
T1.1.1 Knowledge Package AI-based Predictive Maintenance Systems for Manufacturing SM

Contents

Nomenclature	2
1. Introduction	2
1.1. PredMain in Brief	3
1.2. Artificial Intelligence in PdM Domain	3
1.3. PdM Workflow	4
2. Review Results on the State of the Art	5
2.1. State of the Art of PredMain with AI at the Signal Processing Level	5
2.1.1. Narrow AI methods in PredMain	5
2.1.2. General AI methods in PredMain	6
2.2. State of the Art of PredMain with AI at the Decision and Management Level	6
2.3. Survey Analysis among SMEs	7
2.4. State of The Art of General-Purpose AI (via Illustrative Example)	12
2.5. Conclusions on Reviewed State of the Art and the Survey	16
3. Findings and Resulting Knowledge from the Design and Implementation of the Demonstrator 16	
3.1. The Demonstrator with Ball Screws (BSCs)	16
3.2. The Fatigue of BSC	17
3.3. Service Life and Load Cycle	18
3.4. Sensors and Monitored Variables	19
3.5. Setting up Communication with Predictive Maintenance Algorithms	19
3.6. Findings Relevance of the AI at the Signal Processing Level	19
Vibration Fluctuations (Bearing Main STD).....	21
Mean Stiffness.....	22
3.7. Findings Relevant to the AI at the Business Intelligence Level of the Demonstrator	23
4. Analysis of Relevant Projects in the Area	24
4.1. Machine Learning for Anomalous Vibration Detection Under the Noise Level in Nonstationary Environment	24
4.2. PredMain with AI at the Signal Processing Level	27
4.3. Machine Learning Based Anomaly Detection (Learning Entropy) in Long-Term Periodic Processes	30
5. Synthesis for Small and Medium Enterprises	30
5.1. PredMain with AI at the Signal Processing Level	30
5.2. PredMain with AI at the Decision and Business Intelligence Level.....	31
6. Synthesis of Survey Results, Real Demonstrator Achievements, and Research Findings and Potentials	34
References	35
Appendix	37
AI-Generated Code (the Example)	37

Nomenclature

AI	... Artificial Intelligence	HI	... Health Indicators
AP	... Automation Pyramid	LLM	... Large Language Model
BSC	... Ball Screw	LSTM	... Long-Short Term Memory, a popular class of deep neural networks
CMM	... Condition Monitoring based Maintenance (CMM)	ML	... Machine Learning
CNN	... Convolutional Neural Networks	PdM	... Predictive Maintenance
DL	... Deep Learning	PvM	... Preventive Maintenance
DNN	... Deep Neural Networks	RM	... Reactive Maintenance
GAI	... General AI, also General-Purpose AI, (DL based AI trained on incredibly large amount of data, i.e. texts, images)	ROI	... Return of Investment
		RUL	... Remaining Useful Lifetime
		SMEs	... Small and Medium Enterprises

1. Introduction

The document aims to share extensive knowledge and project results on AI-assisted Predictive Maintenance (PdM) with small and medium-sized enterprises (SMEs) across borders. This is intended to facilitate SMEs' access to PDM and implementation of PdM, and to create a competitive advantage resulting from Artificial Intelligence (AI) and Machine Learning (ML) algorithms and methods.

The document aims to provide SMEs with knowledge and project results on AI-assisted Predictive Maintenance (PdM) to facilitate their access and implementation of PdM, and to create a competitive advantage from AI and ML algorithms. The document discusses the evolution of PdM and the use of data-driven condition monitoring for early failure prediction using AI.

In principle, the AI shall be divided into Narrow AI and General AI, with the former focusing on signal processing and including also deep learning tasks and the latter on complex tasks that exceed the 1.5-year PredMain project scope. The document focuses primarily on the achievements of Narrow AI for predictive maintenance, especially at the signal and sensor processing level, and addresses the potential of AI for business intelligence tasks. Section 2 reviews the use of AI for PdM including the survey among SMEs, and briefly mentions the very ongoing general-purpose AI connotations (since GPT-3). Section 3 reports on the findings and knowledge gained from the demonstrator, including the proposed concept of the cost model with its potential for AI in business intelligence. Section 4 discusses relevant findings from the PredMain team's other projects and experiences related to machine learning for feature analysis and signal processing for PdM. Section 5 synthesizes the gained knowledge with current theoretical achievements and points out the need for focus on both machine learning-based methods at the signal processing level and their interconnection to higher (business) levels of the automation pyramid via the concept of the proposed cost model. Section 6 concludes the document and the project.

As an aside, it is worth noting the following. Given the current developments in AI at the time of this report's completion, it is very newly publicly claimed that the real revolution may come from OpenAI's recently developed and widely available AI such as ChatGPT [1], which belongs to the Large Language Models (LLM) category. GPT version 3 is a very new but already publicly available tool that falls into the category of so-called general AI (GAI). It is an artificial neural network with about 75,000,000,000 parameters that is incredibly expensive to develop, train, and run [2]. In the near future (this decade), GAI may be used by SMEs as a certain partial knowledge base (if it proves to be trustworthy both technically and legally [2], [3]). However, GAI is far from solving and guaranteeing real solutions for PdM tasks without the knowledge and experience of enterprises, and the involvement of their engineers

for specific and applied tasks is still inevitable (we give a brief basic discussion of current GAI in subsection 2.4))

1.1. PredMain in Brief

PdM is an evolution of preventive maintenance (PvM), reactive maintenance (RM), and condition monitoring-based maintenance (CMM).

Based on data-driven condition monitoring, early failure prediction using AI based on data-driven ML algorithms defines the mainstream in PdM research.

Since October 2021, a consortium is developing a comprehensive knowledge package on AI-based predictive maintenance as part of the Interreg research project PredMAIn. With the right selection of sensors, PdM can detect changes in the operating status of components long before a failure occurs. This gives maintenance personnel enough time to timely order components and perform replacement at a proper time.

1.2. Artificial Intelligence in PdM Domain

The artificial intelligence we are looking at can be divided into two main categories, namely the so-called:

- Narrow AI, and
- General AI or General Purpose AI (GAI).

'Narrow AI' can be defined as those machine learning techniques that focus on a single purpose or less complex tasks (as if these tasks could be easily solved by a human but are challenging for traditional algorithms and methods). Narrow AI can include deep learning and deep neural networks, which are the most popular AI algorithms today (also used for general AI, e.g., at the sensory level (vibration analysis, image processing)). An example of a narrow AI application could be the detection of an anomaly or even a specific error in vibration data.

Narrow AI methods that use shallow neural networks, i.e., less complicated neural networks that can be analyzed mathematically, may also be a good option for a task with limited training data. General AI involves complex algorithms that learn from an extremely large amount of data (i.e., observations) and solve complex tasks that are generally non-trivial even for a human. The most common tool for general AI methods today is deep neural networks.

Training general AI would, of course, require very large amounts of training data. In industrial applications, the AI techniques in the domain of PdM and its applications extend from the sensor level up to business-level decision-making. Thus, it makes sense to distinguish, at least, between (various) AI implementations, specifically at:

- the signal processing level (typically up to SCADA systems, processing sensor data via application of narrow AI) (general AI applications are not known), or at
- the business intelligence level (MES to ERP systems)

where we can refer to the general concept of automation pyramid (AP) as sketched in Fig. 1 which shows the span of problems and areas where AI can be applied.

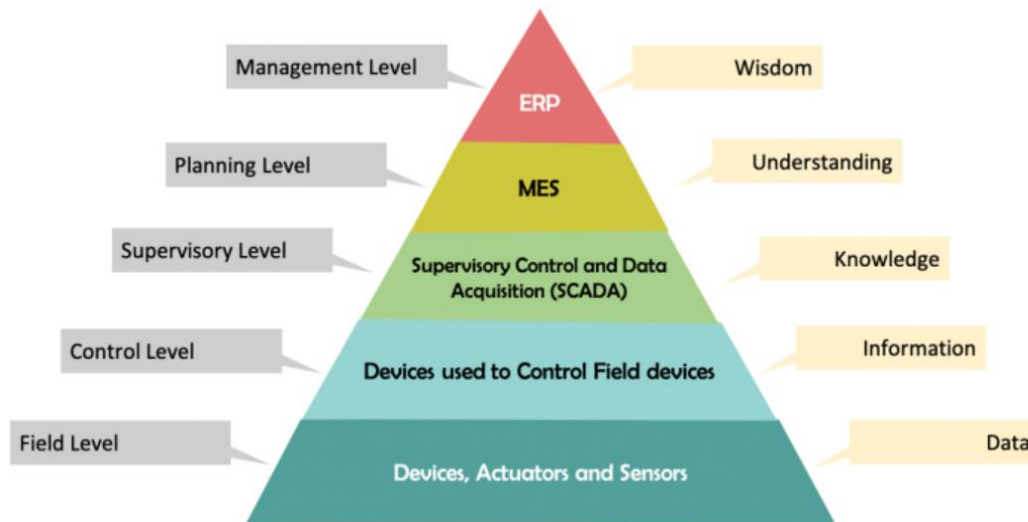


Fig. 1: Automation Pyramid (AP) (adopted from “Automation Pyramid [4].”)

Of course, the distinction of automation levels according to the automation pyramid is not strictly followed in practice, and different SW solutions may span different levels.

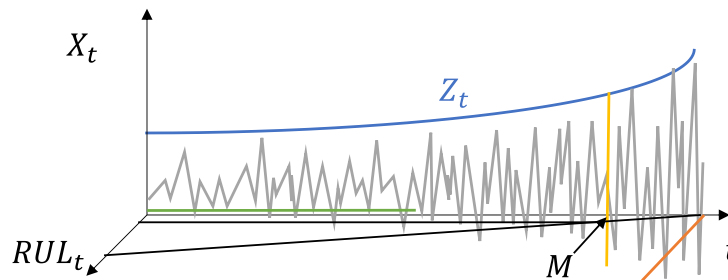
Many machine learning and AI solutions today, in the narrow sense, address only a single level (or a limited number of levels) of the automation pyramid. However, more general AI that would cover a variety of levels of the automation pyramid, or AI solutions that generally cover one level of the pyramid but for multiple problem domains and multiple machines, and that rely solely on training data and machine learning, still seem to be mainly at the academic level.

Given the current state of the art, the shortened duration of the PredMain project, and the readiness of industry (SMEs), we focus primarily on the achievements of Narrow AI for predictive maintenance, especially at the signal and sensor processing level, which corresponds to the lower levels of AP (and of course, it makes sense to start there). However, we also briefly address the potential of AI for business intelligence tasks (upper levels of AP), esp. in sense of the PdM cost model's concept.

This knowledge package is organized as follows. Section 2 reviews the use of AI for PdM in view of its use at the signal processing level, then at the higher (management) level, and these are concluded in contrast to the feedback of carried survey on SMEs. Section 3 reports the finding and knowledge gained by the demonstrator, and the proposed concept of the cost model with its potential for AI at business intelligence is discussed too. Section 4 concludes the knowledge pack for SMEs including relevant findings from PredMain-related projects and Section 5 synthesizes the findings on PdM for SMEs

1.3. PdM Workflow

Initially, it was necessary to produce a synopsis of how typical predictive maintenance applications are organized in terms of the tasks they perform, individually, and in which order. The literature search performed in the context of publication [5] culminated in a workflow that identifies these tasks in a general, data-independent way (Fig. 2).



1. Failures/Faults	2. Regular Behavior	3. Anomalous Behavior	4. Health Indicators	5. Remaining Useful Lifetime
<ul style="list-style-type: none"> Given, or not given, then obtain by anomaly detection 	Identify by: <ul style="list-style-type: none"> characterising process measure regularity criteria 	<ul style="list-style-type: none"> Feature Extraction Known: Supervised anomaly detection Unknown: Use unsupervised AD 	<ul style="list-style-type: none"> Feature selection New labels Z_t Health Indicator (HI) 	<ul style="list-style-type: none"> Prediction Problem Supervised: (Z_t, RUL_t) From this decision about maintenance time M

Fig. 2: PdM workflow, a sketch of the typical progression of tasks (the true one may deviate from case to case), each task 1-5 may be assisted by Machine Learning, critical features are selected, classification to 'healthy' / 'unhealthy', regression for remaining useful lifetime (Z_t, RUL_t) ('Many labels' even though few anomalies!)

Using the experiences from former projects and insights from current literature ([6], see the discussion in Sect. 4.1), we observed the typical workflow of PdM shown in Fig. 2, at the beginning of the project. After making sure that there is labeled data (Step 1), existing data is analyzed concerning regular (Step 2) and irregular (anomalous, Step 3) behavior using feature extraction. This leads to the ability to train and use health state classifiers (Step 4) and remaining useful lifetime forecasts (RUL, Step 5), with the help of AI-learning models.

2. Review Results on the State of the Art

The summary of the academically reviewed state of the art is presented in two subsections 2.1 and 2.2 that are synergically mirrored by reality via survey results from SMEs in subsection 2.3.

2.1. State of the Art of PredMain with AI at the Signal Processing Level

A systematic and extensive review of the state of the art of predictive maintenance with machine learning and AI has been carried out within the PredMain project [5]

One important conclusion from reviewed work [7] is given by the following citation: “*Utilization of Domain Knowledge: deep learning is not a skeleton key to all machine health monitoring problems. Domain knowledge can contribute to the success of applying DL models on machine health monitoring. For example, extracting discriminative features can reduce the size of the followed DL models and appropriate task-specific regularization term can boost the final performance*“. This important necessity of an expert (field engineer) input is further discussed and reported also via findings in our project PredMain in a further section of this document (subsection 3.6) and further supported by discussion in subsection 4.2.

2.1.1. Narrow AI methods in PredMain

Well achievable and positive factors:

The field of machine learning for signal processing, including neural networks (as a fundamental tool of AI), has developed extremely well, starting with real-time learning filters, supervised and (especially) unsupervised anomaly detectors, time-frequency analysis of vibrations by image processing using deep convolutional neural networks with automatic feature extraction, denoising autoencoders, and much more.

With the right technological knowledge about a particular process, i.e., by involving experts, the solutions can be best optimized, false positives can be minimized, and the solutions can be computationally efficient and not overcomplicated.

Limiting and negative factors:

Following [5], the most pressing problem is the unavailability of data, both in terms of required volume and qualitative importance (lack of labeled errors). Financial resources for the provision of adequate sensory equipment and massive training data management play some role, and the use of cloud solutions to manage adequate high-capacity data transmission is increasingly accepted, as noted in particular in the 2020 survey [8].

The reliable and general use of narrow AI for various applications is not possible without the involvement of an expert (field engineer) with the technological skills and knowledge of the process and technology in question, see for example Fig. 22 in this document.

2.1.2. General AI methods in PredMain

Well achievable and positive factors:

There are well-developing and promising scientific methodologies based on deep neural networks, transformer networks, etc., that can extract knowledge from documents, process data, and from engineers in a very advanced and automated way.

Limiting and negative factors:

However, these general AI concepts are still far from real applicability to PredMain for SMEs. Reasons include the need for too large training data, the need for relatively overwhelming computational power (for general AI), and currently, the UE legislation about such general AI decisioning is still to come (e.g. [3]) with its potential consequences to SW and AI developers that are not known yet.

2.2. State of the Art of PredMain with AI at the Decision and Management Level

From our research in [5], it results that the reasons for barriers to PdM are lack of awareness, lack of communication, and lack of adherence to the technology being used. SMEs have limited or no knowledge of the latest solutions, innovations, and technologies offered by researchers and industry IT. There is also a lack of qualified personnel who can handle new data-driven software and solutions. The main reason for the unavailability of data is that the equipment and systems used in SMEs are almost obsolete and cannot generate data due to the lack of sensors. The outdated equipment in SMEs also has issues with monitoring, as PdM works in a real-time environment to scan the data and generate alerts for expected failures. Implementing PdM in SMEs is a barrier due to the lack of skilled technicians. The technical challenges are most desperate after implementing PdM because the company is sure that our work will not be stopped by an outage; however, a technical problem in the PdM solution would stop the processes, which would slow down production.

Most of the selected papers have described the challenges of technical methods, unavailability of data, and lack of qualified personnel for PdM in SMEs, as shown in the systematic map in Fig. 4 with the corresponding support level. It is also described that the devices installed in SMEs lack sensor technologies to collect data to monitor specific device conditions, leading to the problem of unavailability of data analysis. Device dependency is also a challenge for SME PdM solution providers. These cross-functional barriers in manufacturing lead to limited development of AI-based methods. To overcome these barriers to PdM implementation in SMEs, researchers must first focus on cost models,

such as how much it would cost for an ordinary SME to upgrade its equipment to sensor technologies to implement the smart factory concept.

One important finding is that current research lacks cost models, which could be the biggest hurdle for small and medium-sized enterprises to adopt PdM. This is because SMEs do not know the costs of PdM implementation and are uncertain about the return on investment (ROI). However, for large enterprises, PdM offers tenfold ROI, 25 to 30 percent lower maintenance costs, 70 to 75 percent fewer failures, 35 to 45 percent less downtime, and 20 to 25 percent more production [9]. The unique features of PdM that are attractive to SMEs are reducing maintenance costs and downtime, improving product quality and production speed, increasing overall equipment effectiveness, and extending equipment life and availability.

2.3. Survey Analysis among SMEs

The PredMAIn survey collected data on the status quo of implementation of predictive maintenance procedures and processes in small and medium-sized enterprises (SMEs) in the Austria-Czech Republic Interreg program area. To incorporate the view from practice into the study, Software Competence Center Hagenberg GmbH together with Intelligent Predictive Networks GmbH surveyed to determine the status quo of where small and medium-sized enterprises from the project region stand today regarding "Predictive Maintenance".

The most important results of this survey are summarized as follows.

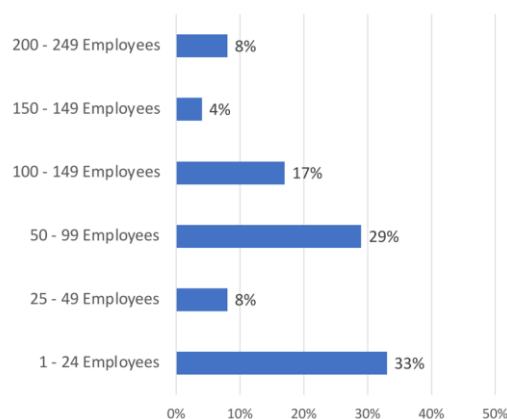


Fig. 3: In summer 2022, an online survey with 40 companies participated, 24 companies matching the SME criteria were included in the study(# of employees < 250 or revenue < €50 Million) , 67% of the companies are from Austria, 33% are from Czech Republic, “Metal”, ”IT” & “Automotive” with 21% each have been most dominant among the participants

In the survey, type of machines and equipment maintenance is being conducted by the participating companies as follows:

- Optical quality control equipment
- Food production equipment
- Metalworking machines
- Air humidification equipment, water treatment equipment
- Mining machinery, motors, conveyor belts, vibratory rollers
- Automated teller machines and kiosk systems
- lasers, press brakes, welding machines
- CNC machines, compressors
- Hydroelectric power plants
- 3D printers, CNC milling machines, 3D scanners

Satisfaction with the results of PdM has been surveyed together with a self-assessment of the corresponding maturity level:

Tab. 1: Self-assessment of the maturity level of surveyed SMEs

Low Maturity Level	Medium Maturity Level	High Maturity Level
<ul style="list-style-type: none"> • company has not yet dealt with PdM • First "idea" to do Industry 4.0, but "no idea" what PdM can do for us. • No database and no know-how in the field of predictive analytics yet. 	<ul style="list-style-type: none"> • First considerations regarding the use cases of PdM made. • Unstructured database distributed across various data sources available for initial analyses. • Little or no know-how in the field of predictive analytics. First analyses were carried out with excerpts from the existing databases. 	<ul style="list-style-type: none"> • Specific use cases for PdM identified and documented • Database available and continuous data collection started. • 46% n = 24 54% (Basic) know-how in the field of predictive analytics available. • Conducting analyses on a regular basis and first attempts with prognosis models started

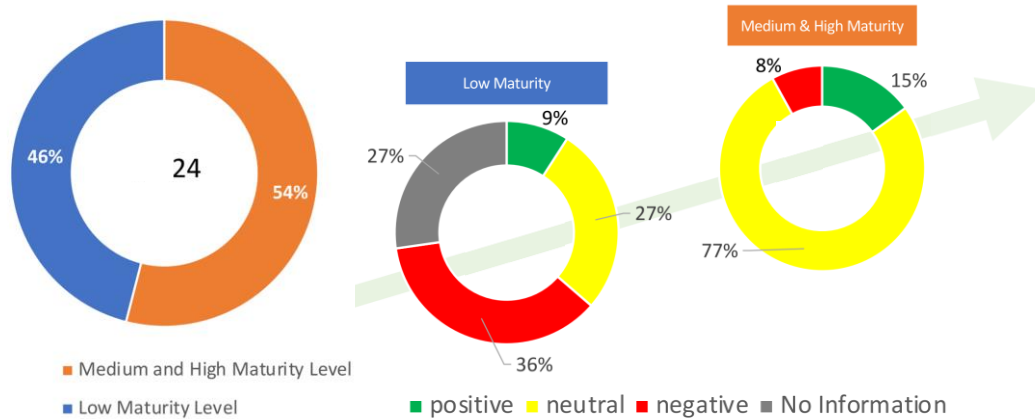


Fig. 4: Although satisfaction with PdM increases with maturity of SMEs in PdM, the ‘neutral’ rating predominates, which indicates the need for further developments in the field of PdM.

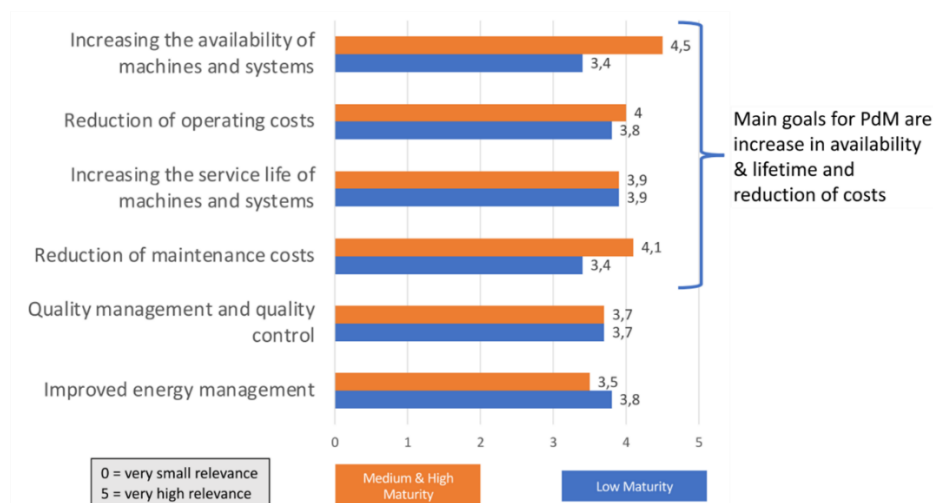


Fig. 5: Cost and availability targets prevail while other (higher level) aspects have not been reported so crucial yet.

The survey on barriers to PdM adoption yielded ambiguous results, as the opposite was actually expected - high scores for barriers among low-maturity companies and low scores among medium- and high-maturity companies Fig. 6.

Medium and high maturity companies have more experience and therefore have a more realistic view of obstacles, as in all categories low maturity companies seem to be less affected by obstacles than their medium and high maturity counterparts.

Particularly "lamented" are the lack of know-how in the approach and the limited technical resources for implementing PdM. In addition, the availability of relevant data is also a problem. Another problem that frequently arises in practice is also reflected in the survey results: the lack of an assessment of the financial impact of a PdM initiative, or put more simply, the lack of a "solid" business case.

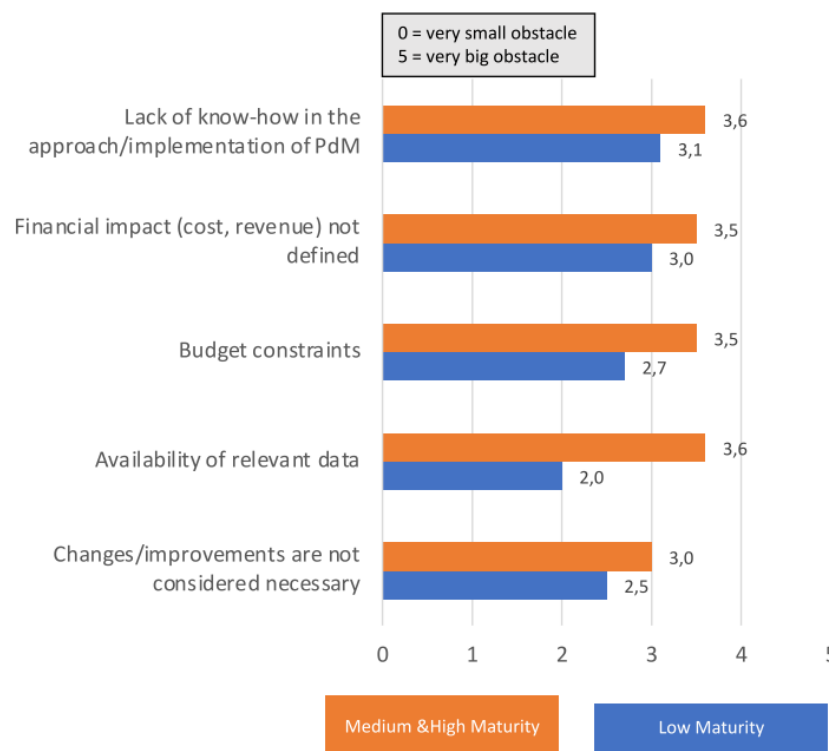


Fig. 6: The biggest obstacle seems to be the way of handling PdM and the costs.

Further, the type of data collected for PdM was surveyed. While the 2018 Austria Study (Predictive Maintenance in Austrian Companies - Österreich -Studie 2018) showed that the higher the maturity level, the more diverse the data sources used for PdM, this is not the case in the SME environment. Everyone seems to use the same "data sources."

A key factor in determining the functionality of a ball screw is positioning accuracy. This is also closely related to the preload in the ball screw nut, whereby the preload is gradually reduced during the operation of the ball screw until it is completely lost and a slight backlash occurs. At this point, the ball screw ceases to be usable in many applications and therefore the end of the life of the ball screw will be considered here. The expected wear development is shown in Figure 3.

Compared to the other data classes, the collection of product and process-related quality data is rather low in all maturity levels. Yet it is precisely this data that, depending on the production process, has a high to very high information content with regard to the condition of the machine and/or the production equipment used on it.

In the 2018 Austria study, companies with a high maturity level used more diverse data sources than companies with a low maturity level.

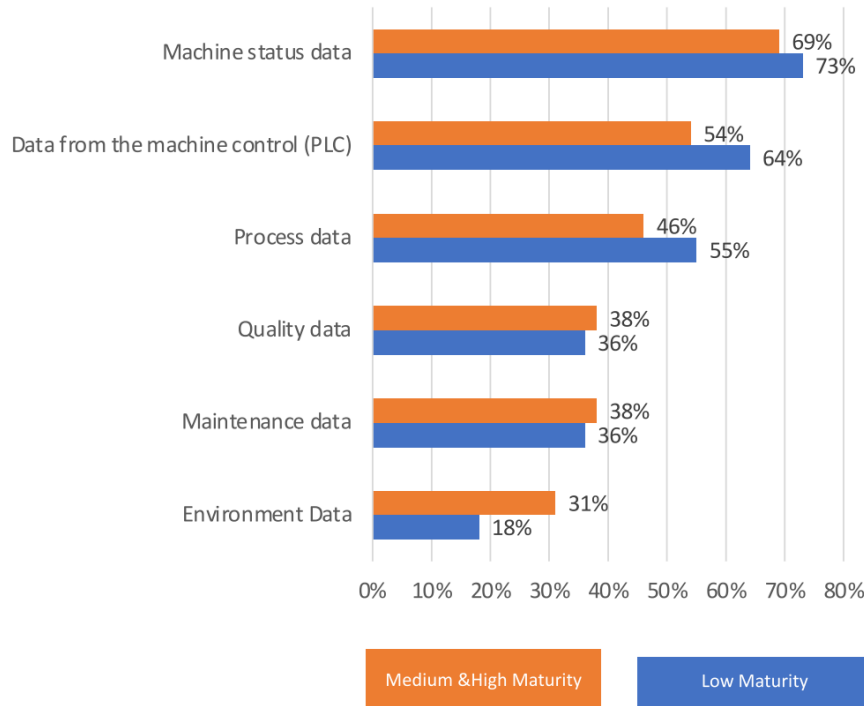


Fig. 7: The use of product and process-related quality data is rather low, indicating that PdM is not yet widespread in SMEs at higher levels of the automation pyramid.

Functions involved in the implementation of predictive maintenance

“PdM initiatives are run by Production and Maintenance”

For PdM mainly specialists from the functional areas of maintenance & repair, technology, and IT are used.

A discrepancy between the medium and high maturity assessment and operational practice is somehow evident. Data Scientists are available in only 54% of companies with Medium & High Maturity Level. This means that only every second of the companies which attest to a medium & high degree of maturity can fall back on corresponding "data specialists". Whether this specialist function provided by data scientists can be covered by specialists from other areas is questionable.

It is possible that companies make use of external specialists who have in-depth know-how in the area of predictive analytics.

Technologies for condition detection - currently in use

“PdM initiatives are run by Production and Maintenance”

The Austria 2018 study* had shown that the higher the maturity level, the more technologies are used for recording the condition of machinery and equipment. In the SME environment, these differences are not so clear cut.

The "classics" electrical condition measurement, vibration, acoustics, lubricant and infrared analysis prevail.

Interesting is the low rate of adoption of electrical condition measurement within the companies with Low Maturity. Current and voltage are usually easy to detect and are therefore included first in the analyses.

What IT infrastructure is used for PdM in your company?

Medium & High Maturity companies have a more specialized IT infrastructure than their Low Maturity counterparts.

No surprise is the use of Microsoft products especially Excel, which is still the leading data science platform in unofficial "studies". The high proportion of in-house server and database platforms is in all likelihood because companies still do not want to take production data "off-site" and many issues relating to IT security and compliance have not yet been resolved when connecting cloud environments to production IT.

The same applies to the high proportion of software developed specifically for the company. Another reason for this could be that the standard software products available do not offer the flexibility that companies consider necessary for implementing PdM.

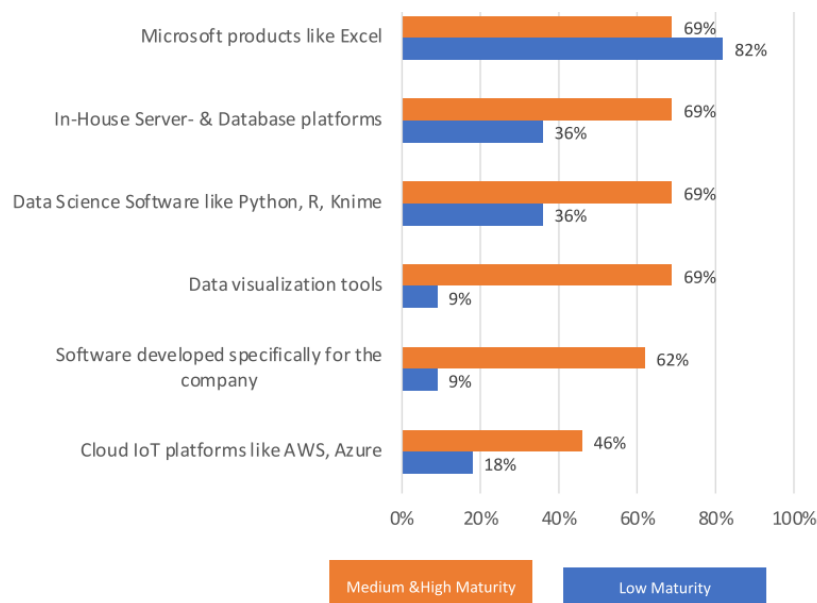


Fig. 8: Medium & High Maturity companies have a more specialized IT for PdM

Partners for the implementation of PdM

Technology partners come first in terms of "importance". Frequently, the introduction and enhancement of PdM do not represent the company's core competence and the companies expect their technology partners to make the greatest contribution in the implementation of PdM. The special, customer-specific requirements must be supported by appropriate technologies and it is therefore hardly surprising that the technology partners are given the highest priority.

At this point, it should also be pointed out that such partnerships go beyond pure technology. Especially in the customer-supplier chain, it is important to clarify how the "data ownership" between the companies is structured. Each of the partners has a (legitimate) interest in accessing the data generated in such projects and using the insights gained to better understand and optimize their own products and develop new services.

It is also encouraging to see that universities are at the forefront of important partners. As it seems, universities are "close to the business community" and these teaching and research facilities are readily

called upon by companies for PdM projects due to the concentration of knowledge and methodological expertise.

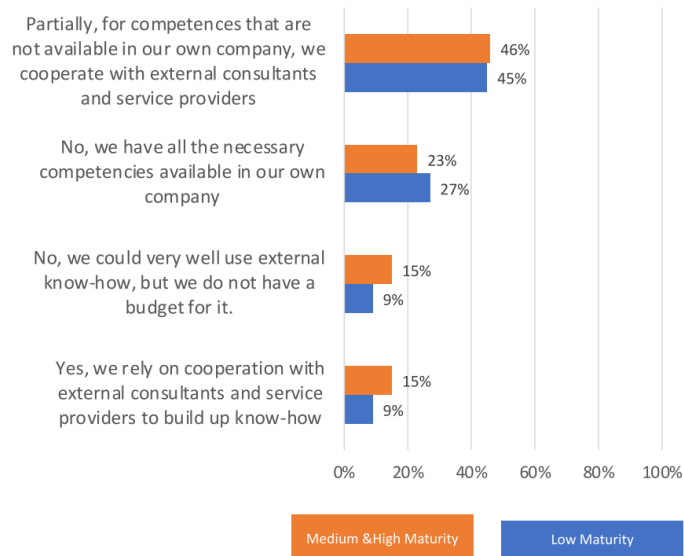


Fig. 9: Customer-specific PdM requirements are best supported by technology partners

To enforce cooperation of companies with universities is suggested to increase maturity in machine learning algorithms and to assure future workers be qualified and available.

Do you work with outside consultants in your PdM program?

“PdM initiatives are mainly done with internal resources”

The data do not really show a difference between Low and Medium/High Maturity companies

Internal and external collaboration is important. This requires companies to be aware of their own strengths and equally aware of their weaknesses. In the case of the latter, it is important to develop further with external help in order to exploit the potential that presents itself.

IPN's experience also shows that initial internal initiatives are often launched and external specialists are only brought on board when the project is critical or on the verge of failure.

The 2018 Austria Study* showed that companies with a high level of maturity collaborate more with external partners

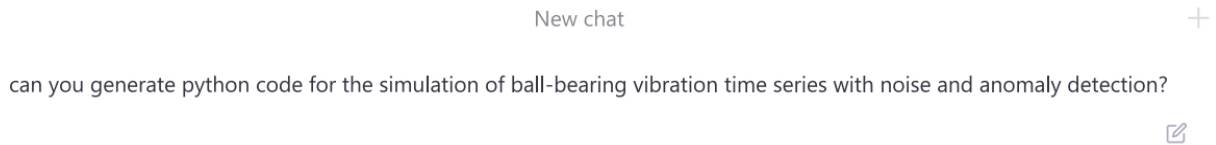
2.4. State of The Art of General-Purpose AI (via Illustrative Example)

At the ending stage of our project, a milestone emerged in general purpose AI worldwide that is to affect the industry and research, and we become to stand at the edge of revolutionary general-purpose AIs; however, we shall not forget that about the importance of the humans in the (AI)loop. We include it here because of its revolutionary importance and very rapid development.

Since November 2022, a general purpose AI (GAI) has been publicly presented, and the open.ai CHAT GPT [1] is just one of the currently available tools, of which new versions are being released these days. In this subsection, we present the current potential of GAI for SMEs as a tool that can be used as a knowledge base that can help SME engineers find new methods and be introduced to machine learning algorithms, even through simple codes that GAI generates. Thus, the maturity of companies in using machine learning algorithms for predictive maintenance could be accelerated if GAI is used with

appropriate critical thinking and caution and by engineers who have the right field experience and educational background, as discussed below.

To demonstrate the importance of the very newest new trends in AI , a question was asked to ChatGPT Feb Version [1] as follows:



At the first trial, the ChatGPT returned its answer including about 50 lines of clear and well-commented Python code as indicated in Fig. 10.

Sure! Here's an example Python code that generates a time series of ball-bearing vibrations with added noise, and then applies anomaly detection using the Isolation Forest algorithm from the Scikit-learn library.

```
python Copy code  
  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
from sklearn.ensemble import IsolationForest  
  
# Generate ball-bearing vibration time series  
time = np.arange(0, 10, 0.01) # time in seconds  
freq = 10 # vibration frequency in Hz  
amp = 0.2 # vibration amplitude  
vibration = amp * np.sin(2 * np.pi * freq * time)  
  
# Add noise to the time series  
noise = np.random.normal(0, 0.05, len(time))  
vibration += noise
```

Fig. 10: The answer and part of ChatGPT generated code, where only one error (see Fig. 11) was to be fixed by setting longer time of simulation, i.e., time=np. arrange (0,100,0.01) and then code run (you can see the full code in Appendix).

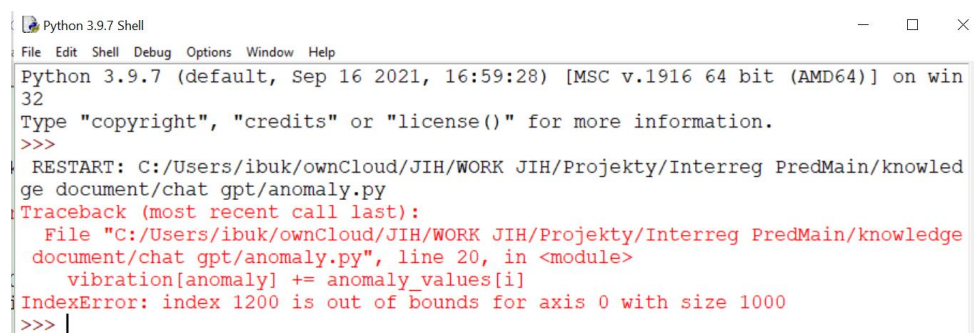


Fig. 11: The only bug in the AI-generated code (ChatGPT) was easy to fix to make the code work (the time of simulation had to be extended to 100 instead 10, see Fig. 10)

Then, the working AI-generated code created a time series, made some data samples anomalous as simple outliers, chose a highly cited method of Isolation Forest [10] for anomaly detection, and applied it via the existing Python module and plotted results as shown in Fig. 12 and Fig. 13.

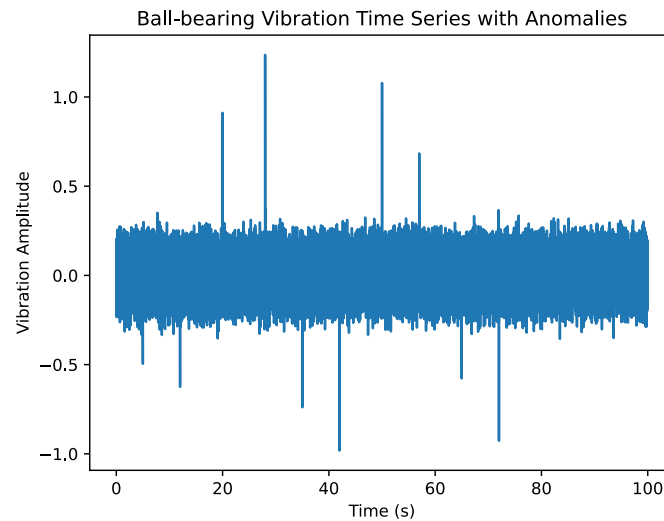


Fig. 12: The General-Purpose AI-generated vibration-like time series with anomalies as simple outliers.

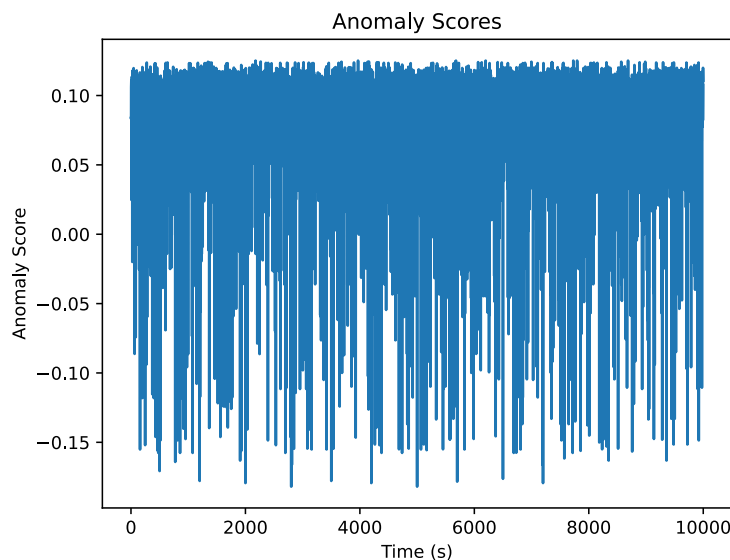


Fig. 13: The anomaly scores calculated by the code purely generated by GAI (Fig. 10-Fig. 12), the outlier anomalies are indeed somehow observable as most negative values (the AI suggested a good method though too complex one for a very simple task).

The AI system (ChatGPT) generated a code for vibration-like data with simple anomalies and anomaly detection as requested. Then the AI system applied a known anomaly detection method and recorded the results. A user can learn from a single code and learn a known and relevant method that exists and what are its use and principle. The code had to be corrected manually, but this was very easy for a user who is experienced in the relevant programming language and understands the task (the user should know what is required and what to expect). In addition, the AI proposed to solve a very simple problem with an unnecessarily complicated method. The problem could have been solved using the most common approach, namely the outlier distance from the mean. Therefore, the request to the AI was modified and submitted as follows:

can you generate python code for the simulation of ball-bearing vibration time series with noise and anomaly detection via a very simple method?

Then, the AI suggested correctly the simplest method (followed by changed code) and the AI described the update as follows:

After simulating the signal, the code uses a simple moving window method for anomaly detection. It iterates over a sliding window of a specified size and calculates the mean and standard deviation of the window. If the value at the center of the window is above the mean plus a specified threshold times the standard deviation, an anomaly is detected and its time is printed.

This brief example is intended to illustrate our conclusions on available general AI (GAI) for SMEs for predictive maintenance (with general connotations) as follows:

- Predictive maintenance using artificial intelligence (AI) is becoming increasingly attractive among small and medium-sized enterprises (SMEs). In particular, the general AI (GAI) provides a revolutionary knowledge base that can help SME engineers think through maintenance problems, generate proposals for potential solutions, and search for modern methods.
- However, it's important to note that GAI should be used properly by users with appropriate backgrounds, analytical skills, and critical thinking abilities. When used by individuals without the skills and critical thinking, GAI can lead to inappropriate methods or solutions.
- It is difficult for SMEs to find and deploy employees with real skills and potential, and it is difficult to say whether GAI will have a positive or negative impact in this regard. Certainly, it is important to properly educate and train employees to develop their skills and understanding of both the physical process and the principles of machine learning so that SMEs can benefit from GAI while eliminating the risks of its blind application.
- It shall be highlighted, that GAI, in principle, is a trained neural network, and its output can be sometimes existing facts and sometimes generalized (artificially invented facts due to the generalization ability of neural networks). That means the outputs of GAI must undergo thorough analysis and verification before being used for a real solution.
- Despite these challenges, the impact of GAI (as a knowledge and skills base) on predictive maintenance in SMEs cannot be underestimated. As GAI technology evolves revolutionary fast, it has the potential to significantly increase a SMEs maturity and enable SMEs to better understand and apply modern machine learning methods to predictive maintenance.
- Considering that a GAI like GPT-3 is an extremely expensive and demanding technology to develop, train, and operate [2] (including the (non-)availability of training data in the industry, demanding HW and wast energy consumption to train and run GAI), most SMEs will not be able to develop and operate their own similarly powerful GAI HW soon (this decade and perhaps later either).
- However, the lack of industrial training data for ML solutions in SMEs can be in the close future partially compensated by newly developing methods of collaborative learning, e.g., Federated Learning [11].
- The use of neural networks and machine learning for targeted analysis, such as automatic analysis of documents and extraction of knowledge by combining neural networks with more deterministic methods (clustering, Bayesian networks, random decision forests, knowledge graphs, graph networks) can be likely expected as a practice for the needs of SMEs. Today, there are many recognized tools (and open-source modules), including pre-trained localized language models, whose performance is acceptably reliable for real-world applications.
- We encountered attempts by SMEs to gain special advantages in solving some programming problems by GAI. For more complicated problems, the use of GAI resulted in taking much

longer to fix the problems that arose afterward. Thus, the use of GAI requires the caution and critical thinking of an experienced engineer.

- The human supervision and evaluation of GAI by experienced engineers in SMEs is inevitable and blind transfer of generated codes and methods into predictive maintenance solutions is necessary.
- The full extent of GAI's impact on industrial solutions and society as a whole remains to be seen.

2.5. Conclusions on Reviewed State of the Art and the Survey

Overall, the particular tasks of Predictive Maintenance are not to be solved by GAI which is nowadays typically represented by large language models (Transformer Neural Networks [12]).

The Predictive Maintenance tasks are rather focused, and when AI solutions are suitable, then narrow AI tools (deep or non-deep) at the level of signal processing are suitable, and such relevant ones are primarily focused in our relatively short-term project PredMain and this document.

3. Findings and Resulting Knowledge from the Design and Implementation of the Demonstrator

3.1. The Demonstrator with Ball Screws (BSCs)

Ball screws are often used in industrial applications such as machine tools and manufacturing equipment, and they must function properly. This can be achieved through predictive maintenance, which is an approach to equipment maintenance where we try to predict and schedule the ideal time for maintenance to prevent production downtime or equipment damage.

With predictive maintenance of ball screws, key parameters such as motor currents, temperature, and vibration can be monitored to assess the condition of the ball screw. This information is then used to predict potential problems and determine the best time for maintenance. This increases maintenance efficiency and reduces the risk of downtime, leading to cost savings and improved performance. Predictive maintenance of ball screws can also help reduce maintenance costs as it is only carried out when necessary and not based on a fixed maintenance schedule. This results in better use of resources and more efficient maintenance.

At the beginning of the predictive maintenance design of a ball screw, like any other piece of equipment or component, many questions and issues need to be addressed. How do we determine the RUL (remaining useful lifetime), or how will the end-of-life be reflected in service? What quantities and variables are we able to monitor? Which ones are useful for predictive maintenance?

The ball screw (BSC) is used when there is a need for precise positioning with high efficiency. The ball screw converts rotary motion to rectilinear motion, usually with the drive motor connected to a shaft that rotates and the rectilinear motion being performed by the ball screw nut.

There are several different ball screw designs and manufacturing processes. From a design point of view, the main issues are whether the ball screw has a defined clearance, what its load capacity is, and how the transfer of the balls between the threads is handled. Accuracy is determined by the manufacturing technology used - a rolled thread has poorer accuracy, while a ground thread achieves higher accuracy. All these parameters have an impact on the production cost and therefore the final price of the ball screw.

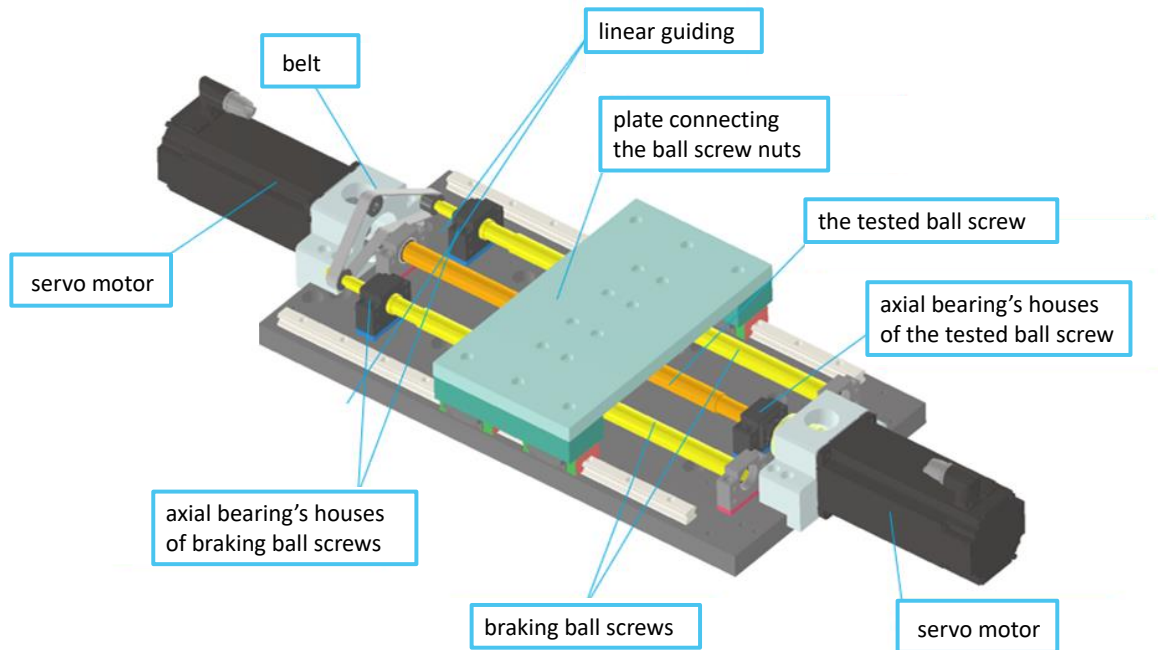


Fig. 14: Visual model of the demonstrator

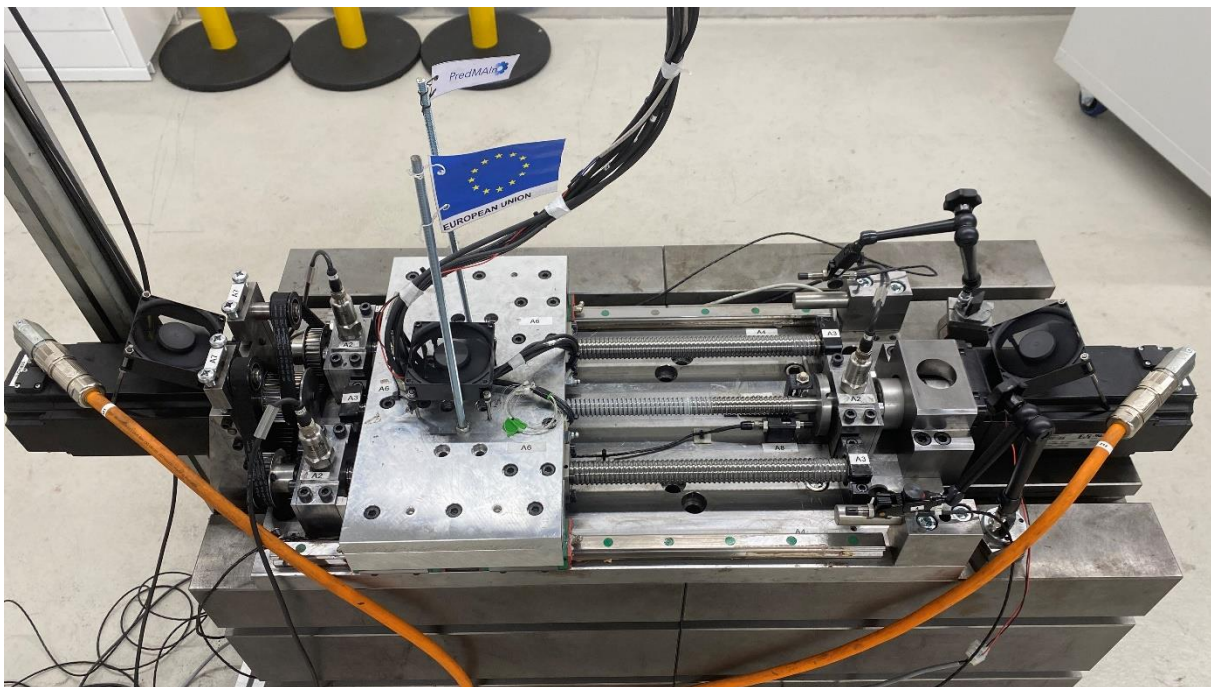


Fig. 15: The demonstrator.

3.2. The Fatigue of BSC

Contact fatigue wear occurs due to cyclic loading. First, the first fatigue cracks form in the thin surface layer. Their subsequent expansion leads to the development of fatigue wear until parts of the surface are broken off. This process, which results in the formation of pits, usually circular, is called pitting. Pitting is greatly influenced by the lubricant which is forced into the surface cracks by the action of pressures, where it is sealed during the passage of the ball, and the pressure of the lubricant increases, contributing to further crack propagation.

As soon as pitting occurs during the wear process, the damaged components should be replaced as soon as possible, as from this point onwards the damage develops very progressively (even exponentially) and leads to rapid component failure. This is mainly due to abrasive wear caused by loose metal particles in the lubricant. This is compounded by other influences such as microscopic shocks (vibrations) caused by pits in the surface and increasing backlash.

During the experiment, the assumption that the damage first occurs in the nut threads, from where it is then transmitted to the balls and ultimately to the ball screw shaft, was confirmed. This is because the balls are hardened and therefore very hard and the shaft thread has a much greater distance than the thread in the nut; therefore, the applied load is distributed over the entire length used.



Fig. 16:: Pitting in the ball screw nut and damaged balls at the end of the experiment

A key factor in determining the functionality of a ball screw is positioning accuracy. This is also closely related to the preload in the ball screw nut, whereby the preload is gradually reduced during the operation of the ball screw until it is completely lost and a slight backlash occurs. At this point, the ball screw ceases to be usable in many applications and therefore the end of the life of the ball screw will be considered here. The expected wear development is shown in Fig. 16.

3.3. Service Life and Load Cycle

Due to time constraints (project duration only 15 months), it was decided to test smaller ball screws because smaller dimensions mean a lower maximum load and also a shorter lifetime. This also ensured that the demonstration was financially economical, as compared to larger ball screws, there are lower costs both for the purchase of the ball screws themselves and for the design and construction of the demonstrator, which can be sized for the lower load. After consultation with the manufacturer, KSK Precise Motion, Inc., a 20 mm diameter ball screw was selected. At the same time, a load cycle was designed by the manufacturer. For an axial load force of 4000 N, 1000 rpm, and a symmetrical test cycle, a lifetime of approximately 720 h was calculated.

The test cycle was programmed according to the manufacturer's design. First, the ball screw under test is loaded with torque from the servomotor, with the braking servomotor holding the ball screw nuts in the same position. Once the desired load is achieved, the braking servomotor adjusts the braking torque so that the connecting plate will move to the other side of the demonstrator at the given acceleration and speed. Here, the orientation of the load torque is changed on the drive motor (e.g. from 3.185 Nm to -3.185 Nm), and when this is reached, the braking motor is used to perform a backward movement.

3.4. Sensors and Monitored Variables

As already mentioned, at the end of the service life, the nut loses tension and backlash occurs. The nut preload cannot be measured directly - when converting the preload force to pressure in the spacer ring, the pressure value was found to be too small and not within the measurable range of any sensors (e.g. strain gauges). However, the preload is related to the ball screw stiffness, which can be measured indirectly, as well as the clearance (or positioning accuracy).

Based on previous experience, it was decided to use primarily vibration analysis, supplemented by information from temperature sensors and motors (torque). Ultrasonic sensors are currently a less explored area and were only developed during the project. Conversely, microphones were not used at all, as they work on the same principle as vibration analysis (sound is vibration in the air), but due to the high environmental disturbance, they will warn of impending end-of-life much later. Tribometers were not used because of the complexity of measurement and evaluation - the ball screws were lubricated with grease, not oil.

Although the number of sensors used is high (11 sensors in total), in practice this number is not necessary. When monitoring the lifetime of a ball screw in real operation, it is sufficient to use two accelerometers in the radial direction, measuring in mutually perpendicular directions (similar to those on the demonstrator ball screw nut). The other accelerometers on the bearing houses were used as additional measurements in a case bearing degradation and failure occurred. In that case, the largest increase in vibration will be on these accelerometers. Capacitive position sensors are not needed in actual operation, because with predictive maintenance algorithms, we will be able to determine the condition of the ball screw using vibration. There is no need for temperature sensors because, as we already know, deterioration of the CS will be reflected in the vibration signal long before the temperature increase that occurs at the very end of the lifetime.

3.5. Setting up Communication with Predictive Maintenance Algorithms

The AI-based predictive maintenance algorithms obtained in Task T1.4.1 were implemented in the COMES Maintenance Management System (T1.5.1). Communication with COMES Maintenance was chosen via the OPC UA communication protocol. The OPC UA server is running in the PLC demonstrator, in which variables were set to allow the AI algorithms to determine the state of the ball screw (Table 8) - the raw vibration data recorded throughout one cycle was selected, as well as the braking torque data (the driving torque is the same throughout the experiment).

3.6. Findings Relevance of the AI at the Signal Processing Level

Already during the first test, it became apparent that the calculated lifetime did not correspond to reality. Therefore, it was decided to change the parameters - the rotational speed was increased to 1300 rpm and the load force was set to 5000 N. This resulted in a reduction of the life of the angular contact bearings located on the tested ball screw to an average of 75 h and they had to be changed very frequently. In addition, problems with data storage were encountered during the first test - storage was unreliable, with frequent storage failures.

Testing continued until the ball screw failed, i.e. until the threads and balls were so badly damaged internally that the screw seized and was unable to move.

Despite the increase in rotational speed and load force, the actual life of the ball screw was 2107 h, almost three times longer than expected. Due to the short duration of the project (15 months), it was necessary to change the parameters very significantly to test the ball screws in time. In connection with the change of parameters, the design of the demonstrator had to be modified - larger bearings with higher

life had to be used, for which at the same time a new bearing house had to be manufactured. Lastly, errors in the algorithm had to be corrected so that data storage problems no longer occurred.

After the design change, SKF 7302BEP bearings were used. Due to the need to speed up the testing, it was necessary to shorten the test time to a maximum of 20 days (480 h). Therefore, the rotational speed was set to 1300 rpm and the load force to 6500 N. A higher rotational speed could not be set due to the significantly increasing vibration and noise. For these defined parameters, the lifetime was calculated to be 127 h (≈ 5.3 days). Taking into account almost three times the actual lifetime compared to the calculation in the first test, it was assumed that the actual lifetime should be 381 h (≈ 16 days). The lifetime of the axial-radial bearings was determined to be 527 h. Considering the assumed test length, the measurement intervals for all recorded data were set to 200 cycles (≈ 17 min).

Unfortunately, during the test, it was again found that the predicted lifetime did not correspond to reality, so the load force was increased again - first to 8,000 N and later to 9,500 N. Despite this significant increase in load force, the total test duration was 427 h. Only in retrospect, after all the tests were completed, it was found that there were large differences in lifetime between the individual ball screws and the lifetime of this second screw was well above average. Because the load was varied throughout the test, the data from this test could not be used for learning AI algorithms.

For the remaining tests, it was decided to increase the load force one last time to 10,000 N, which corresponds to a torque of 7.96 Nm. This value is at the limit of the technical capabilities of the designed demonstrator - for example, the motors used have a rated torque of 7.7 Nm (maximum 10 Nm). At the same time, this load force is still below the dynamic load capacity of the ball screw and bearings and is therefore guaranteed not to affect the relevance of the data and findings obtained. All other parameters remained unchanged, the calculated lifetime was reduced to 35 h (≈ 1.5 days), and the actual lifetime due to previous experience was assumed to be approximately 200 h (≈ 8.3 days). The calculated lifetime of the axial-radial bearings was 146 h.

With this defined stress cycle, the manufacturer calculated a life expectancy of 27219 cycles (test time 35 h). The actual lifetime was almost 4 times longer, with an average of 107850 cycles (test time 138 h), which means that if we followed the manufacturer's calculated lifetime for maintenance, we would have used on average only 25.2% of the total lifetime of the ball screw. At the same time, however, the average actual lifetime was shorter than expected - probably because the second tested bolt had a significantly above-average lifetime. The actual lifetime of the tested screws was found of a large variance, for example, the lifetime of one screw was more than twice as long as that of another screw.

There are several likely causes for the very significant increase in actual lifetime expectancy compared to calculations. Here is a list of the most likely:

1. The manufacturer does not have calculations that have been verified by experimental lifetime tests. In practice, ball screws are loaded with a much more complex range of load parameters, so a comparison with the calculation is not possible.
2. The geometry of the contact between the ball and the thread, particularly the relationship between the radius of the thread and the ball, is a very important factor in service life - even a small change in the radius of the ground thread has a large effect on the service life of the ball screw. If the radius of the thread is smaller than the design (the curvature of the ball and the thread is more similar and therefore the contact area is larger), the service life increases significantly, but at the same time rolling friction increases. On the other hand, if the radius of the thread is larger than designed, the service life decreases and the rolling friction between the ball and the thread decreases.
3. The tests were carried out under laboratory conditions, i.e. without the effects of forces and impurities, which in practice significantly reduce the service life of a ball screw.
4. Manufacturers' service life calculations for ball screws were made more than 20 years ago. Manufacturing quality and precision have likely improved since then, resulting in longer lifetime.

5. Less important is the fact that we cannot immediately change the load orientation compared to the proposed test cycle. If we calculate the average absolute value of the load force during the test, this value is $F_a = 9844$ N, i.e. about 156 N less than indicated in the calculation. If we correct the calculation for this value of the load, we obtain a life of 28529 cycles (36 h).
6. The actual load on the ball screw nut is reduced by the losses in the angular contact bearing pair compared to the load generated by the servo motor. Unfortunately, these losses cannot be measured accurately. However, from the braking torque, it can be estimated that the losses in the bearings are about 1 kN for constant motion and a load force of 10 kN.

Vibration Fluctuations (Bearing Main STD)

Illustrated via Fig. 17- Fig. 19, Fig. 20 shows the fluctuation of Bearing Main standard deviation over the performed test cycles on the demonstrator. In Fig. 20, the X-axis represents the cycles of the experiment, and the Y-axis shows a normalized standard deviation. In overall, the variable has a little change when it represents the healthy state of the system. This change is always not higher than 10% of the maximum possible value. At the same time, it grows dramatically at the end of the experiment when the system is getting its broken state. These facts are very helpful for the classification model.

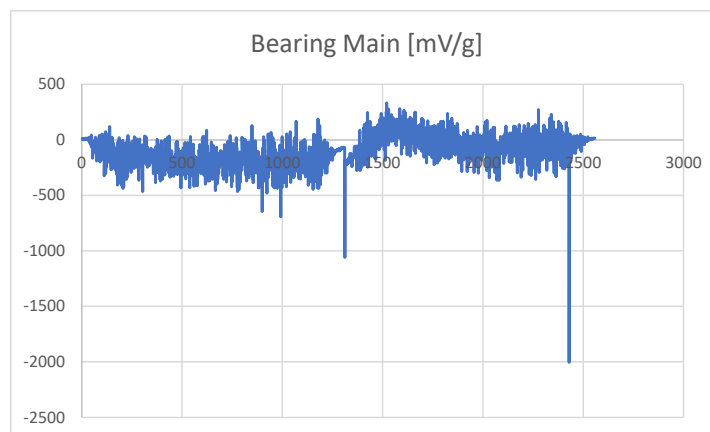


Fig. 17: Example of vibrations measured on the demonstrator at the early operation stage (Raw_2022_08_30_23_49_27)

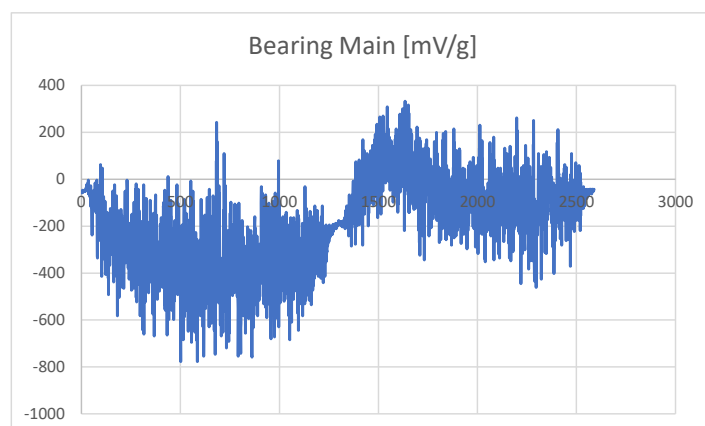


Fig. 18: Example of vibrations measured on the demonstrator at mid-operation stage (Raw_2022_09_06_04_56_08)

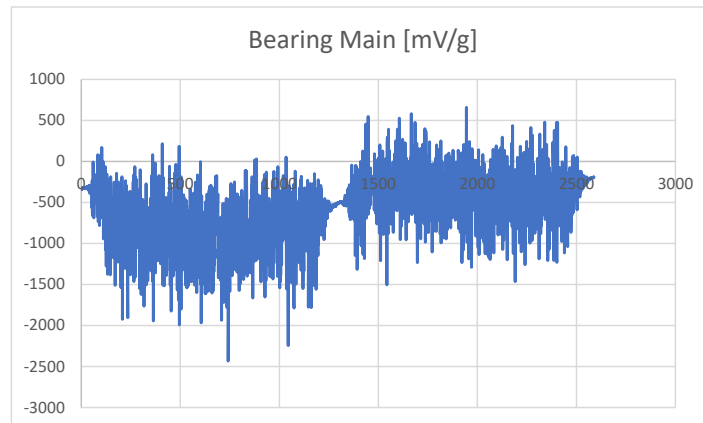


Fig. 19: Example of vibrations measured on the demonstrator at later operation stage (Raw_2022_09_10_18_59_01)

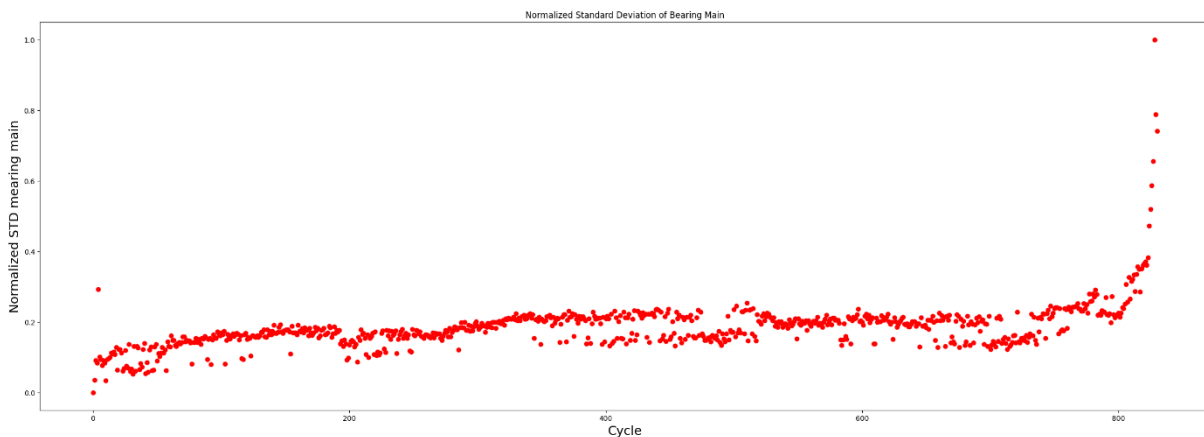


Fig. 20: Fluctuation of Bearing Main standard deviation of vibrations over the performed test cycles (normalized to $<0,1>$) (such as illustrated via Fig. 17- Fig. 19).

Mean Stiffness

Fig. 21 displays the variations of the mean stiffness of the calculated variable over the experiment cycles. The X-axis represents the number of experiment cycles, while the Y-axis depicts the normalized mean stiffness. Although this variable exhibits a similar trend in the change of its value during system failure, its fluctuations over the "healthy" cycles introduce an unacceptable level of noise that may lead the classification model to identify false-positive signals. Both variables were normalized to ensure that their values lie within the range of 0 to 1.

Initially, the main health indicator was believed to be reliable. However, it became evident that any change in stiffness, even when it was very close to the end of the experiment, was not detectable. Consequently, it was reconsidered which feature to monitor to detect any early signs of failure. A search for alternative features led to the identification of good health indicators (HIs). However, since the system had not failed, it was not possible to train the Remaining Useful Life (RUL) predictor.

After the system failed, the demonstrator was disassembled, and deteriorated balls of the ball screw were found be the cause of the failure, as illustrated in Fig. 16.

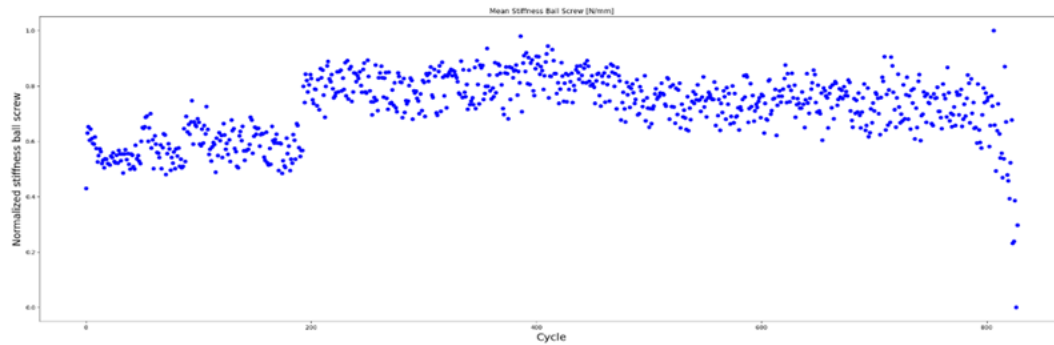


Fig. 21: The change of the calculated variable mean stiffness over the experiment cycles (normalized to $\langle 0,1 \rangle$)

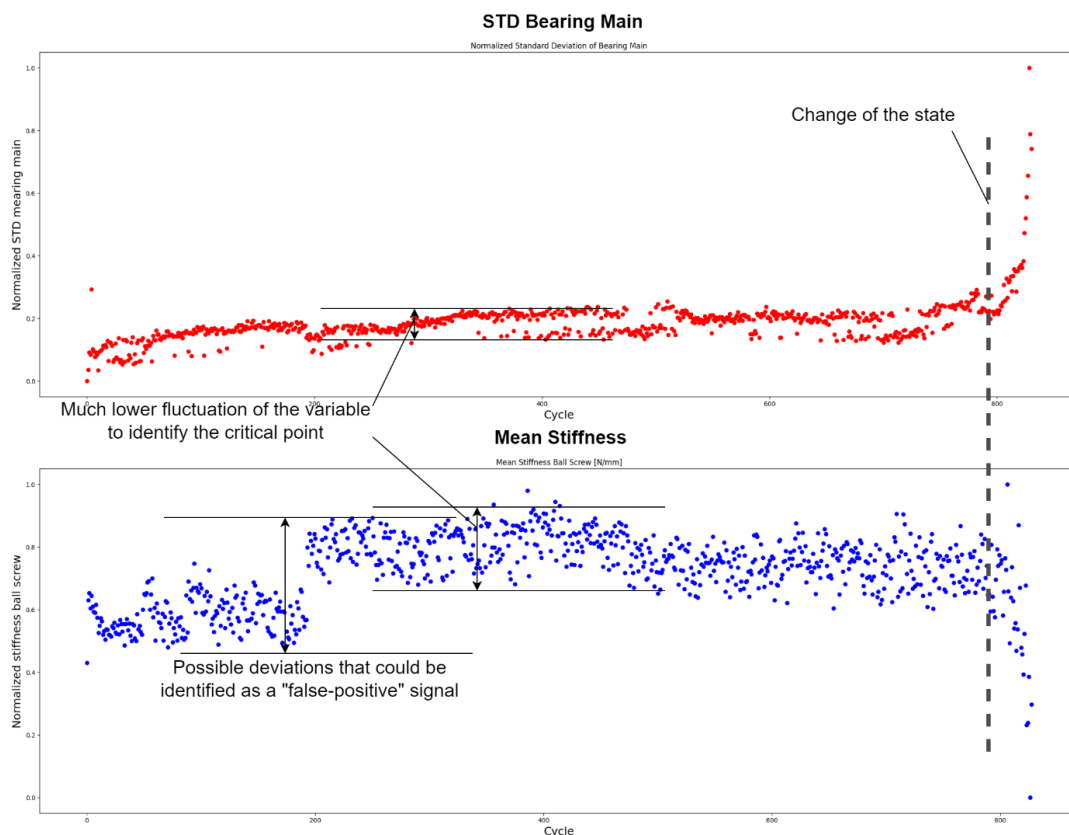


Fig. 22: The relationship of fluctuations and stiffness, the potential „false-positive“ detection has to be prevented by the expert (field engineer) input either before as the known existing phenomena in the process or by the immediate feedback from the field engineer that has the technological competency of the process.

3.7. Findings Relevant to the AI at the Business Intelligence Level of the Demonstrator

Unfortunately, due to limited data resources from the demonstrator, we were unable to create and validate a realistic cost model of PdM supported by AI. As a result, our prediction of remaining useful life (RUL) is inaccurate, and we are unable to create a robust profitable PdM scheduling strategy based on our developed concept of the cost model.

During the 1.5-year PredMain project, the main focus was on developing AI algorithms for signal processing, with the task of AI at the business intelligence level considered an additional objective.

However, it became apparent that business intelligence research was a significant undertaking that required additional time and capacity, and thus, it was decided that it deserved a separate project work package. This finding was consistent with a survey that revealed that small and medium-sized enterprises (SMEs) lack the capacity and tools to address this overarching aspect of predictive maintenance (PdM) adequately. Considering this aspect in further research is crucial, especially given that companies often find the cost of PdM hard to approach.

4. Analysis of Relevant Projects in the Area

4.1. Machine Learning for Anomalous Vibration Detection Under the Noise Level in Nonstationary Environment

In a collaborative research effort between the University of South Bohemia in Ceske Budejovice, Czech Technical University in Prague, and mySCADA Technologies s.r.o., it was determined that a crucial area of interest is the detection of anomalies and faults in non-stationary and extremely noisy environments. Non-stationarity can be introduced through varying setups of shaft rotations, while extreme mechanical noise can be caused by system inputs, such as the varying raw stone input into stone crushers. The granularity and other parameters of the raw stone change in time, and the measured vibrations are heavily affected by such noise. The varying shaft speeds and high level of nonstationary mechanical noise make it challenging for users to evaluate particular frequencies. Therefore, unsupervised anomaly detection is carried out first to help machine learning methods to classify and predict the remaining useful life (RUL) of the system.

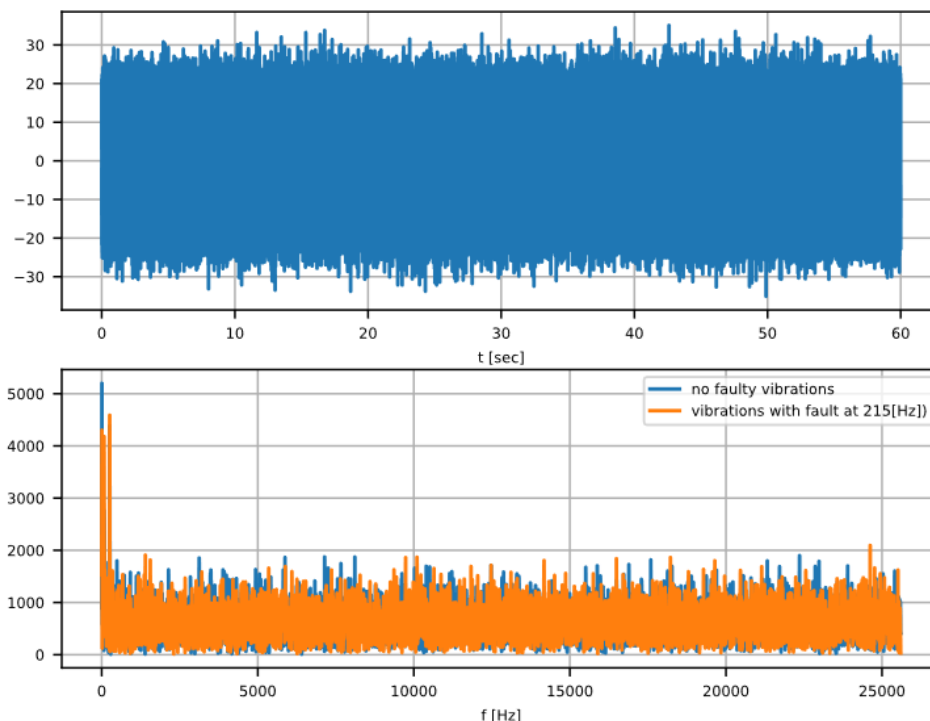


Fig. 23: (Top axes) A Gaussian noise where a new frequency appears for $t > 40$ [sec] and the magnitude of the new vibration is not seen (bottom axes) as it is under the noise level (see Fig. 24).

Previous research experience shows that multiscale techniques can be employed to detect unknown anomalous vibrations under noise level with nonstationarity. Applying the multiscale approach from [13] can detect and allocate newly appeared frequency under noise level, as demonstrated in Fig. 23 and

Fig. 24. Such detection can be a useful feature extraction technique for early bearing fault detection or damage detection in noisy environments.

The project PredMain partially supported the research of an original method for machine-learning-based anomaly detection [14]. The results showed that this machine-learning-based method has the potential for applications in nonstationary and noisy environments. The theoretical potentials of the method are demonstrated via Fig. 25- Fig. 27, while the real application of the machine learning method in another relevant project is discussed in subsection 4.2

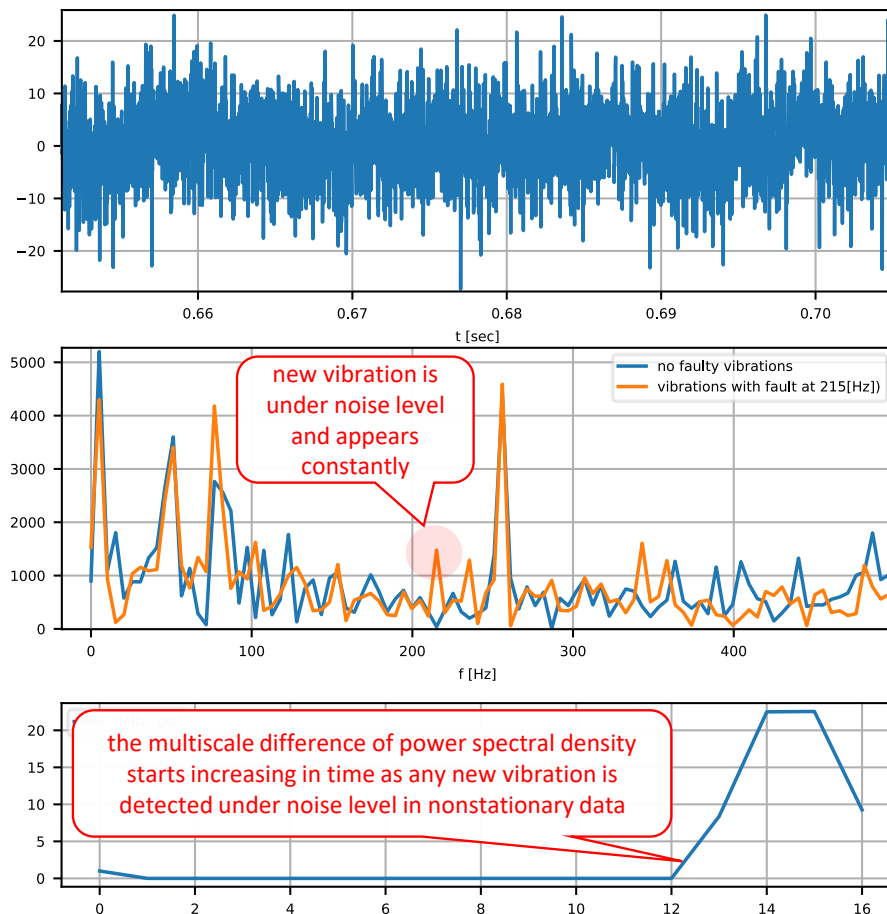


Fig. 24: For noisy signal (top axes), the new frequency at 215 [Hz] (mid axes, yellow spectrum) is under the noise level and appears in the last 2/3 of measured time; the windowing multiscale difference of spectra increases in time as it detects the newly appeared frequency under the noise level (bottom axes) that makes it a non-AI-based method for detection of anomalous persistent vibrations under the noise level in nonstationary environments (first real tests by mySCADA Technologies in 2022).

It can be mentioned that the non-AI-based feature as in Fig. 24 (bottom) can be further used as input to machine learning methods or future AI-based decision systems. Fig. 25-Fig. 27 shows a concept of purely machine-learning-based concept of a feature (as already discussed in the previous paragraph).

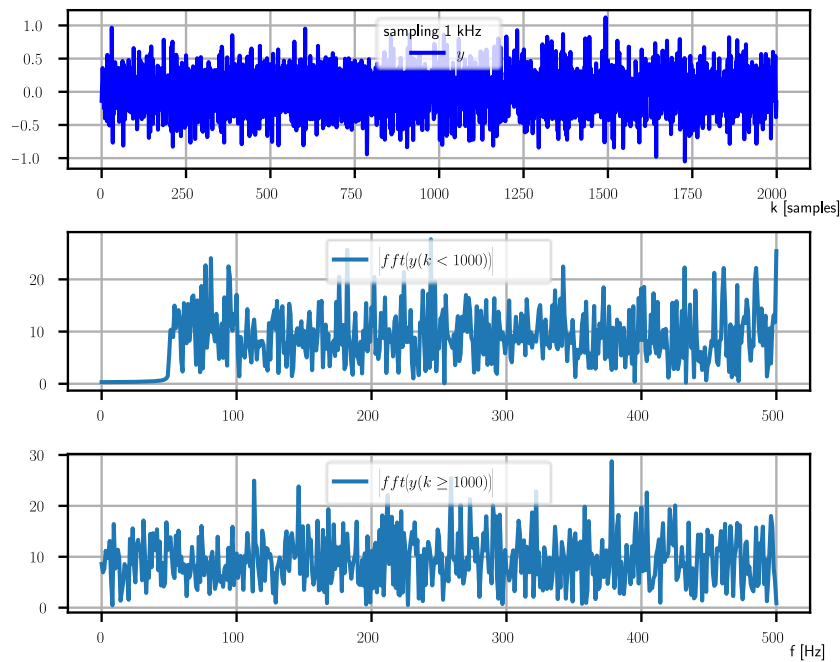


Fig. 25: A Gaussian noise as the example of heavily noisy vibration measurement, where the signal instantly changes the frequency spectra at $k=1000$ where real-time machine-learning detection technique [14] instantly detects the change of frequency spectra in real time (see Fig. 26 , Fig. 27)

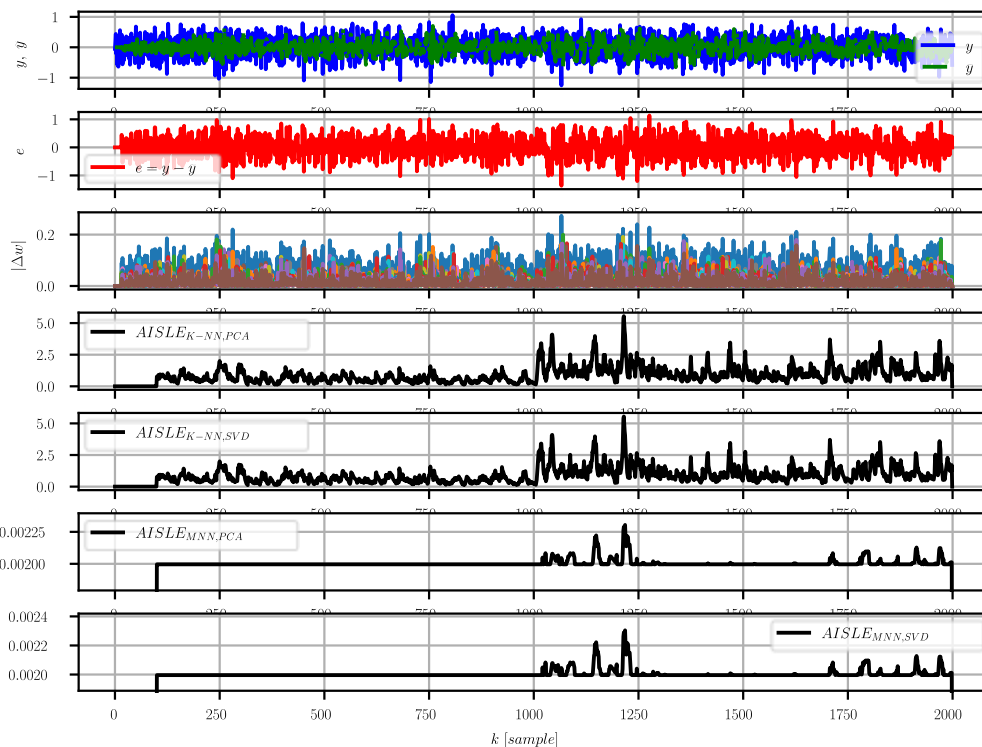


Fig. 26: The real-time learning neural predictor can not learn the noisy data (top axes) and the prediction error does not indicate any anomaly (second axes, see second axes in Fig. 27) ; then real-time learning neural weight are at 3rd axes from top, and the bottom four axes show the instant detection of change of frequency spectra in real time based on neural weights (see Fig. 27).

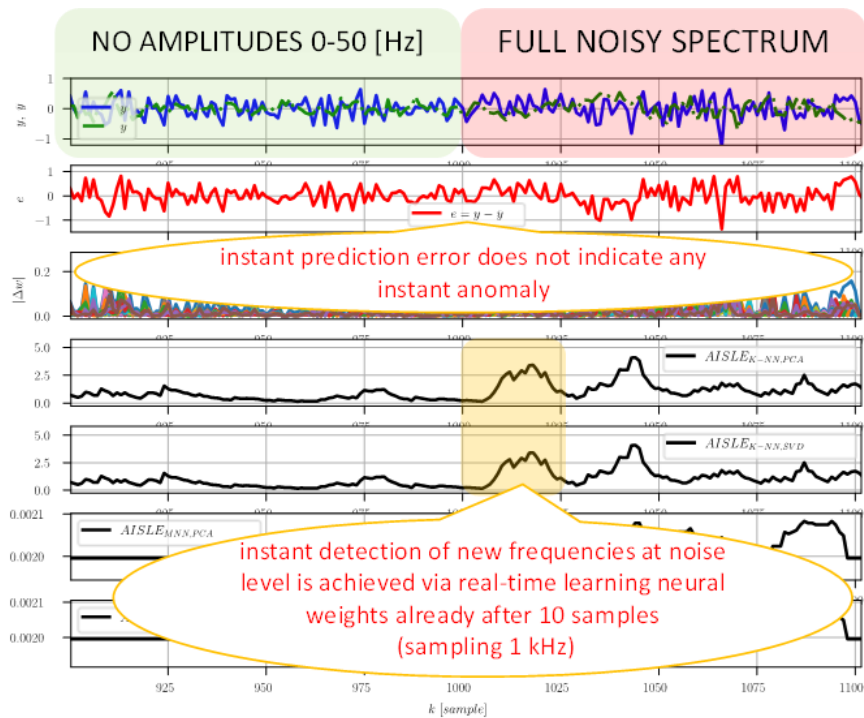


Fig. 27: Detail and explanation for Fig. 25 and Fig. 26.

4.2. PredMain with AI at the Signal Processing Level

The PdM workflow (Sect. 1.3) requests to generate labeled data in Step 1 if there is none to start out with. Our data consisted of raw feature time series gathered from the ball-screw demonstrator for several lifetime cycles under heavy load. While some of these raw features seemed promising to qualify as Health Indicators (HI) in Step 4 of the Workflow, it was necessary to conduct a systematic Features extraction search (Step 2 and Step 3) to find high-quality latent features on which the predictive analysis (training and testing of the learning model) could be carried out.

We will describe this feature extraction technique here. Before doing so, however, we mention some relevant previous publications that inform us about the nature of the HI features.

The characteristic cycle nature of our data, namely the back and forward motion of the demonstrator's tray carrying the accelerometer sensors at the bearings and the ball screw is symptomatic of the way health-indicating features are found, namely by aggregation across single cycles. A single aggregated value is attached to each cycle and monitored over long periods. Research for a similar situation is reported in [15] where the following result was seen in cycles over different instances within the full life cycle of the demonstrator (Fig. 16):

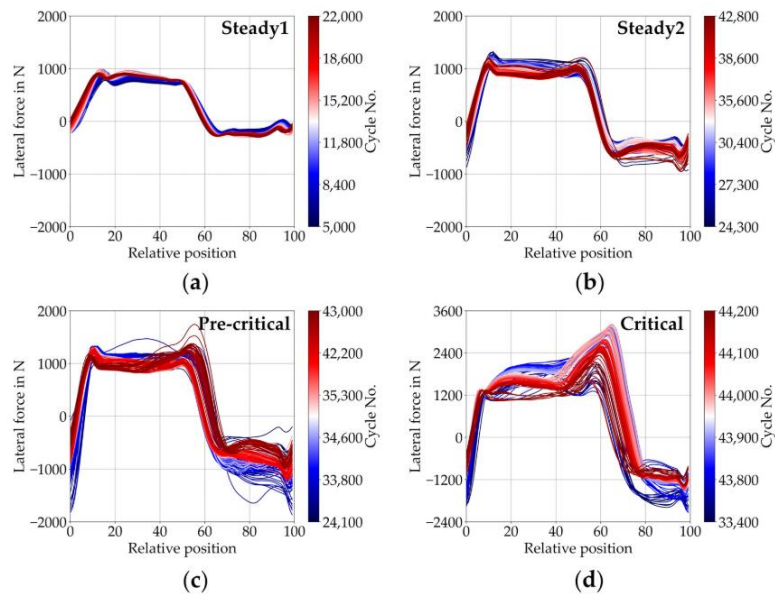


Fig. 28: Characteristic cycle shapes of the four operation states of oscillating tribometer setup, picture adopted from [15]. The increase of the lateral force at the beginning of the return movement is the most significant indicator of the approaching critical phase.

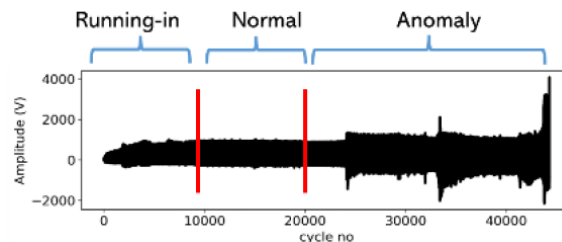


Fig. 29: Feature extraction [15], selection, and health indication in one model: An LSTM Auto-Encoder is identifying the health indicators in the latent space.

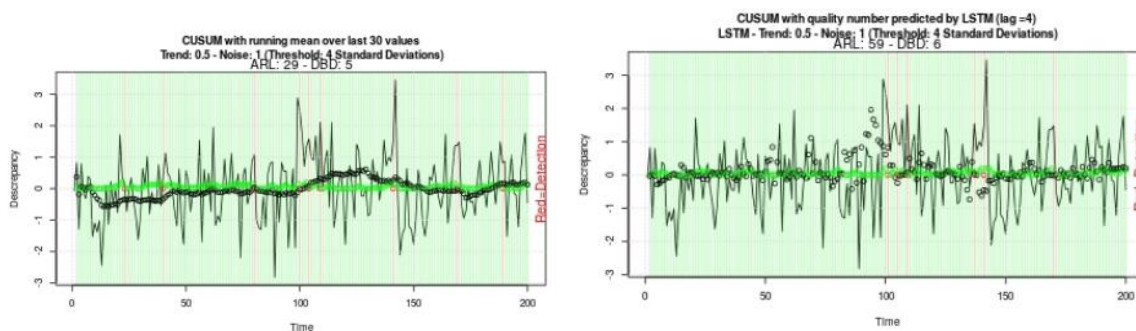


Fig. 30: From [16] it is shown that the well-known classical CUSUM rule for change point detection can be improved by replacing the target parameter, which usually is taken to be the running mean (left), by a predictive model (right: an LSTM): There is a distinct improvement concerning false positive detections. The ‘average run length’ (ARL) is the number of time steps in between two false positives. The ‘delay before detection’ (DBD) is the time before a real anomaly’s detection.

In [15], the per-cycle-aggregated mean amplitude of the transversal force is recorded over the full life cycle of a self-lubricated bearing, and the three stages of wear (run-in, normal, anomalous) are identified.

After generating the time series of aggregated values, it is possible to identify the three stages of different wear, the last of which defines the state which eventually needs to be predicted, i.e., which defines the RUL. In [16], it is shown how a recurrent neural network can predict trends that are considered ‘regular’ and which should not be flagged as anomalous, while detection of anomalies (appearing discrepant from the trend) are detected and labeled as such in the health state classification.

The used case in [16], an air pressure-powered grating filter of a powder separator in food grade packaging, clearly shows the benefits of a more precise predictive model in terms of the accuracy of the RUL of a single cycle before replacement of the filter must occur.

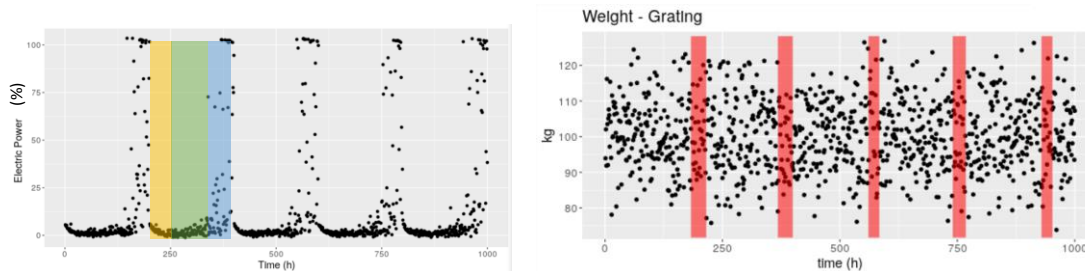


Fig. 31: Left - Original data of the powder filter used for detecting run-in (yellow), steady-state (green), and anomalous (blue) wear behavior. Right – corresponding assigned labels as a result of feature extraction for best health indicator.

- Classification into Regular/Anomalous vs. Regression of RUL

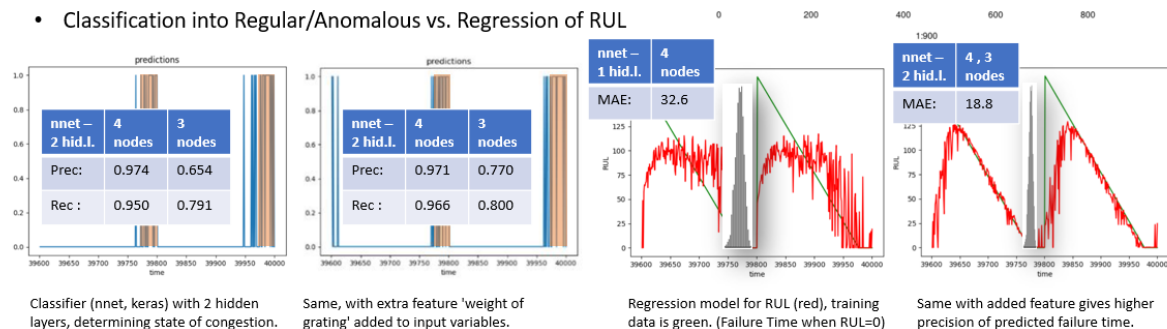


Fig. 32: Predictive classification models trained and used for powder filter: 2-hidden-layer feed-forward neural network with indicated numbers of nodes and corresponding results in terms of precision and recall (left). Regression models trained and used indicated the distribution of resulting predicted RULs with the error function chosen to be the mean absolute error (right).

Using these prior experiences and insights, we proceeded to label the raw data, by applying a systematic thresholding scheme applied to a library (‘tsfeatures’ [6]) of different time series features, leading to optimal health indicators (shown in Fig. 33). The characteristic form of a ‘good health indicator’ fulfills two purposes:

- Indicating the time of the final (anomalous) stage of wear as accurately as possible;
- Showing a characteristic feature of each moment in time before the onset of the anomalous stage.

These results, shown in 3.1 reveal (in the case of our data) the following characteristic shape of the best indicators:

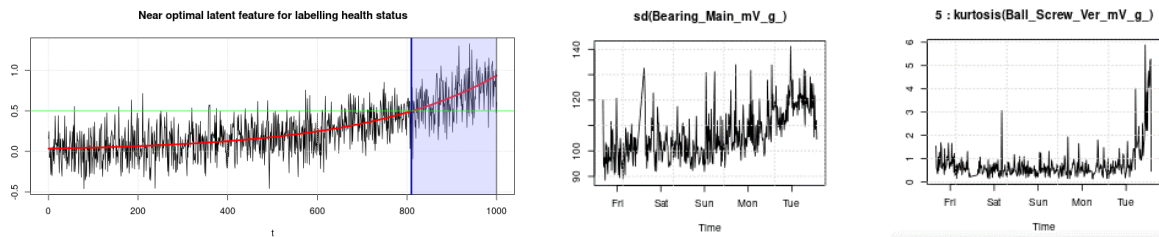


Fig. 33: Left - Near-optimal feature: red line defined by local regression and thresholding (here: 50% of the observed mean divergence at the end of the life cycle at $t=1000$) defining the anomalous stage (blue). The feature is well suited as an input for RUL prediction because it serves both purposes a.) and b.) - it grows slowly before the first anomalous value occurs, and progressively thereafter. Middle - Real example of near-optimal features (see also Fig. 17-Fig. 22 from the demonstration experiment). Right - Feature satisfying criterion a), but not b): the data are not suitable for training the RUL model because they are constant before the onset of the anomalous phase.

4.3. Machine Learning Based Anomaly Detection (Learning Entropy) in Long-Term Periodic Processes

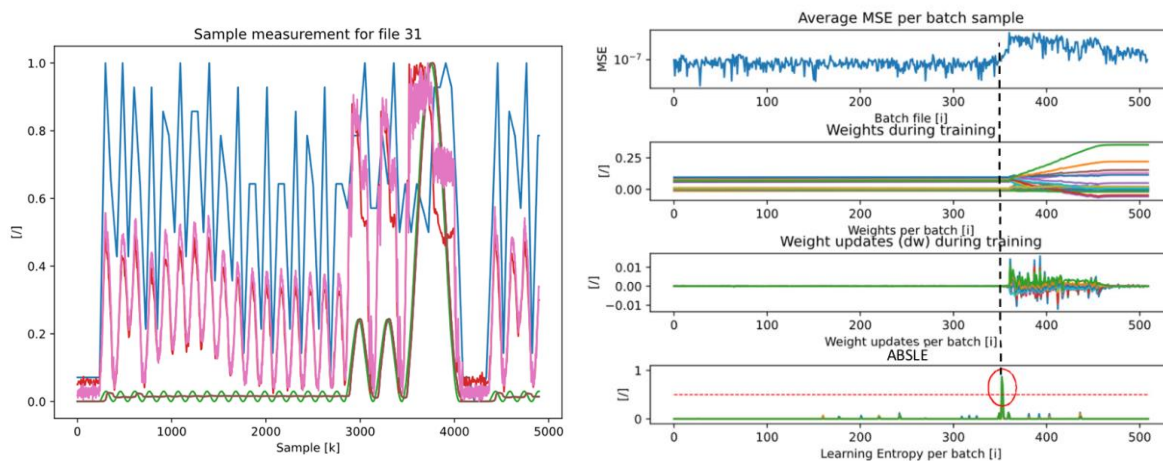


Fig. 34: Machine-learning-based anomaly detection (Learning Entropy) in long-term periodic testing of automotive components indicated that the long-term test will result in component failure [17], [18].

In project [17], [18], we applied unsupervised machine-learning-based anomaly detection via batch learning for the long-term test of automotive components. The method instantly detected the change in temporal correlations among measured variables via observing unusual behavior of neural weights during real-time windowed retraining of a shallow neural architecture.

The method required pretraining only on the data measured on the very own component in real-time itself and did not use any other data sources.

5. Synthesis for Small and Medium Enterprises

5.1. PredMain with AI at the Signal Processing Level

Predictive maintenance with AI can be highly beneficial for small and medium-sized enterprises (SMEs); however, to improve the reliability of the decision system, it is important to merge data analysis with expert knowledge. This can be achieved through a human-in-the-loop approach, where the user can

create their own labels instead of relying on external labels. While automating the labeling process can be difficult, incorporating the operator's expert knowledge about monitored sensor values and their significance for the state of wear can enhance the effectiveness of the proposed features.

Implementing one's own instrumentation for data acquisition can be feasible, but it depends on the specific case. Experiments conducted by INTEMAC have shown that it is beneficial to experiment on one's own to gather experience, but basic mistakes such as changing conditions during experiments can mess up the results. Low-cost solutions for data acquisition are possible, but collecting enough data with faults that are eventually to be predicted can be challenging.

In terms of data collection, it is crucial to collect enough data with faults, as well as enough regular data. One key insight is to include the operator's expert knowledge comprehensively, as this can significantly enhance the effectiveness of the proposed features. To reduce the number of requests addressed to the user in deciding how certain features are to be used in the predictive learning model, a rule set developed by a project partner can be incorporated into the user interface. This information can then be handed over to the predictive maintenance module, which automates the labeling process. The type of AI implemented will determine the resulting knowledge and consequences, and SMEs should carefully consider the most appropriate AI for their needs.

Using a simple stochastic cost-model (see description in subsection 5.2, Fig. 35, Fig. 36), different maintenance strategies can be evaluated in terms of their cost-effectiveness. The most important examples of such strategies are the reactive, the preventive, and the predictive maintenance strategy.

Well achievable and positive factors:

The cost model described in 5.2 focuses on determining the rate at which – during a long-time use of the same maintenance strategy – a production process including the downtime during the maintenance cycle renders profit. The positive result observed (see Fig. 37, right) states that as long as there is a possibility to reduce either the idle time, or the repair time, or cost of repair, there is a theoretical bound on the minimal accuracy of the predictive model which – if achieved – always leads to a higher profit rate.

This benefit can be achieved if by the aid of the predictive model for the RUL, the time of breakdown is forecasted with the degree of precision necessary to make the production time large in relation to the downtime. As a lack of accuracy of the chosen algorithm links directly to a decrease of the production interval, the quality of the predictive model is thus seen to be an imminent factor in the achievement of the intended rise in the production rate.

Limiting and negative factors:

A negative result can be read off from Fig. 37, on the left: If alongside with a change from reactive maintenance to predictive maintenance no further change of the influencing parameters repair-cost, and length of downtime is undertaken, then the resulting profit rate is always less than that of reactive maintenance.

This implies that a successful strategy employing predictive maintenance versus reactive maintenance always makes positive use of the fact that the maintenance occurs before the breakdown. This can be achieved either by reducing repair costs or idle time by scheduling the maintenance, in advance.

5.2. PredMain with AI at the Decision and Business Intelligence Level

An important finding is that very often the research for PdM is disconnected from any monetary considerations, which may be the biggest hurdle for SMEs to adopt PdM. This is because SMEs do not know the costs of PdM implementation and are uncertain about the return on investment (ROI). We

propose a cost model addressing the question of the principal usefulness of PdM versus other maintenance strategies, such as reactive, or preventive maintenance.

- There is production time, idle time, repair time.
- The downtime is the sum of idle- and repair time.
- There is a gain-rate, and a cost-rate (for repair).

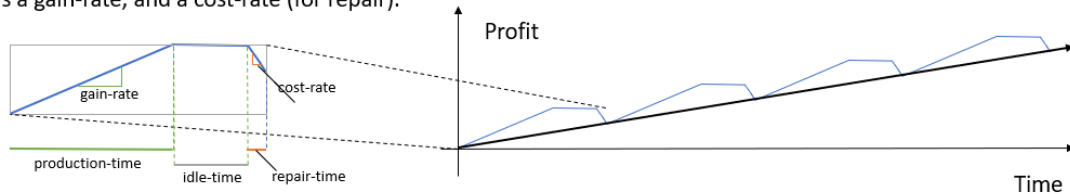


Fig. 35: Cost-Model based on the five variables of production-, idle-, and repair-time, as well as production gain-rate, and repair-cost-rate: This implies that the model specifically accounts for the variability in the cycle length, the idle and repair cycles, as well as the individual cost and gain rates. It is of interest to understand the resulting expected overall gain rate (asymptotic slope of the black line). Standard arguments using the ergodic theorem for stationary processes yield conditions for the usefulness of the Pdm if compared with either reactive maintenance (use until breakdown) or preventive maintenance (fixed maintenance schedule).

There is a lack of commonly accepted natural principles concerning predictive maintenance assisted by AI. The main question is if it will increase the profit estimates of its users (SMEs). AI predictions should help SMEs understand and analyze the removal of uncertainty in costs and revenues, and it can significantly impact estimating expected profit outcomes and assessing the risk of loss.

If AI forecasting methods reduce uncertainties, which can be understood as probability distributions of, e.g., total downtime, idle time, repair time, downtime costs, revenue per unit production time, and remaining useful life (RUL), the expected gain of PdM over other maintenance strategies and uncertainty in estimating loss risks or profit opportunities can be reduced, as demonstrated in Fig. 32-Fig. 36.

As seen in Fig. 32, the idea of a cost-model (submitted to the industrial statistics conference ENBIS2023) is to describe the individual maintenance cycles with five random variables (production-gain-rate, repair-cost-rate, production-time, idle-time, and repair-time) and assume stationarity for the repeated appliance of the maintenance. Under these assumptions, the corresponding identified long-term slope of the average profit rate (considering all five factors) is obtained by standard renewal theoretic arguments (Fig. 33).

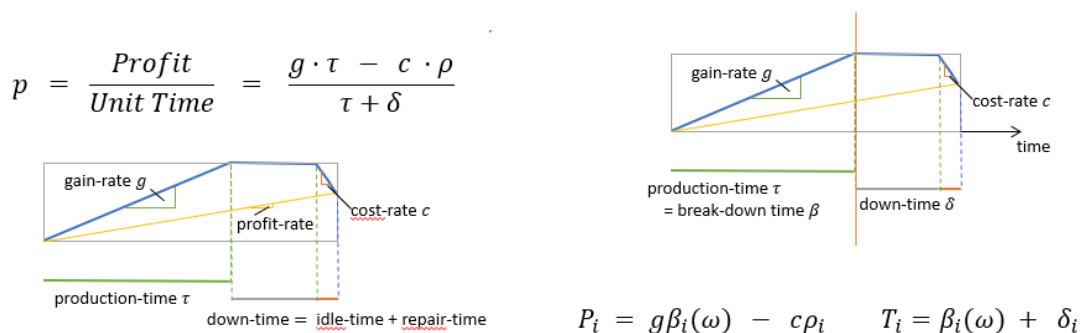


Fig. 36: Left – Formula for profit rate of a single cycle. All five variables (g, τ, c, ρ, δ) can but don't have to be random (indicated by the dependence on ω). Right – shows the case of reactive maintenance, in which the production time equals the break-down time. The fact that production remains uninterrupted for the longest possible period is the main asset of reactive maintenance. Disadvantages, such as lack of preparedness for the breakdown (leading to long idle times), and higher repair costs due to damage to other machine parts often make this strategy less profitable.

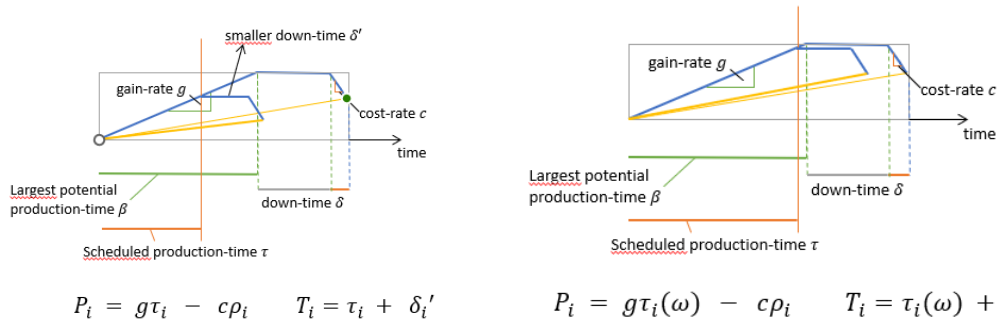


Fig. 37: Left – Preventive Maintenance characterized by an early enough scheduling to guarantee maintenance before break-down. The disadvantage of this strategy, overcome by PdM, is not being able to select the maintenance time dependent on the case, i.e., on the workpiece in the specific individual cycle. Right – PdM applied such that maintenance occurs just before breakdown, therefore maximizing the uninterrupted production interval.

Finally, the analysis of the expected profit rate alone is not enough to sufficiently inform about the certainty with which an observed rate is close to the expected value. A variance analysis (Fig. 38 Fig. 39) reveals this dispersive quality and therefore belongs to the complete analysis of the PdM cost reduction.

revenues per time unit of remaining useful lifetime by machine learning (Artificial/Business Intelligence, AI/BI) tools
remaining useful life time (machine learning, signal processing, AI, smart instrumentation)
cost per time unit of downtime by machine learning (AI/BI) tools
downtime time by machine learning (AI/BI) tools

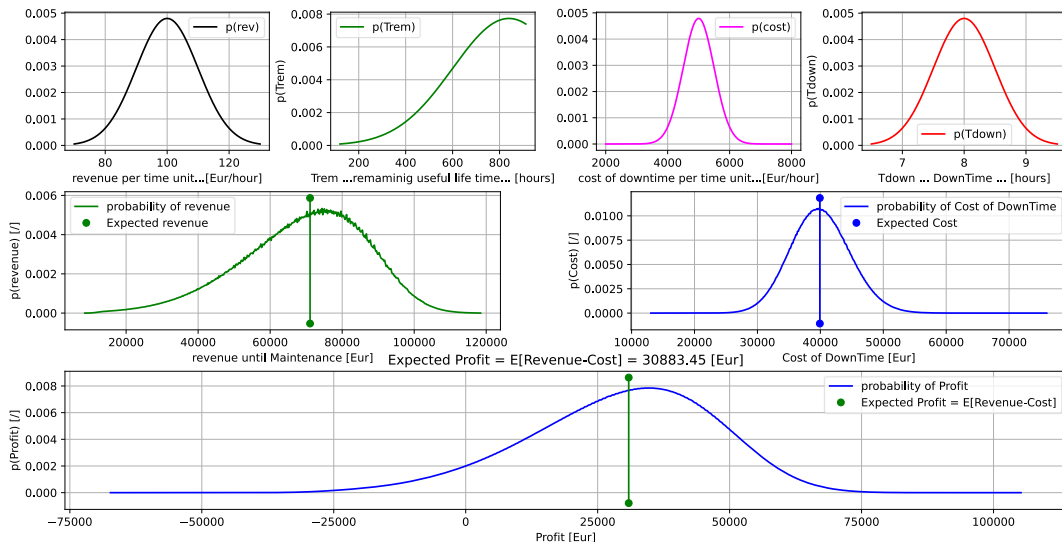


Fig. 38: Illustration of probability distributions as uncertainties in revenue per unit time, remaining useful life, cost of downtime, and duration of downtime (upper diagrams): It can be seen (lower plot) that the probability of profit prediction indicates a profit of more than 25,000; however, the probability of loss is not zero (see Fig. 39 for how lower uncertainties, e.g., via AI predictions, may affect the profit estimate and the risk of loss).

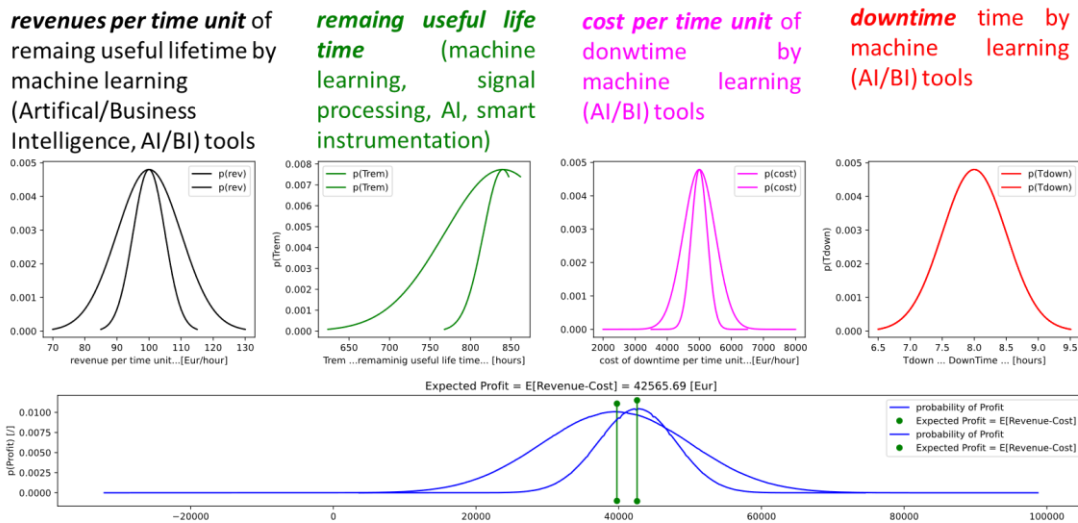


Fig. 39: The different possible predictions of the probability distributions, i.e., uncertainties, of the production and maintenance parameters (upper graphs) lead to a different width of the profit probability distribution, i.e., to a different estimate of the uncertainty profit or loss (lower graph). This determines the risk of achieving the intended profit rate. The main contributor to the dispersive nature of the distribution is the accuracy of the predictive learning model, e.g., as shown in Fig. 32 on the right. We identify this connection as the main link between the quality of the AI solutions and the monetary risk for SMEs investing in PdM.

6. Synthesis of Survey Results, Real Demonstrator Achievements, and Research Findings and Potentials

The survey covered the type of machines and equipment maintenance, self-assessment of maturity, barriers to PdM adoption, type of data collected, functions involved in the implementation of PdM, condition detection technologies, and IT infrastructure used for PdM in SMEs. The results show that the lack of knowledge, technical resources, relevant data, and financial assessment were some of the barriers to PdM adoption. SMEs with a low maturity level have a low adoption rate for electrical condition measurement compared to SMEs with a high maturity level. In terms of IT infrastructure, medium and high maturity SMEs have more specialized IT infrastructure than their low maturity counterparts, with Microsoft Excel being the leading data science platform. The survey also shows that PdM initiatives are led by Production and Maintenance, and companies may be using external specialists with deep expertise in predictive analytics.

Predictive maintenance, an approach to an equipment maintenance that uses machine learning and AI techniques, can be used to optimize ball screw maintenance. By monitoring key parameters such as motor current, temperature and vibration, predictive maintenance algorithms can assess the condition of the ball screw and predict potential problems so that maintenance can be scheduled at the optimal time. This approach not only increases maintenance efficiency, but also reduces the risk of downtime, resulting in cost savings and improved performance. In addition, predictive maintenance of ball screws can reduce maintenance costs by ensuring that maintenance is performed only when needed, rather than based on a fixed maintenance schedule.

In developing algorithms for predictive maintenance of ball screws, a number of questions and issues need to be addressed, such as how to determine the remaining useful life (RUL) and which variables and quantities are most useful for predictive maintenance. Machine learning techniques can be used to analyze substantial amounts of sensor data and predict potential failures, helping to optimize maintenance schedules and minimize costs.

The design and manufacturing processes of ball screws also play an important role in their functionality and longevity. Different ball screws have different load capacities, clearances and accuracy levels, which affect their cost and performance. Predictive maintenance algorithms can be customized to the specific characteristics of a given ball screw design, improving the accuracy of predictions and ensuring that maintenance is optimized for the application in question.

Overall, predictive maintenance using machine learning and AI techniques is a powerful tool for optimizing the maintenance of ball screws and other industrial machinery. By leveraging sensor data and machine learning algorithms, manufacturers can reduce maintenance costs, minimize downtime, and improve performance, leading to increased efficiency and profitability.

The PredMain project focused on developing AI algorithms for signal processing to support PdM in industrial systems. While limited data resources prevented the creation of a realistic cost model for AI-supported PdM, the project successfully identified promising health indicators (HI) through a systematic feature extraction technique. These HIs were found to accurately indicate the onset of abnormal wear and allow predictive analysis to identify potential failures before they occur.

The results of the project also demonstrate the importance of addressing the business intelligence aspect of PdM, especially for SMEs that lack the necessary capacity and tools. Most importantly, the time and cost savings achieved by sufficiently accurate predictive modeling need to be evaluated and compared to the investment costs for data and knowledge acquisition. Despite this challenge, the successful identification of HIs demonstrates the potential of AI to support more accurate and robust predictive models, leading to improved maintenance scheduling and reduced costs associated with system failures.

References

- [1] “ChatGPT: Optimizing Language Models for Dialogue,” *OpenAI*, Nov. 30, 2022. <https://openai.com/blog/chatgpt/>
- [2] B. Dickson, “Don’t expect large language models like the next GPT to be democratized,” *TNW / Deep-Tech*, May 21, 2022. <https://thenextweb.com/news/dont-expect-large-language-models-like-the-next-gpt-to-be-democratized>
- [3] N. A. Smuha, “The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence,” *Computer Law Review International*, vol. 20, no. 4, pp. 97–106, Aug. 2019, doi: 10.9785/cri-2019-200402.
- [4] “Automation Pyramid [DIDO Wiki].” https://www.omgwiki.org/dido/doku.php?id=dido:public:ra:xapend:xapend.a_glossary:a:autopyramid
- [5] M. Khan, A. Ahmad, F. Sobieczky, M. Pichler, B. A. Moser, and I. Bukovsky, “A Systematic Mapping Study of Predictive Maintenance in SMEs,” *IEEE Access*, pp. 1–1, 2022, doi: 10.1109/ACCESS.2022.3200694.
- [6] Y. Kang, R. J. Hyndman, and K. Smith-Miles, “Visualising forecasting algorithm performance using time series instance spaces,” *International Journal of Forecasting*, vol. 33, no. 2, pp. 345–358, Apr. 2017, doi: 10.1016/j.ijforecast.2016.09.004.
- [7] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring,” *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, Jan. 2019, doi: 10.1016/j.ymsp.2018.05.050.
- [8] A.-I. Neicu, A.-C. Radu, G. Zaman, I. Stoica, and F. Răpan, “Cloud Computing Usage in SMEs. An Empirical Study Based on SMEs Employees Perceptions,” *Sustainability*, vol. 12, no. 12, p. 4960, Jun. 2020, doi: 10.3390/su12124960.
- [9] G. Sullivan, R. Pugh, A. Melendez, and W. Hunt, “Chapter 5: Types of Maintenance Programs,” *Operations & Maintenance Best Practices—A Guide to Achieving Operational Efficiency (Release 3)*, pp. 1–9, 2010.

-
- [10]F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *2008 Eighth IEEE International Conference on Data Mining*, Dec. 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.
- [11]W. Zellinger *et al.*, “Beyond federated learning: On confidentiality-critical machine learning applications in industry,” *Procedia Computer Science*, vol. 180, pp. 734–743, 2021, doi: 10.1016/j.procs.2021.01.296.
- [12]P. Xu, X. Zhu, and D. A. Clifton, “Multimodal Learning with Transformers: A Survey,” 2022, doi: 10.48550/ARXIV.2206.06488.
- [13]I. Bukovsky, W. Kinsner, V. Maly, and K. Krehlik, “Multiscale Analysis of False Neighbors for state space reconstruction of complicated systems,” in *Merging Fields of Computational Intelligence and Sensor Technology (CompSens)*, 2011 IEEE Workshop On, Apr. 2011, pp. 65–72. doi: 10.1109/MFCIST.2011.5949517.
- [14]I. Bukovsky and O. Budik, “Learning Entropy: On Shannon vs. Machine-Learning-Based Information in Time Series,” in *Database and Expert Systems Applications - DEXA 2022 Workshops*, vol. 1633, G. Kotsis, A. M. Tjoa, I. Khalil, B. Moser, A. Taudes, A. Mashkooor, J. Sametinger, J. Martinez-Gil, F. Sobieczky, L. Fischer, R. Ramler, M. Khan, and G. Czech, Eds. Cham: Springer International Publishing, 2022, pp. 402–415. doi: 10.1007/978-3-031-14343-4_38.
- [15]Josef Prost and *et al.*, “Semi-Supervised Classification of the State of Operation in Self-Lubricating Journal Bearings Using a Random Forest Classifier,” vol. 9, no. 50, 2021, doi: doi.org/10.3390/lubricants9050050.
- [16]S. Mahmoud, J. Martinez-Gil, P. Praher, B. Freudenthaler, and A. Girkingner, “Deep Learning Rule for Efficient Changepoint Detection in the Presence of Non-Linear Trends,” in *Database and Expert Systems Applications - DEXA 2021 Workshops*, vol. 1479, G. Kotsis, A. M. Tjoa, I. Khalil, B. Moser, A. Mashkooor, J. Sametinger, A. Fensel, J. Martinez-Gil, L. Fischer, G. Czech, F. Sobieczky, and S. Khan, Eds. Cham: Springer International Publishing, 2021, pp. 184–191. doi: 10.1007/978-3-030-87101-7_18.
- [17]Ondrej Budik, Daniel Malachov, and Ivo Bukovský, “Batch algorithm of Learning Entropy in SW for Test data Analysis,” in *28th Workshop of Applied Mechanics 2020*, Prague, Czech Republic, Nov. 2020.
- [18]“National Competence Centre of Mechatronics and Smart Technologies for Mechanical Engineering.” <https://starfos.tacr.cz/en/project/TN01000071> , 2019-2022.

Appendix

AI-Generated Code (the Example)

„can you generate python code for the simulation of ball-bearing vibration time series with noise and anomaly detection?“

⇒ *„Sure! Here's an example Python code that generates a time series of ball-bearing vibrations with added noise, and then applies anomaly detection using the Isolation Forest algorithm from the Scikit-learn library.“* (ChatGPT)

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.ensemble import IsolationForest

# Generate ball-bearing vibration time series
time = np.arange(0, 10, 0.01) # time in seconds # the only needed MANUAL FIX of DATA LENGTH
freq = 10 # vibration frequency in Hz
amp = 0.2 # vibration amplitude
vibration = amp * np.sin(2 * np.pi * freq * time)

# Add noise to the time series
noise = np.random.normal(0, 0.05, len(time))
vibration += noise

# Add anomalies to the time series
anomalies = [500, 1200, 2000, 2800, 3500, 4200, 5000, 5700, 6500, 7200]
anomaly_values = [-0.5, -0.7, 0.9, 1.2, -0.8, -1.0, 1.1, 0.7, -0.6, -0.9]
for i, anomaly in enumerate(anomalies):
    vibration[anomaly] += anomaly_values[i]

# Plot the time series with anomalies
plt.plot(time, vibration)
plt.title("Ball-bearing Vibration Time Series with Anomalies")
plt.xlabel("Time (s)")
plt.ylabel("Vibration Amplitude")
plt.show()

# Perform anomaly detection using Isolation Forest
vibration_df = pd.DataFrame(vibration, columns=["vibration"])
isolation_forest = IsolationForest(n_estimators=100, contamination=0.05, random_state=42)
isolation_forest.fit(vibration_df)
vibration_df["anomaly_score"] = isolation_forest.decision_function(vibration_df)

# Plot the anomaly scores
plt.plot(vibration_df["anomaly_score"])
plt.title("Anomaly Scores")
plt.xlabel("Time (s)")
```

```
plt.ylabel("Anomaly Score")
plt.show()

# Identify the anomalies using a threshold
threshold = -0.4
vibration_df["anomaly"] = vibration_df["anomaly_score"] < threshold
anomaly_indices = vibration_df[vibration_df["anomaly"] == True].index.tolist()

# Plot the time series with detected anomalies
plt.plot(time, vibration)
plt.scatter(time[anomaly_indices], vibration[anomaly_indices],c="r")
plt.title("Ball-bearing Vibration Time Series with Detected Anomalies")
plt.xlabel("Time (s)")
plt.ylabel("Vibration Amplitude")
plt.show()
```