

Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies

**Bc. Miroslav Hájek**

**Machinery Vibrodiagnostics  
with the Industrial Internet of Things**

Progress report in Master's thesis project II

Thesis Supervisor: Ing. Marcel Baláž, PhD.  
January 2024



Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies

**Bc. Miroslav Hájek**

**Machinery Vibrodiagnostics  
with the Industrial Internet of Things**

Progress report in Master's thesis project II

Study programme:	Intelligent Software Systems
Study field:	Computer Science
Training workplace:	Institute of Computer Engineering and Applied Informatics
Thesis supervisor:	Ing. Marcel Baláž, PhD.
Departmental advisor:	Ing. Jakub Findura
Consultant:	Ing. Lukáš Doubravský

January 2024



# Návrh zadania diplomovej práce

Finálna verzia do diplomovej práce<sup>1</sup>

## Študent:

**Meno, priezvisko, tituly:** Miroslav Hájek, Bc.

**Študijný program:** Inteligentné softvérové systémy

**Kontakt:** xhajekm@stuba.sk

## Výskumník:

**Meno, priezvisko, tituly:** Marcel Baláž, Ing. PhD.

## Projekt:

**Názov:** Vibrodiagnostika strojov s priemyselným internetom vecí

Machinery vibrodiagnostics with the industrial internet of things

**Miesto vypracovania:** Ústav počítačového inžinierstva a aplikovanej informatiky,  
FIIT STU, Bratislava

**Oblast problematiky:** Internet vecí, Spracovanie signálov, Feature engineering

## Text návrhu zadania<sup>2</sup>

Monitorovanie prevádzkového stavu rotačných strojov za účelom včasného odhalenia poškodení je dôležité pre plynulý priebeh priemyselných procesov bez náhleho zlyhania kľúčového technického vybavenia. Nadmerné vibrácie alebo graduálna či náhla zmena ich charakteru sú spoľahlivými indikátormi opotrebenia dielcov. V mnohých prípadoch bývajú zavedené iba pravidelné pôchodzkové merania s následným vyhodnotením časových a frekvenčných priebehov kvalifikovaným personálom. Kontinuálna diagnostika a prediktívna údržba rozširujúca sa so zariadeniami IIoT spôsobuje enormný nárast objemu zaznamenaných dát. Sledovanie výchyiek operátorom a manuálna identifikácia súčastok vyžadujúcich údržbu v celom závode sa tak stáva prakticky nerealizovateľná.

Preskúmajte spôsoby zisťovania bežných poškodení strojov z vibračných signálov a analyzujte algoritmy na redukciu množstva posielaných dát zo senzorov vzhľadom na osobitosti aplikačnej domény. Navrhnite reprezentáciu údajov na základe typických črt signálu, ktorá zníži výpočtové nároky na zvyšok komunikačného reťazca. Zvolený spôsob predspracovania má zároveň umožniť diagnostiku poškodení zvoleného stroja. Implementuje vaše riešenie s ohľadom na možné nasadenie na prostriedkami limitovanú senzorovú jednotku. Následne posúďte efektívnosť, porovnajte dosiahnuté presnosti diagnostiky a verifikujte voči zaužívaným postupom.

<sup>1</sup> Vytlačiť obojstranne na jeden list papiera

<sup>2</sup> 150-200 slov (1200-1700 znakov), ktoré opisujú výskumný problém v kontexte súčasného stavu vrátane motivácie a smerov riešenia

## Literatúra<sup>3</sup>

- NANDI, Asoke Kumar; AHMED, Hosameldin. Condition monitoring with vibration signals: compressive sampling and learning algorithms for rotating machines. Hoboken, NJ, USA: Wiley-IEEE Press, 2019. ISBN 978-1-119-54462-3.
- YU, Gang. A Concentrated Time-Frequency Analysis Tool for Bearing Fault Diagnosis. IEEE Transactions on Instrumentation and Measurement. 2020, vol. 69, no. 2, pp. 371-381. ISSN 1557-9662. DOI: 10.1109/TIM.2019.2901514. Conference Name: IEEE Transactions on Instrumentation and Measurement.

Vyššie je uvedený návrh diplomového projektu, ktorý vypracoval(a) Bc. Miroslav Hájek, konzultoval(a) a osvojil(a) si ho Ing. Marcel Baláž, PhD. a súhlasí, že bude takýto projekt viest v prípade, že bude pridelený tomuto študentovi.

V Bratislave dňa 22.2.2023

---

Podpis študenta

---

Podpis výskumníka

## **Vyjadrenie garanta predmetov Diplomový projekt I, II, III**

Návrh zadania schválený: áno / nie<sup>4</sup>

Dňa: .....

---

Podpis garanta predmetov

---

<sup>3</sup> 2 vedecké zdroje, každý v samostatnej rubrike a s údajmi zodpovedajúcimi bibliografickým odkazom podľa normy STN ISO 690, ktoré sa viažu k téme zadania a preukazujú výskumnú povahu problému a jeho aktuálnosť (uveďte všetky potrebné údaje na identifikáciu zdroja, pričom uprednostnite vedecké príspevky v časopisoch a medzinárodných konferenciách)

<sup>4</sup> Nehodiace sa prečiarknite



## **Declaration of Honor**

I hereby declare on my honor that I wrote this thesis independently under the supervision of Dr. Marcel Baláž, after consultations and with use of cited literature.

Bratislava, 4 January 2024

.....

Bc. Miroslav Hájek





# Annotation

Slovak University of Technology in Bratislava

Faculty of Informatics and Information Technologies

Degree course: Intelligent Software Systems

Author: Miroslav Hájek

Master's Thesis: Machinery Vibrodiagnostics with the Industrial Internet of Things

Thesis supervisor: Dr. Marcel Baláž

Departmental advisor: Jakub Findura

Consultant: Lukáš Doubravský

January 2024

The progress report in Master's thesis project II deals with condition monitoring and predictive maintenance of rotating machinery by employing feature discovery and statistical classification.

Trend indicators are derived from vibration signals to form summarizing numerical attributes. The similarity of features to a dependent variable describing fault types is assessed by supervised feature selection metrics individually and in the ensemble. The accuracy of fault detection using the k-nearest neighbor is evaluated on the best feature subsets to conserve data rates of the monitoring solution. The incremental learning model is compared to its batch counterpart when truthful labels are delayed or missing.

Based on the MaFaulDa dataset, classification metrics are evaluated in several scenarios, and lossy compression ratios are expressed. The measurement plan and sensor device are designed to enable collection of novel dataset of machinery behavior for compressors and pumps. The contribution of this work lies in identifying the viability of using a few quantities to describe not just the presence of the machine defect but also its cause.



# Anotácia

Slovenská technická univerzita v Bratislave

Fakulta informatiky a informačných technológií

Študijný program: Inteligentné softvérové systémy

Autor: Bc. Miroslav Hájek

Diplomová práca: Vibrodiagnostika strojov s priemyselným internetom vecí

Vedúci diplomovej práce: Ing. Marcel Baláž, PhD.

Pedagogický vedúci: Ing. Jakub Findura

Konzultant: Ing. Lukáš Doubravský

Január 2024

Priebežná správa o riešení DP2 sa zaobrá monitorovaním stavu rotačných strojov a ich prediktívou údržbou prostredníctvom techník výberu atribútov a štatistickej klasifikácie.

Z vibračných signálov sú odvodené trendové indikátory a tak vytvárajú summarizujúce číselné atribúty. Podobnosť atribútov so závislou premennou popisujúcou typy porúch sa posudzuje metrikami výberu atribútov v učení s učiteľom. Metriky sú hodnotené individuálne a súborovo. Presnosť detektie porúch algoritmom k-najbližších susedov sa vyhodnocuje na podmnožinách najlepších atribútov. Tým sa znížia požiadavky na objem prenosu dát v monitorovacom riešení. Model prírastkového učenia sa porovnáva s učením v dávkach, v situáciach kedy je expertné označenie pre-vádzkového stavu oneskorené alebo chýbajúce.

Na základe súboru údajov MaFaulDa sú klasifikačné metriky vyhodnotené podľa viacerých scenárov a tiež sú vyjdrené stratové kompresné pomery. Plán merania a senzorové zariadenie sú navrhnuté na zber nového súboru údajov o správaní sa strojov ako sú kompresory a čerpadlá. Prínos tejto práce spočíva v identifikácii možnosti použitia iba pár veličín na opis nielen prítomnosti poruchy, ale aj jej príčiny.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Problem analysis</b>	<b>3</b>
2.1	Condition monitoring . . . . .	3
2.1.1	Maintenance strategies . . . . .	3
2.1.2	Vibration fault types . . . . .	6
2.1.3	Technical standards . . . . .	8
2.2	Signal preprocessing . . . . .	11
2.2.1	Detrending . . . . .	12
2.2.2	Adaptive noise cancellation . . . . .	13
2.2.3	Time synchronous averaging . . . . .	13
2.3	Feature extraction . . . . .	14
2.3.1	Temporal domain features . . . . .	15
2.3.2	Spectral domain features . . . . .	15
2.3.3	Time-frequency domain features . . . . .	18
2.3.4	Wavelet domain features . . . . .	19
2.3.5	Empirical wavelet features . . . . .	23
2.4	Feature transformations . . . . .	25
2.4.1	Feature normalization . . . . .	25
2.4.2	Principal components analysis . . . . .	25
2.5	Feature selection . . . . .	26
2.5.1	Filtering methods . . . . .	26
2.5.2	Scores for ranking . . . . .	27
2.6	Diagnostics techniques . . . . .	29
2.6.1	Novelty detection . . . . .	29
2.6.2	Classification . . . . .	31
2.6.3	Incremental learning . . . . .	33

2.7	Evaluation datasets . . . . .	34
2.7.1	Machinery Fault Database . . . . .	34
2.7.2	CWRU bearings dataset . . . . .	36
2.7.3	Unbalance of the rotating shaft . . . . .	37
<b>3</b>	<b>Design</b>	<b>39</b>
3.1	Research questions . . . . .	39
3.2	Dataset exploration . . . . .	41
3.2.1	Fault annotations . . . . .	42
3.2.2	Signal filters . . . . .	43
3.2.3	Statistical tests . . . . .	44
3.3	Feature relevance . . . . .	44
3.3.1	Feature extraction . . . . .	44
3.3.2	Data volume savings . . . . .	47
3.3.3	Feature selection . . . . .	49
3.4	K-nearest neighbor classifier . . . . .	50
3.4.1	Batch models . . . . .	51
3.4.2	Online models . . . . .	56
<b>4</b>	<b>Implementation</b>	<b>61</b>
4.1	Machinery for monitoring . . . . .	61
4.2	Sensor hardware and drivers . . . . .	63
4.3	Preliminary measurements . . . . .	64
<b>5</b>	<b>Conclusion</b>	<b>67</b>
<b>Bibliography</b>		<b>69</b>
<b>A</b>	<b>Resumé</b>	
<b>B</b>	<b>Work plan</b>	
<b>C</b>	<b>Technical documentation</b>	

# List of Figures

2.1	P-F curve represents the evolution of the asset's health [7] . . . . .	5
2.2	Bathtub curve [1] . . . . .	5
2.3	Complex machinery vibrations [6] . . . . .	6
2.4	The transducer linear response and resonance in tolerance intervals [11]	11
2.5	Transfer function of 1st order DC blocker filters [13] . . . . .	12
2.6	Adaptive noise cancellation filter diagram . . . . .	14
2.7	Comparison of time-frequency transform spectrograms [27] . . . . .	21
2.8	Dyadic filter banks for discrete wavelet transform [16] . . . . .	22
2.9	Empirical wavelet segmentation of Fourier spectrum [34] . . . . .	23
2.10	Illustration of improved EWT spectral segmentations . . . . .	24
2.11	DenStream [48] . . . . .	30
2.12	Half-space tree [51] . . . . .	31
2.13	Nearest neighbors classification algorithm [54] . . . . .	32
2.14	Incremental learning deployment architecture [62] . . . . .	34
2.15	Machinery fault simulator for MaFaulDa . . . . .	35
2.16	CWRU machine apparatus . . . . .	37
2.17	Motor driving shaft in unbalance measurement [68] . . . . .	37
3.1	Activity diagram of MaFaulDa dataset preprocessing . . . . .	41
3.2	Inner bearing vibrations (A) for each fault category with the highest fault severity at 2500 rpm . . . . .	43
3.3	Feature value ranges in inner bearing position (A) . . . . .	45
3.4	Number of principal components to cumulative explained variance percentage in the inner bearing position . . . . .	47
3.5	PCA loading plots for min-max scaled features from the inner bearing position . . . . .	47

3.6	Approval rating of the best triplet of features to target variable “fault” scored by rank product of correlation, F statistic, and mutual information. . . . .	50
3.7	Confusion matrix of KNN predictions on bearing position A . . . . .	51
3.8	Range of prediction accuracy in all KNN models with subset of 3 features. Subplots show three different predicted variables. . . . .	53
3.9	Cross sections of 3-dimensional feature spaces with ground truth labels for best attributes. Features are chosen by rank product method. . . . .	55
3.10	Batch KNN algorithm prediction accuracy with various feature sets. . . . .	56
3.11	Label sequencing for progressive evaluation . . . . .	57
3.12	Classification labels in incremental learning ( <i>from left</i> ): true labels, predicted labels, mistakes in predictions. . . . .	58
3.13	Incremental learning on all extracted feature with delayed reveal of labels in sliding windows . . . . .	58
3.14	Incremental learning on all extracted feature with delayed reveal of labels at regular intervals in tumbling windows . . . . .	59
3.15	Incremental learning with missing true labels and sliding window delay of 10 observations . . . . .	59
4.1	Machines dedicated for vibration measurements . . . . .	61
4.2	KSB Guard cloud monitoring for pumps . . . . .	62
4.3	Sensor unit hardware block diagram . . . . .	64
4.4	Temporal domain waveforms in all spatial directions with 300 ms duration at 1 s timestamp . . . . .	65
4.5	Orbital and spectral methods of vibration signal analysis . . . . .	65

# List of Tables

2.1	Expert observed likely vibration causes (based on [6, 5, 11]) . . . . .	7
2.2	ISO 20816 vibration severity chart with typical magnitudes [12] . . . . .	9
2.3	Temporal domain features . . . . .	15
2.4	Frequency domain features . . . . .	16
2.5	Correlation coefficients . . . . .	28
2.6	Distance metrics for KNN . . . . .	32
2.7	MaFaulDa description of columns . . . . .	35
2.8	CWRU dataset description of columns . . . . .	37
2.9	“Unbalance on the rotating shaft” dataset description of columns . .	38
3.1	Label count for whole MaFaulDa dataset with recordings split to 5 chunks . . . . .	46
3.2	KNN model performance trained on all extracted features . . . . .	52
3.3	Features with the highest accuracies on the training set found com- binatorically . . . . .	54
3.4	Three chosen features with rank product of correlation, F statistic, mutual information and their associated KNN accuracies. . . . .	54
4.1	Hardware parameters of previous and proposed sensor units . . . . .	63



# List of Equations

2.0	DC blocker IIR filter of 1st order . . . . .	12
2.1	Adaptive noise filtering LMS update formula . . . . .	13
2.2	Adaptive noise filtering NLMS update formula . . . . .	13
2.3	Time synchronous averaging . . . . .	13
2.4	MMS algorithm: Local maxima identification equality . . . . .	17
2.5	MMS algorithm: Local minima identification equality . . . . .	17
2.6	Harmonic series search criterion . . . . .	18
2.7	Teager–Kaiser energy operator . . . . .	18
2.8	Transient-extracting transform . . . . .	19
2.9	Mother wavelet function . . . . .	19
2.10	Continuous Wavelet transform . . . . .	19
2.11	Representation of components for Synchrosqueezing Wavelet transform .	20
2.12	Wavelet packet coefficient . . . . .	21
2.13	Feature euclidian norm . . . . .	25
2.14	Min-max feature scaling . . . . .	25
2.15	Feature standardization . . . . .	25
2.16	Singular Value Decomposition . . . . .	26
2.17	Pearson correlation coefficient . . . . .	27
2.18	Fisher score . . . . .	28
2.19	Mutal information . . . . .	28
3.1	Lossy compression ratio with features . . . . .	48



# 1 Introduction

The industry experiences a shift in traditional asset operational status evaluation and utilization. The rise of Industry 4.0 means greater automation and robotization of the production halls to achieve optimal usage of available resources. The secondary aspect in the enterprises' endeavor, but not less important, is to keep track of the equipment's wear and tear. The corrective action, be it repair or replacement, should be done on time in response to the key indicators.

The goal is to preserve required safety and production efficiency while extending the useful life of rotating mechanical parts. In the factories and logistics where this sort of equipment is vital, there is a rising interest in the ability to watch in real time the machine's health status. Proactive fault diagnosis is imperative to initiate a repair without adding unnecessary costs.

Vibrations are a non-intrusive way to sense and record eventually fatal deficiencies right at the onset. The experts use it to distinguish faulty states and to identify the malfunction's root causes. In critical circumstances such as is the case of the large turbines in power plants, the precautions leading to regular machinery check-ups are already in place. The monitoring solution has to be sufficiently independent, reliable, and accurate to achieve wider acceptance and spread.

The main issue to consider in large-scale machinery monitoring with vibrations is that there are lots of uninformative streams of samples not directly useful for the production line operator. The dashboard must aggregate these flows into trend variables with severity levels categorized based on industrial standards. The majority of signals are viewed once at the maximum. Therefore, to store or even transmit them from the edge device in its entirety would be wasteful. The complex overview of the mechanical equipment status is attainable only when agent devices and sensors are cheap enough with a long lifespan powered out of the battery pack. Preferably these devices should also remain physically small to reduce the additional clutter in the factory.

Attempted machine learning and deep learning approaches have the crucial impediment that the construction of every single machine is unique to some extent because of tolerances and variable load during regular operation. The model has to be trained for the target environment to achieve the ideal performance. In addition, failures are relatively rare events that usually occur several months apart. In these circumstances, it is hard to quickly obtain a large enough sample of fault events. Novelty detection is a technique that can be applied in this case.

This progress report on the master's thesis II is organized followingly. The problem analysis chapter 2 begins by exploring the mechanical maintenance approaches and industry standards on common fault identification. Vibration monitoring starts with a section on preprocessing that summarizes basic filters. The process of transforming raw samples into features and their meaningful selection is covered in automatic fault pattern recognition with machine learning models trained in offline and online contexts. An overview follows of the most common machinery fault databases.

Chapter 3 on solution design starts by posing research questions and establishing goals. Then, the MaFaulDa dataset statistical properties and labeling procedure are described on the original samples as well as on extracted attributes. Data rates with feature selection are derived based on the MaFaulDa. In the section about classification models, we compare the performance for various feature sets in the k-nearest neighbors algorithm. Finally, in preparation for the evaluation part, the sensor device is proposed which is suitable for measurement of designated machines in the industrial environment.

# 2 Problem analysis

In the problem analysis chapter we explore the feature engineering methods and machine learning algorithms for fault diagnostics. The basis we build upon is the domain knowledge of mechanical engineers in vibration signal measurement and its evaluation.

## 2.1 Condition monitoring

All rotating machinery eventually fails because of the long-term strain on the individual parts, inadequate workmanship, installation, or operational procedures. In the end, these factors cause the equipment not to fulfill its intended functionality. Many instrumentation methods are practiced to reveal evolving faults: vibration and acoustic noise monitoring, electric supply line measurements, thermography, oil and particle analysis, ultrasonic testing, etc. Vibration signals are the preferred tool for rotating machinery monitoring [1].

The defect needs to be repaired or replaced, preferably without significant production downtime, further damage to the other attached elements, or any endangerment of the responsible personnel. The maintenance strategies are chosen according to the machine's importance as a result of potential failure's impact on the system. The guide to set appropriate maintenance procedures as outlined in the IEC 60706-2 standard, and involves reliability-centered maintenance analysis [2].

### 2.1.1 Maintenance strategies

There are three different approaches to maintenance across the industry: **reactive, preventive, and predictive** [3]. In general, the more sophisticated methods are beneficial in a high-stakes environment. The unexpected machine shutdown can have a negative economic impact on the enterprise, resulting in decreased product

quality and demands for spare parts to be ready in the supply inventory at all times. In certain situations, it suffices to utilize a simpler maintenance program, but predictive maintenance gains attraction in Industry 4.0 to optimize usage of assets [4].

**Reactive maintenance** allows machinery to run until a complete failure. This is the most inappropriate way to maintain the production line, but it is straightforward enough. It requires a large stock of replacement parts on-site and breakage inflicts a “crisis management mode” upon the plant [3]. On-demand repairs are justified when short downtime is acceptable, full and swift replacement of a broken machine with a backup is possible, or there is a negligible threat to the surrounding environment from failure [5].

**Preventive maintenance** is performed before any issue is detected. Maintenance occurs at regular intervals derived from a predetermined period in the calendar or expected machine running time (MTTF - Mean Time To Failure). The schedule is crucial but can result in components being replaced in good condition, creating waste. The parts can occasionally stay in operation too long, and the machine breaks as a result. Conservative planning is usually the norm to keep the machines always in a perfect state, and therefore more frequent interventions are required [1].

**Predictive maintenance** known as condition-based maintenance (CbM), improves the predictability of reactive maintenance and eliminates waste in overall resource utilization of cautious prevention. The machine downtime is scheduled after the detection of unhealthy trends in fault monitoring with sensors and the identification of troublesome components.

A measurable decrease in effectiveness allows us to order necessary parts in advance and organize repairs of several machines at a convenient time. The missed detection leads to increased costs compared to previous methods and raises the expectation that faults are distinguishable among themselves [6].

The *P-F curve* is a widespread representation of equipment degradation over time based on historical records (Fig. 2.1). Corrective action should be taken between the event of potential failure (P), when the fault detection is activated, and functional failure point (F) in the P-F interval [8]. These division points are not exactly set but have a statistical distribution to them.

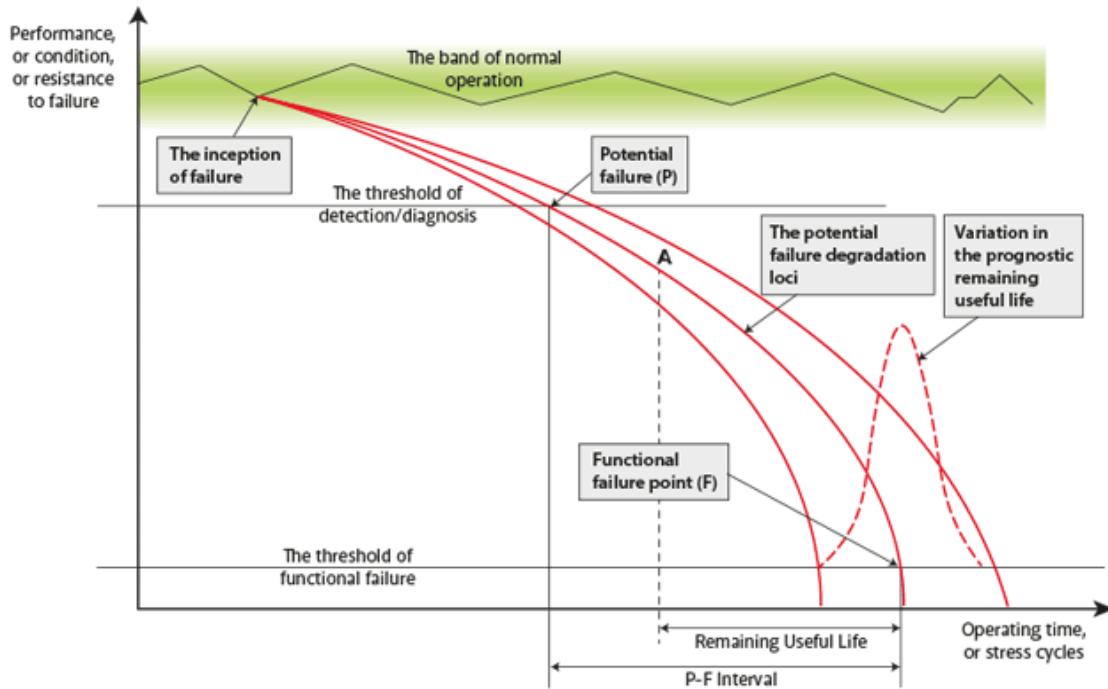


Figure 2.1: P-F curve represents the evolution of the asset's health [7]

The *Remaining Useful Life* (RUL) of the specific running machine in the given instance can be merely estimated analytically, with the survival probabilities of the individual components, based on the model of the “run-to-failure” histories, and usage parameters [9]. Predictive condition monitoring aims to extend lifespan to the maximum by predicting expected RUL.

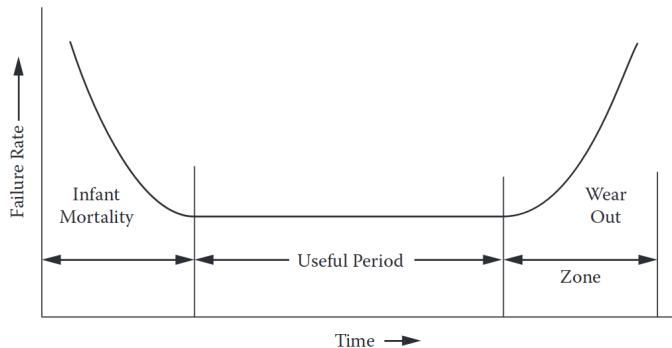


Figure 2.2: Bathtub curve [1]

A high failure rate is present not only at the worn-out stage when the parts are fatigued or corroded, but also in the early stages soon after assembly. Manufacturing or material defects, inadequate installation, or improper start-up procedures, are all suspected causes. During the stable middle phase, malfunction can occur after the

machine's excessive overload. The time plot to failure rate is known as the bathtub curve (Fig. 2.2).

### 2.1.2 Vibration fault types

Mechanical problems during machinery operation bring about vibrations in many instances. Therefore, vibroacoustic diagnostics is considered as one of the most important methods in early component fault identification [5].

The cause of vibration comes out of the changing force in its magnitude or direction. The most emerging defects can be encompassed by explaining the deficiencies of the mechanical structure. These defects are broadly categorized as **unbalance**, **misalignment**, **looseness**, **eccentricity**, **deformation**, **crack**, and **influence of the external force** (friction) [6]. It is important to stress that our concern is not the underlying deformities in mechanical parts, but the correct fault classification based on the signal waveform.

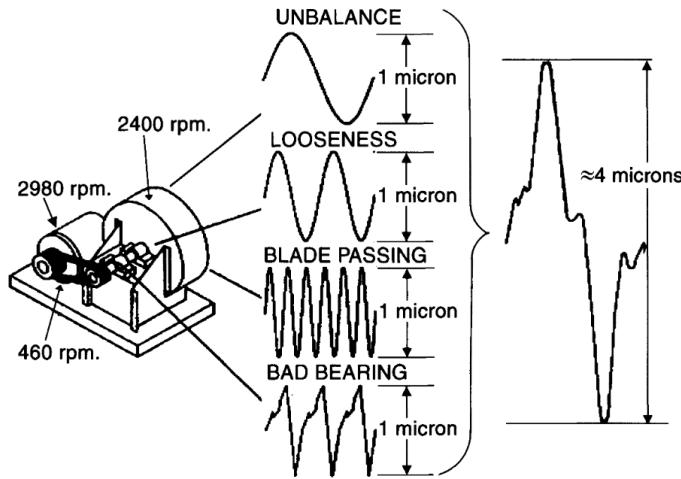


Figure 2.3: Complex machinery vibrations [6]

Rotating machine disorders exhibit frequency signatures at various ranges in the frequency spectrum with supplementary symptoms carried in phase signal. Most of the occurring faults can be tied to the main rotational speed of the component under investigation (Fig. 2.3) [6]. Imbalance, misalignment, and looseness normally appear at frequencies up to 300 Hz. These low-frequency faults are associated with the movement of the shaft and primarily coincide with revolution speed and its harmonics. Bearing and gearbox defects in the late stages of development, show up

in the range between 300 Hz and 1 kHz. Higher frequencies, measured traditionally to a limit of 10 kHz, help notice the faults in bearings even sooner [10].

One of the methods vibration experts utilize in the identification of the damaged part, according to the frequency spectrum, is **order analysis**. The excessive peaks at harmonic frequencies are of interest. Harmonics are integer multiples of fundamental frequency (1x rpm) (Tab. 2.1):

Frequency content		Likely reason	Other causes
Synchronous	1 x rpm	Imbalance	Eccentric journals Bent shaft / Misalignment (high axial vibration) Bad belt (if rpm of belt)
	2 x rpm	Looseness	Misalignment (high axial vibration) Cracked rotor Bad belt (if rpm of belt)
	3 x rpm	Misalignment	and axial looseness
	Many x rpm	Bad gears Severe looseness	Gear teeth x rpm Fan blade count x rpm
Sub-synchronous	<1 x rpm	Oil whirl	Bad drive belt Background Resonance
Non-synchronous	-	Electrical problems (x 50 Hz) Reciprocating forces Aerodynamic forces Bad antifriction bearings	Rubbing

Table 2.1: Expert observed likely vibration causes (based on [6, 5, 11])

Because of inherent tolerances in machine manufacturing and assembly, the rotational frequency always manifests itself, even in baseline signature [6, 11]. In the most likely scenario, some faults appear as compared to rotational frequency solely in **synchronous, subsynchronous, or non-synchronous components**. The defects can occur in a predictable combination of the ones mentioned. Other common patterns experts look for are modulation sidebands typical for bearings and gears extractable with cepstrum analysis [5]. Procedures relying on elimination narrow down unrelated causes effectively.

### 2.1.3 Technical standards

Vibration-based condition monitoring practices adopted in the factory's predictive maintenance management must comply with normative guidelines formalized in ISO international standards. Standards are concerned with each step in the process that originates with transducer placements and data acquisition. They prescribe conventions for setting fault severity levels and provide empirically observed vibration characteristics of common defects. Relevant standards for IoT diagnostics systems are *ISO 20816* (updated from ISO 10816) and *ISO 13373*.

**ISO 20816-1:2016** establishes the approaches of vibration measurement and evaluation on non-rotating housing of machinery parts [12]. The measurement units are agreed upon for kinematic quantities of vibrations. Acceleration is measured in meters per second squared ( $m/s^2$ ), velocity in millimeters per second ( $mm/s$ ), and displacement in micrometers ( $\mu m$ ). It is customary to evaluate broad-band vibration velocity in terms of root mean square value (rms), as it is related to signal energy. No simple direct relationship is expressible among these physical quantities, except in stationary signals.

The vibration severity is the maximum magnitude measured in two radial directions (horizontal, and vertical) or supplemented with a third direction along the shaft on the axial axis. Multiple measurement locations, i.e. on different bearings or couplings, should be assessed independently.

Criteria introduced to judge vibration severity are absolute vibration magnitude, change in the magnitude vector, and rate of change. In terms of maximal magnitudes, the machines of varying sizes are split into four severity zones defined in the chart (Tab. 2.2). The values in this table serve as guidelines towards realistic requirements between machine manufacturers and their customers.

Zone A is reserved for newly commissioned machines. Zone B signifies suitability for long-term operation. In zone C the machine is deemed in unsatisfactory condition and corrective action should be taken soon. Finally, in zone D vibrations can cause damage to the machine. The span of acceptable values differs with the machine class from I through to IV and their output power of 15 kW (class I), 75 kW (class II), 10 MW (class III), or greater.

Vibration velocity RMS [mm/s]	Class I Small machines	Class II Medium machines	Class III Large machines Rigid supports	Class IV Large machines Flexible support
0.28	Good (A)	Good (A)	Good (A)	Good (A)
0.45				
0.71				
1.12				
1.8		Satisfactory (B)		
2.8		Satisfactory (B)	Satisfactory (B)	Satisfactory (B)
4.5				
7.1	Unacceptable (D)	Unacceptable (D)	Unacceptable (D)	Unacceptable (D)
11.2				
18				
28				
45				

Table 2.2: ISO 20816 vibration severity chart with typical magnitudes [12]

The operational limits in the form of *alarms* and *trips* are usually established on the zone boundaries or close to them. Alarms are placed between zones B and C providing a warning about reaching the threshold significant for noticeable change. Trips in between zones C and D urge immediate action or machine shut down. Both limits should not exceed 1.25 times the upper boundary of the lower zones and initially are set based on previous experience with the machine [11].

**ISO 13373-1:2002** delves into further nuances of vibration monitoring and expands on procedures outlined in terms of the vocabulary in ISO 20186. According to the standard, the data collection operates in continuous or periodic observation modes which follow an event or interval. Both designs can be permanently mounted. In continuous collection, it is recommended to have a “multiplexing rate sufficiently rapid, so there is no significant data or trends lost” [11]. When channels are scanned in succession with gaps between data points, the system is known as “scanning”.

The condition monitoring program is run according to a flowchart adapted from one designed by the standard specifically to best benefit the plant. Those steps can be summarized as follows [11]:

1. Review machinery history and establish failure modes.
2. When vibration monitoring is not applicable, check for other condition-monitoring techniques or resort to preventive maintenance.
3. Select monitoring points and take preliminary vibration measurements.

4. Select vibration monitoring techniques: broadband, frequency analysis, or special techniques, and set parameters of measurement units.
5. Take baseline measurements.
6. Change levels that would warrant investigation.
7. Carry out routine condition monitoring.
8. If an alarm was exceeded, notify appropriate personnel to review data and trends, perform diagnostic evaluation, and repair as necessary. In case a new baseline is needed, continue in the step of taking baseline measurements.
9. Shut down the machine when the trip level is exceeded. Then proceed the same as after the alarm is triggered.

Measurement of vibrations should be accompanied by a description of the machine and its operating conditions. The machine description includes its identifier and type, power source, rated rotation speed and power, configuration (shaft or belt driven), and machine support. Measurement parameters are to be recorded alongside the measurement value itself, such as timestamp, transducer type, sensor location and orientation in **MIMOSA convention** (Machinery Information Management Open System Alliance), measurement units and units qualifier (p-p, rms), and other processing options (filters, number of averages, etc.) [11].

The transducer of choice for condition monitoring is the accelerometer, which can provide the acceleration value of the body and velocity after signal integration. However, standard advise against double integrating for displacement. The recommended frequency range for an accelerometer is 0.1 Hz to 30 kHz. The choice of transducer mount significantly lowers its resonance frequency, which is least influenced by the stud mount and stiff cement mount. The resonance is reduced to around 8 kHz with the use of soft epoxy or permanent magnet.

Broadband measurement requires “frequency ranges of 0.2 times the lowest rotational frequency to the highest frequency of interest” [11], not exceeding 10 kHz, with RMS velocity 0.1 - 100 mm/s. Bearings and gear diagnosis may push the upper-frequency limit even higher. The tolerances of amplitude and frequency calibrations fall into two types with allowable tolerances of  $\pm 5\%$  or  $\pm 10\%$  (Fig. 2.4).

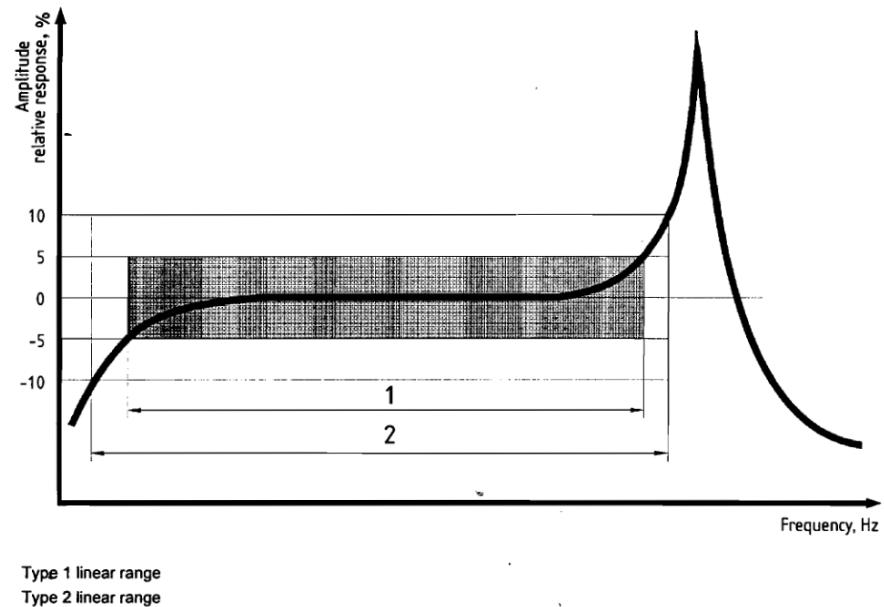


Figure 2.4: The transducer linear response and resonance in tolerance intervals [11]

Equipment’s “health” can be mischaracterized when there are significant differences in the machine’s normal operating conditions. Baseline measurements in all acceptable conditions are to be acquired to reduce the error in vibration evaluation. According to the bathtub curve (Fig. 2.2) reference signatures should be obtained after the initial part wear-in period. The reference spectral mask of the baseline condition is designed if maximal acceptance amplitudes are different for each significant frequency band [5].

The vibration baseline is defined by broad-band magnitudes and phases of motion vectors, the waveform in the time and frequency domain, the rotational speed of the machine as well as its frequency response to different speeds during start-up and coast-down captured in the Bode plot and waterfall plot. Changes during the machine’s operation are then depicted in value trends. Trends can be shown by overall amplitudes or limited to frequency bands.

## 2.2 Signal preprocessing

The vibration signals in a factory environment are inherently full of disturbances. Nearby equipment operation and handling of heavy objects in the surroundings can all contribute to the unwanted chaotic movement in otherwise mostly pure oscillatory

motion. In addition, accelerometers suffer from systematic measurement errors in the form of thermal noise, zero-g offset as a result of slight miscalibration, and bias originating from a constant force of gravity. These unavoidable distortions are to some extent suppressable by digital filters. In the preprocessing stage, we consider trend removal, noise reduction with adaptive filters, and time synchronous averaging to eliminate external interference.

### 2.2.1 Detrending

The oscillatory motion should be centered around the zero level for further manipulation. The constant offset is eliminated simply by subtracting the overall mean from the signal. Moreover, the high pass DC blocker infinite impulse response (IIR) filter of 1st order can adjust to shifts of the average value over time (Equation 2.1). The transition band depends upon the choice of corner frequency  $f_{3dB}$  (Fig. 2.5).

$$y_k = \left(1 - \frac{\omega}{2}\right) \cdot (x_k - x_{k-1}) + (1 - \omega) \cdot y_{k-1}; \quad \omega = 2\pi \cdot \frac{f_{3dB}}{f_s} \quad (2.1)$$

A steeper 3 dB attenuation band can be achieved by increasing the order of the filter. Then the cutoff frequency should be such that filter coefficients are fractions to counteract rounding errors [13].

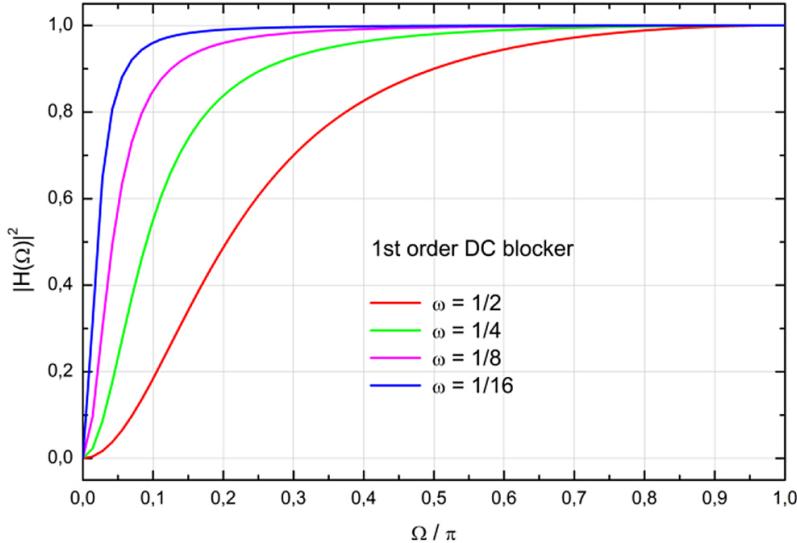


Figure 2.5: Transfer function of 1st order DC blocker filters [13]

The finite impulse response (FIR) filter is not recommended for DC component

removal because of the undesirable ripple effect with the small number of taps. Cascaded-integrator-comb (CIC) filters are proposed as an alternative instead [14].

### 2.2.2 Adaptive noise cancellation

Adaptive noise cancellation (ANC) involves an adaptive filter that self-adjusts coefficients through an update algorithm in response to the reference noise signal. The objective of this filter is to minimize the mean square error (MSE) cost function in the error signal  $e_k$  between signal contaminated with Gaussian noise  $d_k$  and filter output  $y_k$ . Additive noise  $n_k$  is assumed to be correlated with noise signal  $\mathbf{X}_k$  [15].

Wiener-Hopf equations solve the optimal gradient of the MSE function and provide us with FIR filter coefficient vector  $\mathbf{W}_k$ . The least mean squares (LMS) algorithm recursively approximates this analytical solution with the method of steepest descent (Equation 2.2) [15]. The multiple parameters are to be considered in the evaluation of filter performance: convergence rate, estimated error, and signal-to-noise ratio (SNR).

$$\mathbf{W}_{k+1} = \mathbf{W}_k + 2\mu \mathbf{X}_k e_k \quad (2.2)$$

The convergence stability is affected by step size  $\mu$  which is bounded from above with the inverse of the maximal eigenvalue of input covariance matrix  $\lambda_{max}$ . The normalized least mean squares (NLMS) can handle input of varying scales (Equation 2.3).

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \frac{\mu}{\|\mathbf{X}_k\|^2} \mathbf{X}_k e_k \quad (2.3)$$

### 2.2.3 Time synchronous averaging

Time synchronous averaging (TSA) diminishes the impact of vibration sources unrelated to the rotational frequency and its harmonics. TSA averages time-domain waveform over  $N$  points and aligns it to a synchronization pulse with period  $T$  (Equation 2.4).

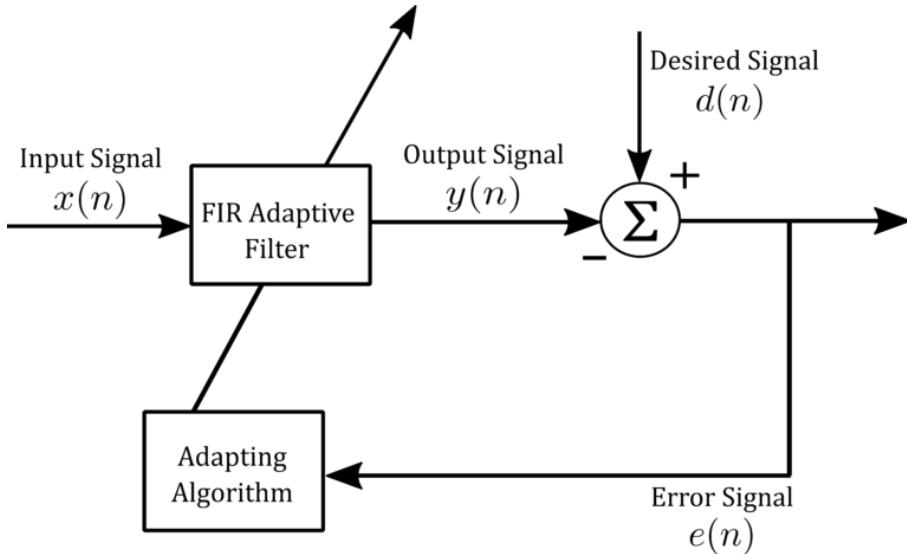


Figure 2.6: Adaptive noise cancellation filter diagram

$$x_{TSA} = \frac{1}{N} \sum_{n=0}^{N-1} x(t + nT) \quad (2.4)$$

This technique has been successfully applied to the gearbox and bearing fault diagnosis [6, 16].

## 2.3 Feature extraction

After preprocessing, raw numerical vectors are merely low-level descriptors of the underlying physical phenomena. At first, these incomprehensible sequences of numbers are reduced to summary attributes called features in the process of feature extraction or feature discovery. Features can be hand-crafted (as is our case), learned implicitly within the model representation, or explicitly from an optimization problem solution.

Predictive maintenance has ideal prerequisites for the application of feature engineering because the signal is usually pseudo-stationary, and the trend monitoring variables come out of extensive domain expertise in mechanics. The advantages of the add-in extraction effort, as opposed to processing samples unmodified, are to gain better classification precision, reduce computational burden and storage capacity downstream with dimensionality reduction [17].

It is important to note that the design of features is not a standalone step in the

machine learning pipeline, but it should be done iteratively to improve the target model. Signal features are computed in the temporal and spectral domain [18].

### 2.3.1 Temporal domain features

The most widely found features in the literature are rudimentary statistical measures of the central moment: mean, variance, standard deviation, skewness, and kurtosis (Tab. 2.3). Statistics can be calculated in any domain, but the mean value should not be used in detrended data. The vibration severity metrics out of technical standards are also highly regarded. The characteristics of amplitude include root-mean-square, peak-to-peak distance, and maximum [19].

The other significant time-domain attributes are derived as ratios of previous simpler ones. These ratios are crest factor, margin factor, impulse factor, and shape factor (Tab. 2.3) [16]. Many articles have been successful in fault detection of bearings out of transients in impulsive signals with kurtosis, crest factor, and margin indicators [18]. It is also suggested that the shape factor can signify unbalance and misalignment faults [16].

Feature	Equation	Feature	Equation
Standard deviation	$X_\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$	Crest factor	$X_{cf} = \frac{\max( x_i )}{X_{rms}}$
Skewness	$X_{sv} = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^3$	Margin factor	$X_{mf} = \frac{\max( x_i )}{\left( \frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2}$
Kurtosis	$X_{kv} = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^4$	Impulse factor	$X_{if} = \frac{\max( x_i )}{\frac{1}{N} \sum_{i=1}^N  x_i }$
Root mean square	$X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	Shape factor	$X_{sf} = \frac{X_{rms}}{\frac{1}{N} \sum_{i=1}^N  x_i }$
Peak-to-peak	$X_{ppv} = \max(x_i) - \min(x_i)$	Maximum	$X_{max} = \max( x_i )$

Table 2.3: Temporal domain features

### 2.3.2 Spectral domain features

The mechanical faults present themselves as oscillatory patterns which are combinations of frequencies with various amplitudes. The Fourier transform is one of the most prominent strategies in power spectral density estimation. Experts on vibro-

diagnostics utilize it as a primary signal processing technique for data analysis as it is recommended in ISO 13373-2 standard [20].

The inherent symmetries in the Fourier matrix made it possible to implement the Fast Fourier transform (FFT) algorithm with time complexity  $O(n \log n)$ . The drawback of the plain spectral analysis is the lack of resolution for events that occurred at distant time instants, and so their spectral components might adversely blend in together.

Feature	Equation
<i>Spectral centroid</i>	$X_{fc} = \frac{\sum_{i=0}^{N-1} f_i \cdot a(f_i)}{\sum_{i=0}^{N-1} a(f_i)}$
<i>Energy</i>	$E(N) = \sum_{i=1}^N a^2(t)$
<i>Energy ratio</i>	$E_r = E_i / \sum_{i=1}^N E_i$
<i>Spectral roll-on</i>	$E(f_c) = 0.05 \cdot E(f_s/2)$
<i>Spectral roll-off</i>	$E(f_c) = 0.95 \cdot E(f_s/2)$
<i>Spectral flux</i>	$X_{\text{flux}} = 1 - \text{corr}(A_{t-1}, A_t)$
<i>Noisiness</i>	$X_{\text{noise}} = \text{SNR}(X) = \mu / \sigma$
<i>Shannon entropy</i>	$H(X) = -\sum_i P(X = x_i) \cdot \ln(P(X = x_i))$
<i>Spectral negentropy</i>	$\Delta I_E(f, \Delta f) = \sum_{k=0}^{N-1} \frac{E(k, f, \Delta f)^2}{\frac{1}{N} \sum_{k=0}^{N-1} E(k, f, \Delta f)^2} \cdot \ln \left( \frac{E(k, f, \Delta f)^2}{\frac{1}{N} \sum_{k=0}^{N-1} E(k, f, \Delta f)^2} \right)$

Table 2.4: Frequency domain features

In the spectral domain, we can obtain the usual statistical properties of the distribution being spectral centroid, skew and kurtosis. Additionally, spectral roll-on and roll-off, fundamental frequency, entropy, negentropy, spectral flux, signal-to-noise ratio (noisiness), energy in frequency bands, and energy ratio are extracted (Tab. 2.4) [21].

In geometric terms, the spectral centroid represents the barycenter of the frequency magnitude plot. Spectral roll-off gives a notion about the spectral distribution because it identifies the frequency  $f_c$  below which 95% of the signal energy is contained. Complementary, the roll-on frequency is chosen so that 5% of the signal

energy is below this value. According to the definition, spectral flux is normalized cross-correlation between two successive amplitude spectra a value of one means the spectra are the most dissimilar.

“Negentropy measures the inclination of a system to increase its level of organization” [22]. The larger *spectral negentropy* suggests more fault-induced impulses.  $E(k, f, \Delta f)$  denotes the squared envelope spectrum (SES) which is an envelope in the Fourier domain:  $\mathcal{F}\{|y(k, f, \Delta f)|^2\}$ .

If a single principal frequency exists, it can be determined by maximum likelihood estimation. Such frequency would explain the signal spectrum the best [21]. The frequency spectrum is a discrete set of amplitudes where peaks have to be reliably identified to create representative attributes.

The essential peak-finding approaches are based either on magnitude or gradient. All found extrema are commonly filtered with the magnitude of prominences and the width at half prominence. In the magnitude-based method, the middle point  $x_i$  is compared to two neighboring points and the peak is then:  $x_{i-1} < x_i > x_{i+1}$ . The gradient-based method evaluates the first derivative at the point which is equal to zero if the point is a local maximum, local minimum, or inflection point [23].

A substantial improvement is a robust non-parametric peak identification named *MMS* based on the sum of terms in an arithmetic progression based on maximum, minimum, and sum. MMS max-min finder in the elementary form processes points in the window of length 3, it advances one point and deems its middle point as a local extremum if it satisfies equalities below. Equation 2.5 is for the hill and Equation 2.6 is for the valley. The filtration techniques are incorporated in the adaptations of the MMS algorithm: MMS-WBF, MMS-SG, MMS-LH [23].

$$\frac{a_{max} - a_{min}}{S_3 - a_{min} \cdot 3} = \frac{a_{mid} - a_{min}}{S_3 - a_{min} \cdot 3} \quad (2.5)$$

$$\frac{a_{max} - a_{min}}{a_{max} \cdot 3 - S_3} = \frac{a_{max} - a_{mid}}{a_{max} \cdot 3 - S_3} \quad (2.6)$$

Multiple harmonic series and the sidebands can be separated into a discrete set of frequency components, each with a central frequency, uncertainty, and amplitude  $C_i(v_i, \Delta v_i, A_i)$  by an exhaustive search algorithm. Harmonic family identification

is a non-trivial problem because of spectrum estimation errors. The criterion is proposed to select harmonics at the minimal distance from the true fundamental frequency multiple (Equation 2.7). Two series with the same fundamental frequency are merged and thought of as a modulation series [24].

$$v_i^{(r)} = \frac{v_j}{\min |v_j - r \cdot v_i|} \quad (2.7)$$

### 2.3.3 Time-frequency domain features

The **Short-time Fourier transform** (STFT) splits the time-domain signal into equal-length intervals. Individual chunks have a 50% overlap and are multiplied with weights of window function to balance scalloping loss and spectral leakage due to Fourier transform periodicity assumption. Traditionally, the Hann window is commonly used instead of the rectangular window [5, 20].

In the time-frequency domain, the same features can be derived as in the frequency domain, but in addition, the attributes are time-localized in this way. The STFT has a considerable flaw for implementation in a self-adaptable system and that is fixed resolution. The optimal window size has to be set beforehand or chosen after performing multiple transformations on chunks out of the range of lengths. Welch's method averages multiple consecutive blocks to better estimate the spectrum. We have already researched the suitability of STFT for online detection of constant frequencies [25].

The alternative to isolating weak impacts with high time resolution is a **Teager-Kaiser energy operator** (TKEO) (Equation 2.8). It is a tool for envelope analysis to demodulate characteristic AM-FM signals present during bearing faults. Energy operator output can be utilized as a standalone feature attribute.

$$\psi[x(n)] = [x(n)]^2 - x(n-1)x(n+1) \quad (2.8)$$

Improved TKEO is necessary to prevent analysis from suffering from the noisy source. The key idea is to perform TKEO after signal decomposition into narrowband components with different center frequencies. The extracted modes are reconstructed with weights assigned based on their correlations to the original signal [26].

Time-frequency spectrum modification preserving localization of abrupt wide-band spikes at time  $t_0$  and simultaneously reducing energy smear over the larger region is a goal of **Transient-extracting transform** (TET). Post-processing of STFT window  $G(t, \omega)$  involves multiplication of the spectrum with the Transient extracting operator (TEO). This operator is expressed in the form of a Dirac delta function  $\delta(t)$  (Equation 2.9).

$$\text{Te}(t, \omega) = G(t, \omega) \cdot \delta(t - t_0) \quad (2.9)$$

TET representation retains non-zero coefficients where the absolute value of the ratio between two STFTs,  $G^{tg}[n, k] / G[n, k]$  is less than half the sampling interval  $T$ . These two transforms use distinct windows  $g[n]$  and  $n \cdot g[n]$ . Decomposition of a signal with TET is proved to produce significantly larger kurtosis (around 38 in TET, 4 in other methods) and hence better discriminate the transient fault [27].

### 2.3.4 Wavelet domain features

The bands for lower frequencies should be longer in time than for higher frequencies. The **Wavelet transform** (WT) possesses such a multiscale discrimination property, effectively increasing resolution in time-frequency plain. Wavelet basis functions are constructed for that purpose (Equation 2.10). There are several wavelet families. For example, Haar, Daubechies, Coiflets, Symlets, Morlet, and Meyer [16].

$$\psi_{s,\tau} = \frac{1}{\sqrt{s}} \psi \left( \frac{t - \tau}{s} \right) \quad (2.10)$$

**Continuous Wavelet transform** (CWT) (Equation 2.11) is done by scaling and translating the mother wavelet  $\psi$  picked out of the appropriate family [16]. The scale factor is denoted with  $s$  and time position with  $\tau$ . The choice of wavelet type is data-driven because distinct wavelet shapes have an impact on the response and ultimately contribute to filter length. The decision lies between recognition abilities for impulse-like signals or the inclusion of wider surrounding space.

$$W_{x(t)}(s, \tau) = \frac{1}{\sqrt{s}} \int x(t) \cdot \psi^* \left( \frac{t - \tau}{s} \right) dt \quad (2.11)$$

The CWT is computationally intensive when a highly detailed scale resolution is required because each wavelet scale convolves with the entire signal. The fCWT algorithm allows 100 times higher spectral resolution than previous implementations at the same speed. It increases performance 122 times compared to Wavelib and 34 times in comparison with PyWavelets [28].

The fast CWT algorithm reaches compelling improvement by applying Parseval's theorem to the wavelet transform formula that removes the dependence on the time offset parameter [28]. The convolution takes place with the mother wavelet in the Fourier base. Then, the inverse FFT produces the coefficients for individual scales.

**Synchrosqueezing Wavelet transform** (SST) is a modification of CWT attempting to sharpen the representation of frequency components by coefficient reassignments from around the central frequencies towards the middle of the bands. The justification for these reallocations is rooted in the signal approximation as amplitude-modulated oscillating modes with additive noise  $\eta(t)$  (Equation 2.12) [29].

$$s(t) = \sum_{k=1}^K A_k(t) \cos(\theta_k(t)) + \eta(t) \quad (2.12)$$

The components are defined by their instantaneous amplitudes  $A_k(t)$  and instantaneous phases  $\theta_k(t)$ . The energy spread to adjacent bins can be effectively squeezed only in regions with constant phase and large enough component separation. Despite the promising properties of this transform, white noise causes severe interference in the resulting time-frequency map.

The spectrograms to illustrate the difference in the ability of the Fourier transform, the continuous Wavelet transform, and their modifications to pick up underlying patterns in bearing faults are shown in Fig. 2.7.

The dyadic filter bank is another signal decomposition technique that generates subbands at multiple granularity levels. The practical realization of the multiscale description is **Discrete Wavelet Transform** (DWT). The DWT behaves as a quadrature mirror filter and splits waveforms using a wavelet filter to detail coefficients (D1) and approximation coefficients (A1) (Fig. 2.8a) [16]. The low-pass filter  $h(k)$  creates approximation coefficients further decomposed at the successive levels. Detail coefficients represent the result of the high-pass filter  $g(k) = (-1)^k h(1 - k)$

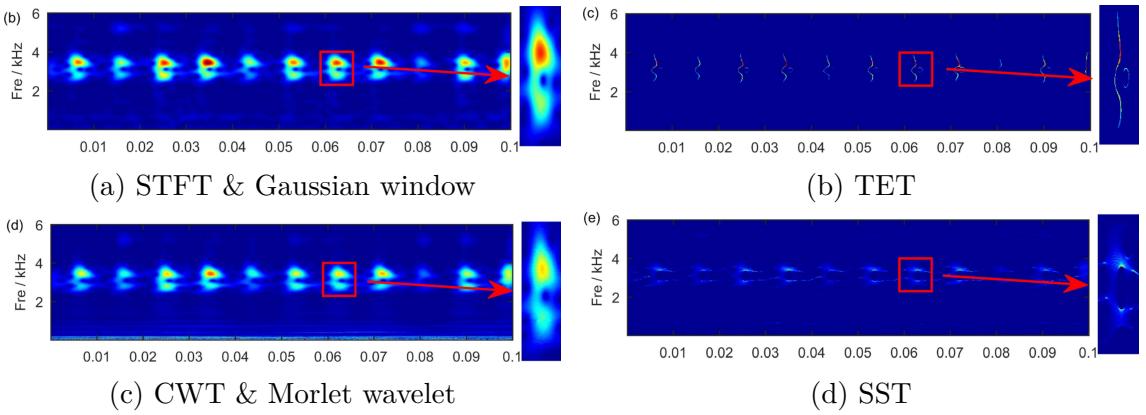


Figure 2.7: Comparison of time-frequency transform spectrograms [27]

after decimation by a factor of 2.

The maximum depth of the decomposition tree is  $\log_2 n$  where  $n$  is the number of input samples. Energy, energy ratio, and entropy are prevalent features that succinctly encode the wavelet coefficients. Otherwise, the additional extracted levels raise the total number of data dimensions.

In washing machine status classification, the discrete wavelet transform with Daubechies wavelet (db4) and fifth-level decomposition provide features with a combination of approximation (cA5) and detail coefficients (cD1, ..., cD5). Washing machines belong to three categories: no fault, electric motor clamping screws problem, and a loose or broken counterweight. Extracted measures were sample mean and sample variances over autocorrelation functions of coefficients (AcDn) and smoothed coefficients cD1, cD2 by moving average filter [30].

**Wavelet Packet Decomposition** (WPD) applies filters to split detail coefficients identically as approximation ones (Fig. 2.8b) thus increasing the resolution in the high-frequency bands and providing uniform spectrum partitioning.

$$w_{j,n,k} = \langle f, W_{j,k}^n \rangle = \langle f, 2^{j/2} W^n(2^j t - k) \rangle \quad (2.13)$$

Each wavelet packet coefficient  $w_{j,n,k}$  captures subband frequency content around time instant  $2^j k$  (Equation 2.13) [31]. This measure is an inner product of the source and scaled wavelet packet function. The aforementioned feature extraction established in DWT can be applied. For example calculation of the wavelet packet node energy.

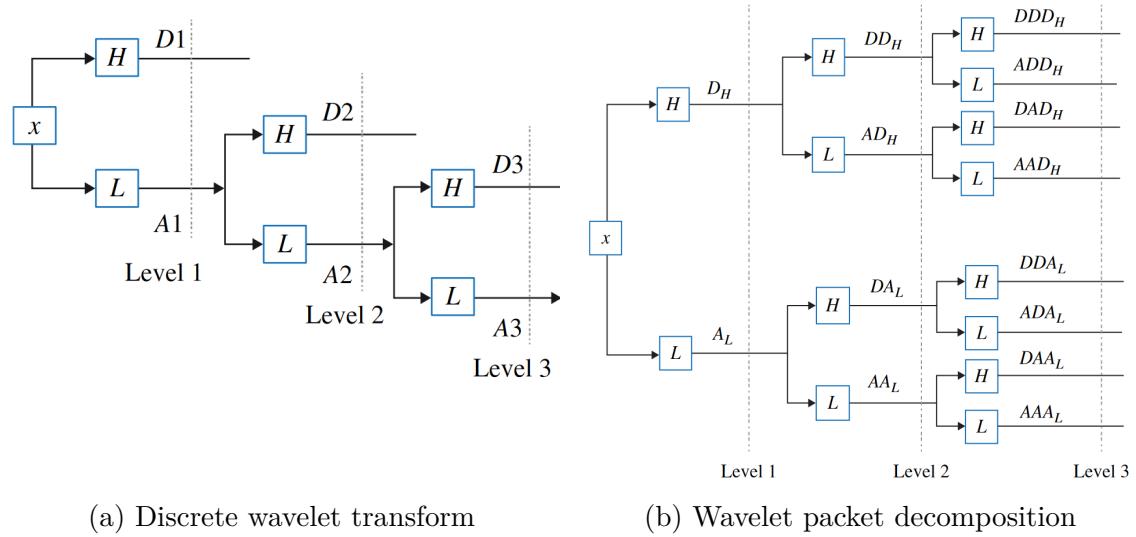


Figure 2.8: Dyadic filter banks for discrete wavelet transform [16]

The wavelet packet energy ratio for weak feature extraction has been incorporated into the method of multiple frequency band demodulation (MFBD). The highest  $n$  energy coefficients are selected out of 8 narrow frequency bands to subsequently affect the principal components. Demodulation principle for the frequency bandwidth prescribes  $n$  to satisfy condition:  $1 / 2^n > f_{\text{modulation}} / f_s$ . The first few eigenvectors that explain together more than 80% of energy are retained to reconstruct the signal with a Fourier transform and retain the weak fault components [32].

Tool wear diagnosis based on acoustic emission (AE) signal considers *wavelet packet energy* in bands  $E_8$ ,  $E_{10}$ ,  $E_{12}$ , *energy ratios*  $P_8$ ,  $P_{13}$ , and *energy entropy* as having high correlation ( $|r| > 0.8$ ) with the band saw flank face width. The acoustic signal is decomposed into three layers using Daubechies db3 wavelet. The bottom layer contains bands numbered 7 through 14, each with a bandwidth of 62.5 kHz, because of the 1 MHz sampling frequency. The feature vector constructed in the article includes other statistical metrics out of power spectral density that has reached a notable correlation with evolving wear. The statistics are *skewness*, *kurtosis*, *shape factor*, and *centroid frequency* [33].

Discrete wavelet transform and wavelet packets partition the spectrum into predefined frequency bands that do not always adequately capture individual elementary oscillations. Adaptive spectral segmentation is needed to extract separate intrinsic mode functions (IMF).

### 2.3.5 Empirical wavelet features

**Empirical Wavelet Transform** (EWT) constructs adaptive bandpass filters with Meyer wavelet. The inner product of the signal with the scaling function  $\hat{\phi}_n$  obtains the approximation coefficients and the inner product with the wavelet function results in detail coefficients.

The normalized frequency axis in range  $\omega \in [0, \pi]$  is divided by split points  $\omega_0, \dots, \omega_N$  where  $\omega_n = f_n \cdot 2\pi / f_s$  (Fig. 2.9). Each segment is bounded between  $[\omega_{n-1}, \omega_n]$  with a transition phase of width  $2\tau_n$  and polynomial transition function  $\beta(x)$ . A tight frame set of empirical wavelets is built by setting the transition phase proportional to the band boundary:  $\tau_n = \lambda\omega_n$ , and proportional constant must obey constrain:  $\lambda = \min\left(\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n}\right)$ . The  $N$  boundaries defining different portions of the Fourier spectrum are placed at the center between two consecutive local maxima [34].

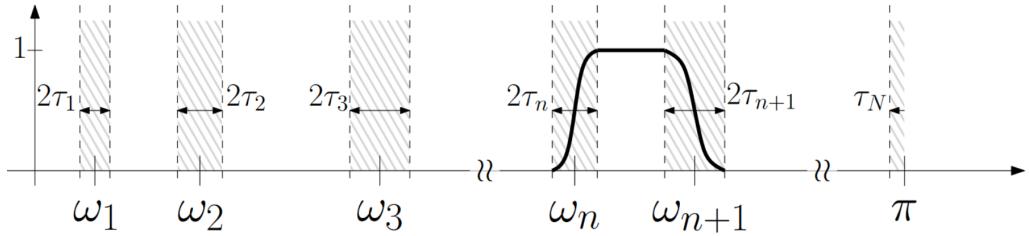


Figure 2.9: Empirical wavelet segmentation of Fourier spectrum [34]

The drawback of EWT is improper segmentation in noisy and non-stationary signals, producing too many uninformative partitions of the spectrum. The solutions we discuss fall into two groups that apply improvements before or after the wavelet transformation step.

A combination of **Maximum Correlated Kurtosis Deconvolution** and improved EWT (MKCD-EWT) favors periodic impacts by dynamically constructing an optimal FIR filter that maximizes the correlation kurtosis of the signal. The envelope curve smooths the amplitude spectrum by linear interpolation (Fig. 2.10a). The threshold accounting for desired SNR modifies the envelope when more than  $n$  components are discovered ( $A_h$  is maximal amplitude and  $A_l$  is minimal amplitude):  $\lambda = A_l + \frac{C}{SNR}(A_h - A_l)$ . *Squared envelope spectrum* and *Teager energy operator spectrum* are computed as fault features for the IMF with the highest kurtosis [35].

The segmentation boundary detection can also be accomplished on the spectral

envelope consisting of **Piecewise Cubic Hermite Interpolating Polynomial** (PCHIP-EWT) instead of scanning the Fourier spectrum directly (Fig. 2.10b). Divisions of the upper cut-off frequency are kept if local power indicates a subband with useful information ( $p_i(f) \geq \lambda$ ) as in the case of an abrupt level change. Local power is the ratio of  $n$ th local envelope maximum and the difference between indexes of its adjacent local minima:  $p_i(f) = K_{max(i)}(f) / (f_{k_{min}(i+1)} - f_{k_{min}(i)})$ . The best  $\lambda$  is set experimentally [36].

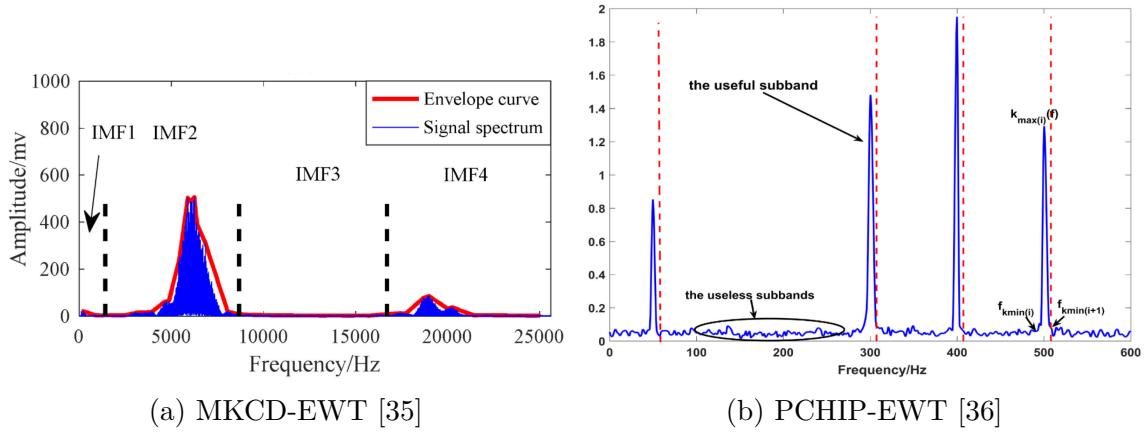


Figure 2.10: Illustration of improved EWT spectral segmentations

The inspiration for filtering on an adaptive basis comes from **Empirical mode decomposition** (EMD). This procedure locates local maxima and minima in the time waveform that are interpolated with cubic splines to form signal envelopes. The average of the upper and lower envelope is repeatedly subtracted from the residuals recovering the higher order and lower frequency IMFs [37].

The EMD is not recommended for practical applications because it suffers from mode mixing problems and lacks rigid theoretical foundations. Many improvements based on the mode sifting process, such as Local Mean Decomposition (LMD), Ensemble Empirical Mode Decomposition (EEMD), Concealed Component Decomposition (CCD) [38], and Empirical Wavelet Transform, have been devised as being capable of estimating more reasonable modes.

Wavelet domain features and associated spectral segmentation appear to be a powerful tool in conserving data bandwidth. These variables turn out to be difficult to communicate and comprehend. Therefore, we will use more standardized metrics in further studies.

## 2.4 Feature transformations

Numerical features from the feature extraction phase have non-normal distributions and span wide ranges. Broad differences among features skew the spread on a particular axis. Inevitably, it can degrade the discernment of fault diagnostics models that map input onto a smooth function as regression does [39]. The feature scaling, power transform, and principal component analysis modify attribute values to gain more meaningful predictors, but one must be cautious in model interpretation.

### 2.4.1 Feature normalization

Vibrations are measured with an accelerometer in the form of 3D vectors where each axis has its own component. Models have to be resilient to any slight inclination or sensor orientation. To satisfy this prerequisite, feature  $f$  extracted from all three dimensions is first composed into a single vector and the Euclidian norm is computed [40].

$$\tilde{f} = \sqrt{f_x^2 + f_y^2 + f_z^2} \quad (2.14)$$

Feature normalization most often takes two forms. **Min-max scaling** changes the original range of values into an interval  $[0, 1]$  (Equation 2.15). **Standardization** (Equation 2.16) constrains the mean of the variable to 0 with a variance of 1 [39].

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.15)$$

$$\tilde{x} = \frac{x - \bar{x}}{\sigma_x} \quad (2.16)$$

### 2.4.2 Principal components analysis

Highly correlated features generated from a relatively small original column space are redundant as they do not provide any additional information for diagnostics. **Principal Component Analysis** (PCA) solves this problem by projecting potentially linearly dependent features into a new feature space where the incoming

information is preserved in a smaller number of features [39]. The threshold of how many principal components are picked depends on the amount of explained variance and desired quantity of data reduction.

$$\mathbf{C} = \mathbf{U}\Sigma\mathbf{V}^T \quad (2.17)$$

PCA consists of taking the Singular Value Decomposition (SVD) (Equation 2.17) of the mean-centered input matrix. The disadvantage of this method is the loss of explainability in transformed space though it generally outperforms the model working with hand-crafted features [18]. The signal samples can be processed directly by PCA without going through an intermediate step of calculating statistical measures.

## 2.5 Feature selection

Features do not contribute to the predictive power of the model with an even share. A certain subset can achieve better results than others. Choosing the optimal feature subset is the NP-hard combinatorial problem.

### 2.5.1 Filtering methods

The features can be chosen intrinsically as a part of a model by *embedded methods* or by machine learning search algorithm at a serious computational expense in *wrapper methods*. However, we will focus on *filtering methods* which rank the predictors in order of their importance for the problem at hand and separate the best performing group [17]. The most common strategy is *K-Best Selection*.

The general steps in selecting the appropriate predictors are as follows [16]:

1. **Subset generation** - sets of features are generated in different search directions and with various strategies. Attributes are either appended to an empty set or pruned away from a universal set, sequentially or randomly.
2. **Subset evaluation** - comparison of subset quality is assessed with relevance measures, some of which are discussed below.

3. **Stopping criteria** - search is exhausted when the specified number of features has been found, subset metrics cannot be improved further, or satisfactory model performance is achieved. Subset generation and evaluation can be performed multiple times until the stopping criteria are met.
4. **Validation** - the resulting subset is tested for the specific model on synthetic and real-world datasets against well-known results.

Filter-based feature selection is preprocessing step independent of model choice with small computational requirements. Measures of information, correlation, similarity, and interdependence output the relevancy rating. Predictors are rated individually or in interacting congregations.

### 2.5.2 Scores for ranking

Most of the scores are based on supervised learning, so they expect true class labels to apportion the measurements respectively. After the scores are assigned to the first  $n$  features, those below a threshold are removed.

The frequently used scores upon which the feature relevance is ordered are [16]:

- **Variance threshold** - removes low-variance features below the set threshold.
- **Correlation coefficient** - expresses the linear relationship between two variables. The codependent variables are of three sorts: quantitative, ordinal, and nominal. The choice of coefficient calculation is determined by the type of variables under consideration shown in Table 2.5.

*Pearson correlation* coefficient expresses similarity between two quantitative features:  $f_i$  and  $f_j$ . In the classification setting, the correlation of feature  $f$  makes sense only with dichotomous target class label  $c$  using *point biserial coefficient*. Mathematically, this is equivalent to the Pearson coefficient. Rank correspondence is quantified with either Spearman rho or Kendall's Tau [41].

Attributes are ranked in descending order according to the absolute value of their correlation coefficient. We seek the highest correlation to the class label.

$$r(i) = \frac{\text{cov}(f, c)}{\sqrt{\text{var}(f) \cdot \text{var}(c)}} \quad (2.18)$$

Variable Y\X	Quantitative X	Ordinal X	Nominal X
Quantitative Y	Pearson $r$	Biserial $r_b$	Point Biserial $r_{pb}$
Ordinal Y	Biserial $r_b$	Spearman $\rho$	Rank Biserial $r_{rb}$
Nominal Y	Point Biserial $r_{pb}$	Rank Biserial $r_{rb}$	Phi, L, C, Lambda

Table 2.5: Correlation coefficients

- **Fisher score** - measures the difference between the means of the classes. It is interchangeable with ANOVA F-value, but it is evaluated for each feature  $X^j$  separately. Ideally, the features in the subset have large distances between samples of various classes in  $C$  and distances within a class are the smallest possible. In the formula (2.19),  $n_j$  is the sample size of  $j$ th feature,  $\mu^j$  is its sample mean, and  $\mu$  is the overall mean.

$$FS(X^j) = \frac{\sum_{i=1}^C n_i(\mu_i^j - \mu)^2}{\sum_{i=1}^C (n_i - 1) \cdot (\sigma_i^j)^2} \quad (2.19)$$

- **Mutual information** - quantifies the dependence between features, or between features and class labels. It is almost identical to Information Gain. The probability distribution of proximity of variables derives from the relative entropy known as the Kullback-Leibler distance. Mutual information presumes variables are discrete. In the case of quantitative variables, mutual information is estimated by binning or nearest neighbors methods [42].

Probabilities  $P(x)$ ,  $P(y)$ ,  $P(x, y)$  are estimated in the contingency table from event occurrence count to all sample population  $|x| / N$ . Joint probability  $P(x, y)$  represents samples of feature  $x$  simultaneously in class  $y$ .

$$MI(X, Y) = \sum_{y \in Y} \sum_{x \in X} P(x, y) \cdot \log \left( \frac{P(x, y)}{P(x)P(y)} \right) \quad (2.20)$$

Multiple subsets of predictors produced by each evaluation metric can train several variants of a classification model. Sets of attributes can be combined into an ensemble by *electoral system*. One such example is **majority voting** which chooses the best feature out of the group. **Rank product** unifies several feature

orderings by computing the geometric mean of feature rank in every experiment realization [43].

## 2.6 Diagnostics techniques

Fault identification in the rotating machinery is a one-class or multi-class classification problem acting in a semi-supervised manner because labels for degraded conditions are scarce in practice. The automation goals in monitoring can be broadly categorized as anomaly detection and recognizing the momentary fault type.

The guiding principles for algorithm selection are simplicity in terms of their straightforward visual explanation for the production managers, and the ability to progressively improve the model on the streaming data to address peculiarities in individual machine constructions.

### 2.6.1 Novelty detection

Anomaly, novelty, or outlier detection determines whether a health status deviates considerably from the baseline profile. The expert can then step in and diagnose the machine after the notice. Anomaly is a rare observation different from the others, raising suspicion that it was created by unrelated behavior [44]. The observations get assigned anomaly scores, and those over the threshold are novelties.

The measurements coming in the steaming fashion have to be processed in a single pass. The detection model must deal with the minimal admissible assumptions about the nature of the input events. The outliers are derived based on non-parametric statistical models, nearest-neighbor clustering, and isolation-based approaches [45].

**DenStream** is a density-based algorithm adapted from DBSCAN to cluster streaming data of arbitrarily shaped groups. Samples it includes in the first step into coherent clusters are core data points in each other's neighborhoods. Core points have at least  $MinPts$  ( $\mu$ ) points in their neighborhood of radius  $Eps$  ( $\varepsilon$ ) units. Then non-core points in the proximity area of the core point are attached to the cluster containing it [46].

Quality of clustering results is evaluated by *Silhouette score* in the range [-1; 1]. It demands points within clusters to have high cohesion and at the same time to have large separation from other clusters. The low score indicates a too small or too large number of clusters [47].

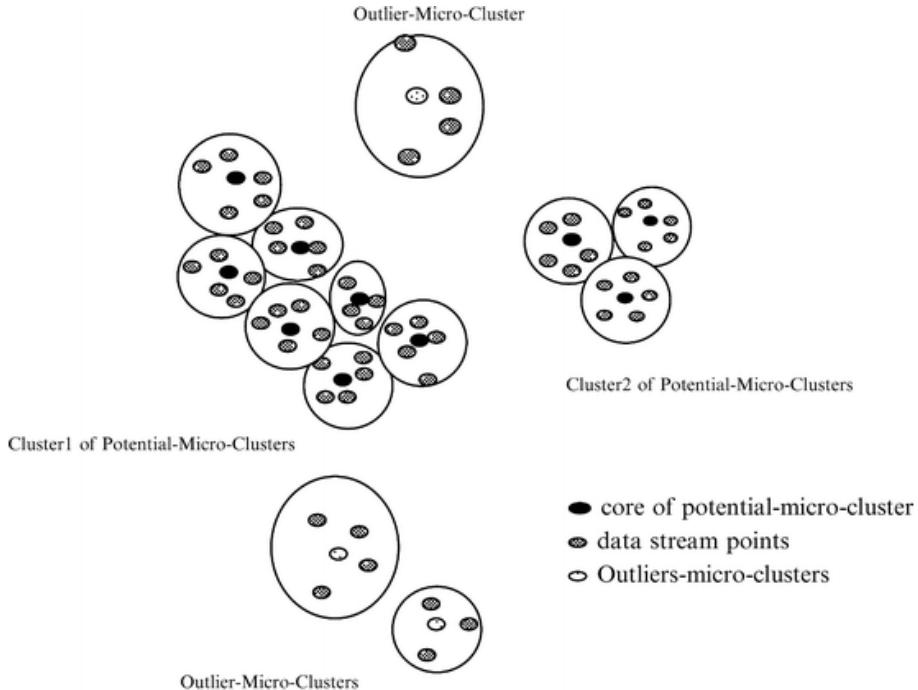


Figure 2.11: DenStream [48]

In the online maintenance phase, DenStream summarizes the nearby observations into core *micro-clusters* that can be potential micro-clusters or outlier *micro-clusters* (Fig. 2.11) [49]. The (outlier) *o-micro-clusters* can grow into (potential) *p-micro-clusters* when they encompass  $\beta\mu$  points. The outliers are discounted after some time in accordance to the decay function:  $f(t) = 2^{-\lambda t}$  or below the lower weight limit  $\xi$ . The on-demand offline stage runs DBSCAN over the approximate representation in micro-clusters to deliver final apportionment [50].

**Half-Space Tree** (HS-Tree) stands upon the concept of Isolation forest. It assumes that random splitting of each axis in the feature space will isolate outliers to their separate divisions sooner than non-deviant observations [45, 10]. This ensemble of trees is better suited for batch setting. HS-Tree stands out in adapting to changing streams because it is trained solely on normal data, requires constant memory, and is faster than density-based methods [51].

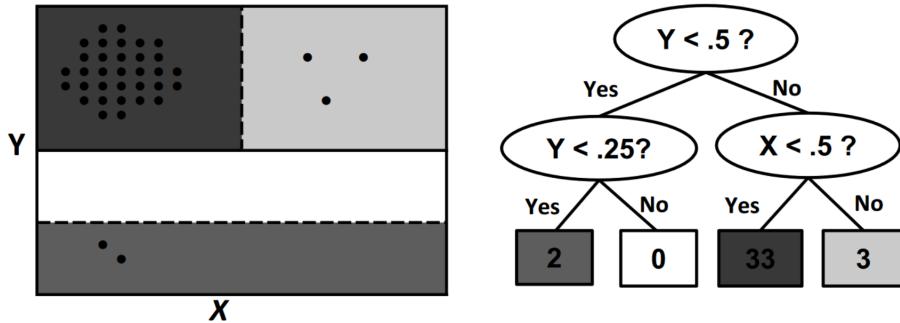


Figure 2.12: Half-space tree [51]

A full binary tree is built before the novelty detection begins by splitting tree nodes along the divisions in the randomly chosen perpendicular planes. The node stores its depth, value limits of the axis bisection (half-space), count of contained data points (mass) in two consecutive windows, and links to both child nodes [51].

The anomaly profile in the latest window is always compared to the predecessor reference window. After the latest window is filled up, it replaces the reference window. It suffices to use a window size of 250 and an ensemble of 25 trees [51].

### 2.6.2 Classification

Accurate multi-class classification of machine fault causes according to the characteristics of known ones is a much more difficult task than novelty detection. Fault combinations have to be recorded and transformed into feature space. Interactions among fault root causes have to be considered. We are aware of rapid advances in knowledge transfer for deep neural networks [52]. So far, solutions seem not production-ready and hard to explain. Therefore, we opt to use a simpler model.

Performance of classification is estimated by several metrics on the validation set obtained using hold-out or cross-validation techniques. Frequently used quantities for classifier model evaluation include accuracy, precision, recall, f1 score, area under the ROC curve, and counts of hits and misses in a confusion matrix.

**K-nearest neighbors** (KNN) assigns the data point to the class where the majority of  $k$  closest instances belong (Fig. 2.13a). It means it can work in a semi-supervised environment because it can infer labels just from knowing a few annotations. The major drawback of KNN is a preference for the majority class

in imbalanced class-size datasets [53]. The issue is mitigated with class weights or resampling classes by oversampling or undersampling. The KNN algorithm requires features to be normalized to assign the same importance to each predictor.

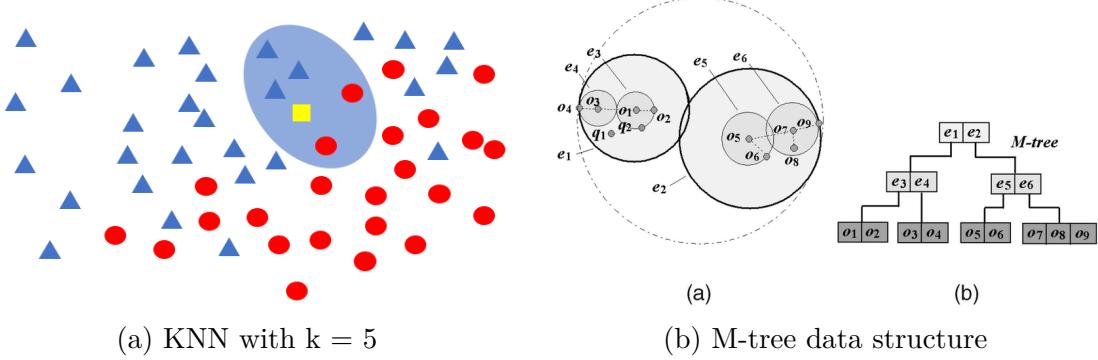


Figure 2.13: Nearest neighbors classification algorithm [54]

The sense of distance between feature vectors  $\mathbf{x}, \mathbf{y}$  have to be defined, so several metrics are available like *Euclidian distance*, *Mahalanobis distance*, *Manhatann distance* etc. (Tab. 2.6) [55, 56]. The optimal  $k$  parameter is set by supervised learning according to the breaking point in the elbow curve that plots choices of  $k$  against the error rate. The demanding neighborhood queries are sped up using spatial index in spatial databases that utilizes search tree such as kd-tree, R-tree, or M-tree.

Distance	$d(\mathbf{x}, \mathbf{y})$
Manhatann distance	$ x_i - y_i $
Euclidian distance	$\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Mahalanobis distance	$(\mathbf{x} - \mathbf{y})^T C^{-1} (\mathbf{x} - \mathbf{y})$

Table 2.6: Distance metrics for KNN

The nearest-neighbor classifier has been successfully applied in machinery fault diagnostics. On the CWRU bearing dataset, the KNN with the accuracy of 96.2% slightly outperformed SVM (95%) on the combination of time and frequency-domain features, time-domain features - KNN 91.2%, SVM 88.8%, and frequency domain features - KNN (98.8%), SVM (96.2%) [57].

Comparison of KNN and KLDA on a feature set consisting of average, kurtosis, skewness, and standard deviation vectors in each domain has been conducted,

achieving a data reduction rate of 95%. Best accuracy was reached for PSD features with 99.13% with KLDA and 96.64% with KNN classifiers and Mahalanobis metric. The sampling frequency was set at 40 kHz [58]. Despite KNN lagging in accuracy, we have to keep in mind annotations for faults were complete, and machine learning was not tested in a streaming context.

### 2.6.3 Incremental learning

Online or incremental machine learning operates on the streaming data, updating the model parameters with each new incoming event or in mini-batches. This approach finds its use in big data processing when the whole dataset is not available in advance or cannot be processed at once because of memory limitations.

There are some additional obstacles to watch out for with incremental learning in comparison to batch learning [59]:

1. **Concept drift** is defined as the change in data distribution function over time. The two types of concept drift are virtual and real. In virtual drift, changes occur only in the input distribution. Real drift means that the alteration comes to underlying functionality. Concept shift occurs with an abrupt change.
2. **Stability-plasticity tradeoff** concerns the speed with which the model adapts to new information. The model can react quickly, making it less stable, or retain patterns for longer but become irresponsive to sudden shifts.
3. **Model complexity** should be adjustable to ensure flexibility in unforeseen circumstances. Simpler models in the ensemble can also further increase prediction robustness. Resource limitation bound complexity from above.
4. **Memory model** can store aggregates from seen observations and its typical examples or finite window of latest samples with forgetting factor.

**Model benchmarking** in incremental learning is achieved by comparing models to their batch counterparts or using progressive validation [60, 61]. It has been shown that incremental clustering algorithms have overall worse accuracy than batch versions [59]. In the validation process, precautions should be taken to prevent data leakage from future events into the past.

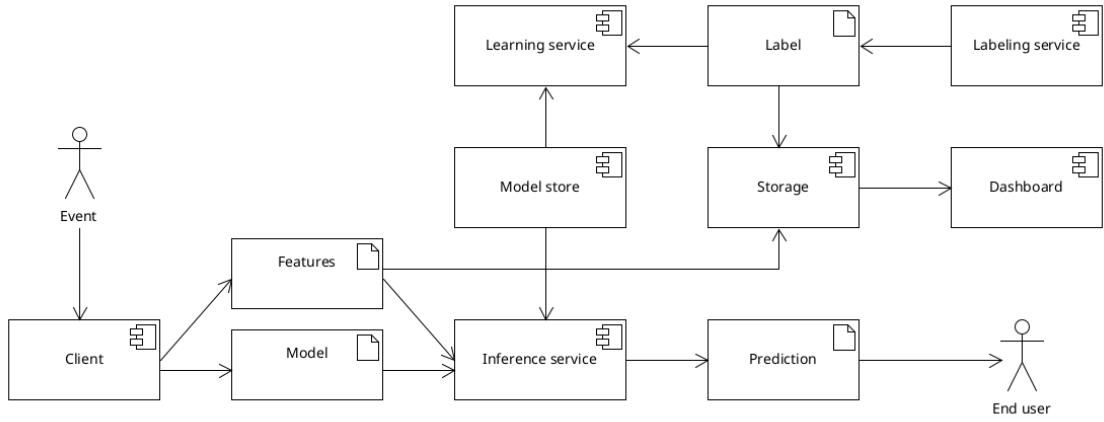


Figure 2.14: Incremental learning deployment architecture [62]

Deployment of an online machine learning model alongside supporting services is different from in established MLOps processes. Model store and Inference service components are supplemented with Labelling and Learning service [62]. Their goal is to tune model parameters gradually as additional ground truth labels are provided. Labels can be provided later in a scheme called “log and wait”, but data features are stored until such time.

## 2.7 Evaluation datasets

The experimentally designed features’ relevancy is first proven in comparison to comprehensive benchmark datasets. There are a few standardized datasets used in the related work, e.g. [63].

*MaFaulDa* dataset combines vibration and acoustic measurements of the shaft in deviating positions and bearing abnormalities. *CWRU* dataset focuses solely on faults in ball bearings. Another less known dataset concerns shaft unbalance, but compared to the previous two, it demonstrates behavior during speed up.

### 2.7.1 Machinery Fault Database

*MaFaulDa*<sup>1</sup> is a collection of 1951 multivariate time series for 4 different operational conditions on rotor kit Alignment Balance Vibration Trainer (ABVT) (Fig. 2.15).

---

<sup>1</sup>[https://www02.smt.ufrj.br/~offshore/mfs/page\\_01.html](https://www02.smt.ufrj.br/~offshore/mfs/page_01.html)

Each series has 5 seconds in duration and is captured at 50 kHz. Vibration signals were obtained with piezoelectric accelerometers with a linear response up to 10 kHz, amplitude range to  $\pm 490 \text{ m/s}^2$ , and resolution step of 10.2 mV per  $\text{m/s}^2$ .

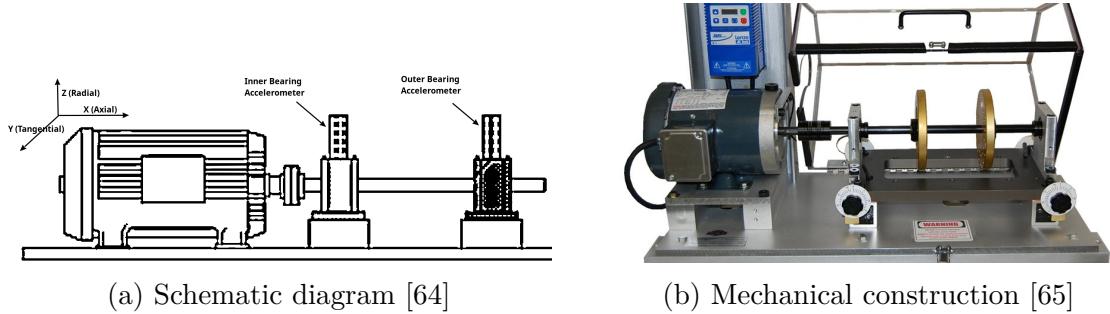


Figure 2.15: Machinery fault simulator for MaFaulDa

Observations were conducted on three cardinal axes simultaneously with 2 sets of accelerometers, each one associated with one bearing (inner and outer bearings) (Fig. 2.15). Additionally, a magnetic speedometer produced a pulse on shaft turn. The cardioid condenser microphone recorded sound emissions at a frequency range 20 Hz - 20 kHz. Sensors were fed into a four-channel dynamic signal acquisition module.

Columns in the dataset are organized as depicted in table 2.7. Machine rotational speeds were kept constant during a particular measurement, but covered a range from 737 to 3686 rpm with steps of approximately 60 rpm (equiv. 10 Hz - 60 Hz) [64]. The maximal rotational frequency achieved with a high unbalance load is 3300 rpm.

Columns	Description
1.	Pulse with modulation of speedometer signal to estimate rotation frequency (in TTL levels)
2., 3., 4.	Underhang bearing accelerometer (inner - between the rotor and motor) - axial, radial, tangential direction
5., 6., 7.	Overhang bearing accelerometer (outer - outside most position after the rotor) - axial, radial, tangential direction
8.	Microphone

Table 2.7: MaFaulDa description of columns

This database contains normal operating conditions, faults out of unbalance,

horizontal and vertical shaft misalignment, and three types of faulty bearings in inner and outer positions: outer track, inner track, rolling elements [64].

- **Normal** conditions are baseline without the adverse effect of fault at 49 different rotation speeds.
- **Unbalance** shaft time series uses 8 unbalancing weights from 6 to 35 grams and varying 45 - 49 speeds for each weight, adding to 333 mass unbalance loads.
- **Vertical misalignment** set comprises 50 signals each (or 51 in one instance) obtained under displacements: 0.51, 0.63, 1.40, 1.90, 1.27, 1.78 mm.
- **Horizontal misalignment** signals were recorded under displacements: 0.50, 1.00, 1.50, 2.00 mm, each with 49 different speeds (or 50 in one instance) [64].
- **Bearing faults** are unnoticeable without unbalance. Therefore, weights of 6, 20, and 35 grams were attached to induce a detectable effect. Each mass was combined with cage, outer race, and ball faults at multiple rotation speeds, usually at 50 different speeds.

### 2.7.2 CWRU bearings dataset

In Case Western Reserve University (CWRU) bearing dataset<sup>2</sup> recordings were made of a fan end and drive end bearings under motor loads of 0, 1, 2, and 3 Horsepower (equivalently 0, 0.75, 1.49, 2.24 kW). Shaft speed was unaltered in all experiments, but it fluctuated between 1720 and 1797 rpm (approx. 29 Hz).

Single point defects were created with diameters of 0.007, 0.014, 0.021, 0.028, and 0.040 inches (equivalently 0.18, 0.36, 0.72, 1.02 mm). Fault locations on bearings are in the inner raceway, in the outer raceway directly and orthogonally relative to the load zone, and on rolling ball elements (Fig. 2.16) [57].

The sampling frequency during baseline set, drive end, and fan end bearing capture is 12 kHz. For drive end bearings, samples were taken at 48 kHz. The duration of the time series varies from 5 to 40 seconds. Drive end and fan end

---

<sup>2</sup><https://engineering.case.edu/bearingdatacenter/download-data-file>

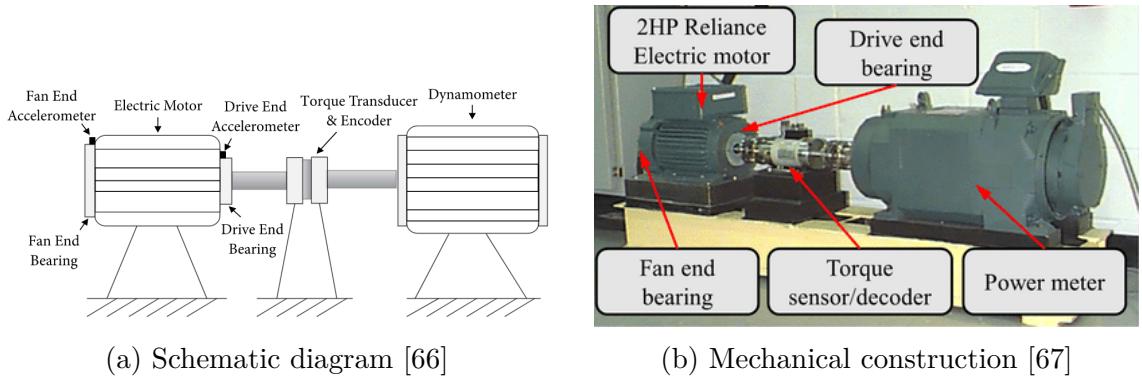


Figure 2.16: CWRU machine apparatus

Columns	Description
1. DE	Drive end accelerometer samples
2. FE	Fan end accelerometer samples
3. BA	Base accelerometer samples (optional)
4. RPM	Rotation speed of the motor in rpm

Table 2.8: CWRU dataset description of columns

bearing signals are measured in each experiment. Accelerometer was sometimes mounted on the supporting baseplate.

### 2.7.3 Unbalance of the rotating shaft

Unbalance Detection of a Rotating Shaft<sup>3</sup> is a Kaggle dataset that simulates 4 different unbalance strengths.

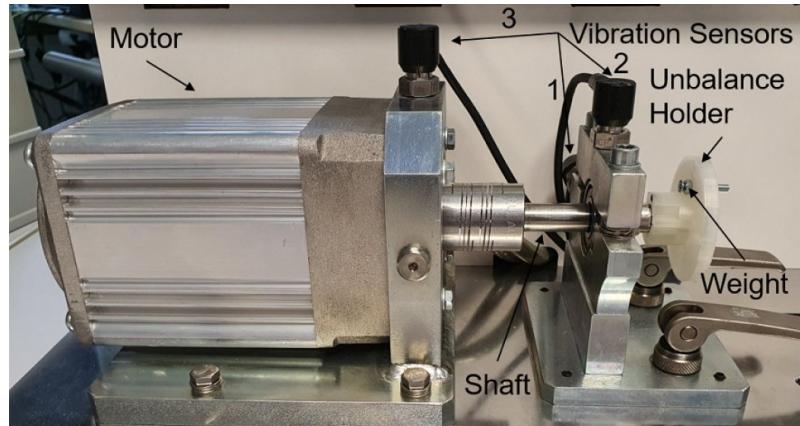


Figure 2.17: Motor driving shaft in unbalance measurement [68]

<sup>3</sup><https://www.kaggle.com/datasets/jishnukoliyadan/vibration-analysis-on-rotating-shaft>

The setup is shown in Fig. 2.17. A mass of 3.28 grams (or 6.61 grams during the severe unbalance test) is attached to the unbalance holder in 5 sets (numbered 0 - 4) on the radii 0, 14, 18.5, 23, 23 mm. The rotation speed of the motor is perpetually rising between 630 and 2330 rpm in development datasets (marked with suffix D) and speeds from 1060 to 1900 rpm in the evaluation datasets (suffix E). The vibrations were recorded at a sampling rate of 4 kHz [68].

Columns	Description
1. V_in	Input voltage to the motor controller (V)
2. Measured_RPM	Rotation speed of the motor (rpm)
3. Vibration_1	1. Vibration sensor (samples)
4. Vibration_2	2. Vibration sensor (samples)
5. Vibration_3	3. Vibration sensor (samples)

Table 2.9: “Unbalance on the rotating shaft” dataset description of columns

The accelerometers used are piezoelectric and have a frequency range of up to 10 kHz, dynamic range of  $\pm 490 \text{ m/s}^2$ , and resolution step of 10.2 mV per  $\text{m/s}^2$ . These sensor parameters are the same as in the case of MaFaulDa. In total, three different uniaxial accelerometers are mounted on the motor housing.

# 3 Design

The design phase sets out to elaborate on the scenarios for experiments with IoT device. In these experiments, we compare and combine approaches from industry standards and papers presented in the analysis section. Conclusions are taken into consideration in establishing the plan of measurements and in the construction of the sensor unit.

## 3.1 Research questions

This thesis aims to provide answers to four research questions. The focus is primarily on making data flow more efficient in an industrial sensor network that monitors rotating machines. The **research questions** are:

- RQ1.** Which temporal and spectral features can be extracted from vibration signals to provide the most accurate record of machinery faults?
- RQ2.** What is the reduction in transmission goodput when chosen signal features are used?
- RQ3.** What accuracy of prediction models can be achieved with various feature subsets?
- RQ4.** How can machinery faults be continuously identified and predicted based upon collected events?

In accomplishing the objectives of our research we propose several **goals**:

- Statistically and visually describe vibration signals from the Machinery fault database (MauFaulDa).
- Establish a list of conditions that should be later investigated in the experimental setting.

- Prepare dataset to be used in conjunction with machine learning models, namely by identifying labels and balancing classes.
- Find the best subsets of features in the temporal and spectral domain with previously analyzed feature extraction and selection methods.
- Evaluate the performance of models described in the diagnostics section with a significant focus on the k-nearest neighbor algorithm.
- Combine feature selection with an online machine learning model.
- Acquire measurements of vibrations from machines in the real environment to form a novel dataset of machinery behavior.
- Develop hardware and implement its firmware to obtain such measurements of the quality demanded by vibrodiagnostics standards.

However, we leave out from our efforts experiments on the data features calculated from wavelets and peaks in the spectral domain. The reason is that we did not find a way to represent extracted features more succinctly as a single number. We also did not discover a strategy for choosing only the relevant frequency bins. The assembled description is retained to lead further research on that topic.

The goals are impacted by certain **risks** that are assessed and tracked:

**R-1.** Machines in the real environment will not be available for vibration measurements.

*Mitigation:* Contact and establish collaboration with alternative partners.

**R-2.** Repeated measurements will not be consistent, and a lack of data is obtained from different classes.

*Mitigation:* Prepare a plan for measurements and photo document sensor placements.

**R-3.** Fault modes can not be reliably differentiated and labeled in the dataset.

*Mitigation:* Consult domain experts and machine maintenance staff.

**R-4.** Suggestions made by exploring the MaFaulDa dataset will not be applicable in practice.

*Mitigation:* When applicable, apply the same procedures as in MaFaulDa to measured data.

## 3.2 Dataset exploration

In establishing the viability of methods to be deployed on the sensor node, we explore the MaFaulDa dataset. It is the largest known machinery fault collection, so it is possible to create multiple subsets based on required conditions.

One representative recording is first selected in each available fault category. The sample is visualized and statistically described in both temporal and spectral domains. A whole step-by-step procedure is outlined in the activity diagram (Fig. 3.1).

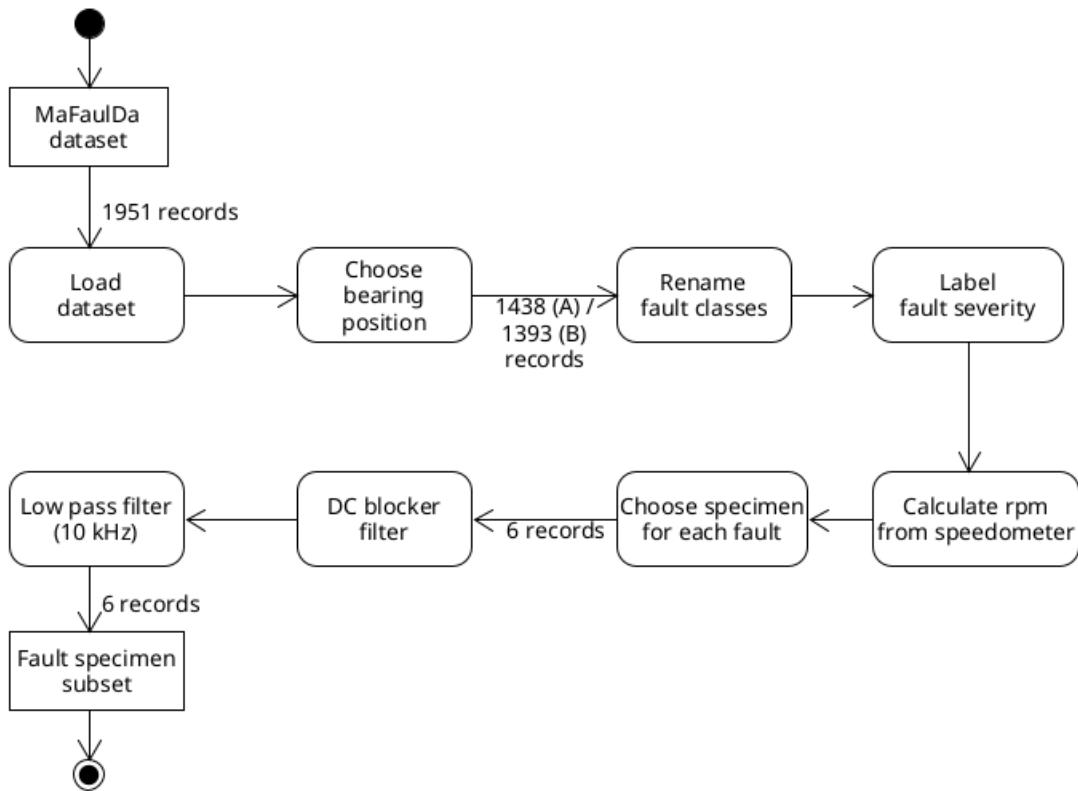


Figure 3.1: Activity diagram of MaFaulDa dataset preprocessing

MaFaulda contains 1951 records labeled with induced faults of increasing severity. The defects were set up on the machine simulator as is mentioned in the part about datasets. Time series of the triaxial piezoelectric accelerometers in separate files have a sampling frequency of 50 kHz.

These vibration sensors are placed in two positions. The first placement is around the inner underhang bearing named *A* which is closer to the motor. The second location is around the outer overhang bearing denoted as *B* position.

### 3.2.1 Fault annotations

The MaFaulDa has annotations altogether for 10 classes of faults, of which there is 1 class for fault-free baseline operation, 3 classes for shaft defects, 3 classes for inner bearing defects, and 3 for outer bearing defects. Some categories are redundant or irrelevant for a given sensor position.

Therefore, rotor shaft misalignment in vertical and horizontal directions are merged into one joint group. Depending on the chosen bearing position, only records having relevant labels are considered.

This means that fault classification solely concerns bearing in direct contact and shaft mechanically passing through it. The bearings affect each other, but the effect on the opposite side should appear via a common interconnection shaft.

In the end, that leaves **6 types of labels**: baseline, two shaft faults are imbalance and misalignment, and three bearing faults are cage fault, ball fault, and outer race fault. In the step of choosing the accelerometer location, records pointing to the other bearing as a source of the malfunction are discarded. Then the next action renames the labels to be better recognizable and unite the same phenomena.

Groups of identified machine defects are additionally characterized by altered masses attached or motor shaft displacement shifts. The set amount is sorted in ascending order, separating **multiple event severities**. However, the count of severity levels is not identical in every group. Levels are hence scaled into the range between zero and one using a min-max scaler. Scaling is applied to classes separately.

The strength of the recorded response to the underlying defect is also dependent on the shaft **rotational speed**. Speed in rpm units is calculated from pulsed speedometer output. It is the average distance between two successive rising edges:

$$\text{rpm} = 60 / \overline{\Delta t} \quad (3.1)$$

**One specimen waveform** is picked from each fault class to illustrate their superficial differences. Recordings are filtered to get the highest severity levels and around a mean rotation speed of 2500 rpm (42 Hz) to see the patterns most pronounced. The baseline class sample is chosen according to fixed rotor speed.

### 3.2.2 Signal filters

The DC component in the three-dimensional vibration signal is removed by subtracting the global mean. Immediately follows a digital IIR Butterworth **low pass filter** of 5<sup>th</sup> order with cutoff frequency 10 kHz at -3 dB.

Before the low pass filter usage, the peak at 20 kHz with a sideband was present as an unwanted artifact. It could not have been reliably recorded due to the linear frequency response of the sensor up to 10 kHz. At the same time, such a frequency is outside the range of any feasible MEMS accelerometer.

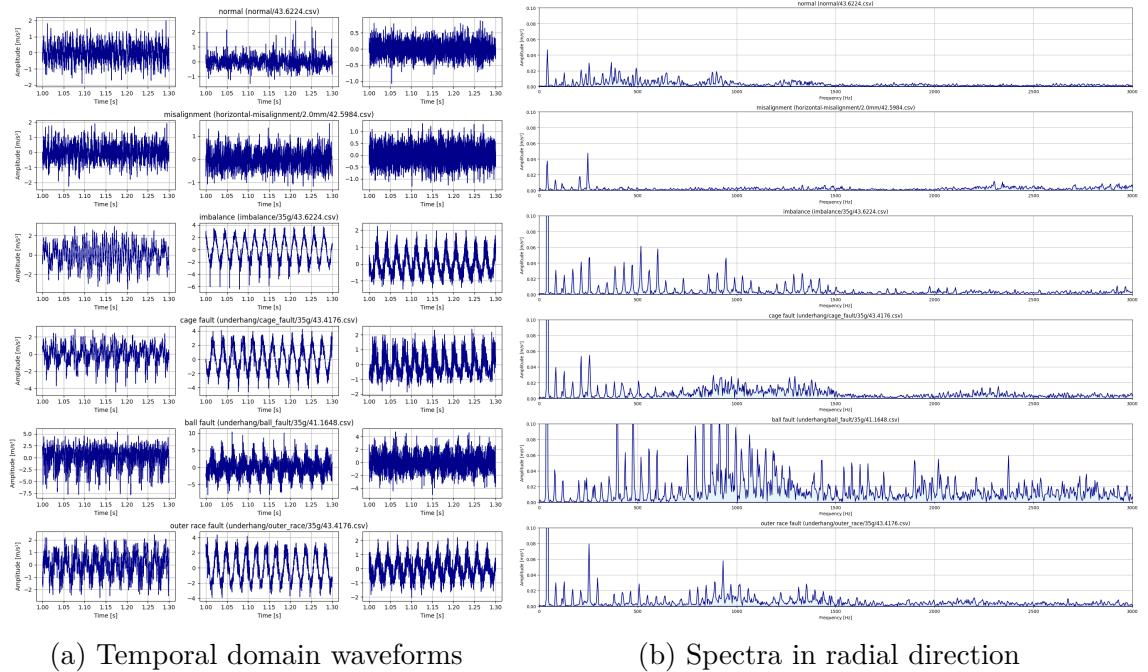


Figure 3.2: Inner bearing vibrations (A) for each fault category with the highest fault severity at 2500 rpm

Temporal domain waveforms of the 300 ms signal slice are shown in the graphs in Figure 3.2a. Subplots for radial, tangential, and axial directions are laid out in columns from left to right. Amplitudes vary with limits from  $\pm 3 \text{ m/s}^2$  in baseline and misalignment time series up to  $\pm 11 \text{ m/s}^2$  in case of severe bearing faults.

The frequency spectrum in Figure 3.2b is obtained by FFT and Hann window of length  $2^{14}$ . The signal chunk represents an uncertainty box with a duration of approximately 328 ms and a spectral resolution of little over 3 Hz. The graph has been cropped in both axes to make the most important peaks visible.

### 3.2.3 Statistical tests

The statistical tests and visual checks are conducted to assess **normality and stationary** of time series. Half a second of amplitude samples are used from every sensor channel. These 25 thousand observations are downsampled tenfold to 2500.

**Shapiro-Wilk's test** rejects the null hypothesis ( $p < 0.05$ ) that data is drawn from a normal distribution under most circumstances. The signal has normal distribution when it resembles pink noise lacking a regular pattern or weak exhibition of fault symptoms. **Quantile-quantile plots** confirm non-normal distribution because of the striking samples tilt to the diagonal line.

**Augmented Dickey-Fuller test** rejects the null hypothesis of unit root ( $p < 0.001$ ). The same is confirmed with the **autocorrelation** function shape. It denotes that the stochastic process is stationary as the oscillation is bounded.

## 3.3 Feature relevance

Attributes described in section 2.3 are independently summarized from each sensor position and direction. Then similarity score of features is ascertained in relation to a predicted variable. The subset of the strongest predictors is chosen based on the order of their perceived importance.

### 3.3.1 Feature extraction

Each file in the MaFaulDa contains six accelerometer channels. In the feature extraction process, the sequence of samples is split into five parts or 1-second intervals. The portions are passed through the same DC removal and low pass filters as previously. The rotational speed is derived from speedometer pulses within the chunk.

Signal chunks are converted afterward into **10 temporal and 11 spectral features**. Welch's method for spectrum density estimation averaging over  $2^{14}$  long FFT segments after Hann windowing is the source for spectral features. The reference implementation of feature calculation is crafted according to mathematical formulas atop of Python packages *SciPy*<sup>1</sup> and *Time Series Feature Extraction Library*

---

<sup>1</sup>SciPy: <https://scipy.org/>

$(\text{TSFEL})^2$ .

The Euclidean norm of feature in the triaxial vector eliminates reliance on the direction of measurement. Value ranges are depicted in Figure 3.3.

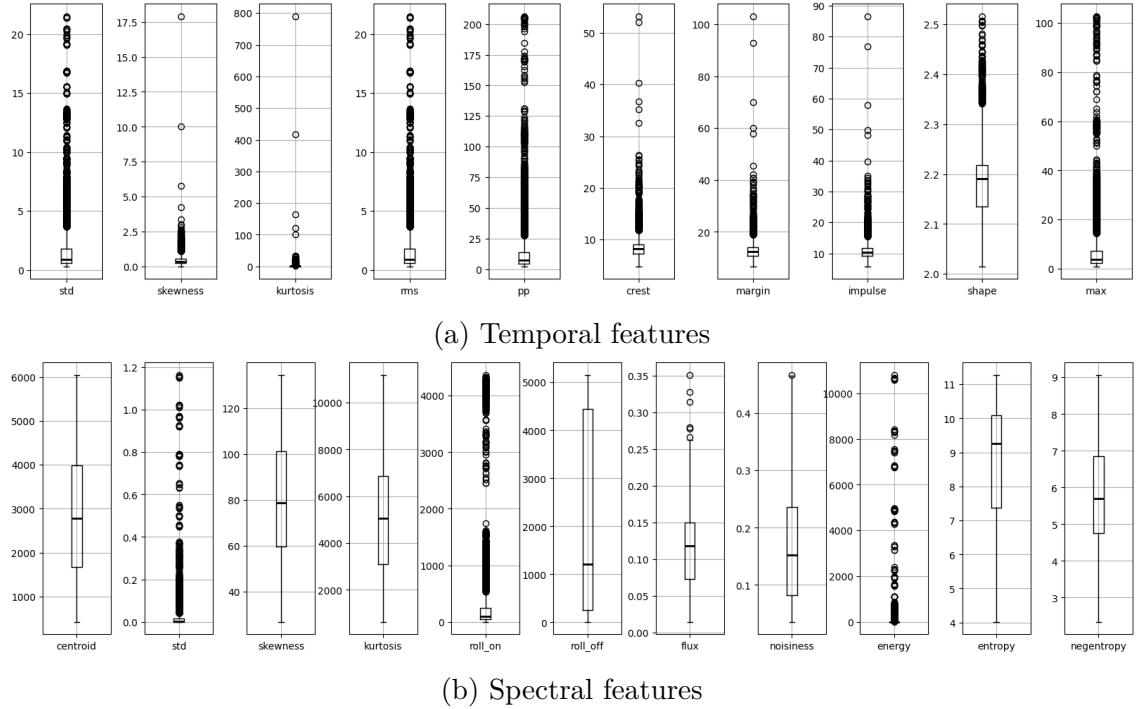


Figure 3.3: Feature value ranges in inner bearing position (A)

Fault labels and their severities come from the directory structure within the dataset. The binary target variable indicating whether to initiate a warning is called an anomaly. Anomalies are labeled according to relative fault severity level. We decided to investigate two fault severities having levels above 0.6 and 0.9. The quantity of observation by fault and anomaly severity is shown in Table 3.1. The dataset is substantially unbalanced.

Pearson's correlation of features to rpm is very low in the whole dataset. In the temporal domain, the correlation coefficient is within an interval of -0.08 to 0.26. In the spectral domain, the correlation to rpm for all FFT window sizes from  $2^8$  up to  $2^{14}$  is mostly very low from -0.14 up to 0.26, except for the centroid being around 0.35.

The correlation among features can reduce prediction power if a pair is elected where  $|\text{corr}| > 0.95$ . The feature is not added to the subset when the threshold

---

<sup>2</sup>TSFEL: <https://tsfel.readthedocs.io/>

Fault	Inner bearing (A)	Outer bearing (B)
<b>normal</b>	49 (3%)	49 (4%)
<b>misalignment</b>	498 (35%)	498 (36%)
<b>imbalance</b>	333 (23%)	333 (24%)
<b>cage fault</b>	188 (13%)	188 (13%)
<b>ball fault</b>	186 (13%)	137 (10%)
<b>outer race fault</b>	184 (13%)	188 (13%)
<b>Anomaly (&gt; 0.6)</b>	<b>Inner bearing (A)</b>	<b>Outer bearing (B)</b>
<b>False</b>	837 (58%)	831 (60%)
<b>True</b>	601 (42%)	562 (40%)
<b>Anomaly (&gt; 0.9)</b>	<b>Inner bearing (A)</b>	<b>Outer bearing (B)</b>
<b>False</b>	1227 (85%)	1197 (86%)
<b>True</b>	211 (15%)	196 (14%)
<b>Total</b>	<b>1438 (100%)</b>	<b>1393 (100%)</b>

Table 3.1: Label count for whole MaFaulDa dataset with recordings split to 5 chunks

is exceeded. High correlations are more substantial in the temporal domain than are present in these pairs (ordered from the most correlated): {std, rms}, {pp, max}, {crest, margin}, {impulse, std}, {impulse, rms}. In the spectral domain set {skewness, kurtosis} has a strong correlation.

It is assumed that data points spread throughout each dimension of the feature space could distinguish groups well. The variables that are the best explained after min-max scaling total dataset variance in the temporal domain are shape (29%), rms, std, max, and p-p (each around 15%). In the spectral domain, variance is best explained by roll-off (28%), entropy, skewness, centroid, and kurtosis (each around 12%).

The variables are also more inter-correlated in the temporal domain shown by principal component analysis. For 95% of the explained variance of PCA 3 components (98.69%) are needed in the temporal domain whereas 4 in the spectral domain (95.26%). Figure 3.4 visualizes cumulative explained variance with an increasing number of principal components.

PCA efficiently expresses attributes in less dimensional space, but the resulting linear combination is hard to comprehend for explaining decisions. Loading plots of PCA (Fig. 3.5) illustrate correlations of features to two principal components. The

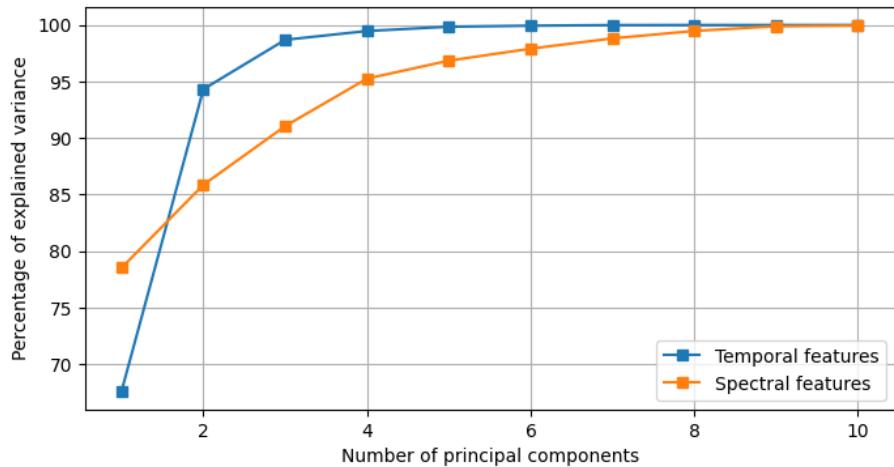


Figure 3.4: Number of principal components to cumulative explained variance percentage in the inner bearing position

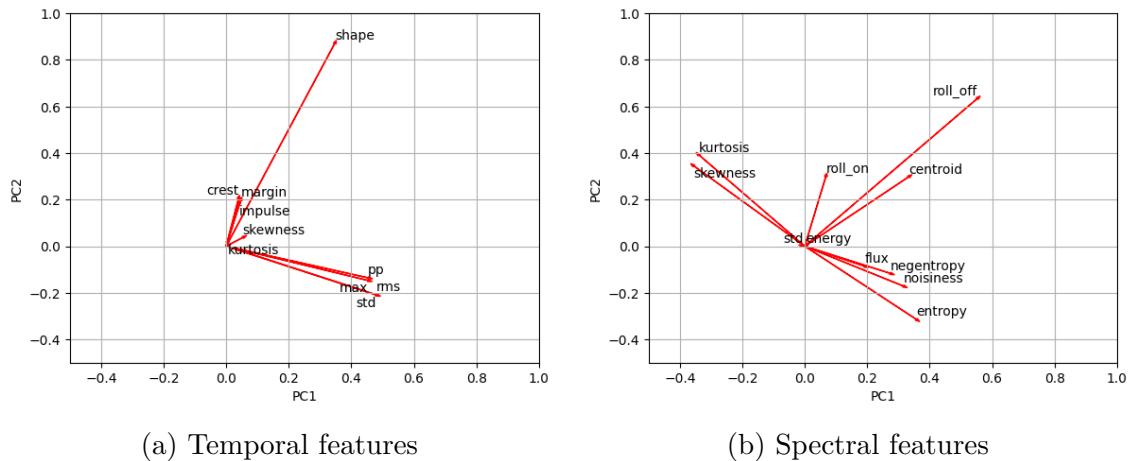


Figure 3.5: PCA loading plots for min-max scaled features from the inner bearing position

first PC in the temporal domain focuses more on the amplitude range: *max*, *rms*, *pp*, *std*. The second PC mainly describes the impulsiveness of the waveform: *shape*, *impulse*, *crest*, *margin*.

However, the groups are not as clear-cut for spectral features. Overall chaos in spectra can be attributed to PC1: *flux*, *entropy*, *negentropy*, *noisiness*, and the shape of frequency distribution to PC2: *roll-on*, *roll-off*, *centroid*.

### 3.3.2 Data volume savings

The apparent advantage of feature discovery is reducing the amount of data downstream. Data compression must occur on edge devices to enable the utilization of

wireless low-power wide area networks (LPWAN). The protocol stack may differ, so goodput is compared without node configuration metadata and keepalive messages.

The machinery monitoring system relies on determining several parameters:

- **Number of source channels ( $S$ ):** comprises the number of monitored machines, measurement locations for sensors, and active sensor axes.
- **Sampling frequency ( $f_s$ ):** is set based on the linear response of the accelerometer, the types of faults intended for detection, and how soon they should be noticed after they arise. The higher required sensitivity means a higher sampling rate derived according to the Nyquist theorem. At a minimum, it should be 15 kHz to 20 kHz.
- **Interval between successive measurements ( $T$ ):** specifies the minimal response time to sudden failure. The more critical the machine is, the interval should be shorter. The bigger the machine parts, the slower the defect evolves.
- **Duration of valid recording ( $D$ ):** is the captured snapshot of machine unaltered behavior associated with a timestamp. Duration should cover at least 3 windows for spectral estimation. The spectral resolution of 1 Hz amounts to 3 seconds of signal under such assumptions.
- **Number of extracted features ( $F$ ):** are ideally key trend indicators pointing to symptoms of common malfunctions. We aim for a total of 6 features.

Equation 3.2 expresses the lossy compression ratio ( $\mathcal{C}$ ) formula if trend indicators are stored instead of full recording. The number of raw channels ( $S_{\text{in}}$ ) can differ from those extracted in features ( $S_{\text{out}}$ ). Parameter  $D = 0.5$  when we use frequency bins with 1 Hz resolution.

$$\mathcal{C} = \frac{D \cdot f_s \cdot S_{\text{in}}}{F \cdot S_{\text{out}}} \quad (3.2)$$

**Compression ratio** for MaFaulDa dataset compared to all 21 extracted features in 3 dimensions is 2381:1. If 6 features are kept, the compression is 25000:1, which is a saving of the original data by 99.996%.

As an example to approximate required network goodput and storage in practice, we consider continuous vibration **monitoring for municipal water pumping**

**station.** The station has 3 pumps and 3 electric motors.

A pump and motor pair have 4 bearings together for drive end and non-drive end positions. Each position has a sensor mounted in 3 directions that makes a total of *36 source channels*. The sampling frequency at each position is set to *20 kHz*. The recordings have *duration of 5 seconds* and are triggered regularly every 1 hour (*8760 times per year*).

In a year, the system gathers 31.54 Gs (gigasamples) which is 58.74 GiB with a 16-bit ADC resolution. Reasonably precise spectral estimation with 10 thousand bins needs 3.15 Gs per year. On the other hand, 6 features out of each channel keep only 1.89 Ms per year for a lossy compression ratio of 16667:1. Low data volumes potentially enable feature selection and models to be offloaded directly to edge devices. The entire machine’s history can be kept in a small flash memory module.

### 3.3.3 Feature selection

The objective of feature selection is to find a subset of the most relevant predictors in each domain. As a starting point, the set consists of 3 non-correlated attributes under diverse conditions. Ultimately, a number of chosen features are to be tweaked to increase prediction accuracy. Four criteria are combined to put together 24 scenarios that filter rows from MaFaulDa:

- **online:** a boolean selector that influences whether metrics are calculated in batch or incremental fashion,
- **placement:** chooses between bearing “A” and bearing “B”,
- **predicted variable:** select a column out of “fault”, “anomaly60”, “anomaly90” against which similarity is judged,
- **rotational speed limit:** chooses whether only observations occurring within predetermined range  $2500 \pm 500$  rpm are filtered.

In feature evaluation for fault prediction, we use hold-out validation in batch models and progressive valuation in online models. Classes in observations for batch models are rebalanced to the majority class with a random oversampling strategy.

The training and testing set split ratio is 80 to 20. Classes for incremental learning are ordered by relative severity level and shuffled within levels.

The best group of attributes is elected based on a training set with multiple methods. We compute the mean of the absolute value from point-biserial correlation, F statistic, and mutual information to the predicted variable. The features are then ordered in descending order by the received score. These individual ranks are combined by rank product to create an ensemble out of the metrics. Reference implementation of feature selection metrics uses Python packages Scikit Learn<sup>3</sup> in batch learning and RiverML<sup>4</sup> for online learning.

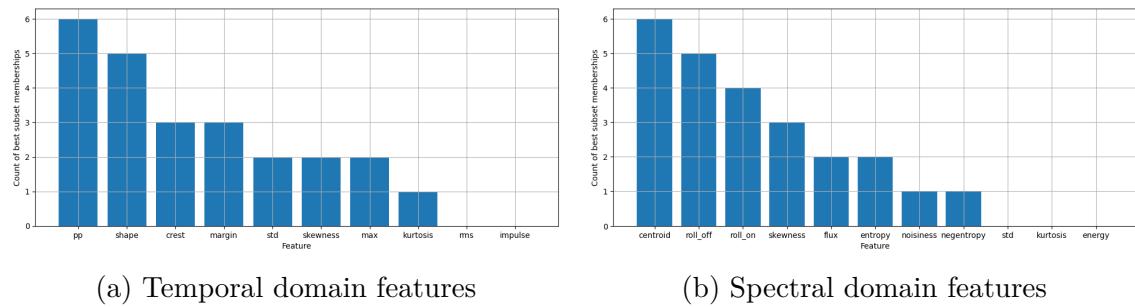


Figure 3.6: Approval rating of the best triplet of features to target variable “fault” scored by rank product of correlation, F statistic, and mutual information.

The choice of features is very sensitive to experimental conditions. Tendencies are demonstrated by counting how many triplets the indicator appears in across all possible situations. The results of approval voting are shown in Figure 3.6. The most occurring attributes in the temporal domain are peak-to-peak, shape, and crest, and in the spectral domain, those are centroid, roll-off, and roll-on.

### 3.4 K-nearest neighbor classifier

We utilize the k-nearest neighbor algorithm to check machinery diagnostics abilities with reduced feature sets. The KNN classifier being a lazy learner means it can be adapted easily from offline to online context. Training labels are min-max scaled and establish nearest-neighbor decision boundaries for attribute values. Voronoi diagrams can display these regions and explain the model in that way. Batch learning

<sup>3</sup>SciKit Learn: <https://scikit-learn.org/>

<sup>4</sup>RiverML: <https://riverml.xyz/>

in KNN serves here as a target performance whose attainment is desirable with online learning.

### 3.4.1 Batch models

Three types of KNN model experiments run with hold-out validation in a batch setting. First, the model learns all extracted features, so no feature selection occurs. Then, the brute-force method searches for a combination of three features with the highest training accuracy. In the end, the model performance for three attributes chosen by feature selection techniques is compared to principal components.

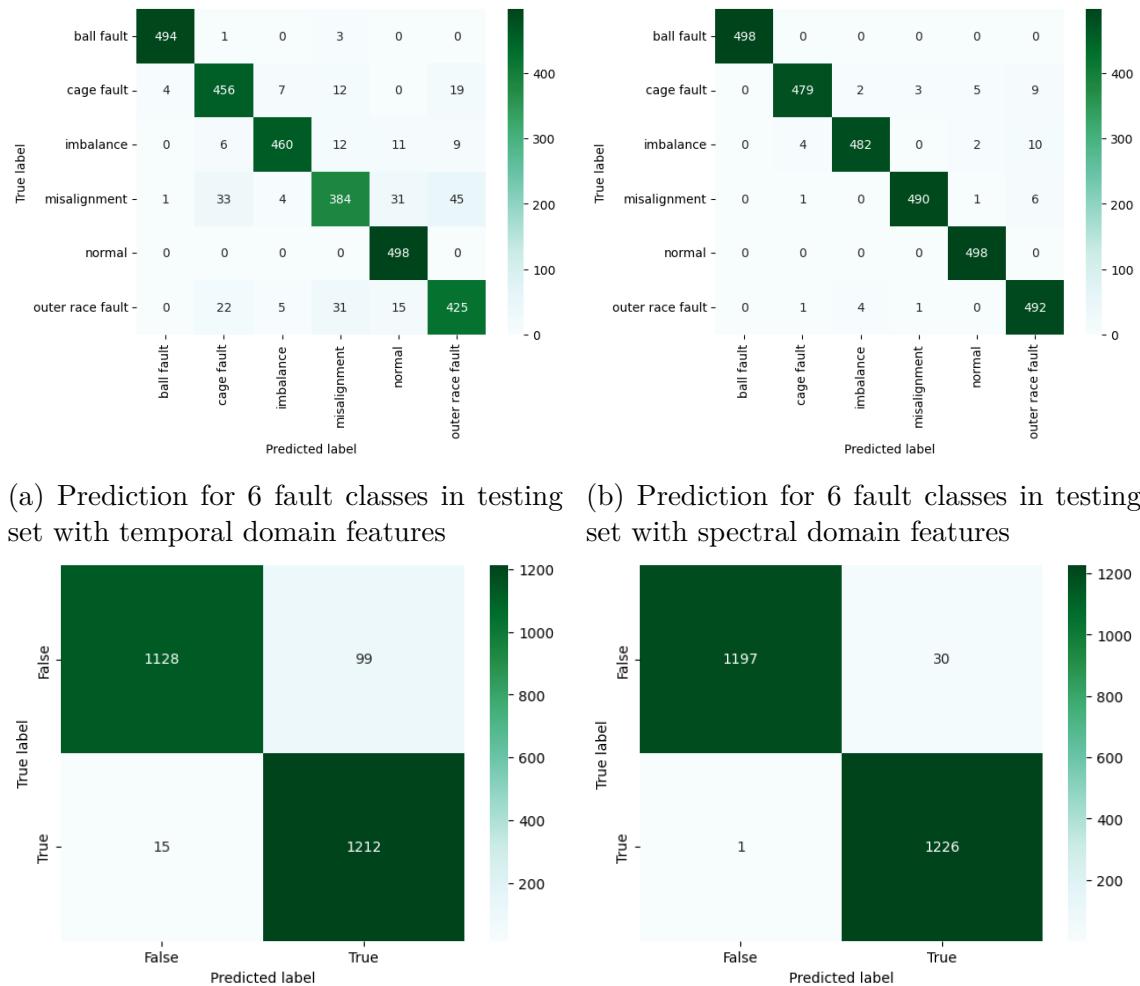


Figure 3.7: Confusion matrix of KNN predictions on bearing position A

Two models are created for the target variable in classification with all 10 temporal and 11 spectral features. The subset of records includes all rotational speeds

on bearing position A. Either six fault labels or anomalies above 0.9 severity level are guessed by the model with the same hold-out validation split as before.

Confusion matrices for fault prediction and high severity anomaly are shown in Figure 3.7. In this example, the number of k neighbors is 5, the distance metric is the Euclidian norm, and the algorithm for proximity queries is a k-d tree. The most inaccuracies in the temporal domain are between misalignment and bearing race faults. In the spectral domain, the model confuses imbalance and bearing faults because mass to unbalance is hung onto the shaft to cause bearing defects.

In anomaly prediction, the error of the first degree is 7 times more prevalent than the error of the second degree. False positives are preferable, since we do not want the machine to fail prematurely and not know about it. In all cases, spectral features maintain better prediction metrics than all temporal features because of less interdependency among features. Bearing B exhibits overall worse classification performance because of more noise in the original signal.

Bearing position	Target variable	Observations	Temporal domain			Spectral domain		
			Training accuracy	Testing accuracy	Training F1 score	Training accuracy	Testing accuracy	Training F1 score
A	Fault	$\sum 2998$ 498 per class	0.9413	0.9093	0.91	0.9917	0.9836	0.98
	Anomaly (0.9)	$\sum 2454$ 1227 per class	0.9668	0.9535	0.96	0.9921	0.9874	0.99
B	Fault	$\sum 2998$ 498 per class	0.8634	0.7935	0.79	0.9207	0.8765	0.87
	Anomaly (0.9)	$\sum 2394$ 1197 pre class	0.9148	0.8668	0.87	0.9428	0.9156	0.92

Table 3.2: KNN model performance trained on all extracted features

Accuracies and F1 scores in both bearing positions with all extracted features are shown in Table 3.2. Models are overtrained, especially for temporal features, because of the substantial difference between accuracy on training and validation sets. Binary classification of anomalies is unsurprisingly more precise in general than the multi-class case. Defect-type detection with all features reaches accuracy on the testing set above 98% for bearing A, and above 87% for bearing B.

Next, we exhaustively list all combinations of three features and look at the range of model performances generated. There are  $\binom{n}{3}$  combinations of attribute triplets that entail 120 separate KNN models for temporal domain features and 165

for the spectral domain. The feature set with the best evaluation scores serves as a benchmark for attributes picked by selection techniques. The accuracy of the KNN algorithm for different target variables is in Figure 3.8.

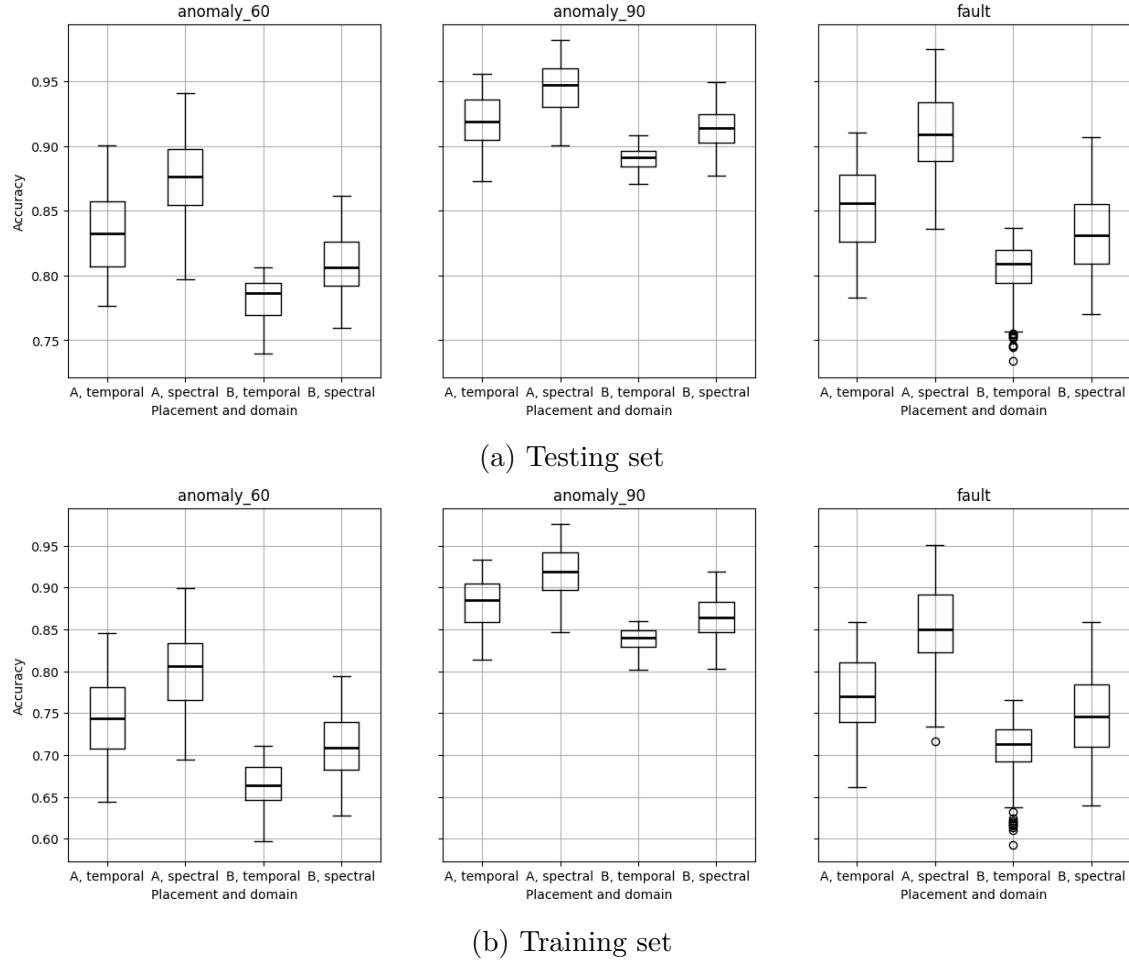


Figure 3.8: Range of prediction accuracy in all KNN models with subset of 3 features. Subplots show three different predicted variables.

Features picked combinatorically for predictions are listed in Table 3.3. The worst accuracy occurs for temporal features in multi-class classification, 86% and 77% depending on bearing position, as compared to 98% and 91% accuracy for spectral features. An improvement would come with increasing the number of features, using a different base feature set with greater discriminatory power, or using a more sophisticated model.

The features picked using the rank product method are shown in Table 3.4. Their prediction accuracies in KNN models are subpar to optimal sets. However, the advantage is that the feature election process is not so computationally taxing.

Place	Target variable	Domain	Best feature triplet	Training accuracy	Testing accuracy
A	anomaly (0.9)	temporal	{std, skewness, shape}	0.9559	0.9328
		spectral	{centroid, noisiness, entropy}	0.9818	0.9707
	fault	temporal	{std, skewness, shape}	0.9106	0.8591
		spectral	{centroid, kurtosis, entropy}	0.9750	0.9505
B	anomaly (0.9)	temporal	{std, kurtosis, pp}	0.9085	0.8551
		spectral	{centroid, std, roll-on}	0.9495	0.9185
	fault	temporal	{std, skewness, kurtosis}	0.8369	0.7654
		spectral	{centroid, std, roll-off}	0.9067	0.8584

Table 3.3: Features with the highest accuracies on the training set found combinatorically

On bearing A, validation set accuracies for fault diagnostics are 85% and 92%, for each of the domains. Because of the low  $k$  hyperparameter value, models are overtrained either way.

Place	Target variable	Domain	Best feature triplet	Train accuracy	Test accuracy
A	anomaly (0.9)	temporal	{shape, std, rms}	0.9333	0.8916
		spectral	{centroid, flux, entropy}	0.9654	0.9474
	fault	temporal	{std, shape, max}	0.9048	0.8544
		spectral	{roll-off, centroid, skewness}	0.9504	0.9210
B	anomaly (0.9)	temporal	{std, shape, crest}	0.8961	0.8505
		spectral	{std, noisiness, entropy}	0.9265	0.8843
	fault	temporal	{pp, crest, skewness}	0.8194	0.7380
		spectral	{centroid, roll-on, roll-off}	0.8914	0.8390

Table 3.4: Three chosen features with rank product of correlation, F statistic, mutual information and their associated KNN accuracies.

Lowering the number of attributes allows visualization of labels in planar scatter plots. The data points in 3-dimensional cross-sections are colored by true fault classes of all severities in Figure 3.9. We observe the inability to separate faults by linear boundaries in given feature spaces and the noncompactness of clusters.

Cluster overlaps of various defect types are also confirmed by silhouette scores around zero. Therefore, the  $k$  hyperparameter cannot be increased much, as it would cause the merge of mutually confused categories. Low-severity faults should be relabeled as a class without fault to watch out just for alert conditions.

### 3.4. K-NEAREST NEIGHBOR CLASSIFIER

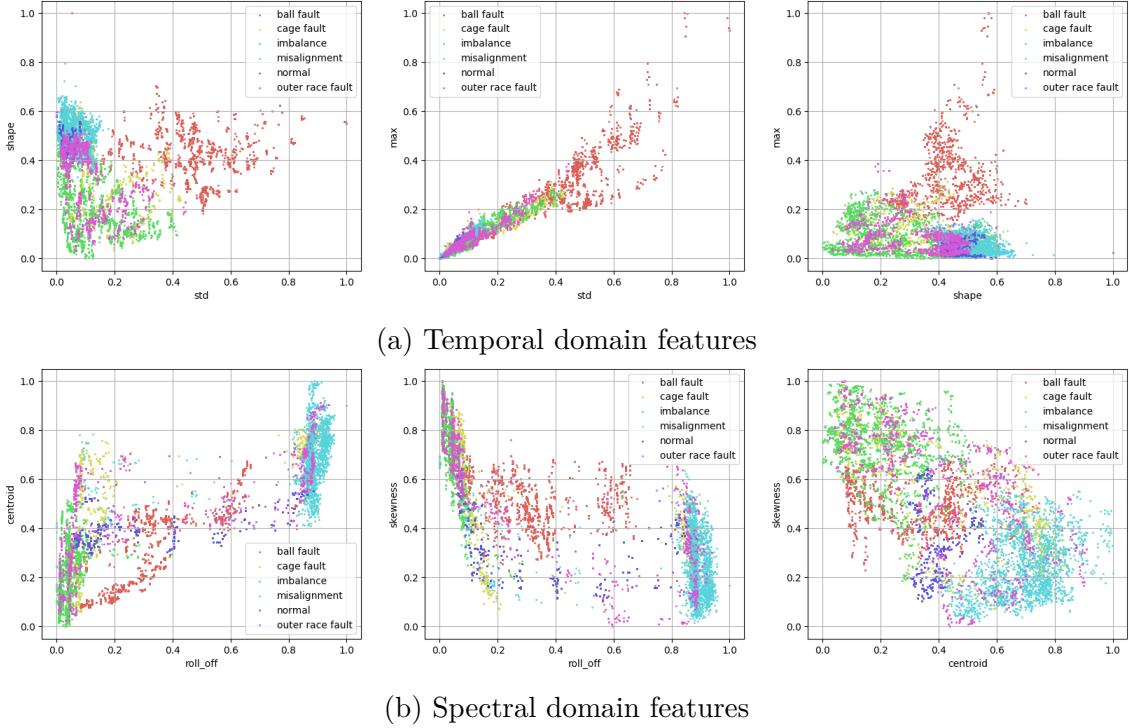


Figure 3.9: Cross sections of 3-dimensional feature spaces with ground truth labels for best attributes. Features are chosen by rank product method.

Model performance of the KNN algorithm preceded by a variety of feature selection methods is compared in Figure 3.10. The best combination of 3 features always achieves better than when the model is trained on all features. This has to do with curse of dimensionality phenomenon.

We suspected that the PCA property of maximizing variances in each of the few dimensions would support more separation of clustered groups and would be beneficial in classification. The bar chart supports this assumption. PCA always produces greater model accuracy than any other method of feature selection. The rank product of the three scoring techniques is not better every time, but it balances out aggregated methods to achieve more stable results across testing sets.

In every instance, a subset of spectral features is responsible for better accuracy than a subset of temporal features. Presumably, it is so because of many correlated pairs of attributes in the temporal domain. Further tests should be made to determine optimal feature subsets and the impact of altering the feature election process.

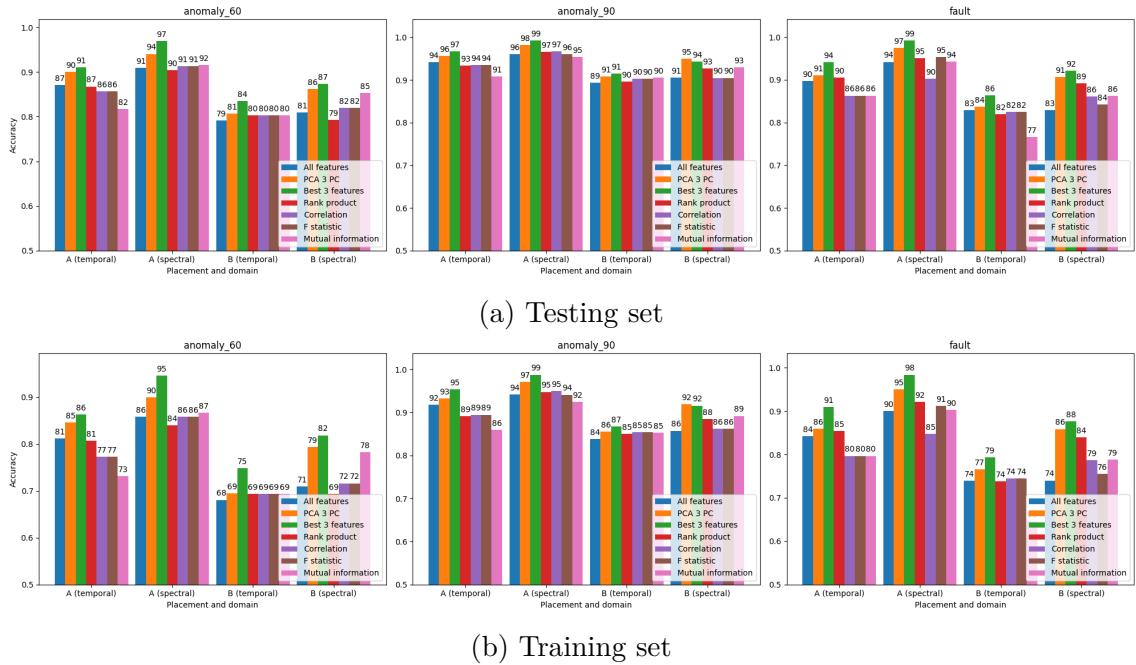


Figure 3.10: Batch KNN algorithm prediction accuracy with various feature sets.

### 3.4.2 Online models

Online learning imitates hardened conditions for machinery diagnostics that appear in deployment. Delayed provision or omission of actual labels undoubtedly degrades the reliability of the classification. The question is how quickly the accuracy approaches the optimal one from the nearest neighbors trained in batch and what the effect of routine difficulties in the ongoing labeling process is.

The KNN models in incremental learning experiments learn on the same base training dataset for bearing position A and with all extracted features as in a batch context. In this manner, we can compare the training accuracies in the last sample for both models. KNN is set to 5 nearest neighbors and a proximity metric of Euclidian distance. Online learning metrics are evaluated by progressive valuation on a dataset that is left unbalanced.

The **stream of events is sorted** by rising severity levels (Fig. 3.11b) which ensures steady increments in label counts throughout the whole duration of the simulation (Fig. 3.11a). This constructed event sequence is a bit unrealistic because all types of faults never begin to appear simultaneously with equal strengths. It is meant to approximate the gradual overall degradation of the machine.

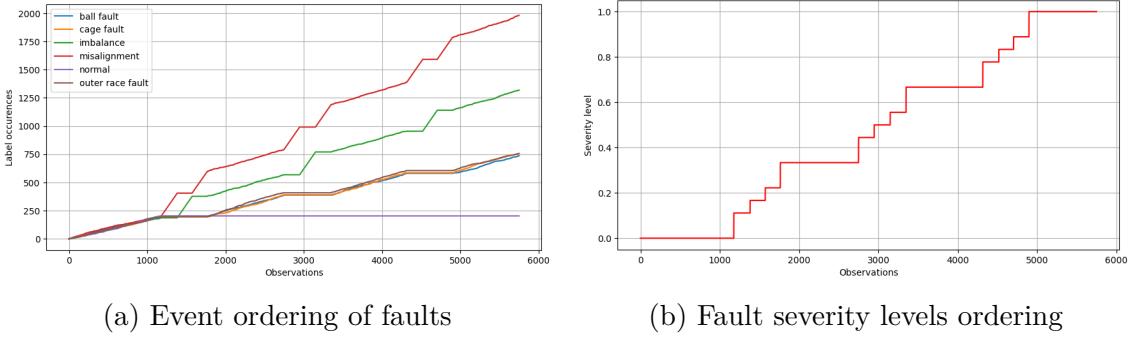


Figure 3.11: Label sequencing for progressive evaluation

Major breaking points in the stream are after 1171 observations out of 5751, where all 203 normal conditions are consumed in the training process. Counters of other faults show that model predictions are skewed towards more represented classes of imbalance and misalignment. The uneven evolution of category counts in a stream impacts the development of accuracy in the remaining experiments. The training accuracies of comparable batch models are 94.13% (*temporal features*) and 99.17% (*spectral features*).

During gradual learning, the correct label is supplied after a fixed period passes after its prediction. This **sliding window** simulation examines accuracy every 100 iterations. Wait times before revealing the actual class associated with the sample are 1, 50, 100, or 250 steps (Fig. 3.13).

Accuracies after sequentially seeing all samples are 82.92% (*temporal*) and 89.11% (*spectral*) when labels are shown instantly. Scatter plots in Figure 3.12 visualize mistakes in predictions projected onto two principal components. Labeling delay of 250 observations causes accuracy to drop to 70.51% (*temporal*) and 78.81% (*spectral*). Accuracy better than 50% is achieved 200 steps sooner in the spectral domain than in the temporal domain.

The **tumbling window** simulates regular expert visits annotating observations recorded until that moment. Labels for the whole previous window are supplied at once. In the final sample, the accuracies are 83.41% (*temporal*) and 90.14% (*spectral*) with immediate feedback, and 77.95% (*temporal*) and 85.90% (*spectral*) with window length of 250 samples (Fig. 3.14).

A **tumbling window** is more accurate according to progressive valuation because the labeling delay decreases towards the window's end. Initial 0% accuracy is

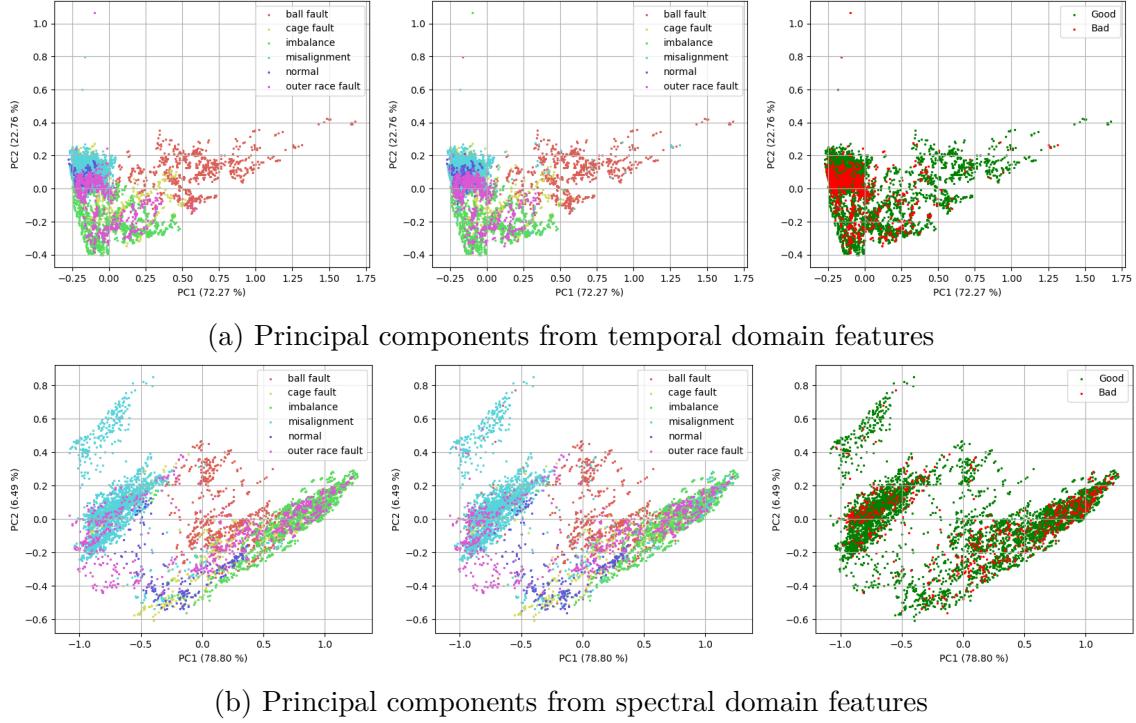


Figure 3.12: Classification labels in incremental learning (*from left*): true labels, predicted labels, mistakes in predictions.

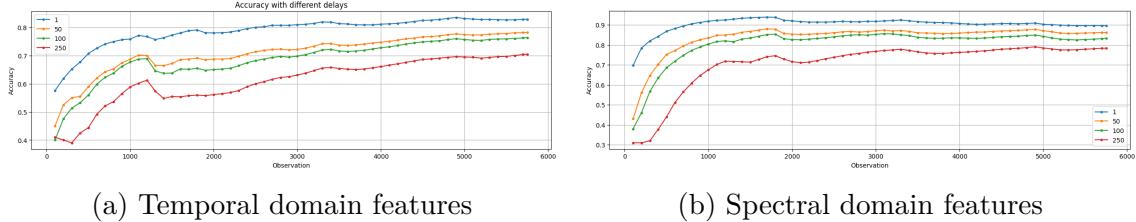


Figure 3.13: Incremental learning on all extracted feature with delayed reveal of labels in sliding windows

caused by a warming-up period in data collection during the span of the first few windows. The true labels are unknown. After just a handful of windows in the beginning, accuracy jumps above 60% and stabilizes after 1000 observations.

Another common problem with online learning is **missing annotations** due to the size of the dataset. In the simulation of missing labels, the equal-length gaps are skipped before another observation is annotated. This approach of choosing samples to annotate without considering their representativeness results in severe harm to predictions.

Labeling just every 5<sup>th</sup> sample (25% of the total dataset) with a sliding window delay of 10 samples reduces accuracy for the model out of temporal features by 8.77%

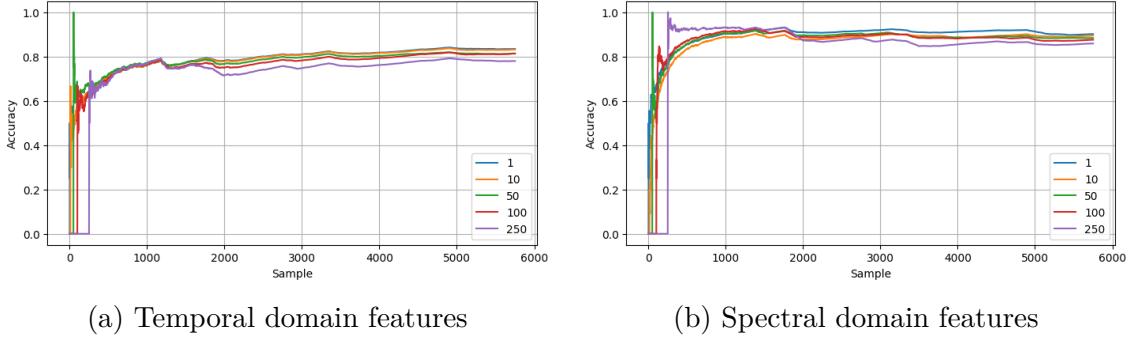


Figure 3.14: Incremental learning on all extracted feature with delayed reveal of labels at regular intervals in tumbling windows

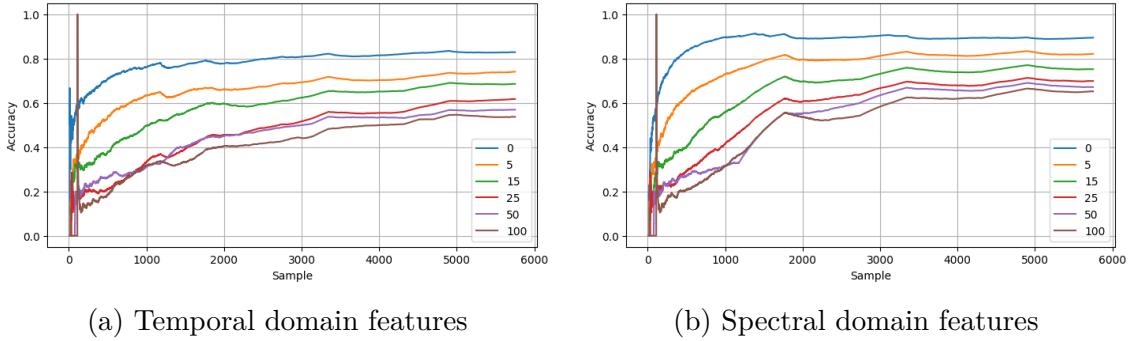


Figure 3.15: Incremental learning with missing true labels and sliding window delay of 10 observations

to 74.07%, and by 7.33% for spectral features to accuracy of 82.01% (Fig. 3.15). Even if only 1% of the dataset is annotated (every 100<sup>th</sup> sample), the model out of spectral features retained an accuracy of 65.17%. The same cannot be said about the model from temporal features with very poor accuracy of 53.68%. More missing labels require recording more observations before the equivalent accuracy is reached.



# 4 Implementation

The methodology applied so far for the dataset recorded in a laboratory is to be tested on vibration signals from the industrial environment. Established guidelines for introducing a monitoring system have been described in the section on technical standards (Section 2.1.3).

We employ a slightly customized validation procedure that involves the selection of machines to be monitored, identifying positions for measurements according to technical specifications, taking preliminary readings, and developing a device capable of capturing the usual failure modes. The collection of the novel dataset is accompanied by an agreed upon schedule.

## 4.1 Machinery for monitoring

Three kinds of rotating machines are at disposal for vibration measurements. Those are a standing fan, scroll compressor, and water pump.

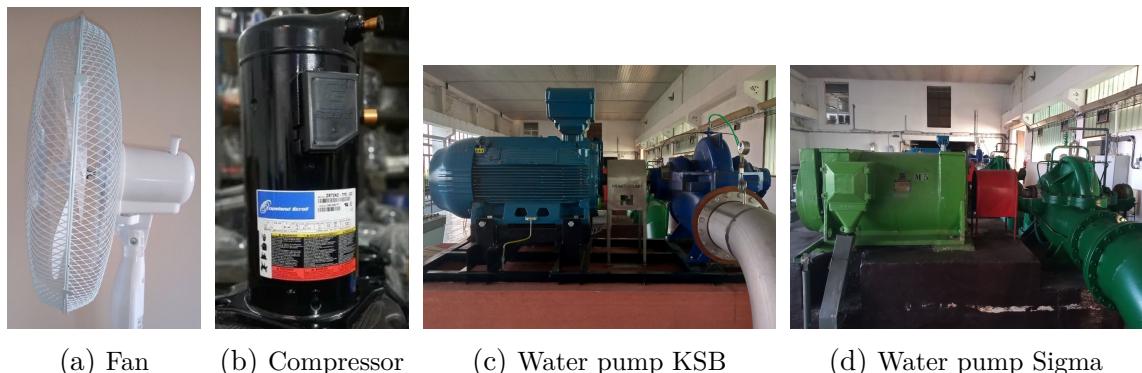


Figure 4.1: Machines dedicated for vibration measurements

**Standing fan** is model *Kalorik TKG VT1037* (Fig. 4.1a) and one unit is available to us. It serves as a test bed during the sensor unit development. The sensor is placed on the plastic casing at the back of the drive motor. The fan has a 45 cm diameter with 3 propellers and a power of 45 W (class I). It has a switch for 3 ro-

tational speeds, which are approximately 18.5 Hz (1100 rpm), 20.4 Hz (1200 rpm), and 22 Hz (1300 rpm). Acceleration observed is  $\pm 5 \text{ m/s}^2$  in radial and tangential directions, and  $\pm 1 \text{ m/s}^2$  in axial direction.

**Scroll compressor** is model *Copeland ZR16* (Fig. 4.1b). Altogether, two units are available in two independent air conditioning units for the data center. The compressor has 9.7 kW of power (class I) and rotates at 2900 rpm (48.3 Hz). Two possible measurement locations are located on the sides on top of the bearings, just above the base and above the scroll. Accelerations picked up by the sensor are  $\pm 3 \text{ m/s}^2$  tops.

**Water pumps** are available as 3 units in municipal drinking water pumping station. The apparatus consists of a single-stage axially split volute casing pump and an attached electric induction motor.

The newer primary pumps with bundled wireless cloud monitoring system are two units of *KSB Omega 300-560* (Fig. 4.1c). The pumps were commissioned in 2018, and they rotate at 1493 rpm (24.9 Hz). The electric motor provides 400 kW of power (class III). The outer bearing on the pump vibrates with an acceleration of  $\pm 5 \text{ m/s}^2$ .

The secondary pump is one unit named *Sigma 300-OVD-600* (Fig. 4.1d) installed in 1986. It rotates at 1485 rpm (24.75 Hz), and its electric motor has power of 450 kW (class III). Each pump and motor have 2 bearings. Therefore, there are 4 measurement positions in total.

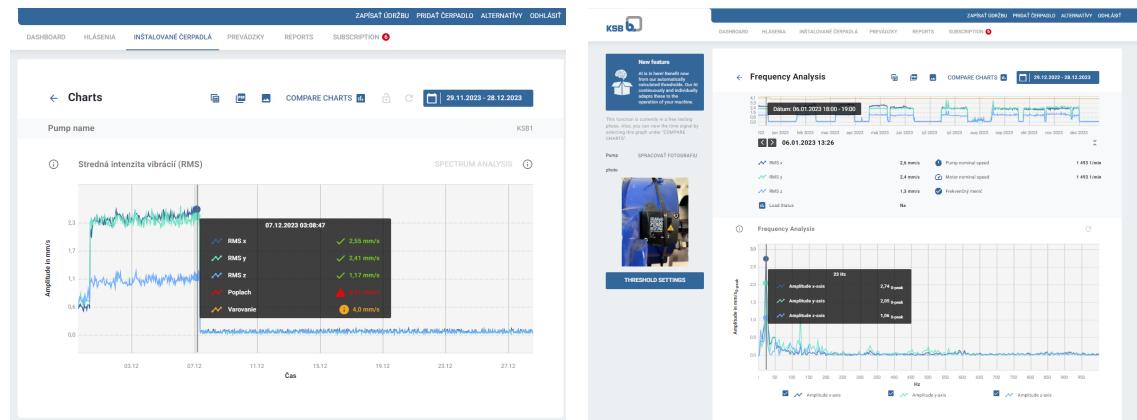


Figure 4.2: KSB Guard cloud monitoring for pumps

**Complete schedule** for planned vibration data gathering on the designated machines and measurement procedure is described in Appendix C. In addition, we are able to export the entire historical record at hourly intervals. Logs from **KSB Guard** cloud monitoring tool contain vibration rms velocities and frequency spectra for two monitored KSB water pumps (Figure 4.2).

## 4.2 Sensor hardware and drivers

In preliminary exploratory vibration measurements of machines, the ADXL335 accelerometer is connected to BeagleBone Black microcontroller. The bandwidth, sensitivity, and sampling frequency of the makeshift solution do not appear to be adequate (Table 4.1).

Accelerometer	ADXL335	IIS3DWB
<b>Vendor</b>	Analog Devices	STMicroelectronics
<b>Bus</b>	Analog	SPI
<b>Axis</b>	3	1 or 3
<b>Range (g)</b>	$\pm 3$	$\pm 2$ to 16
<b>Bandwidth (kHz)</b>	0.55 - 1.6	5 - 6.3
<b>Sensitivity (mg/LSB)</b>	2.930	0.061
<b>Noise density (<math>\mu g/\sqrt{Hz}</math> rms)</b>	150 - 300	75
<b>Microcontroller</b>	Beaglebone Black	ESP32-PoE-ISO
<b>CPU SoC</b>	TI Sitara AM3358	ESP32-WROOM-32
<b>Output data rate (kHz)</b>	2.5	26.7
<b>ADC resolution (bit)</b>	12	16
<b>FIFO</b>	-	3 kB (512 samples)

Table 4.1: Hardware parameters of previous and proposed sensor units

The IIS3DWB accelerometer is a cheap enough MEMS accelerometer that approaches industrial standards for vibration monitoring. ESP32-PoE-ISO is chosen as a microcontroller development kit because it has an SD card slot connected via SD/MMC bus. It enables us to store all samples from FIFO. The accelerometer uses an SPI bus with a maximum speed of 10 MHz that can be connected to any physical GPIO pin. The block diagram of the designed hardware device is in Figure 4.3.

The drivers for necessary low-level interfaces and FAT32 filesystem are imple-

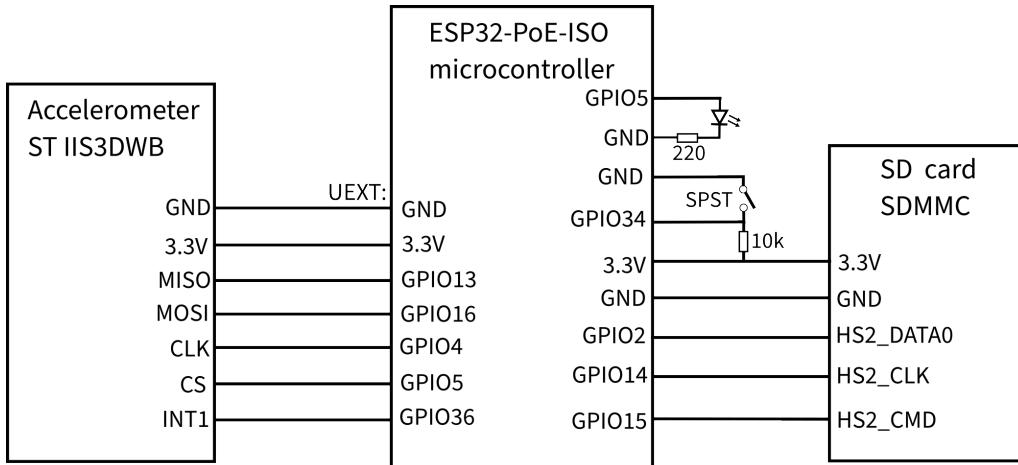


Figure 4.3: Sensor unit hardware block diagram

mented already in ESP-IDF SDK<sup>1</sup>. The accelerometer driver is also made available by the vendor<sup>2</sup>. Activity diagrams of required firmware functionality are in Appendix C.

### 4.3 Preliminary measurements

During the initial on-site inspections of machinery, we carried out preliminary measurements of vibration signals. Raw waveforms of acceleration for the fan, two different compressors, and the outer bearing on a pump and motor in three axes are plotted in Figure 4.4a. The acceleration rarely exceeds 1 g.

In frequency spectra in Figure 4.5b behavior of machines can be clearly discerned below the resonance frequency of the accelerometer (550 Hz). The compressor suspected to be faulty has a stronger 4<sup>th</sup> harmonic and above than the other one.

Figure 4.4b documents the effort to obtain vibration velocity (as per ISO 20186) by integration with the composite trapezoidal rule. The significant drift after integration is eliminated by subtracting the mean of the envelope constructed using cubic interpolation of peaks. Envelopes must be tweaked further to achieve more accurate values. The orbital plot of the first harmonic frequency  $f_{max} \pm 5$  Hz in Figure 4.5a allows us to check the sensor deviation out of the axis of rotation.

<sup>1</sup>ESP-IDF SDK: <https://docs.espressif.com/projects/esp-idf/en/latest/esp32/api-reference/index.html>

<sup>2</sup>IIS3DWB driver: <https://github.com/STMicroelectronics/iis3dwb-pid>

### 4.3. PRELIMINARY MEASUREMENTS

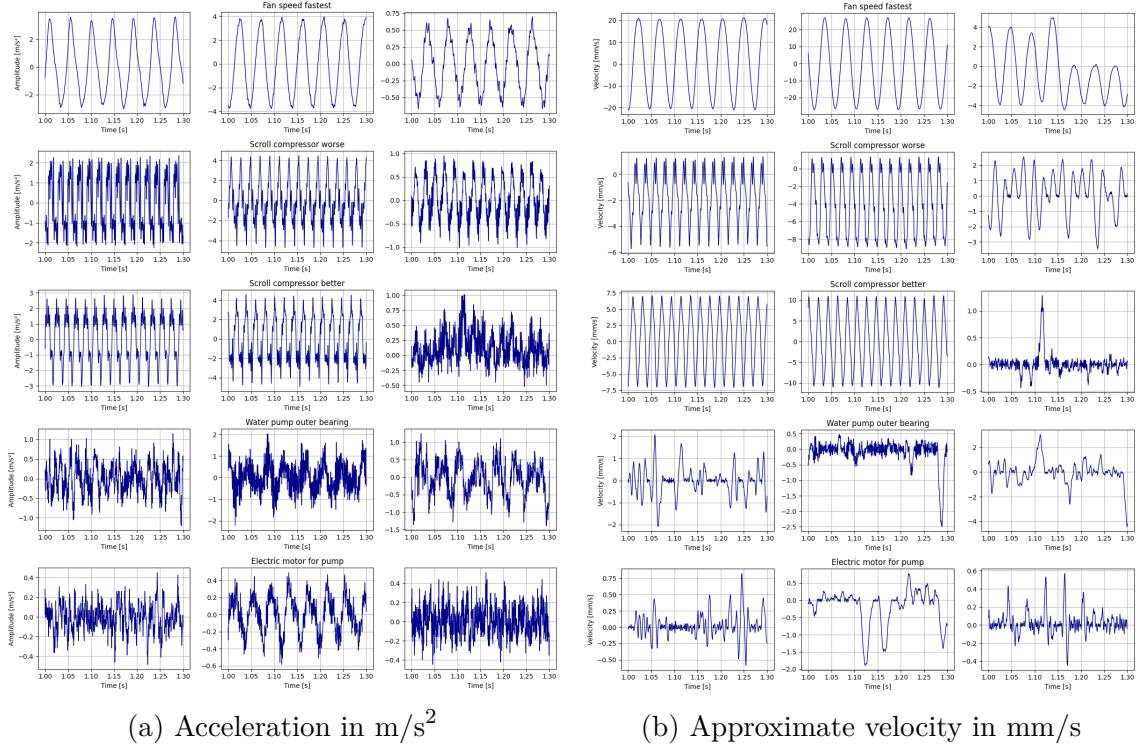
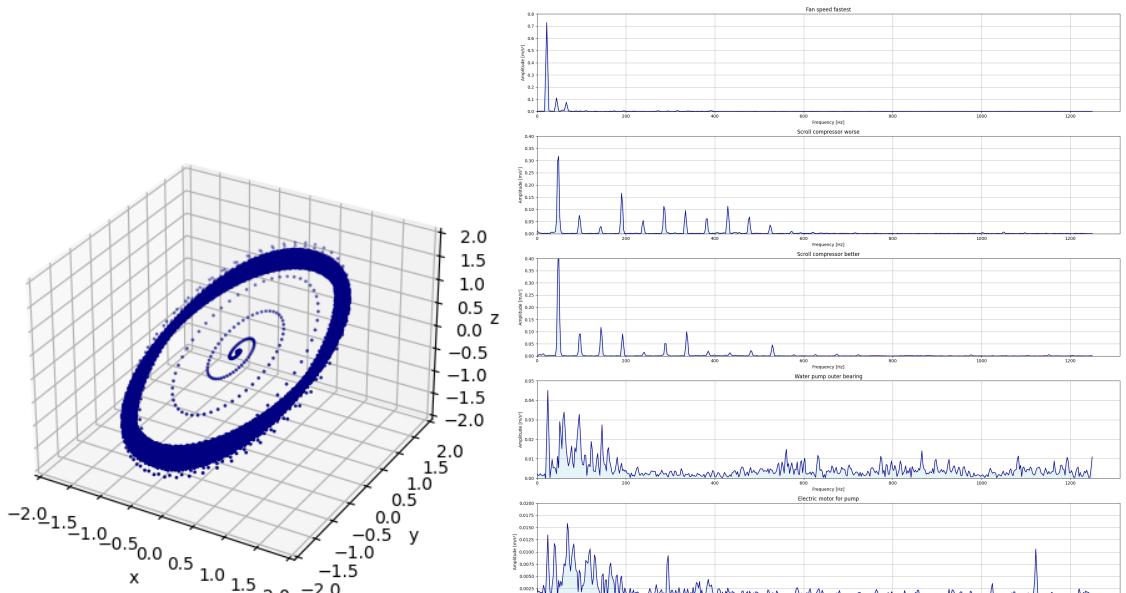


Figure 4.4: Temporal domain waveforms in all spatial directions with 300 ms duration at 1 s timestamp



(a) Orbital analysis of compressor's 1<sup>st</sup> harmonic frequency (b) FFT with Hann window of length  $2^{10}$  ( $\approx 409$  ms)

Figure 4.5: Orbital and spectral methods of vibration signal analysis



# 5 Conclusion

In the thesis we focused on trend indicator selection for an inexpensive industrial condition monitoring solution from vibration signals. The goal is to enable timely fault detection of machinery parts with as little input data as possible. For that purpose, we answer four research questions.

Attributes extracted to describe machine behavior come mainly from descriptive statistics, audio signal processing studies, and vibrodiagnostics technical standards (ISO 20186 and ISO 13343). These formulas compute 10 features summarizing the waveform in the temporal domain and 11 features characterizing spectral density estimation in 3 spatial directions (**RQ1**).

In order to achieve more pronounced data savings, we choose the subset of 3 features in each domain by keeping the ones with the most similarity to the target variable. Feature selection metrics of the correlation coefficient, F statistic, mutual information, and their rank product are applicable in supervised learning. The features are squished from multiple dimensions using the Euclidian norm. Lossy compression ratios attained are 2381:1 for all features and 25000:1 for 6 features in the MaFaulDa dataset. We managed to discard more than 99.995% of irrelevant data (**RQ2**).

Feature subsets are subjected after normalization to a k-nearest neighbor classifier that ascertains their relative fault detection power. Because of model overtraining with small k, the feature triplets equal or slightly outperformed the whole set of features on the validation set with accuracy up to 10%. The spectral features reach higher accuracies than temporal because of their smaller interdependency (**RQ3**).

The ensemble of feature selection with rank product produces the best model performance out of three combined in the majority of situations. No approach could find a triplet of predictors with an accuracy close to optimal one discovered exhaustively. Training on three principal components produced better accuracy than filtering feature selection, but PC mapping onto original trend indicators is unclear

even in loading plots (**RQ3**).

The considerable obstacle in an autonomous fault detection system deployment is the availability of labels for target variables. Annotations can be assigned belatedly or even never. Incremental learning kNN model on an unbalanced dataset on the whole feature set achieves at best 90% accuracy with immediate feedback, 85% with labels coming in 250 long tumbling windows, and 82% with just 25% of observations associated with the label. The comparable model trained in batch reaches an accuracy of 98% (**RQ4**).

Conclusions so far have been made on the MaFaulDa dataset imitating the realistic conditions. Therefore, in proper validation of the proposed solution in practice, we will compare it with the custom-made dataset.

The vibration signals will be gathered on compressors in air conditioning units and water pumps in municipal pumping stations. We will develop firmware for sensor units capable of sampling accelerometers and saving measurements onto SD cards. The challenge awaits us in labeling samples themselves as it requires substantial expert knowledge.

Feature selection methods should be incorporated into incremental learning. Hyperparameters shall be adjusted, and the balancing method used to increase model performance. The outlined tasks awaiting completion are planned for the DP III phase.

# Bibliography

1. MOHANTY, Amiya Ranjan. *Machinery Condition Monitoring: Principles and Practices*. CRC Press, 2015. ISBN 978-1-4665-9305-3.
2. EL-THALJI, Idriss. Predictive Maintenance (PdM) Analysis Matrix: A Tool to Determine Technical Specifications for PdM Ready-Equipment. *IOP Conference Series: Materials Science and Engineering*. 2019, vol. 700, p. 012033. Available from DOI: 10.1088/1757-899X/700/1/012033.
3. SCHEFFER, C.; GIRDHAR, P. *Practical Machinery Vibration Analysis and Predictive Maintenance*. IDC Technologies, Elsevier, 2004. ISBN 0-7506-6275-1.
4. ÇINAR, Zeki Murat; ABDUSSALAM NUHU, Abubakar; ZEESHAN, Qasim; KORHAN, Orhan; ASMAEL, Mohammed; SAFAEI, Babak. Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability*. 2020, vol. 12, no. 19, p. 8211. ISSN 2071-1050. Available from DOI: 10.3390/su12198211.
5. ŽIARAN, Stanislav. *Technická diagnostika*. 1. vyd. Bratislava: Vydatelstvo STU, 2013. ISBN 978-80-227-4051-7.
6. DAVIES, A. *Handbook of Condition Monitoring: Techniques and Methodology*. Dordrecht: Springer Netherlands, 2012. ISBN 978-94-011-4924-2.
7. JENNIONS, Ian K. *Integrated Vehicle Health Management: Perspectives on an Emerging Field*. SAE International, 2011. ISBN 978-0-7680-6432-2. Available from Google Books: PII i7pwAACAAJ.
8. BOUSDEKIS, Alexandros; MENTZAS, Gregoris. Enterprise Integration and Interoperability for Big Data-Driven Processes in the Frame of Industry 4.0. *Frontiers in Big Data*. 2021, vol. 4. Available from DOI: 10.3389/fdata.2021.644651.
9. OKOH, C.; ROY, R.; MEHNEN, J.; REDDING, L. Overview of Remaining Useful Life Prediction Techniques in Through-life Engineering Services. *Procedia CIRP*. 2014, vol. 16, pp. 158–163. ISSN 2212-8271. Available from DOI: 10.1016/j.procir.2014.02.006.

## BIBLIOGRAPHY

---

10. TORRES, Pedro; RAMALHO, Armando; CORREIA, Luis. Automatic Anomaly Detection in Vibration Analysis Based on Machine Learning Algorithms. In: MACHADO, José; SOARES, Filomena; TROJANOWSKA, Justyna; YILDIRIM, Sahin; VOJTĚŠEK, Jiří; REA, Pierluigi; GRAMESCU, Bogdan; HRYBIUK, Olena O. (eds.). *Innovations in Mechatronics Engineering II*. Cham: Springer International Publishing, 2022, pp. 13–23. Lecture Notes in Mechanical Engineering. ISBN 978-3-031-09385-2. Available from DOI: [10.1007/978-3-031-09385-2\\_2](https://doi.org/10.1007/978-3-031-09385-2_2).
11. ISO 13373-1:2002 - *Condition Monitoring and Diagnostics of Machines - Vibration Condition Monitoring - Part 1: General Procedures*. International Organization for Standardization, 2002.
12. ISO 20816-1:2016 - *Mechanical Vibration - Measurement and Evaluation of Machine Vibration - Part 1: General Guidelines*. International Organization for Standardization, 2016.
13. TITTELBACH-HELMRICH, Klaus. Digital DC Blocker Filters. *Frequenz*. 2021, vol. 75, no. 9-10, pp. 331–339. ISSN 2191-6349. Available from DOI: [10.1515/freq-2020-0177](https://doi.org/10.1515/freq-2020-0177).
14. LYONS, Richard G. *Understanding Digital Signal Processing*. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2011. ISBN 978-0-13-702741-5.
15. DINIZ, Paulo S. R. *Adaptive Filtering: Algorithms and Practical Implementation*. Cham: Springer International Publishing, 2020. ISBN 978-3-030-29056-6. Available from DOI: [10.1007/978-3-030-29057-3](https://doi.org/10.1007/978-3-030-29057-3).
16. NANDI, Asoke Kumar; AHMED, Hosameldin. *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*. Hoboken, NJ, USA: Wiley-IEEE Press, 2019. ISBN 978-1-119-54462-3.
17. JOHNSON, Max Kuhn and Kjell. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. 2019.
18. BRITO, Lucas Costa; SUSTO, Gian Antonio; BRITO, Jorge Nei; DUARTE, Marcus Antonio Viana. Fault Detection of Bearing: An Unsupervised Machine Learning Approach Exploiting Feature Extraction and Dimensionality Reduction. *Informatics*. 2021, vol. 8, no. 4, p. 85. ISSN 2227-9709. Available from DOI: [10.3390/informatics8040085](https://doi.org/10.3390/informatics8040085).

19. MOSTAFAVI, Alireza; SADIGHI, Ali. A Novel Online Machine Learning Approach for Real-Time Condition Monitoring of Rotating Machines. In: *2021 9th RSI International Conference on Robotics and Mechatronics (ICRoM)*. 2021, pp. 267–273. ISSN 2572-6889. Available from DOI: [10.1109/ICRoM54204.2021.9663495](https://doi.org/10.1109/ICRoM54204.2021.9663495).
20. ISO 13373-2:2016 - *Condition Monitoring and Diagnostics of Machines - Vibration Condition Monitoring - Part 2: Processing, analysis and presentation of vibration data*. International Organization for Standardization, 2016.
21. PEETERS, Geoffroy. A Large Set of Audio Features for Sound Description. 2004.
22. AVOCI, Moise. Spectral Negentropy and Kurtogram Performance Comparison for Bearing Fault Diagnosis. In: Dubrovnik, Croatia: International Measurement Confederation (IMEKO), 2020.
23. ADIKARAM, K.K. Lasantha Britto; HUSSEIN, Mohamed; EFFENBERGER, Mathias; BECKER, T. Non-Parametric Local Maxima and Minima Finder with Filtering Techniques for Bioprocess. *Journal of Signal and Information Processing*. 2016, vol. 07, pp. 192–213. Available from DOI: [10.4236/jsip.2016.74018](https://doi.org/10.4236/jsip.2016.74018).
24. GERBER, Timothée; MARTIN, Nadine; MAILHES, Corinne. Identification of Harmonics and Sidebands in a Finite Set of Spectral Components. 2013, vol. 1.
25. HÁJEK, Miroslav; BALÁŽ, Marcel. *Spracovanie dát generovaných senzorovou IoT sietou*. Bratislava, 2022. Bachelor's thesis. Faculty of Informatics and Information Technologies, Slovak University of Technology.
26. SHI, Xiangfu; ZHANG, Zhen; XIA, Zhiling; LI, Binhua; GU, Xin; SHI, Tingna. Application of Teager–Kaiser Energy Operator in the Early Fault Diagnosis of Rolling Bearings. *Sensors*. 2022, vol. 22, no. 17, p. 6673. ISSN 1424-8220. Available from DOI: [10.3390/s22176673](https://doi.org/10.3390/s22176673).
27. YU, Gang. A Concentrated Time–Frequency Analysis Tool for Bearing Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*. 2020, vol. 69, no. 2, pp. 371–381. ISSN 1557-9662. Available from DOI: [10.1109/TIM.2019.2901514](https://doi.org/10.1109/TIM.2019.2901514).
28. ARTS, Lukas P. A.; van den BROEK, Egon L. The Fast Continuous Wavelet Transformation (fCWT) for Real-Time, High-Quality, Noise-Resistant Time–Frequency Analysis. *Nature Computational Science*. 2022, vol. 2, no. 1, pp. 47–58. ISSN 2662-8457. Available from DOI: [10.1038/s43588-021-00183-z](https://doi.org/10.1038/s43588-021-00183-z).

## BIBLIOGRAPHY

---

29. HERRERA, Roberto; BAAN, Mirko; HAN, Jiajun. Applications of the Synchrosqueezing Transform in Seismic Time-Frequency Analysis. *Geophysics*. 2014, vol. 79, pp. V55–V64. Available from DOI: [10.1190/geo2013-0204.1](https://doi.org/10.1190/geo2013-0204.1).
30. GOUMAS, Stefanos; ZERVAKIS, Michalis; STAVRAKAKIS, G. Classification of Washing Machines Vibration Signals Using Discrete Wavelet Analysis for Feature Extraction. *IEEE Transactions on Instrumentation and Measurement*. 2002, vol. 51, pp. 497–508. Available from DOI: [10.1109/TIM.2002.1017721](https://doi.org/10.1109/TIM.2002.1017721).
31. YEN, Gary; LIN, K.C. Wavelet Packet Feature Extraction for Vibration Monitoring. *IEEE Transactions on Industrial Electronics*. 2000, vol. 47, pp. 650–667. Available from DOI: [10.1109/41.847906](https://doi.org/10.1109/41.847906).
32. SONG, Yongxing; LIU, Jingting; WU, Dazhuan; ZHANG, Linhua. The MFBD: A Novel Weak Features Extraction Method for Rotating Machinery. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. 2021, vol. 43, no. 12, p. 547. ISSN 1806-3691. Available from DOI: [10.1007/s40430-021-03259-z](https://doi.org/10.1007/s40430-021-03259-z).
33. ZHUO, Rongjin; DENG, Zhaohui; CHEN, Bing; LIU, Tao; GE, Jimin; LIU, Guoyue; BI, Shenghao. Research on Online Intelligent Monitoring System of Band Saw Blade Wear Status Based on Multi-Feature Fusion of Acoustic Emission Signals. *The International Journal of Advanced Manufacturing Technology*. 2022, vol. 121, no. 7, pp. 4533–4548. ISSN 1433-3015. Available from DOI: [10.1007/s00170-022-09515-3](https://doi.org/10.1007/s00170-022-09515-3).
34. GILLES, Jerome. Empirical Wavelet Transform. *IEEE Transactions on Signal Processing*. 2013, vol. 61, p. 3999. Available from DOI: [10.1109/TSP.2013.2265222](https://doi.org/10.1109/TSP.2013.2265222).
35. LI, Zheng; MING, Anbo; ZHANG, Wei; LIU, Tao; CHU, Fulei; LI, Yin. Fault Feature Extraction and Enhancement of Rolling Element Bearings Based on Maximum Correlated Kurtosis Deconvolution and Improved Empirical Wavelet Transform. *Applied Sciences*. 2019, vol. 9, p. 1876. Available from DOI: [10.3390/app9091876](https://doi.org/10.3390/app9091876).
36. ZHUANG, Cuifang; LIAO, Ping. An Improved Empirical Wavelet Transform for Noisy and Non-Stationary Signal Processing. *IEEE Access*. 2020, vol. PP, pp. 1–1. Available from DOI: [10.1109/ACCESS.2020.2968851](https://doi.org/10.1109/ACCESS.2020.2968851).

37. WANG, Yung-Hung; YEH, Chien-Hung; YOUNG, Hsu-Wen Vincent; HU, Kun; LO, Men-Tzung. On the Computational Complexity of the Empirical Mode Decomposition Algorithm. *Physica A: Statistical Mechanics and its Applications*. 2014, vol. 400, pp. 159–167. ISSN 0378-4371. Available from DOI: [10.1016/j.physa.2014.01.020](https://doi.org/10.1016/j.physa.2014.01.020).
38. TIWARI, Prashant; UPADHYAY, S. H. Novel Self-Adaptive Vibration Signal Analysis: Concealed Component Decomposition and Its Application in Bearing Fault Diagnosis. *Journal of Sound and Vibration*. 2021, vol. 502, p. 116079. ISSN 0022-460X. Available from DOI: [10.1016/j.jsv.2021.116079](https://doi.org/10.1016/j.jsv.2021.116079).
39. ZHENG, Alice; CASARI, Amanda. *Feature Engineering for Machine Learning*. O'Reilly Media, 2018. ISBN 978-1-4919-5324-2.
40. KAMMINGA, Jacob W.; LE, Duc V.; MEIJERS, Jan Pieter; BISBY, Helena; MERTENIA, Nirvana; HAVINGA, Paul J.M. Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2018, vol. 2, no. 1, 15:1–15:27. Available from DOI: [10.1145/3191747](https://doi.org/10.1145/3191747).
41. CALKINS, Keith G. *More Correlation Coeficients (Lesson 13)*. 2005.
42. ROSS, Brian C. Mutual Information between Discrete and Continuous Data Sets. *PLOS ONE*. 2014, vol. 9, no. 2, e87357. ISSN 1932-6203. Available from DOI: [10.1371/journal.pone.0087357](https://doi.org/10.1371/journal.pone.0087357). Publisher: Public Library of Science.
43. BREITLING, Rainer; ARMENGAUD, Patrick; AMTMANN, Anna; HERZYK, Pawel. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS letters*. 2004, vol. 573, no. 1-3, pp. 83–92. ISSN 0014-5793. Available from DOI: [10.1016/j.febslet.2004.07.055](https://doi.org/10.1016/j.febslet.2004.07.055).
44. AGGARWAL, Charu C. *Outlier Analysis*. 2nd ed. Springer Publishing Company, Inc., 2016. ISBN 978-3-319-47578-3.
45. TOGBE, Maurras Ulbricht; BARRY, Mariam; BOLY, Aliou; CHABCHOUB, Yousra; CHIKY, Raja; MONTIEL, Jacob; TRAN, Vinh-Thuy. Anomaly Detection for Data Streams Based on Isolation Forest Using Scikit-Multiflow. In: GERVASI, Osvaldo; MURGANTE, Beniamino; MISRA, Sanjay; GARAU, Chiara; BLEČIĆ, Ivan; TANIAR, David; APDUHAN, Bernady O.; ROCHA, Ana Maria A. C.; TARANTINO, Eu-

## BIBLIOGRAPHY

---

- femia; TORRE, Carmelo Maria; KARACA, Yeliz (eds.). *Computational Science and Its Applications – ICCSA 2020*. Cham: Springer International Publishing, 2020, vol. 12252, pp. 15–30. ISBN 978-3-030-58810-6. Available from DOI: 10.1007/978-3-030-58811-3\_2.
46. AGGARWAL, Charu C.; REDDY, Chandan K.. *Data Clustering - Algorithms and Applications*. CRC Press, 2014. ISBN 978-1-4665-5822-9.
47. ROUSSEEUW, Peter. Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Comput. Appl. Math.* 20, 53-65. *Journal of Computational and Applied Mathematics*. 1987, vol. 20, pp. 53–65. Available from DOI: 10.1016/0377-0427(87)90125-7.
48. AMINI, Amineh; WAH, Teh Ying. A Comparative Study of Density-based Clustering Algorithms on Data Streams: Micro-clustering Approaches. In: *Intelligent Control and Innovative Computing*. Ed. by AO, Sio Iong; CASTILLO, Oscar; HUANG, Xu. New York, NY: Springer New York, 2012, pp. 275–287. ISBN 978-1-4614-1695-1. Available from DOI: 10.1007/978-1-4614-1695-1\_21.
49. GHESMOUNE, Mohammed; LEBBAH, Mustapha; AZZAG, Hanene. State-of-the-Art on Clustering Data Streams. *Big Data Analytics*. 2016, vol. 1, no. 1, p. 13. ISSN 2058-6345. Available from DOI: 10.1186/s41044-016-0011-3.
50. CAO, Feng; ESTERT, Martin; QIAN, Weining; ZHOU, Aoying. Density-Based Clustering over an Evolving Data Stream with Noise. In: *Proceedings of the 2006 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2006, pp. 328–339. ISBN 978-0-89871-611-5. Available from DOI: 10.1137/1.9781611972764.29.
51. TAN, Swee Chuan; TING, Kai Ming; LIU, Tony Fei. Fast Anomaly Detection for Streaming Data. 2011.
52. MAURYA, Seetaram; SINGH, Vikas; VERMA, Nishchal K.; MECHEFSKE, Chris K. Condition-Based Monitoring in Variable Machine Running Conditions Using Low-Level Knowledge Transfer With DNN. *IEEE Transactions on Automation Science and Engineering*. 2021, vol. 18, no. 4, pp. 1983–1997. ISSN 1558-3783. Available from DOI: 10.1109/TASE.2020.3028151.

53. SHI, Zhan. Improving k-Nearest Neighbors Algorithm for Imbalanced Data Classification. *IOP Conference Series: Materials Science and Engineering*. 2020, vol. 719, no. 1, p. 012072. ISSN 1757-8981, ISSN 1757-899X. Available from DOI: 10.1088/1757-899X/719/1/012072.
54. CHEN, Lei; LIAN, Xiang. Efficient Processing of Metric Skyline Queries. *IEEE Transactions on Knowledge and Data Engineering*. 2009, vol. 21, pp. 351–365. Available from DOI: 10.1109/TKDE.2008.146.
55. SHENG, Hao; CHEN, Zhongsheng; XIA, Yemei; HE, Jing. Review of Artificial Intelligence-based Bearing Vibration Monitoring. In: *2020 11th International Conference on Prognostics and System Health Management (PHM-2020 Jinan)*. 2020, pp. 58–67. ISSN 2166-5656. Available from DOI: 10.1109/PHM-Jinan48558.2020.00018.
56. ABU ALFEILAT, Haneen Arafat; HASSANAT, Ahmad B.A.; LASASSMEH, Omar; TARAWNEH, Ahmad S.; ALHASANAT, Mahmoud Bashir; EYAL SALMAN, Hamzeh S.; PRASATH, V.B. Surya. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data*. 2019, vol. 7, no. 4, pp. 221–248. ISSN 2167-6461, ISSN 2167-647X. Available from DOI: 10.1089/big.2018.0175.
57. JAMIL, Mohd Atif; KHAN, Md Asif Ali; KHANAM, Sidra. Feature-Based Performance of SVM and KNN Classifiers for Diagnosis of Rolling Element Bearing Faults. *Vibroengineering PROCEDIA*. 2021, vol. 39, pp. 36–42. ISSN 2345-0533. Available from DOI: 10.21595/vp.2021.22307.
58. ALTAF, Muhammad; AKRAM, Tallha; KHAN, Muhammad Attique; IQBAL, Muhammad; CH, M. Munawwar Iqbal; HSU, Ching-Hsien. A New Statistical Features Based Approach for Bearing Fault Diagnosis Using Vibration Signals. *Sensors*. 2022, vol. 22, no. 5, p. 2012. ISSN 1424-8220. Available from DOI: 10.3390/s22052012.
59. GEPPERTH, Alexander; HAMMER, Barbara. Incremental learning algorithms and applications. *European Symposium on Artificial Neural Networks (ESANN)*. 2016.
60. BLUM, Avrim; KALAI, Adam; LANGFORD, John. Beating the hold-out: bounds for K-fold and progressive cross-validation. In: *Proceedings of the twelfth annual conference on Computational learning theory*. Santa Cruz California USA: ACM, 1999, pp. 203–208. ISBN 978-1-58113-167-3. Available from DOI: 10.1145/307400.307439.

61. HALFORD, Max. *The correct way to evaluate online machine learning models*. 2020. Section: blog.
62. GAIA. *Online Machine Learning With RiverML by Max Halford*. 2022.
63. RIBEIRO, Felipe; MARINS, Matheus; NETTO, Sergio; SILVA, Eduardo. Rotating Machinery Fault Diagnosis Using Similarity-Based Models. In: *Anais de XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. Sociedade Brasileira de Telecomunicações, 2017. Available from DOI: [10.14209/sbrt.2017.133](https://doi.org/10.14209/sbrt.2017.133).
64. PESTANA-VIANA, Denys; ZAMBRANO-LOPEZ, Rafael; de LIMA, Amaro A.; DE M. PREGO, Thiago; NETTO, Sergio L.; da SILVA, Eduardo A.B. The Influence of Feature Vector on the Classification of Mechanical Faults Using Neural Networks. In: *2016 IEEE 7th Latin American Symposium on Circuits & Systems (LASCAS)*. Florianopolis: IEEE, 2016, pp. 115–118. ISBN 978-1-4673-7835-2. Available from DOI: [10.1109/LASCAS.2016.7451023](https://doi.org/10.1109/LASCAS.2016.7451023).
65. SpectraQuest Inc.,: *Machinery Fault Simulator - Lite*. Available also from: <https://spectraquest.com/machinery-fault-simulator/details/mfs-lt/>.
66. SONG, Renwang; BAI, Xiaolu; ZHANG, Rui; JIA, You; PAN, Lihu; DONG, Zengshou. Bearing Fault Diagnosis Method Based on Multidomain Heterogeneous Information Entropy Fusion and Model Self-Optimisation. *Shock and Vibration*. 2022, vol. 2022, e7214822. ISSN 1070-9622. Available from DOI: [10.1155/2022/7214822](https://doi.org/10.1155/2022/7214822).
67. YUHONG, Jin; HOU, Lei; CHEN, Yushu. A New Rotating Machinery Fault Diagnosis Method Based on the Time Series Transformer. 2021.
68. MEY, Oliver; NEUDECK, Willi; SCHNEIDER, André; ENGE-ROSENBLATT, Olaf. *Machine Learning-Based Unbalance Detection of a Rotating Shaft Using Vibration Data*. 2020-07-31. Available from DOI: [10.48550/arXiv.2005.12742](https://arxiv.org/abs/2005.12742).

# Appendix A: Resumé

## A.1 Úvod

Vzostup priemyslu 4.0 so sebou prináša väčšiu mieru automatizácie s cieľom dosiahnuť optimálne využitie dostupných zdrojov. Na základe nepretržitého sledovania opotrebenia zariadení v reálnom čase sa majú zabezpečiť nápravné opatrenia na opravu alebo výmenu súčiastok včas, v reakcii na trendové ukazovatele.

Cieľom je zachovať požadovanú bezpečnosť a efektivitu výroby a zároveň predĺžiť životnosť rotujúcich komponentov. Proaktívna diagnostika porúch je nevyhnutná na začatie opráv bez nadbytočných nákladov. Vibrácie predstavujú nerušivý spôsob, ako zistif a zaznamenať prípadne fatálne zlyhania hned v zárodku. Hlavným problémom pri monitorovaní veľa strojov s vibráciami, je to, že vzniká množstvo záznamov, ktoré nie sú priamo užitočné pre operátora výrobnej linky. Väčšina signálov sa zobrazí maximálne raz, preto je zbytočné ich ukladať alebo prenášať vcelku.

Zároveň na dosiahnutie maximálnej presnosti detekcie musí byť model strojového učenia trénovaný pre cieľové prostredie. Poruchy sú navyše pomerne zriedkavé udalosti, ktoré sa zvyčajne vyskytujú s odstupom niekoľkých mesiacov. Za týchto okolností je fažké rýchlo získať dostatočne veľkú vzorku poruchových udalostí.

## A.2 Sledovanie prevádzkového stavu

Existujú tri rôzne prístupy k údržbe strojov: reaktívny, preventívny a prediktívny. Pri reaktívnej údržbe beží stroj až do úplného zlyhania a je prijateľná vtedy, keď je možná úplná a rýchla výmena pokazeného stroja za záložný. Preventívna údržba prebieha v pravidelných intervaloch odvodených od vopred určeného rozvrhu v alebo strednej doby medzi poruchami. Prediktívna údržba zlepšuje predvídateľnosť oproti reaktívnej údržbe a eliminuje plynvanie voči príliš obozretnej prevencii. Odstávka stroja je naplánovaná po zistení kritických hodnôt a po odhalení problematických

komponentov.

Mechanické problémy počas prevádzky strojov spôsobujú v mnohých prípadoch vibrácie. Vibroakustická diagnostika sa preto považuje za jednu z najdôležitejších metód pri včasnej identifikácii porúch komponentov. Najbežnejšie sa vyskytujúcimi poruchami sú nevyváženosť, nesúososť, vôľa, excentricita, deformácia, trhlina a nadmerné trenie.

Symptómy porúch rotačných strojov sa prejavujú rôznymi frekvenčnými pásmami, ale väčšina je závislá od rotačnej rýchlosťi súčiastky. Nevyváženosť, nevyváženosť a vôľa sa bežne objavujú v frekvenciách do 300 Hz. Poruchy ložísk a prevodovky v neskorých štádiach vývoja sa prejavujú v rozsahu medzi 300 Hz a 1 kHz. Vyššie frekvencie do 10 kHz pomáhajú odhaliť poruchy ložísk v skôrších štádiach rozvoja.

Postupy monitorovania stavu založené na vibráciách musia byť v súlade s normatívnymi smernicami ISO 20816 a ISO 13373. Normy sa týkajú umiestnenia meracích zariadení, zberu údajov, konvencií nastavenia úrovni závažnosti porúch.

### A.3 Extraktia a výber atribútov

Prediktívna údržba má ideálne predpoklady na využitie extrakcie atribútov, pretože signál je zvyčajne stacionárny a trendové premenné v časovej a frekvenčnej oblasti vychádzajú z expertných znalostí v oblasti mechaniky. Výhody dodatočného úsilia v porovnaní so spracovaním pôvodných vzoriek spočívajú v dosahovaní lepšej presnosti klasifikácie, znížení výpočtovej záťaže a znížení potreby úložnej kapacity. Výber atribútov nie je samostatným krokom v procese strojového učenia, ale mal by sa vykonávať iteratívne na zlepšenie výsledného modelu.

Najrozšírenejšími používanými atribútmi sú štatistické miery centrálneho momentu: priemer, rozptyl, štandardná odchýlka, šikmosť a špicatosť. Charakteristiky amplitúdy zahŕňajú kvadratický priemer (rms), vzdialenosť špička-špička a maximum. Ostatné významné atribúty časovej oblasti sú odvodene ako pomery a sú nim: faktor výkyvu, faktor rozpätia, faktor impulzu a faktor tvaru.

V spektrálnej oblasti môžeme získať obvyklé štatistické vlastnosti distribúcie, ktorými sú spektrálne ľažisko, šikmosť a špicatosť. Okrem toho sa extrahujú roll-on a roll-off, fundamentálna frekvencia, entropia, negentropia, vzájomná korelácia

spektier, pomer signálu k šum, energia vo frekvenčných pásmach.

Atribúty neprispievajú k prediktívnej sile modelu s rovnakým podielom. Výber ich optimálnej podmnožiny je NP-ťažký kombinatorický problém. Kroky všeobecného postup pri výbere atribútov metódou filtrovania sú generovanie podmnožín, vyhodnotenie podmnožín, ukončovacie kritérium hľadania, a validácia.

Hodnotenie relevancie atribútov sú založené na skórovaní podobnosti s predikovanou premennou. Často používané spôsoby zoradovania dôležitosti atribútov sú prah rozptylu, koeficienty korelácie, ANOVA F štatistika, a vzájomná informácia. Viaceré podmnožiny prediktorov produkovaných každou z výberových metrík môžu slúžiť na trénovanie viacerých variantov klasifikačného modelu. Množiny atribútov je možné kombinovať do súboru volebným systémom ako sú väčšinové hlasovanie alebo súčin poradí.

## A.4 Diagnostické prístupy

Identifikácia porúch v rotujúcich strojoch je binárny alebo viac triedny klasifikačný problém, ktorý pracuje na princípe učenia čiastočne s učiteľom, pretože označenia pre degradované stavy stroja sú v praxi zriedkavé. Ciele automatizácie monitorovania možno rozdeliť na detekciu anomálií a rozpoznanie typu poruchy.

Detekcia anomálií, novostí alebo odľahlých hodnôt určuje, či sa prevádzkový stav stroja výrazne odchyľuje od normálu. Po upozornení môže zasiahnuť odborník a stroj diagnostikovať. Odľahlé hodnoty sú odvodzované na základe neparametrických štatistických modelov, zhlukovania podľa najbližších susedov a prístupov založených na izolácii anomálnych vzoriek. DenStream je algoritmus zhlukovania založený na hustote prispôsobený z DBSCAN na zhlukovanie prúdových dát do ľubovoľne tvarovaných skupín. Half-space strom predpokladá, že náhodné delenie v každej osi v priestore atribútov izoluje odľahlé hodnoty do samostatných oddielov skôr ako nedeviantné pozorovania.

Presná viac triedna klasifikácia príčin porúch stroja podľa vopred známych charakteristík je oveľa náročnejšia úloha ako objavenie anomálií. Algoritmus k-najbližších susedov (KNN) priradí pozorovanie triede, do ktorej patrí väčšina  $k$  bodov v blízkom okolí podľa použitej miery vzdialenosťi. Nachádza uplatnenie aj v učení čiastočne s

učiteľom, pretože dokáže odvodiť označenia len zo znalosti niekoľkých anotácií.

Ďalším prístupom je online alebo postupné učenie, ktoré aktualizuje parametre modelu s každou novou prichádzajúcou udalosťou. Tento prístup je užitočný pri spracovaní veľkých dát, kedy celý súbor údajov nie je k dispozícii vopred alebo ho nemožno spracovať naraz z dôvodu pamäťových obmedzení.

## A.5 Výskumné otázky

Cieľom tejto práce je poskytnúť odpovede na štyri výskumné otázky:

1. Aké atribúty dokážeme extrahovať z vibračných signálov?
2. Akú úsporu dát dosiahneme výberom atribútov?
3. Aké budú presnosti diagnostiky porúch s rôznymi sadami atribútov?
4. Ako môžeme priebežne označovať poruchové stavy?

## A.6 Návrh metód s MaFaulDa

Súbor údajov MaFaulDa používame ako smerodajný pri určovaní metód schopných nasadenia na senzorovú jednotku. MaFaulda obsahuje 1951 záznamov so vzorkovaciu frekvenciu 50 kHz a označenými simulovanými poruchami rôznej závažnosti. Nahrávky obsahujú časové rady z dvoch trojosích piezoelektrických akcelerometrov. Po úprave ponechávame 6 typov značiek: referenčný bezporuchový stav, dve poruchy hriadeľa (nevýváženosť, nesúososť) a tri poruchy ložísk (poruchy klietky, gulôčok, vonkajšieho krúžku).

Zo signálov rozdelených na päť jednosekundových častí sa odstraňuje jednosmerná zložka odčítaním priemernej akcelerácie, po ktorej nasleduje dolnopriepustný 10 kHz filter. Z časti signálu je potom vytvorených 10 časových a 11 spektrálnych premenných. Euklidovská norma trojrozmerných atribútov eliminuje závislosť na smere merania. Pri výbere atribútov nie je do množiny pridaná taká dvojica, ak ich absolútna hodnota korelácie presahuje 0.95.

Predpokladáme, že dátové body rozprestreté v každej dimenzii priestoru môžu dobre rozlísiť skupiny. Ukazuje sa, že premenné sú navzájom viac korelované v

časovej doméne ako ilustruje analýza hlavných komponentov. Pre 95% vysvetleného rozptylu PCA sú v časovej doméne potrebné 3 zložky, zatiaľ čo vo frekvenčnej doméne 4 zložky. PCA efektívne vyjadruje atribúty v menej rozmernom priestore, ale ich výslednou lineárnu kombináciou sa ľahko odôvodňujú rozhodnutia modelu.

Pri posudzovaní všeobecne najdôležitejších atribútov vychádzame z trojíc nekorelovaných atribútov vytvorených v 24 scenároch. Podmienky scenárov vznikli kombináciou štyroch kritérií: dávkové alebo inkrementálne učenie, pozícia ložiska, predikovaná premenná, limit na rotačnú rýchlosť. Na základe schvaľovacieho hlasovania sú najčastejšie sa vyskytujúcimi atribútmi v časovej doméne: špička-špička, faktor tvaru, faktor výkyvu. Vo frekvenčnej doméne sú to spektrálne ľazisko, roll-on a roll-off.

Tri typy experimentov s modelom KNN prebiehajú s validáciou metódou hold-out. Najprv sa model naučí všetky extrahované prvky, takže nedochádza k výberu prvkov. Metódou hrubej sily sa hľadá kombinácia troch atribútov s najvyššou trénovaciou presnosťou. Následne sa porovnáva presnosť modelu pre tri atribúty zvolené technikami výberu atribútov. Dávkový model KNN slúži ako referenčný, podľa ktorého sa posudzuje KNN v postupnom učení.

Online učenie napodobňuje sťažené podmienky diagnostiky strojov, ktoré sa objavujú v praxi. Oneskorené dodanie alebo vynechanie skutočných značiek nepochybne znižuje spoľahlivosť klasifikácie. Modely KNN v experimentoch s postupným učením sú trénované na rovnakom základnom súbore údajov pre ložisko A nad všetkými extrahovanými atribútmi. Týmto spôsobom môžeme porovnať trénovacie presnosti tréningu pre dávkové a postupné učenie. KNN sme nastavili na 5 susedov a euklidovskú vzdialenosť. Metriky online učenia sa vyhodnocujú v metódou progresívnom vyzodnocovaní na ďalšom nevyváženom súbore údajov.

## A.7 Zber vibrácií v priemysle

Doteraz uplatnená metodika pre súbor údajov zaznamenaných v laboratóriu sa aplikuje na vibračných signáloch z priemyselného prostredia. Pri monitorovaní zužitkujeme mierne prispôsobený postup z noriem. Ten zahrňa výber strojov určených na monitorovanie, identifikáciu pozícií na meranie podľa technických štandardov, pred-

bežné merania a vývoj senzorovej jednotky. Zber nového súboru údajov sprevádza vopred dohodnutý harmonogram.

Na zber údajov boli vyčlenené dva špirálové kompresory ako súčasť klimatičačných jednotiek pre dátové centrum a tri čerpadlá s troma elektrometrami v prečerpávacej stanici na pitnú vodu. Dlhodobejšie merania uskutočníme vlastným vnoreným systémom na báze vývojovej dosky ESP32-PoE-ISO so slotom na SD kartu. Ako senzor vibrácií použijeme MEMS akcelerometer IIS3DWB. Vyznačuje sa vysokou šírkou pásma až 6.3 kHz, nízkym šumom, a vysokou výstupným dátovým tokom 26.7 kHz cez SPI zbernicu.

## A.8 Záver

V diplomovej práci sme sa zamerali na výber trendových ukazovateľov pre riešenie monitorovania prevádzkového stavu a odhaľovanie porúch z vibračných signálov. Extrahované premenné pochádzajú hlavne z popisných štatistik, z článkov o spracovaní zvukových signálov a technických noriem vibrodiagnostiky.

Dosiahnuté stratové kompresné pomery pre MaFaulDa sú 2381:1 pre všetky atribúty a 25000:1 pre šest atribútov. Výber atribútov metódou súčinu poradí zabezpečí väčšinou najlepšiu presnosť KNN modelu oproti metrikám samostatne. Žiadny prístup však nedokázal nájsť trojicu prediktorov s presnosťou blízkou optimálnej, ktorá je až 98%. Trénovanie KNN na troch hlavných komponentoch prinieslo lepšiu presnosť ako výber atribútov.

Model postupného učenia kNN dosahuje prinajlepšom 90% presnosť s okamžitou spätnou väzbou, 85% so značkami oneskorenými o 250 pozorovaní a 82% s iba 25% anotovaného súboru údajov. Porovnatelný model trénovaný v dávkach dosahuje presnosť 98%. Doteraz boli urobené závery podľa súboru údajov MaFaulDa, ktoré plánujeme overiť na súbore údajov získaných v priemysle počas DP3.

# Appendix B: Work plan

## B.1 Summer semester - DP1

Period	Work
1 <sup>st</sup> week	Consultation with the supervisor on directions of the future work based on literature review during the previous semester.
2 <sup>nd</sup> week	Outline the key sections of the analysis part in the thesis.
3 <sup>rd</sup> week	Match supporting literature with analysis sections. Further investigation on the feature engineering methodology in CbM.
4 <sup>th</sup> week	Summarize notes from condition monitoring articles and video recordings of tutorials and conferences.
5 <sup>th</sup> week	Research transformation of a vibration signal to feature space using time-frequency, harmonic, and energy statistical metrics. Progress report meeting with the supervisor.
6 <sup>th</sup> week	Find articles and take notes about unsupervised and semi-supervised techniques in streaming data for machinery diagnostics.
7 <sup>th</sup> week	Narrow down a wide variety of applicable methods for signal decomposition.
8 <sup>th</sup> week	Exploratory analysis on evaluation datasets. Progress report meeting with the supervisor on the topic of related work.
9 <sup>th</sup> week	Organize detailed outline out of notes gathered during literature research.
10 <sup>th</sup> week	Write up the problem analysis about condition monitoring and evaluation datasets.
11 <sup>th</sup> week	Write up the analysis section about feature engineering.
12 <sup>th</sup> week	Write up the analysis section about machine learning diagnostics and consult the final choice of methods in the analysis section.

## B.2 Winter semester - DP2

Period	Work
1 <sup>st</sup> week	Semester kickoff meeting to set goals, and experiments and discuss the status of collaborations with partners.
2 <sup>nd</sup> week	Arrange collaboration with an alternative industry partner. Prepare a checklist for the technical inspection of machinery. Construct a device for exploratory measurements.
3 <sup>rd</sup> week	Technical inspection of air conditioning units in the data center. Feature extraction step to calculate features from MaFaulDa.
4 <sup>th</sup> week	Consultation about plan for machine to measure in the data center and better device for measurement.
5 <sup>th</sup> week	In feature selection using various metrics to determine sets of best features in the MaFaulDa.
6 <sup>th</sup> week	Feature selection used in KNN multiclass and binary classifier.
7 <sup>th</sup> week	Explore incremental learning KNN algorithm with MaFaulDa dataset. Consultation and status report.
8 <sup>th</sup> week	Shorten analysis chapter about wavelets. Look at clustering detection incremental learning in MaFaulDa.
9 <sup>th</sup> week	Refactor separate trials in feature selection to integrate them into the pipeline for kNN validation. Consultation to discuss progress.
10 <sup>th</sup> week	Include incremental learning in analysis. Experiment with gradual feature selection in incremental learning. Fix the sampling rate issue in the provisional sensor unit.
11 <sup>th</sup> week	Consultation in preparation for machine inspections. Preliminary measurements of fan, compressors, and pump.
12 <sup>th</sup> - 15 <sup>th</sup> week	Write up design chapter: research questions, visualization export. Finish writing chapter on design and implementation.

The semester for DP2 was split into 3 periods. Feature engineering and ML model evaluation was planned from 1<sup>st</sup> to 4<sup>th</sup> week, and technical inspections and plan of measurement in weeks 4<sup>th</sup> to 8<sup>th</sup>.

In reality, these tasks switched order because between 1<sup>st</sup> to 5<sup>th</sup> week, we needed to acquire alternative partner and check their machineries for viability in our study. Only then from 4<sup>th</sup> to 11<sup>th</sup> week the main focus was on experiments with the MaFaulDa.

Lastly, preliminary measurements were conducted, as was the plan. Requirements were presented for the new sensor device and its firmware.

## B.3 Summer semester - DP3

Period	Work
1 <sup>st</sup> - 2 <sup>nd</sup> week	<ol style="list-style-type: none"><li>1. Incorporate suggestions from DP2 defense.</li><li>2. Develop and test firmware for designed sensor unit.</li></ol>
2 <sup>nd</sup> - 10 <sup>th</sup> week	<ol style="list-style-type: none"><li>1. Long-term manual regular gathering of vibration data from compressors and pumps.</li><li>2. Find optimal number of features.</li><li>3. Try to increase KNN prediction metrics.</li><li>4. Put together incremental learning with feature selection.</li><li>5. Explore incremental clustering in search of method that can help with labeling.</li></ol>
10 <sup>th</sup> - 12 <sup>th</sup> week	Conclude results from measurements in real environment.



# Appendix C: Technical documentation

## C.1 Data collection plan

The vibration measurements will occur monthly for pumps and biweekly for compressors from February 2024 until May 2024. Each measurement will involve 3

Machine	Placement \Month	02/2024	03/2024	04/2024
Compressor AC #3	SFTA001AT000TN			
	SFTA002AT000TN			
	Steel base (noise)			
Compressor AC #5	SFTA001AT000TN			
	SFTA002AT000TN			
	Steel base (noise)			
Pump KSB position #1	MTRA001AT090TN			
	MTRA002AT000TN			
	PMPA003AT000TN			
	PMPA004AT000TN			
	Motor base (noise)			
	Pump base (noise)			
Pump KSB position #7	MTRA001AT090TN			
	MTRA002AT000TN			
	PMPA003AT000TN			
	PMPA004AT000TN			
	Motor base (noise)			
	Pump base (noise)			
Pump Sigma	MTRA001AT000TN			
	MTRA002AT000TN			
	PMPA003AT000TN			
	PMPA004AT000TN			
	Motor base (noise)			
	Pump base (noise)			

trials for each position. The sensor is mounted to the machine with adhesive on double-sided carpet tape (placement notation follows MIMOSA convention from ISO 13373-1). After every trial, the sensor will be detached and attached again.

The plan was consulted and approved by the machinery owners: VNET a.s. and Bratislavská vodárenská spoločnosť, a.s. Triaxial recording has a duration 60 s (at  $f_s = 26.7$  kHz) and a 16-bit resolution. The total space requirements for 108 recordings are 990 MiB.

## C.2 Firmware activities

Activity diagrams describe three basic actions of firmware for logging accelerometer samples:

- begin new recording,
- log block of samples into CSV file,
- finish recording.

