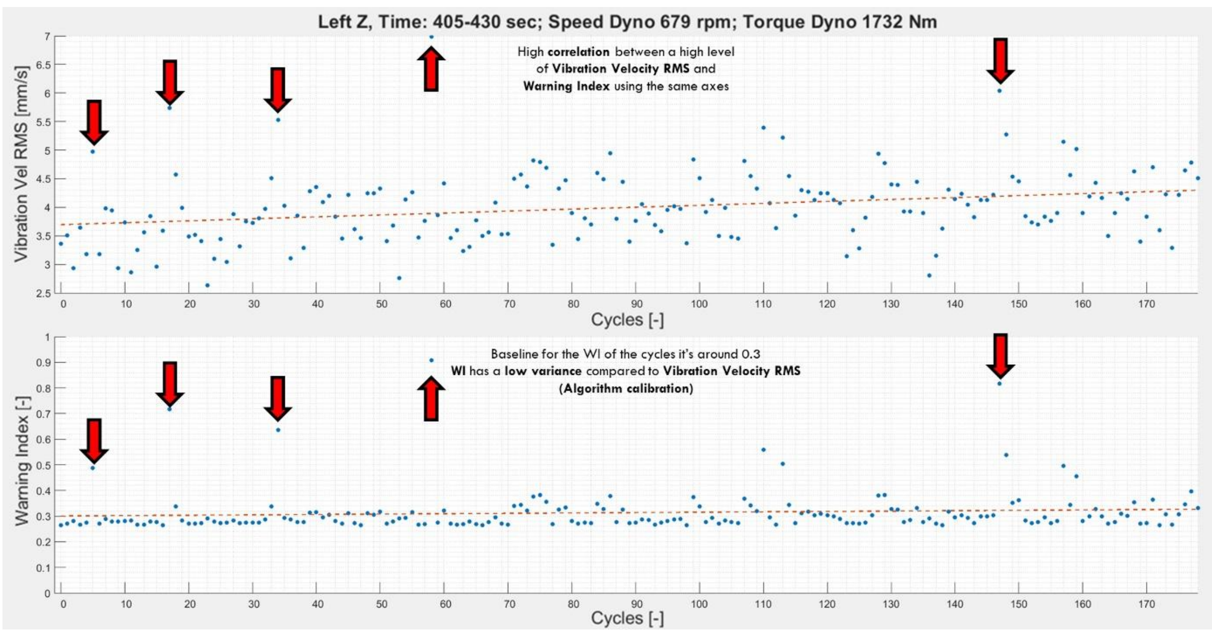


Figure 111 - Mean value for a specific window of interest for the warning index, Left Z.

Figure 112 presents a similar analysis for the right-side Y-direction accelerometer. In this figure, a second subplot is included to provide a quantitative measure of the degradation trend over time. This bar plot illustrates, for each cycle, the cumulative percentage of the total number of cycles to that point that has exceeded a sample warning index threshold of 0.8. This metric serves as a clear indicator of the increasing frequency of high-anomaly events as the durability test progresses, providing a simple yet effective way to track the accumulation of damage or wear over time.

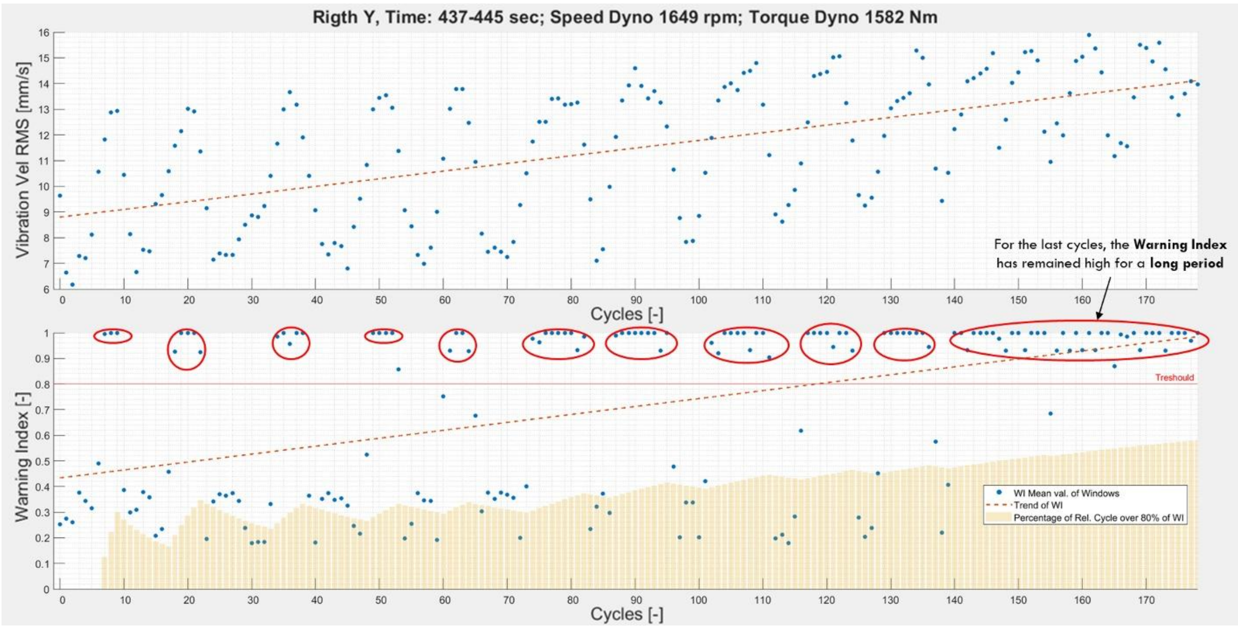


Figure 112 - Mean value for a specific window of interest for the warning index, Right Y.

7.4.3.1 Consideration about Temperature Sensitivity and Fault Progression

The durability test of the powertrain unit designated DVP_SC13T is of particular diagnostic interest as it experienced a sequence of three distinct and significant failure modes over its operational life. The chronological progression of this test, which was conducted across two different test bench facilities (ASAP and STB1), is summarized in Figure 113. This timeline provides a high-level overview of the test phases, their durations, and the specific faults that were identified at the conclusion of each phase.

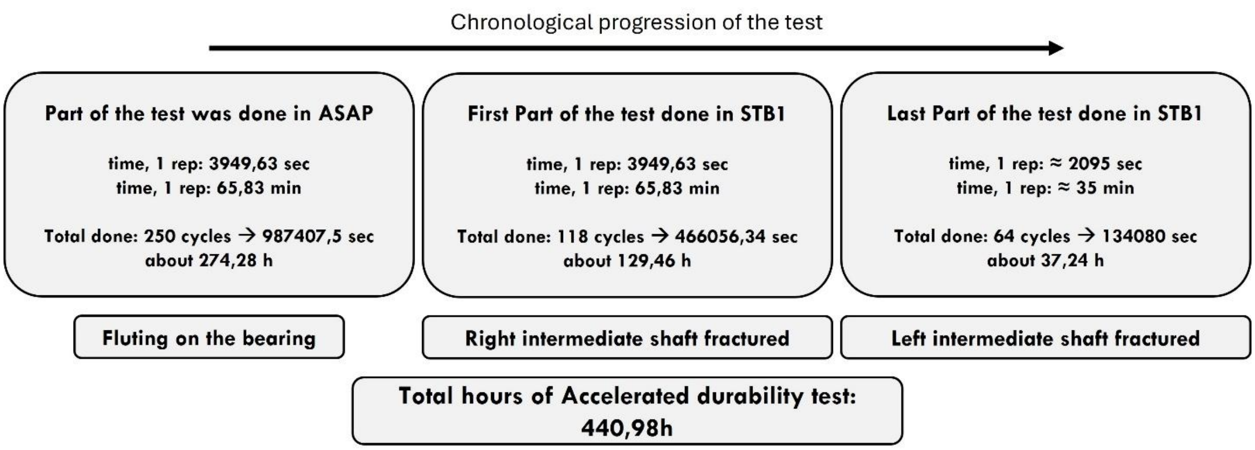


Figure 113 - Chronological progression of the accelerated durability test, DVP_SC13.

As the timeline illustrates, the test was segmented as follows:

1. **Phase 1 (ASAP facility):** The initial part of the test was conducted on the ASAP test bench, where the unit ultimately developed a **bearing fluting fault**.
2. **Phase 2 (STB1 facility - Part 1):** The unit was subsequently moved to the STB1 test bench for continued testing, which concluded with the **fracture of the right-side intermediate shaft**.
3. **Phase 3 (STB1 facility - Part 2):** After repairs, the final phase of testing was also conducted on STB1, ending with the **fracture of the left-side intermediate shaft**.

The following analysis will examine the data from these distinct phases to highlight the different fault signatures and the sensitivity of various sensor modalities to each failure mode.

The first fault, **bearing fluting on the non-drive end (NDE) of the motor shaft**, developed during the test phase at the ASAP facility. It is critical to note that the specific features and data streams calculated and recorded on this bench are not directly comparable to those from the STB1 environment due to differences in instrumentation, software, and potential environmental factors.

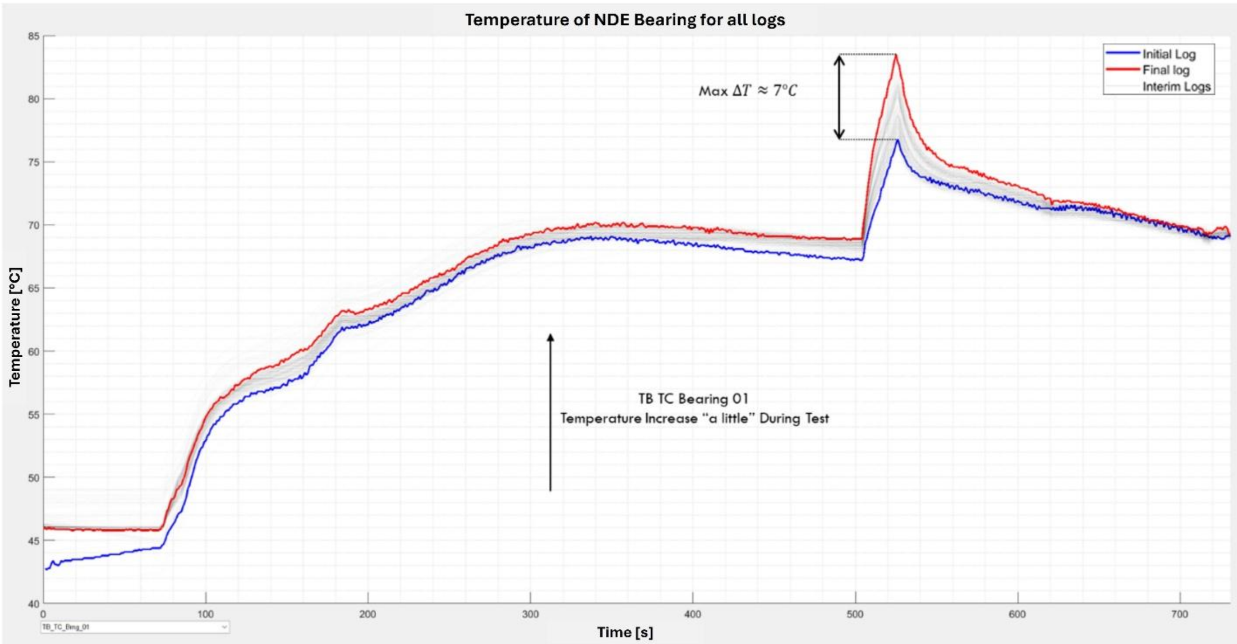


Figure 114 - Temperature in the bearing on the motor shaft NDE.

The primary indicator for this fluting fault was a clear and progressive increase in the bearing's operating temperature, as shown in Figure 114. The figure overlays the temperature profiles from the initial (blue), intermediate (gray), and final (red) logs of this test phase. While the absolute temperature increases between the first and last log is modest (approximately 7°C at its peak), the diagnostic power of this trend lies in the high reproducibility of the test cycle. The fact that the initial and final conditions of each cycle are identical allows for the confident conclusion that the consistent upward drift in temperature is a direct signature of the developing fault, rather than an artifact of operational or environmental variability. Following this diagnosis, the faulty bearing was replaced with a ceramic one to mitigate future electrical-induced wear.

Subsequent to this repair, the unit was moved back to the STB1 test bench for continued durability testing (Phase 2), during which a second, distinct fault began to develop. This new degradation mode was clearly captured by the Vibration Velocity RMS feature. Figure 115 focuses on a specific portion of the work cycle across all logs from this later test phase. The figure demonstrates that the final log exhibits a fundamentally different vibrational signature compared to the stable preceding logs. The amplitude of this vibration index shows a dramatic increase, approximately doubling in magnitude.

This significant step-change indicates that the second failure mode was mechanically more severe and more readily detectable through vibration analysis than the initial fluting fault.

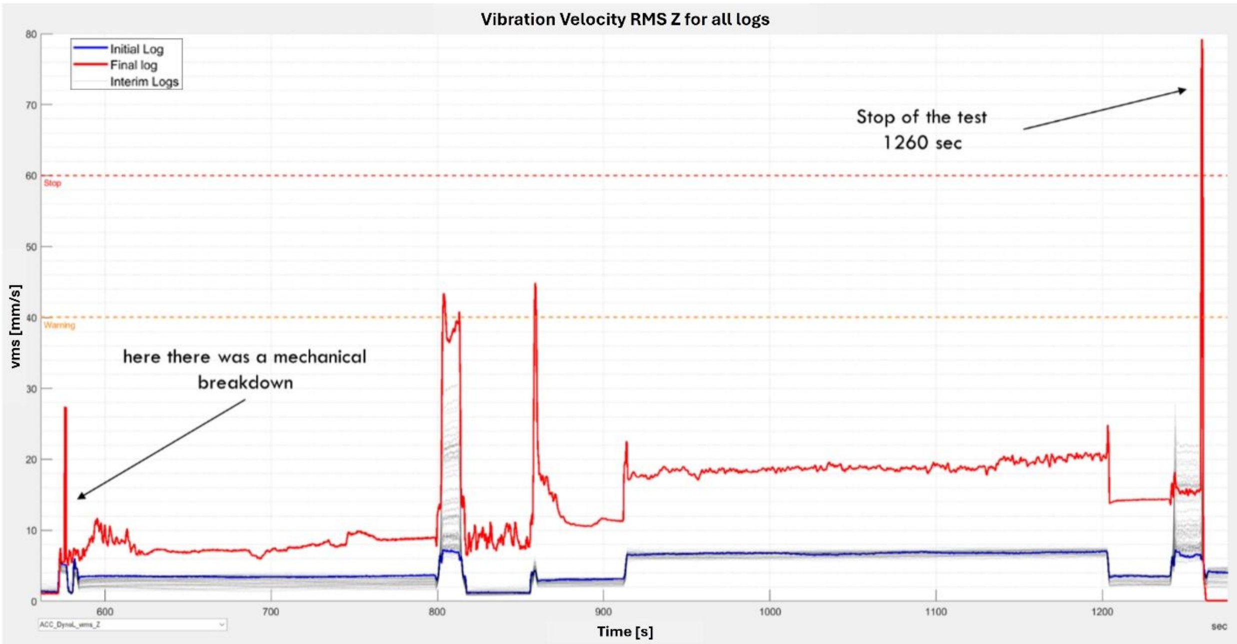


Figure 115 – Vibration Velocity RMS Z for all logs.

In Figure 116, on the other hand, it is possible to perform a thermal analysis of the fault and its thermal release during the last cycle before complete failure and termination of the test.

This dramatic increase in vibrational energy is corroborated by a corresponding thermal analysis of the system in the final cycle before test termination. Figure 116 overlays the gearbox oil temperature for all logs. In the central part of the cycle, highlighted between the two dotted lines, the temperature profile of the final log clearly detaches and rises above the stable thermal band established by all preceding logs.

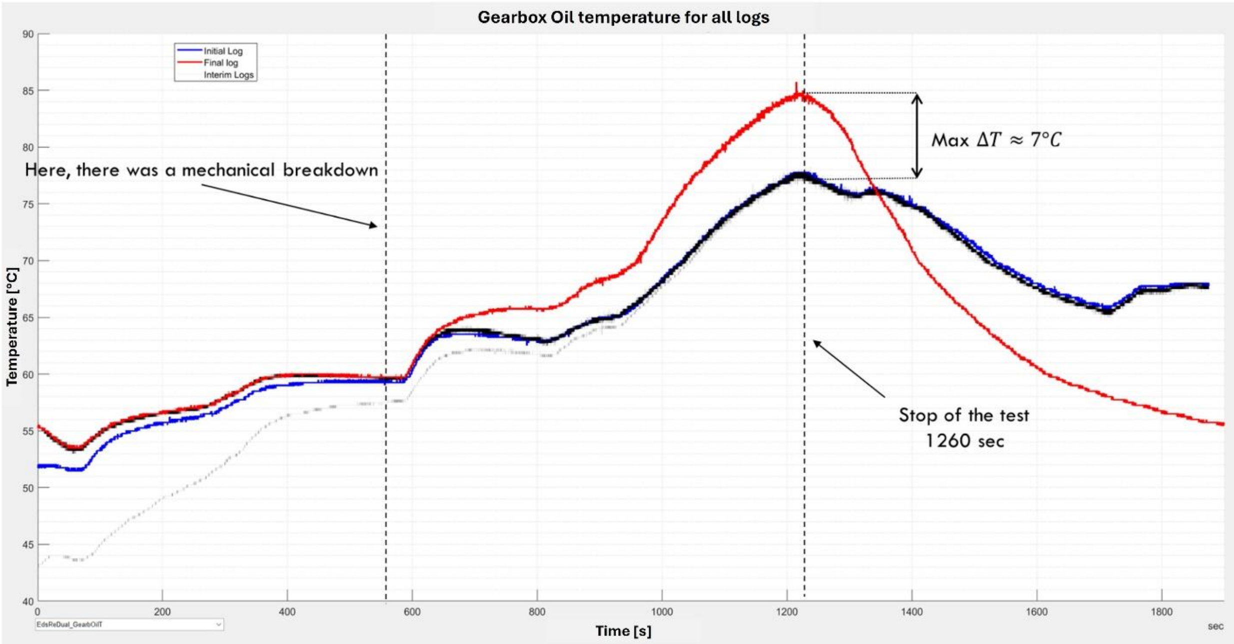


Figure 116 - Gearbox Oil Temperature for all the logs.

This temperature deviation allows for a first-principles estimation of the additional thermal power (ΔQ) generated by the fault. The following calculation is used to estimate the additional heat generated during breakage:

- density of the oil: $\rho \sim 870\ kg/m^3$
- specific heat of oil: $c_{Oil} = 1900\,J/kg * K$
- mass flow rate of oil: $\dot{m} = \frac{Flow[L/min]*\rho}{60} = \frac{22*0.870}{60} = 0.319\ kg/s$

Using the parameters described above, it is possible to estimate the thermal effect of the fault using the following equation:

$$\Delta Q_{est} = \dot{m} * \Delta T * c = 0.319 * 7 * 1900 = 4.243\ kW \qquad 7.1$$

A similar analysis can be performed on the water-glycol coolant circuit for the single conditioning unit. In that case, the parameters used for the estimation are these:

- density of Coolant: $\rho \sim 1065\ kg/m^3$
- specific heat of the coolant: $c_{50\%\ water-glicole} = 3300\,J/kg * K$
- mass flow rate of the Coolant: $\dot{m} = \frac{Flow[L/min]*\rho}{60} = \frac{27.23*1.065}{60} = 0.4824\frac{kg}{s}$

As shown in Figure 117, a temperature increase (ΔT) of approximately 1.5°C is observed.

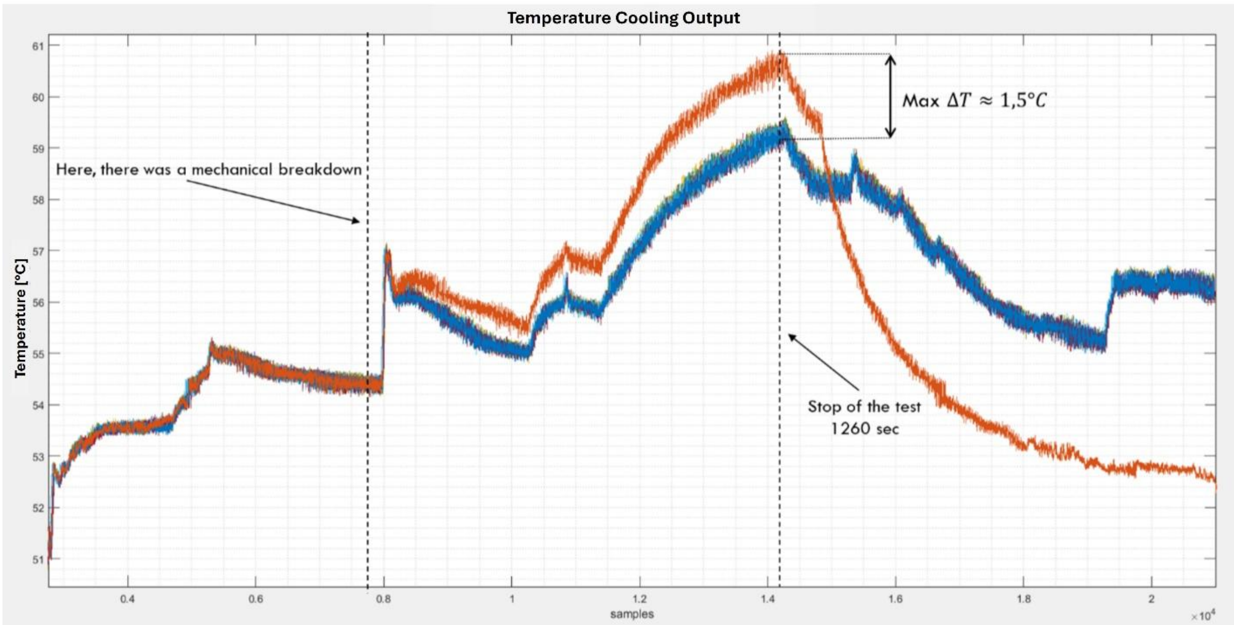


Figure 117 - Temperature Cooling Output.

Applying the same formula again, we obtain the additional heat exchanged through the cooling circuit:

$$\Delta Q_{est} = \dot{m} * \Delta T * c = 0.4824 * 1.5 * 3300 = 2.388\ kW \qquad 7.2$$

The subsequent teardown of the unit confirmed the severity of this second fault. As shown in Figure 118, the root cause was a complete fracture of the right-side intermediate shaft at its connection with the crown gear. The catastrophic nature of this failure is evident from the extensive collateral damage, including completely worn gear teeth on both the crown and pinion, and fractured and broken bearing covers. This physical evidence is consistent with the significant release of vibrational and thermal energy detected in the final cycles of the test.

Figure 118 - Second type of fault, right intermediate shaft fractured, DVP_SC13T.

7.4.3.2 Degradation trends: comparison between vibration acceleration and velocity indices

This section presents a detailed analysis of the degradation trends observed during the final phase (Phase 3) of the durability test for unit DVP_SC13T, which culminated in the fracture of the left intermediate shaft.

The primary condition indicator used for this analysis is the Vibration Velocity RMS, extracted from the six accelerometer channels (X, Y, and Z axes for both the left and right sides of the powertrain). Figure 119 presents the time-series evolution of this index for all six channels across the full duration of the 64 cycles in this test phase. While all axes on the left side exceed the warning threshold in the final cycle, confirming the location of the failure, the left Z-axis channel is particularly informative. It exhibits a clear and progressive upward trend over the final 20 to 25 cycles, providing a distinct signature of the developing degradation.

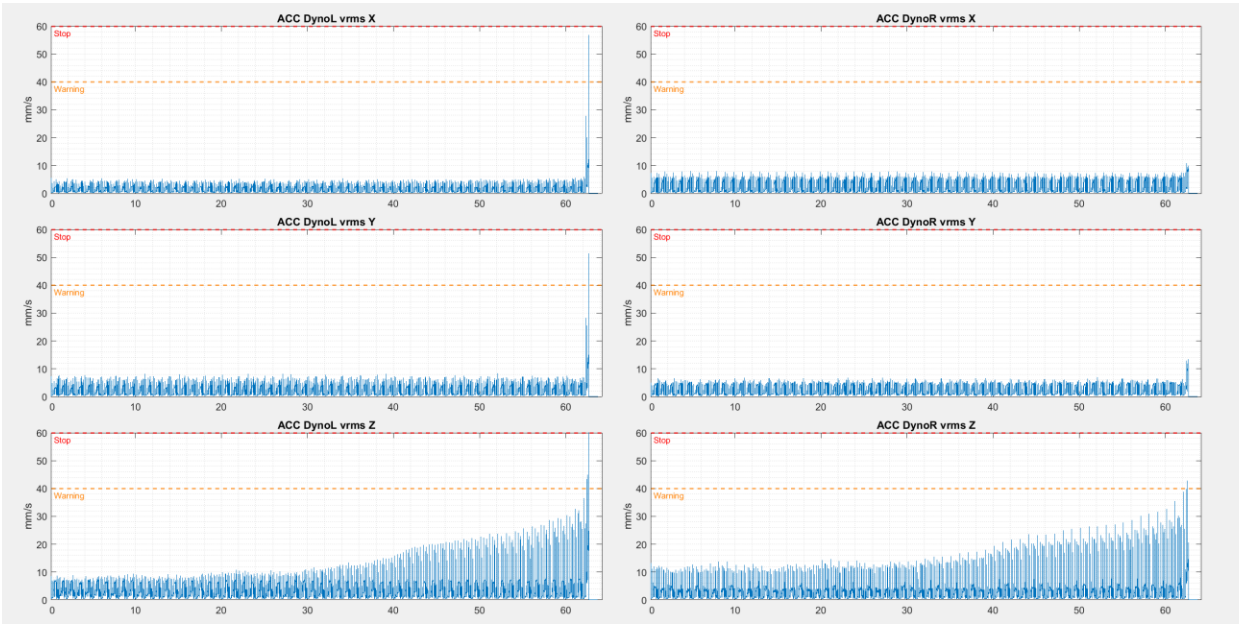


Figure 119 - Trend of all Vibration signal for all the logs of (Phase 3) of Accelerated durability test of DVP_SC13.

To more closely inspect the index's behavior during the most critical portion of the operational cycle, Figure 120 presents the same data but focuses the visualization on a narrow 20-second window (from 800 to 820 seconds) for each of the 64 cycles. This focused view provides a clearer representation of the index's magnitude and stability within this specific high-load segment, making the cycle-to-cycle changes more apparent.

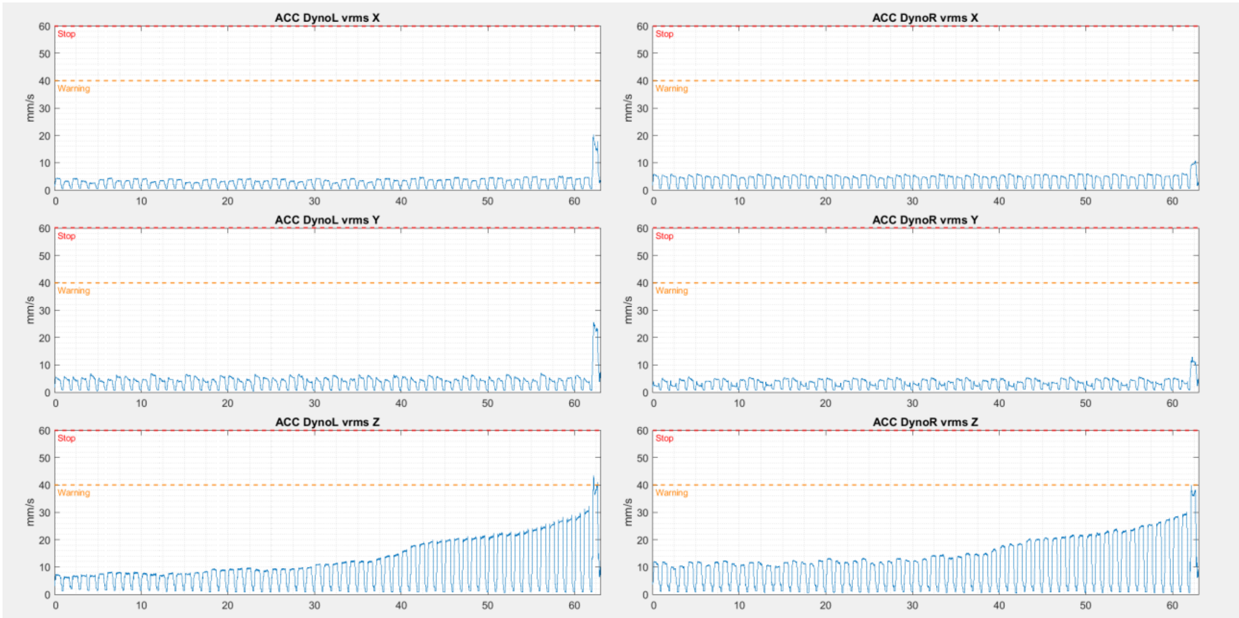


Figure 120 - Trend of all Vibration signal for all the logs of (Phase 3) of Accelerated durability test of DVP_SC13. Range of visualization per every cycles of [800 – 820] sec, for a better focus on the fault.

While the Vibration Velocity RMS is the standard index for the current condition monitoring strategy, this analysis also sought to evaluate an alternative feature based on acceleration. **Figure 121** provides the basis for this comparison. The main plot shows the Vibration Velocity RMS trend, which is the primary synthetic index derived from the accelerometer signal. The inset box at the top right displays the trend of a newly computed Vibration Acceleration Index, which was calculated using a 0.6-second window. The subsequent analysis will compare the sensitivity and effectiveness of these two different indices for tracking the observed fault.

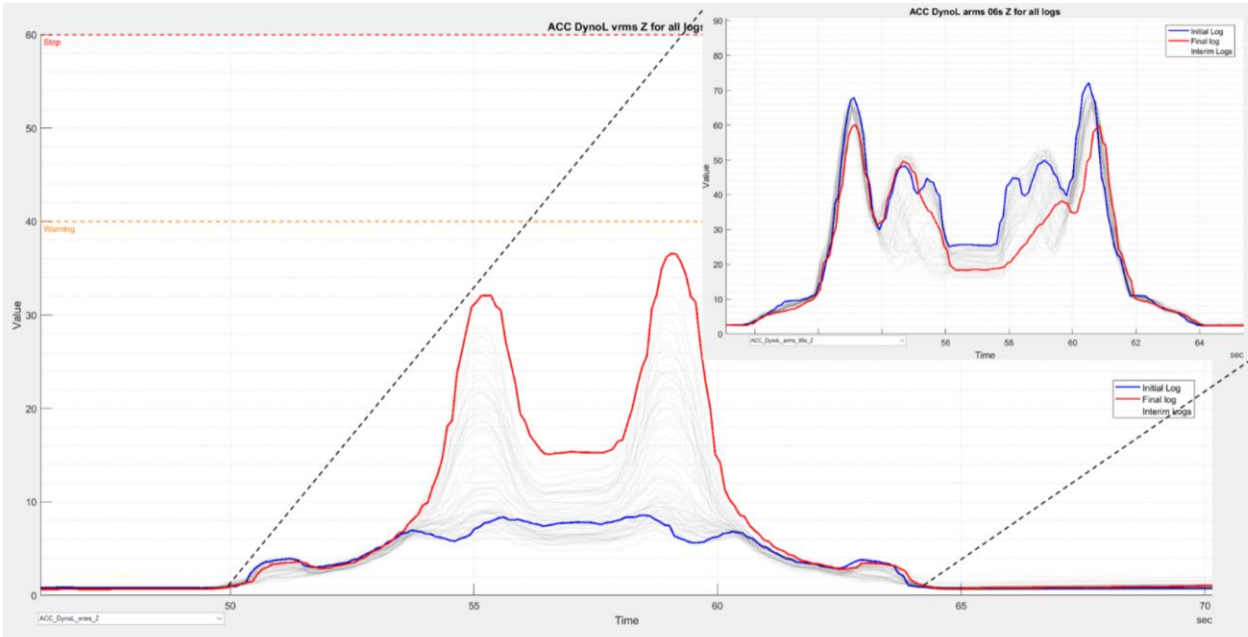


Figure 121 - ACC DynoL Z, comparison vrms. arms 06s.

Given that this phase of the durability test culminates in the catastrophic fracture of the **left intermediate shaft**, a direct comparison can be made between the diagnostic effectiveness of the two synthetic indices shown in Figure 121.

The trend for the Vibration Velocity RMS index is unambiguous. As the test approaches its final cycle, where the failure occurs, the index exhibits a clear and dramatic increase in magnitude, providing a definitive signature of the severe mechanical fault.

In contrast, the Vibration Acceleration Index does not show a similar correlation with the developing fault. Its trend remains relatively flat and does not provide a clear indication of the impending failure.

The conclusion from this direct comparison is that for this specific application and failure mode, the Vibration Velocity RMS is a significantly more sensitive and reliable condition indicator than the Vibration Acceleration Index. This finding underscores the importance of selecting the appropriate physical parameter (velocity vs. acceleration) and feature engineering strategy for a given diagnostic target. This is because, as mentioned in the introduction, the phenomenon is best observed at intermediate frequencies (see Chapter 4.1).

7.4.3.3 Misalignment Detection in Back-to-Back test Configuration

In addition to the durability tests conducted on the STB1 bench, a separate accelerated durability test on a CEDAR unit was performed at the ASAP facility using a back-to-back (B2B) test configuration. In this setup, two powertrain units are mechanically coupled together, with one acting as the motor and the other as the generator (load), thus eliminating the need for an external dynamometer.

Analysis of the raw vibration data from this B2B test, acquired at a sampling frequency of 20 kHz, revealed a clear and significant setup-related issue. Figure 122 presents the order spectrum from this test.

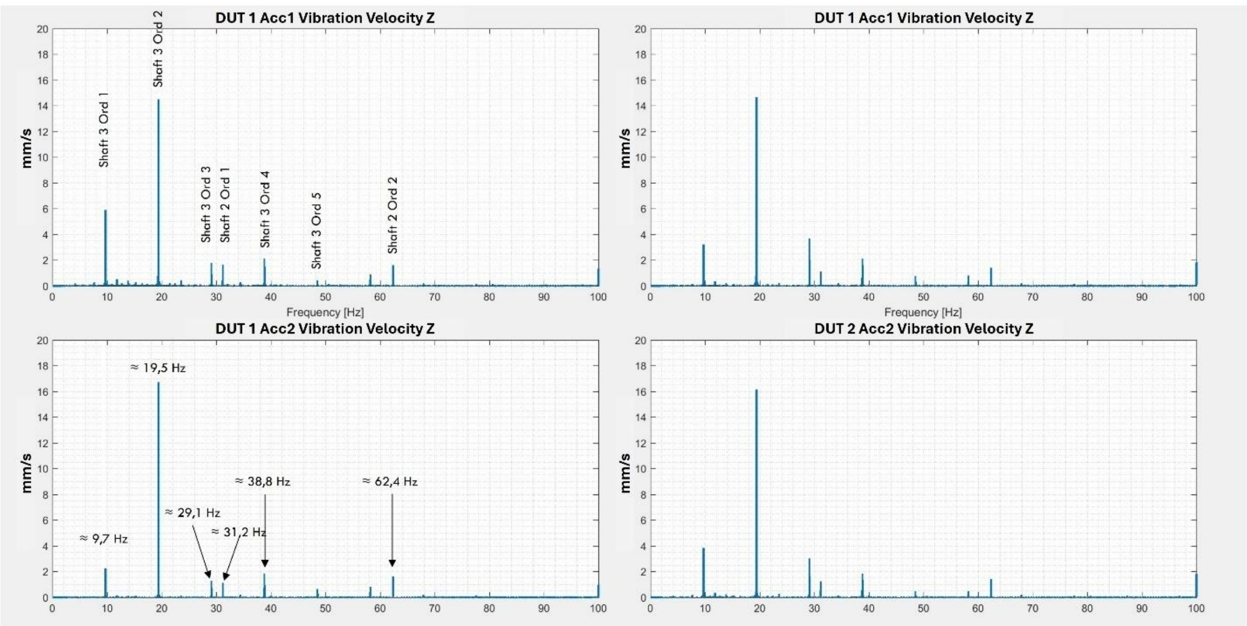


Figure 122 - Vibration Velocity Order Analysis for the Back-to-back test configuration done in ASAP.

The spectrum is dominated by a very high amplitude peak at the **2x order** (twice the fundamental rotational frequency). This is accompanied by significant, though smaller, peaks at the **1x and 3x orders**. This specific spectral signature—a dominant 2x order with supporting 1x and 3x orders—is a classic and well-documented indicator of **parallel misalignment** in rotating machinery [55].

Misalignment occurs when the centerlines of two coupled shafts are not collinear. It is typically categorized into two types, each with a distinct vibrational signature:

- **Parallel Misalignment:** The shaft centerlines are parallel but offset. This condition generates strong radial vibrations, and its characteristic frequency signature is a dominant peak at the 2x order.
- **Angular Misalignment:** The shaft centerlines meet at an angle. This condition typically excites a strong axial vibration at the 1x order, though 2x and other harmonics can also be present.

The clear dominance of the 2x order in the observed spectrum strongly indicates that the B2B test configuration suffered from significant parallel misalignment between the two coupled units. This finding is critical, as such a condition introduces severe, unintended radial loads on the bearings and shafts, which can invalidate the results of a durability test by causing premature failures that are not representative of the component's true lifespan under normal operating conditions.

7.4.3.4 Abnormal Vibrational Trend During Durability Test

During the analysis of the accelerated durability test for the unit designated STNDRD-02, a highly unusual and distinct vibrational pattern was observed. Figure 123 - Trends for all the Vibration velocity index of unit STNDRD-02 during the durability test. presents the trends of the Vibration Velocity RMS indices for all six accelerometer channels. The visualization is specifically focused on a 16-second window (from 729 to 745 seconds) within each of the approximately 15-minute work cycles performed throughout the durability test.

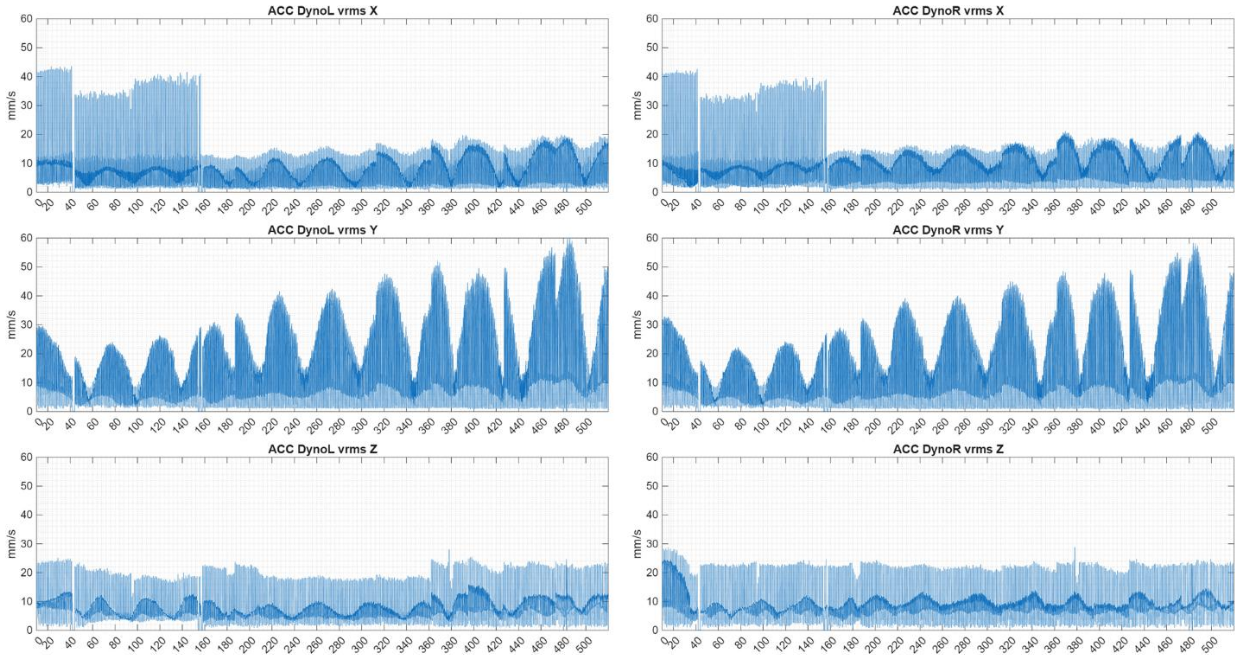


Figure 123 - Trends for all the Vibration velocity index of unit STNDRD-02 during the durability test.

The most striking feature of this data is the prominent cyclical pattern visible in the Y-axis channels. This cyclical variation in the vibration level has a long-term period of approximately 45 work cycles. When analyzing only the trend of a synthetic index like RMS, such a slow, highly regular oscillation could initially suggest a potential issue with the data acquisition chain or a sensor artifact, rather than a true physical phenomenon.

However, despite this unusual oscillation, a clear and consistent long-term upward trend in the overall vibration level is still evident, particularly on the Y-axis. This underlying monotonic increase suggests a genuine, progressive degradation occurring within the system, even as it is modulated by the unexplained cyclical pattern. This case highlights a complex diagnostic scenario where multiple temporal patterns—a slow oscillation and a long-term degradation trend—are superimposed, requiring careful analysis to deconvolve.

7.5 Conclusion

The work presented in this chapter demonstrates the successful development and validation of a modular Condition Monitoring (CM) framework applied across three distinct electric powertrain test campaigns: Accelerated Durability, End-of-Line, and Gearbox Overload. A key outcome of this research was the successful integration of the framework onto a real-time test bench for the durability and EOL tests, while the overload test was analyzed in post-processing. The developed system successfully streams vibration data, computes health indicators, and is capable of raising alarms within the required latency, thereby demonstrating the end-to-end viability of the proposed CM strategy.

However, it is important to note that the current implementation is primarily limited to vibration-based diagnostics. While this approach has proven effective for detecting mechanical degradation, it does not capture faults that manifest predominantly in the torsional dynamics of the powertrain or as electrical anomalies.

Therefore, future work should focus on expanding the framework's sensing modalities to create a more holistic diagnostic system. Key areas for extension include:

1. **Torsional Vibration Analysis:** Deriving synchronous orders from speed or torque ripple measurements to detect gear mesh irregularities and shaft torsion fatigue.
2. **Electrical Signature Analysis (ESA):** Integrating current and voltage signature analysis to identify faults such as rotor eccentricity, inverter switching anomalies, and winding insulation degradation.

The integration of these extensions, contingent upon the availability of high-fidelity torque and electrical measurements, would significantly enrich the CM suite and enable a more comprehensive Prognostic and Health Management (PHM) strategy for electric powertrains.

8 Conclusions

This thesis has presented the development and validation of a systematic framework for Early Fault Diagnosis through the application of Artificial Intelligence techniques to Condition Monitoring data. The primary objective was to define a hybrid methodology, integrating the physical analysis of signals with unsupervised machine learning models, to create a diagnostic system that is robust, adaptable, and applicable to diverse industrial contexts.

Summary of Main Contributions

The original contributions of this research can be summarized in the following points:

1 Development of a Hybrid Framework: A multi-stage methodology was defined and validated, assigning strategic roles to different sensing modalities. Specifically, a novel approach was introduced that uses online oil debris analysis to "supervise" the definition of a healthy operational baseline, significantly increasing the robustness of subsequent unsupervised anomaly detection models.

2 Creation of Multi-level Health Indices: The framework is not limited to a binary classification but generates continuous health indices: a General Health Index (GHI) based on vibrational data for an overall assessment, and component-level Specific Health Indices (SHIs) based on high-frequency ultrasonic data for targeted diagnostics.

3 Experimental Validation on Diverse Industrial Systems: The framework's effectiveness has been demonstrated on two complex industrial case studies: a run-to-failure durability test on a heavy-duty mechanical gearbox, and a series of test campaigns on electric powertrains for automotive applications.

Comparative Discussion and Unitary Vision

The application of the framework to such different contexts (a slow mechanical gearbox and fast electric powertrains) has highlighted its flexibility but also underscored its challenges.

- In the gearbox case study (Chapter 6), where a complete lifecycle until failure was available, the framework demonstrated impressive predictive capability, providing a lead time of over 750 hours. The analysis of ultrasonic data proved to be significantly superior to traditional vibration analysis for the early diagnosis of the specific bearing fault.

- In the electric powertrain case studies (Chapter 7), characterized by accelerated tests and heterogeneous data, the framework was adapted to operate in a context of real-time diagnostics and End-of-Line validation. These applications highlighted the crucial importance of data quality and test design ("data-first approach"), demonstrating how a lack of controlled boundary conditions or adequate sampling frequencies can preclude analysis.

This dual validation shows that, while the methodology is robust, its practical effectiveness is intrinsically linked to the quality and purpose for which the data is acquired.

Limitations and Future Work

Despite the positive results, this work presents certain limitations. The gearbox validation focused on a single failure mode (bearing failure) in a controlled laboratory environment. Similarly, the analyses on the powertrains were limited by the available data.

The future developments of this research will focus on three main directions:

1. Extension to other failure modes: Validating the framework on a broader spectrum of mechanical (e.g., gear pitting, misalignment) and electrical faults.
2. In-field Validation: Implementing and testing the system on machinery operating in real industrial conditions to assess its robustness to operational noise and the variability of working conditions.
3. Exploration of Advanced Models: Investigating the potential of deep learning models, such as autoencoders, within this same multi-index framework to further improve the feature abstraction capability and diagnostic sensitivity.

In conclusion, this thesis has contributed to a practical, validated framework for predictive maintenance, demonstrating an effective path for translating theoretical AI models into concrete solutions for the reliability of mechanical systems.

References

[1] waoo23, “Types of maintenance: a guide to choosing the right solution,” Picomto. Accessed: Oct. 08, 2025. [Online]. Available: https://www.picomto.com/en/types-of-maintenance-industrial-strategy/

[2] C. E. Ebeling, *An Introduction to Reliability and Maintainability Engineering: Third Edition*. Waveland Press, 2019.

[3] “practical-reliability-engineering.pdf.” Accessed: Oct. 24, 2025. [Online]. Available: https://vibadirect.com/koolinks/documents/practical-reliability-engineering.pdf#:~:text=The%20combined%20effect%20generates%20the,similar%20to%20actuarial%20statistical%20models

[4] “8.1.2.4. ‘Bathtub’ curve.” Accessed: Oct. 24, 2025. [Online]. Available: https://www.itl.nist.gov/div898/handbook/apr/section1/apr124.htm

[5] G. Klutke, P. C. Kiessler, and M. A. Wortman, “A critical look at the bathtub curve,” *IEEE Trans. Reliab.*, vol. 52, no. 1, pp. 125–129, Mar. 2003, doi: 10.1109/TR.2002.804492.

[6] wmargolin, “Why the Original Model of the P-F Curve Is the Correct Model,” Aladon. Accessed: Oct. 23, 2025. [Online]. Available: https://www.aladon.com/why-the-original-model-of-the-p-f-curve-is-the-correct-model/

[7] CEN, “Maintenance — Maintenance Terminology,” European Committee for Standardization, Brussels, Belgium, Standard EN 13306:2017, 2017.

[8] admin, “What are the 3 types of industrial maintenance? -.” Accessed: Oct. 08, 2025. [Online]. Available: https://www.amiraltechnologies.com/en/news/blog/3-types-of-industrial-maintenance/

[9] “Types of Maintenance: The Complete Guide,” TRACTIAN. Accessed: Oct. 08, 2025. [Online]. Available: https://tractian.com/en/blog/types-of-maintenance

[10] “predictive-maintenance-ebook-part1.pdf.” Accessed: Oct. 27, 2025. [Online]. Available: https://it.mathworks.com/content/dam/mathworks/ebook/predictive-maintenance-ebook-part1.pdf

[11] P. Tchakoua, R. Wamkeue, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, “New trends and future challenges for wind turbines condition monitoring,” in *2013 International Conference on Control, Automation and Information Sciences (ICCAIS)*, Nha Trang, Vietnam: IEEE, Nov. 2013, pp. 238–245. doi: 10.1109/ICCAIS.2013.6720561.

[12] “What is Prescriptive Maintenance?,” Limble CMMS. Accessed: Oct. 23, 2025. [Online]. Available: https://limblecmms.com/strategies/prescriptive-maintenance/

[13] “The Evolution of Maintenance: Condition Monitoring and Predictive Analytics - Artesis.” Accessed: Oct. 08, 2025. [Online]. Available: https://artesis.com/the-evolution-of-maintenance-condition-monitoring-and-predictive-analytics/

[14] “What is Condition Monitoring? A Guide to Predictive Maintenance | KEYENCE America.” Accessed: Oct. 08, 2025. [Online]. Available: https://www.keyence.com/products/daq/data-loggers/resources/data-logger-resources/what-is-condition-monitoring-a-guide-to-predictive-maintenance.jsp

[15] Engineering Presentations, *What is Predictive Maintenance? - Introduction to Condition Monitoring & Condition Based Maintenance*, (2023). Accessed: Oct. 08, 2025. [Online Video]. Available: https://www.youtube.com/watch?v=pocSio2ExNo

[16] A. Sofiany, "Leveraging Machine Learning to Enhance Anomaly Detection in Rotating Machinery," vol. 65, Sept. 2024, Accessed: Oct. 08, 2025. [Online]. Available: https://www.turbomachinerymag.com/view/leveraging-machine-learning-to-enhance-anomaly-detection-in-rotating-machinery

[17] "Different Types of Industrial Maintenance | HARtech," https://www.har-tech.com/en/. Accessed: Oct. 08, 2025. [Online]. Available: https://www.har-tech.com/en/the-main-types-of-industrial-maintenance/

[18] "What is Predictive Maintenance? | IBM." Accessed: Oct. 08, 2025. [Online]. Available: https://www.ibm.com/think/topics/predictive-maintenance

[19] "Ultrasonic Testing (UT): A Versatile Method for NDT Inspections." Accessed: Oct. 23, 2025. [Online]. Available: https://www.asnt.org/what-is-nondestructive-testing/methods/ultrasonic-testing

[20] R. Tabibian, "7 Benefits of Condition-Based Maintenance," WorkTrek. Accessed: Oct. 24, 2025. [Online]. Available: https://worktrek.com/blog/condition-based-maintenance-benefits/

[21] "Predictive Maintenance Market Size, Trends, Share & Research Report 2030," Mordor Intelligence. Accessed: Oct. 24, 2025. [Online]. Available: https://www.mordorintelligence.com/industry-reports/predictive-maintenance-market

[22] W. Zhang, H. Li, Y. Li, H. Liu, Y. Chen, and X. Ding, "Application of deep learning algorithms in geotechnical engineering: a short critical review," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5633–5673, Dec. 2021, doi: 10.1007/s10462-021-09967-1.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[24] S. Bag, "Which Machine Learning Algorithm Should You Use By Problem Type?," Analytics Vidhya. Accessed: Oct. 27, 2025. [Online]. Available: https://medium.com/analytics-vidhya/which-machine-learning-algorithm-should-you-use-by-problem-type-a53967326566

[25] "Che cos'è la Predictive Maintenance?" Accessed: Oct. 27, 2025. [Online]. Available: https://it.mathworks.com/discovery/predictive-maintenance.html

[26] R. Srivastava and ksolves Team, "The Power of Predictive Maintenance with Machine Learning: Key Benefits," Ksolves. Accessed: Oct. 24, 2025. [Online]. Available: https://www.ksolves.com/blog/artificial-intelligence/leveraging-machine-learning-for-predictive-maintenance-key-benefits

[27] R. Srivastava and ksolves Team, "The Power of Predictive Maintenance with Machine Learning: Key Benefits," Ksolves. Accessed: Oct. 27, 2025. [Online]. Available: https://www.ksolves.com/blog/artificial-intelligence/leveraging-machine-learning-for-predictive-maintenance-key-benefits

[28] D. Grunova, V. Bakratsi, E. Vrochidou, and G. A. Papakostas, "Machine Learning for Anomaly Detection in Industrial Environments," *Eng. Proc.*, vol. 70, no. 1, p. 25, 2024, doi: 10.3390/engproc2024070025.

[29] “Machine Learning for Anomaly Detection in Industrial Processes,” iCoSys. Accessed: Oct. 08, 2025. [Online]. Available: https://icosys.ch/machine-learning-for-anomaly-detection-in-industrial-processes

[30] “2.7. Novelty and Outlier Detection,” scikit-learn. Accessed: Oct. 23, 2025. [Online]. Available: https://scikit-learn/stable/modules/outlier_detection.html

[31] Z. S. Phd, “Difference between Outlier Detection and Novelty Detection,” Medium. Accessed: Oct. 23, 2025. [Online]. Available: https://medium.com/@injure21/difference-between-outlier-detection-and-novelty-detection-f21c21ed0962

[32] C. C. Aggarwal, *Outlier Analysis*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-47578-3.

[33] “Lecture 13: Dual Formulation of Support Vector Machines — Applied ML.” Accessed: Oct. 23, 2025. [Online]. Available: https://kuleshov-group.github.io/aml-book/contents/lecture13-svm-dual.html

[34] Y. Zhao, Z. Nasrullah, and Z. Li, “PyOD: A Python Toolbox for Scalable Outlier Detection”.

[35] V. Spruyt, “The Curse of Dimensionality in Classification,” Computer vision for dummies. Accessed: Oct. 27, 2025. [Online]. Available: https://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/

[36] V. Spruyt, “The Curse of Dimensionality in Classification,” Computer vision for dummies. Accessed: Oct. 24, 2025. [Online]. Available: https://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/

[37] “Viualizing_data_using_t-SNE.”

[38] “Introduction to t-SNE: Nonlinear Dimensionality Reduction and Data Visualization.” Accessed: Oct. 27, 2025. [Online]. Available: https://www.datacamp.com/tutorial/introduction-t-sne

[39] “Understanding Boxplots,” KDnuggets. Accessed: Oct. 27, 2025. [Online]. Available: https://www.kdnuggets.com/understanding-boxplots

[40] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, “Outlier Detection for Temporal Data: A Survey,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2250–2267, Sept. 2014, doi: 10.1109/TKDE.2013.184.

[41] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, July 2009, doi: 10.1145/1541880.1541882.

[42] “STL-A Seasonal-Trend Decomposition Procedure based on Loess.”

[43] C. C, “Anomaly detection for Time Series Analysis,” Medium. Accessed: Oct. 27, 2025. [Online]. Available: https://medium.com/aimonks/anomaly-detection-for-time-series-analysis-eeecd6282f53

[44] R. Pramoditha, “Why is Feature Scaling Important in Machine Learning? Discussing 6 Feature Scaling Techniques,” Towards Data Science. Accessed: Oct. 24, 2025. [Online]. Available: https://towardsdatascience.com/why-is-feature-scaling-important-in-machine-learning-discussing-6-feature-scaling-techniques-2773bda5be30/

[45] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy: IEEE, Dec. 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.

[46] S. Hariri, M. C. Kind, and R. J. Brunner, “Extended Isolation Forest,” *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1479–1489, Apr. 2021, doi: 10.1109/TKDE.2019.2947676.

[47] A. Ph.D, “Isolation Forest: Unveiling the Power of Unsupervised Anomaly Detection,” Medium. Accessed: Oct. 27, 2025. [Online]. Available: https://medium.com/@AyZek_PhD/isolation-forest-unveiling-the-power-of-unsupervised-anomaly-detection-b7c6f7130c42

[48] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying Density-Based Local Outliers”.

[49] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution,” *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, July 2001, doi: 10.1162/089976601750264965.

[50] H. J. Shin, D.-H. Eom, and S.-S. Kim, “One-class support vector machines—an application in machine fault detection and classification,” *Comput. Ind. Eng.*, vol. 48, no. 2, pp. 395–408, Mar. 2005, doi: 10.1016/j.cie.2005.01.009.

[51] P. C. Mahalanobis, “On the Generalized Distance in Statistics,” *Proc. Natl. Inst. Sci. India*, vol. 2, no. 1, pp. 49–55, 1936.

[52] S. Cansiz, “Multivariate Outlier Detection in Python,” TDS Archive. Accessed: Oct. 27, 2025. [Online]. Available: https://medium.com/data-science/multivariate-outlier-detection-in-python-e946cfc843b3

[53] R. B. Randall, “VIBRATION-BASED CONDITION MONITORING,” *SPH P*, 2010.

[54] P. Girdhar and C. Scheffer, *Practical Machinery Vibration Analysis and Predictive Maintenance*. Linacre House, Jordan Hill, Oxford: Newnes, 2004.

[55] V. Wowk, *Machinery Vibration: Measurement and Analysis*. New York: McGraw-Hill Professional, 2000.

[56] Ozi, “Can we represent Simple Harmonic function as triangular waves?,” Physics Stack Exchange. Accessed: Oct. 27, 2025. [Online]. Available: https://physics.stackexchange.com/q/133330

[57] A. Papadopoulos, “PIPING EVALUATION OF FLOW INDUCED VIBRATIONS USING HARMONIC LOADS”.

[58] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.

[59] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, “An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data,” *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016, doi: 10.1109/TIE.2016.2519325.

[60] “Why Measuring in the Time Domain and Frequency Domain Is the Same, but Not - Technical Support Knowledge Center Open.” Accessed: Oct. 27, 2025. [Online]. Available: https://docs.keysight.com/kkbopen/why-measuring-in-the-time-domain-and-frequency-domain-is-the-same-but-not-603167255.html?__cf_chl_tk=4RkxUwwkTRu7NMQY4pDqtrN0WZu.OVniHPylyYzltIw-1761440349-1.0.1.1-YpNg9JMAc0Xb2MmdzmHEIKp.X_bO3QGNAvMbkEuAAxE

[61] J. Antoni, “The spectral kurtosis: a useful tool for characterising non-stationary signals,” *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 282–307, Feb. 2006, doi: 10.1016/j.ymssp.2004.09.001.
[62] L. Cohen, *Time-Frequency Analysis: Theory and Applications*. Upper Saddle River, NJ: Prentice Hall, 1995.
[63] Z. K. Peng and F. L. Chu, “Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography,” *Mech. Syst. Signal Process.*, vol. 18, no. 2, pp. 199–221, Mar. 2004, doi: 10.1016/S0888-3270(03)00075-X.
[64] ISO, “Mechanical Vibration — Measurement and Evaluation of Machine Vibration — Part 3: Industrial Machinery,” International Organization for Standardization, Geneva, Switzerland, Standard ISO 20816-3:2022, 2022.
[65] C. Scheffer and P. Girdhar, *Practical Machinery Vibration Analysis and Predictive Maintenance*. Oxford, UK: Newnes, 2004. [Online]. Available: https://www.sciencedirect.com/book/9780750662758/practical-machinery-vibration-analysis-and-predictive-maintenance
[66] D. Mba and R. B. Rao, “Development of acoustic emission technology for condition monitoring and diagnosis of rotating mechanical equipment—a review,” *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1537–1571, 2006, doi: 10.1016/j.ymssp.2005.09.005.
[67] Mobius Institute, *A Brief Introduction to Heterodyning*, (2022). Accessed: Oct. 27, 2025. [Online Video]. Available: https://www.youtube.com/watch?v=DLdjHJPhFis
[68] “CONMONSense_4-20mA_Contact_Dynamic_Datasheet.”
[69] J. C. Fitch, *The Practical Handbook of Machinery Lubrication*, 4th ed. Tulsa, OK: Noria Corporation, 2018.
[70] “Vibration Signals from Rotating and Reciprocating Machines,” in *Vibration-based Condition Monitoring*, John Wiley & Sons, Ltd, 2011, pp. 25–62. doi: 10.1002/9780470977668.ch2.
[71] S. Nandi, H. A. Toliyat, and X. Li, “Condition Monitoring and Fault Diagnosis of Electrical Motors—A Review,” *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005, doi: 10.1109/TEC.2005.847955.
[72] B. Wang and Y. Wu, “An Oil Debris Analysis Method of Gearbox Condition Monitoring Based on an Improved Multi-Variable Grey Prediction Model,” *Machines*, vol. 13, no. 8, p. 664, July 2025, doi: 10.3390/machines13080664.
[73] M. H. A. Soliman, *Machinery Oil Analysis & Condition Monitoring: A Practical Guide to Sampling and Analyzing Oil to Improve Equipment Reliability*. 2020.
[74] R. Jiang, “A new bathtub curve model with a finite support,” *Reliab. Eng. Syst. Saf.*, vol. 119, pp. 44–51, Nov. 2013, doi: 10.1016/j.ress.2013.05.019.
[75] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. Tameghe, and G. Ekemb, “Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges,” *Energies*, vol. 7, no. 4, pp. 2595–2630, Apr. 2014, doi: 10.3390/en7042595.
[76] A. C. S. V. D. Negreiros, I. D. Lins, M. J. D. C. Moura, and E. L. Droguett, “Reliability data analysis of systems in the wear-out phase using a (corrected) q-Exponential

likelihood," *Reliab. Eng. Syst. Saf.*, vol. 197, p. 106787, May 2020, doi: 10.1016/j.ress.2019.106787.
[77] Y. Yang, "A Signal Theoretic Approach for Envelope Analysis of Real-Valued Signals," *IEEE Access*, vol. 5, pp. 5623–5630, 2017, doi: 10.1109/ACCESS.2017.2688467.
[78] M. A. Goodman and W. Bishop, "System and method for heterodyning an ultrasonic signal," US6707762B1, Mar. 16, 2004 Accessed: Aug. 26, 2025. [Online]. Available: https://patents.google.com/patent/US6707762B1/en
[79] "Ultrasound Monitoring for Predictive Maintenance - SenseGrow." Accessed: Aug. 26, 2025. [Online]. Available: https://www.sensegrow.com/blog/iot-solutions/ultrasound-monitoring-for-predictive-maintenance
[80] A. Bandes, "Ultrasound Condition Monitoring".
[81] "Condition Monitoring through Diagnosis of Anomalies — an LSTM-based Unsupervised ML Approach | by Sabtain Ahmad | Medium." Accessed: Sept. 11, 2025. [Online]. Available: https://sabtainahmadml.medium.com/condition-monitoring-through-diagnosis-of-anomalies-lstm-based-unsupervised-ml-approach-5f0565735dff
[82] S. Ahmad, K. Styp-Rekowski, S. Nedelkoski, and O. Kao, "Autoencoder-based Condition Monitoring and Anomaly Detection Method for Rotating Machines," Jan. 27, 2021, *arXiv*: arXiv:2101.11539. doi: 10.48550/arXiv.2101.11539.
[83] M. Huang, "Anomaly Detection for Condition Monitoring in Robot Systems".
[84] T. Chen, X. Liu, B. Xia, W. Wang, and Y. Lai, "Unsupervised Anomaly Detection of Industrial Robots Using Sliding-Window Convolutional Variational Autoencoder," *IEEE Access*, vol. 8, pp. 47072–47081, 2020, doi: 10.1109/ACCESS.2020.2977892.
[85] O. Kullu and E. Cinar, "A Deep-Learning-Based Multi-Modal Sensor Fusion Approach for Detection of Equipment Faults," *Machines*, vol. 10, no. 11, p. 1105, Nov. 2022, doi: 10.3390/machines10111105.
[86] E. Cinar, "A Sensor Fusion Method Using Transfer Learning Models for Equipment Condition Monitoring," *Sensors*, vol. 22, no. 18, p. 6791, Sept. 2022, doi: 10.3390/s22186791.
[87] Y. Wang, D. Li, L. Li, R. Sun, and S. Wang, "A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model," *Heliyon*, vol. 10, no. 15, p. e35407, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35407.
[88] IEC, "Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems — Part 1: General Requirements," International Electrotechnical Commission, Geneva, Switzerland, Standard IEC 61508-1:2010, 2010.
[89] ISO, "Road Vehicles — Functional Safety — Part 1: Vocabulary," International Organization for Standardization, Geneva, Switzerland, Standard ISO 26262-1:2018, 2018.
[90] ISO, "Condition Monitoring and Diagnostics of Machines — General Guidelines," International Organization for Standardization, Geneva, Switzerland, Standard ISO 17359:2018, 2018.
[91] K. Medjaher, D. A. Tobon-Mejia, and N. Zerhouni, "Remaining Useful Life Estimation of Critical Components with Application to Bearings," *IEEE Trans. Reliab.*, vol. 61, no. 2, pp. 292–302, June 2012, doi: 10.1109/TR.2012.2194175.

[92] A. Widodo and B.-S. Yang, “Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis,” *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.007.
[93] E. Zio, “Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice,” *Reliab. Eng. Syst. Saf.*, vol. 218, p. 108119, Feb. 2022, doi: 10.1016/j.ress.2021.108119.
[94] Y. Lei, Z. He, Y. Zi, and Q. Hu, “Fault Diagnosis of Rotating Machinery Based on Multiple ANFIS Combination with GAs,” *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2280–2294, July 2007, doi: 10.1016/j.ymssp.2006.10.003.
[95] Y. Wu, Z. Cui, W. Yan, H. Huang, and W. Ding, “Accelerated Fatigue Test for Electric Vehicle Reducer Based on the SVR–FDS Method,” *Sensors*, vol. 24, no. 16, p. 5359, Aug. 2024, doi: 10.3390/s24165359.
[96] H. Kaymaz, “Accelerated Life Testing of a Mild Hybrid Powertrain Using Arrhenius Equation,” *J. Name Needed*, Oct. 2022.
[97] D.-H. Ha, S.-C. Park, and C.-W. Shin, “Development of Accelerated Reliability Testing Method for Electric Vehicle Motor and Battery System,” in *SAE Technical Paper Series*, SAE International, 2014. doi: 10.4271/2014-01-0748.
[98] N. R. S. Popsi and others, “Influence of Electric Motor Manufacturing Tolerances on End-of-Line Testing: A Review,” *Energies*, vol. 17, no. 8, Apr. 2024, doi: 10.3390/en17081913.
[99] A. K. M. Shafiullah and C. Q. Wu, “Generation and Validation of Loading Profiles for Highly Accelerated Durability Tests of Ground Vehicle Components,” *Eng. Fail. Anal.*, vol. 33, pp. 1–16, 2013, doi: 10.1016/j.engfailanal.2013.04.008.
[100] “Study on Performance and Continuous Durability Test of Single Input Machinery,” *Mater. Werkst.*, vol. XX, no. YY, p. ZZ-AA, 2020, doi: 10.1002/mawe.202000097.
[101] J. Craven, “Simulation Can Help Resolve NVH Issues in EVs: Jordan Craven Explores the Challenges and Benefits of Modelling Noise, Vibration and Harshness in Electric Vehicles.” Aug. 2022. [Online]. Available: https://www.automotiveworld.com/articles/simulation-can-help-resolve-nvh-issues-in-evs/
[102] D.-Y. Kim, “Analysis of Efficiency and Noise, Vibration, and Hardness Characteristics of Inverter for Electric Vehicles According to Pulse-Width Modulation Technique,” *World Electr. Veh. J.*, vol. 15, no. 12, p. 546, 2024, doi: 10.3390/wevj15120546.

Acknowledgements

First and foremost, I would like to express my sincere and profound gratitude to my PhD supervisor, **Prof. Ing. Enrico Corti**, for his invaluable guidance, unwavering support, and insightful mentorship throughout this entire research journey. His expertise and encouragement were fundamental to the successful completion of this work.

This research would not have been possible without the generous funding and industrial support of **Alma Automotive s.r.l.** I am deeply indebted to my company tutor, **Ing. Paolo Bovicelli**, for his expert guidance during the final and most critical stages of this project. I am also immensely grateful to my former company tutor, **Ing. Marco Abbondanza**, whose foundational support was instrumental in the initial phases of my research. My sincere thanks also go to **Matteo Mami, Alberto Bernardi, Marco Cini, and Michele Proietti**, whose valuable collaboration and technical discussions greatly enriched this work. Their collective insights were instrumental in bridging the gap between academic theory and real-world industrial application.

I would also like to extend my sincere gratitude to **Rimac Technology** for hosting my six-month internship at their facility in Zagreb, Croatia. I am especially thankful to my company supervisor, **Arturo Montufar Arreola**, for the incredible learning opportunity and for providing access to the data that forms the basis of the electric powertrain case studies.

I am grateful to the **Department of Industrial Engineering (DIN)** at the Alma Mater Studiorum University of Bologna for providing an excellent academic environment and the necessary resources for this research. Within this environment, I extend a special thanks to my colleagues and friends in the research group, both past and present, for the countless stimulating discussions and their unwavering support.

Finally, I would like to extend a heartfelt thanks to my family and friends for their endless patience, support, and encouragement throughout my PhD studies.