POLITECHNIKA LUBELSKA

MONOGRAFIE

*Dariusz Mazurkiewicz*

# Smart maintenance with time series modelling and digital twin

Lublin 2021

# Smart maintenance with time series modelling and digital twin

# Monografie – Politechnika Lubelska

Politechnika Lubelska
Wydział Mechaniczny
ul. Nadbystrzycka 36
20-618 LUBLIN

Dariusz Mazurkiewicz

# Smart maintenance with time series modelling and digital twin

Wydawnictwo
Politechniki Lubelskiej

Lublin 2021

Recenzenci:
prof. dr hab. inż. Zbigniew Banaszak, Politechnika Koszalińska
prof. Luis Andrade Ferreira, Porto University (Portugalia)

Skład i łamanie: Dariusz Mazurkiewicz
Tłumaczenie: Dariusz Mazurkiewicz
Korekta językowa: Magdalena Jung

Publikacja wydana za zgodą Rektora Politechniki Lubelskiej



ISBN: 978-83-7947-469-1

Wydawca: Wydawnictwo Politechniki Lubelskiej
www.biblioteka.pollub.pl/wydawnictwa
ul. Nadbystrzycka 36C, 20-618 Lublin
tel. (81) 538-46-59

Druk: DjaF – 30-092 Kraków, ul. Kmietowicza 1/1
www.djaf.pl


Nakład: 60 egz.

Table of Contents

# Introduction

Manufacturing companies are nowadays transforming from mass production to flexible mass customisation, therefore digital manufacturing based on the use of several intelligent systems has become a necessity for the industry. Digitalization means that massive amounts of data are collected by different production process components, with the multiplied amounts of data transferred between them in real time to be analysed. Correctly analysed data make it possible to extract knowledge or information for effective process management. There is no doubt that effective and intelligent manufacturing data source integration, connection, as well as-in particular-intelligent data processing and information exchange with automatically performed executive actions, are highly required by innovative factories. Therefore, many experts point out that future research directions in engineering will focus on the creation of intelligent sensors and their integration by means of digital systems and intelligent platforms. This has consequently led to a growing interest in the concept of digital twin (DT), its capabilities and potential applications, due to the fact that DT is known as a key enabler for digital transformation. All the above also applies to the area of reliability, thus fostering the creation of smart maintenance, i.e. a subset of the smart manufacturing system represented by self-learning and smart machines that predict failure, make diagnoses, and trigger maintenance actions, also with DT support.

Unfortunately, the implementation of the digital twin concept still poses several difficulties and challenges, including a lack of detailed methodology and standards, as well as problems related to processing large amounts of collected data, developing data acquisition systems, data modelling, as well as problems associated with performing executive actions related to design, management or simulation of changes in the system. Another challenge connected with the implementation of DT concerns the effective realisation of cyber-physical fusion. The above difficulties result from the fact that cyber-physical fusion requires the use of many data-related technologies such as data acquisition, data transmission, and data mining. The combination of digital twin and intelligent manufacturing will render all aspects of engineering activity smarter, human-independent, and hence more efficient. It will also have a positive effect on reliability, where smart maintenance is a key industrial need and challenge.

Condition-based and predictive maintenance are two strategies proposing that maintenance decisions be based on information and knowledge extracted from data collected through condition monitoring. Both of them consist of the same three elements: data acquisition, data processing, and, finally, maintenance decision-making for efficient maintenance action. Currently used solutions for predicting failure are based on predictions that do not take qualitative features or non-technical aspects into account. As a result, data-to-information and data-to-knowledge conversions – which are of vital importance for executive actions – are still fragmented, incomplete and not sufficiently effective. For this reason, new solutions supporting fast and effective prediction and decision-making must be produced. This, in turn, poses a great challenge not only for smart maintenance, prognostics, and health management, but also for time series data modelling on which they are based.

Prognostics and health management, together with cyber-physical systems, are considered as the key elements of the Industry 4.0 concept, beginning a new progress phase in manufacturing and maintenance through ICT-driven innovation. For this to become a reality, however, a multi-disciplinary approach must be developed. To that end, this monograph presents, in an extended and organized form, results of individual and team studies from recent years, combined with experience acquired from several visiting professorships and trainings abroad, providing such multi-disciplinary approach. The results cover multi-sensor data collected during the milling process performed using an industrial CNC machine tool, which were then analysed and modelled for smart maintenance implementation via advanced time series modelling with digital twin support.

This work is motivated by the fact that there exist vast amounts of manufacturing and/or maintenance data to find rules and knowledge for smart action, an undertaking that has become the major challenge of smart manufacturing and smart maintenance. The efficient processing of big data collected from physical space is one of the pillars for developing digital twin. Taking into consideration data collected from measurement systems as time series, as well as problems associated with the best possible data transformation model design, one can immediately acknowledge the importance of the main aspects explored in this monograph, namely the modern approach to engineering maintenance and its role in Industry 4.0.

The afore-identified need for DT development, combined with a lack of robust methods for process and machine management via effective virtual and physical convergence, gives rise to poor manufacturing performance with respect to intelligence and predictability. With the above in mind, this work proposes a new approach to data modelling, data-to-knowledge transfer, and smart operation for digital twins in the context of manufacturing. First of all, in Chapter 1 and Chapter 2, opportunities and challenges resulting from big data, time series data analysis, and predictive maintenance problems based thereon are presented. Considering the potential application of cyber-physical systems in reliability analysis, as well as the contemporary prerequisite for sustainability, smart maintenance challenges are discussed in Chapter 2. Smart maintenance of advanced technical systems usually requires the use of multiple sensor data acquisition and modelling, a problem that is covered in Chapter 3. The efficient processing of big data gathered from physical space is one of the pillars of DT development and application. This, in turn, requires selecting adequate models based on different data processing methods (discussed in Chapters 4.3.2-4.3.6), this selection constituting the core element of the proposed multi-criteria decision support model (Chapter 3.3). Big data problems and necessary data reduction methods must also be taken into account, and they are discussed in Chapters 2.2 and 4.3.1. The proposed approach meets several of the current research challenges and industrial needs that are reviewed in Chapter 1 and Chapter 2. Finally, Chapter 5 discusses the problem of a machine tool and its digital twin, based on an analysis of the state of the art in this area. Using the results of DT-oriented sensor data modelling presented in Chapter 4.3, a concept of smart maintenance with digital twin is presented, together with identified challenges for both current and future research.

# List of symbols and abbreviations

| | | |
|---|---|---|
| a | – | absolute component |
| b | – | gradient, drift |
| $c_1, c_2, \dots, c_s$ | – | possible classes for the feature $y$ |
| cp | – | complexity parameter |
| D | – | training dataset |
| i | – | number of samples |
| L,n | – | number of support vectors |
| m | – | number of measurements, number of lags |
| p | – | error probability |
| $p_{mi}$ | – | conditional probability |
| q | – | number of regions |
| R | – | set of real numbers |
| s | – | number of classes |
| T | – | random variable of time to failure |
| t | – | current age |
| $\tau$ | – | time constant |
| $x_0$ | – | absolute component |
| $\theta$ | – | real constant, drift |
| # | – | power of the dataset |

| | | |
|---|---|---|
| ADF test | – | Augmented Dickey-Fuller test |
| ARCH | – | Auto-Regressive Conditional Heteroskedasticity |
| ARIMA | – | Auto-Regressive and Integrated Moving Average |
| CAE | – | Computer-Aided Engineering |
| CBM | – | Condition-based Maintenance |
| CPS | – | Cyber-Physical System |
| CM | – | Corrective Maintenance |
| CMMS | – | Computerized Maintenance Management Systems |
| CNC | – | Computerized Numerical Control |
| CNCMT | – | CNC Machine Tool |
| DT | – | Digital Twin |

| | | |
|---|---|---|
| ES | – | Expert System |
| FHT | – | First Hitting Time |
| FMECA | – | Cause-Effect and Failure Mode and Effect Analysis |
| FN | – | False Negatives |
| FP | – | False Positives |
| GDP | – | Gross Domestic Product |
| GPS | – | Global Positioning System |
| ICT | – | Information and Communication Technology |
| IIoT | – | Industrial Internet of Things |
| IP | – | Internet Protocol |
| LRM | – | Linear Regression Model |
| M2M | – | Machine-to-Machine Communication |
| ML | – | Machine Learning |
| MLE | – | Maximum Likelihood Estimation |
| MT | – | Machine Tool |
| NC | – | Negative Case |
| OUP | – | Ornstein-Uhlenbeck Process |
| PdM | – | Predictive Maintenance |
| PDM | – | Product Data Management |
| PHM | – | Prognostics and Health Management |
| PLM | – | Product Lifecycle Management |
| PvM | – | Preventive Maintenance |
| RUL | – | Remaining Useful Life |
| RULE | – | Remaining Useful Life Estimate |
| SM | – | Smart Maintenance |
| SMDSS | – | Smart Maintenance Decision Support Systems |
| SVM | – | Support Vector Machine |
| TN | – | Ture Negatives |
| TP | – | True Positives |
| TPM | – | Total Productive Maintenance |
| WP | – | Wiener Process |

## Glossary

| | |
|---|---|
| Availability | It relates to the fact that something can be used. This term describes the ability of a technical object to be in a state to perform as and when it is required, under given conditions, assuming that the necessary external resources are provided. |
| Big Data | A data set which is large and complex but includes data which are so voluminous that traditional data processing is not able to manage them. These data are characterized by their volume, velocity, variety, and value. |
| Cyber-Physical System (CPS) | A category of an embedded system that can interact with real-world systems by means of computation, communication, and control. CPSs are capable of making decisions and operating independently. |
| Machine-to-Machine Communication (M2M) | A technology of direct communication between devices. It allows networked devices to exchange data without human interfacing or interaction. |
| Industrial Internet of Things (IIoT) | A concept related to the extension and use of the Internet of Things (IoT) in industry. By using sensors and executive techniques to enhance manufacturing and industrial processes, interconnected devices are able to exchange data and information without human interaction. |

| | |
|---|---|
| Industry 4.0 | A concept describing advanced automation of manufacturing due to the use of smart technology (CPS, IIoT etc.). |
| Smart Maintenance (also known as Maintenance 4.0) | A subset of the smart manufacturing system represented by self-learning and smart machines that predict failure, make diagnosis, and trigger maintenance actions. |
| Sustainable Maintenance | A concept describing the contribution of maintenance to more sustainable manufacturing operations in three aspects of sustainability: economic, environmental, and social. This may include environmental damage prevention, energy consumption reduction, plant and human safety, as well as the elimination of undesired failures, in order to minimize their economic, social, and environmental consequences. |

# 1. Classical and modern approaches to engineering maintenance

Maintenance is a fundamental engineering term. It is defined in the relevant ISO standard [177] as the combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function. The purpose of maintenance is to extend equipment lifetime, or at least the mean time to the next failure whose repair may be costly. It is also expected that effective maintenance policies can help servicing staff to reduce the frequency of service interruptions and their undesirable consequences for the production process. Maintenance requires action that can be performed with the use of inspection, monitoring, diagnosis, prognosis, repair or refurbishment.

Reliability is another fundamental term closely related to maintenance. It is defined [177] as the ability of an item to perform the required function under given conditions for a given time. As briefly put by Gavrilov et al. [45], the reliability theory is a general theory about system failure. It allows researchers to predict the age-related failure kinetics for a system of given architecture (reliability structure) and given reliability of its components. According to reliability theory, even systems that are entirely composed of non-aging elements (with a constant failure rate) will nevertheless deteriorate (fail more often) with age.

It can be claimed that reliability together with maintenance and its main measure, availability, are the key problems associated with the functioning of every company in terms of technological development, automation, and the pursuit of operational efficiency.

The theory of maintenance traditionally distinguishes four general maintenance strategies [26, 62, 177]:

- Corrective maintenance (CM), which is performed after fault recognition and is intended to restore an asset into a state in which it can perform the required function. CM is synonymous with reactive maintenance, also known as breakdown maintenance. CM is the earliest maintenance strategy, also called unplanned maintenance.

- Preventive maintenance (PvM), which is intended to assess and/or mitigate degradation and reduce failure probability. It describes a regular or routine maintenance action taken to keep an asset (technical object) running and, at the same time, to prevent any unplanned downtime from its unexpected failure. This type of maintenance is time-based, i.e. preventive maintenance actions are performed on periodic intervals, regardless of the health status of the asset.
- Condition-based maintenance (CBM), which is a type of preventive maintenance involving the assessment and analysis of physical conditions and the possible ensuing maintenance actions. It is based on continuous monitoring of the asset's actual condition to decide what kind of maintenance action is required. It aims to avoid any unnecessary maintenance tasks by performing maintenance actions only when there is evidence of any abnormal behaviour of the asset, which allows for maintenance to be performed only when it is required.
- Predictive maintenance (PdM), which is condition-based maintenance performed following a forecast (prediction) derived from repeated analysis or known characteristics and evaluation of significant parameters of the asset's degradation. It usually uses advanced data analysis tools and techniques to detect anomalies in operation or possible defects in the asset, in order to ensure that, under given conditions, the asset is in a state to perform as and when required (availability), assuming that the necessary external resources are provided.

CBM is a maintenance programme [2, 75] which recommends that maintenance actions be based on the information collected through condition monitoring process. In CBM, the lifetime of the equipment is monitored through its operating condition, which can be measured via monitoring parameters such as vibration, temperature, and noise levels.

The motivation behind CBM is that 99 percent of equipment failure is preceded by certain signs, conditions or indications that a failure is going to occur. Therefore, CBM is needed to ensure better equipment health management, lower life cycle cost, and catastrophic failure avoidance. In general, the primary goal of CBM is to perform a real-time assessment of equipment condition in order to make maintenance decisions and, consequently, reduce unnecessary maintenance and any other related costs [4, 35, 97, 98, 124].

Condition-based and predictive maintenance are two strategies that recommend taking maintenance decisions based on the information or knowledge extracted from data collected through condition monitoring. Both of them consist of the same three steps, namely [5, 62]: data acquisition (to obtain data relevant to the health of an asset), data processing (to handle and analyse data or signals collected for better understanding and interpretation of the data), and, finally, maintenance decision-making to propose an efficient maintenance action.

A more advanced and effective strategy is prognostics. This strategy deals with fault prediction before it may occur, which makes it possible to achieve zero-downtime performance. Prognostics is defined [100, 156] as the process of predicting the future reliability of a product by assessing the extent of deviation or degradation of the product from its expected normal operating conditions. The predicted time before potential failure is called as remaining useful life (RUL) [11, 140, 141]. Prognostics is usually and naturally extended to cover health management, thereby creating the concept of prognostics and health management (PHM) that combines prediction of future condition with effective recommendations on how to manage the health of an asset. In recent years, PHM has been considered as the leading maintenance technology. According to Lee and Bagheri [89], in future maintenance, PHM methods will have to be applied as the analytical core of a cyber-physical system (CPS). At the current development stage, PHM algorithms are applied to actual in-situ data from assets, but their integration with CPS will help leverage advantages from both parties. In the integrated case, historical life cycle information, which is available through CPS implementation, can significantly improve the performance of these analytical algorithms. PHM and CPS are together considered as the core elements of Industry 4.0 [31, 52, 145, 147], creating a new progress phase in manufacturing and maintenance through ICT-driven innovation, which requires a multi-disciplinary approach.

# 2. Manufacturing in Industry 4.0

## 2.1. Industry 4.0 and its development consequences

Manufacturing companies operating in today's competitive business environment are forced to meet the needs of their customers by offering them the highest quality products while at the same time ensuring the best sustainable results (economic, environmental and social). To achieve this, decision makers who have impact on the functioning of enterprises need to make special efforts to constantly monitor and improve production processes, despite the continuously increasing difficulty with production management. On the one hand, this situation stems from a number of factors that the decision makers should consider in the production management process, external and internal alike. On the other hand, the wealth of information that the decision-makers must take into account renders the decision-making process more difficult, too.

Manufacturing companies are now changing from mass production to flexible customized production, therefore digitalization with several intelligent systems included is an absolute necessity in today’s industry. There is no doubt that some kind of integration and connection of manufacturing data sources, as well as automatic data processing and information exchange with automatically performed executive actions, is highly required. Adequate solutions are offered by Industry 4.0, as a result of which manufacturing factories will [132, 144] become conscious and intelligent enough to predict and maintain the machines, control the production process, and manage the whole factory system. The term “Industry 4.0” refers to the fourth industrial revolution, which is defined [158] as a new level of organization and control over the entire value chain of the life cycle of products. It is geared towards increasingly individualized customer requirements.

When analysing the current manufacturing system development levels, and comparing them with the Industry 4.0 concept, several gaps, challenges, and opportunities can be identified. Some of them are directly related to maintenance and the demand for its being smart. According to [25, 36, 43, 55, 71, 114], smart maintenance (SM) is built on field intelligence that is provided by technology that is either embedded in a product/equipment or facilitated by the use of devices, sensors or any other technology-based tools. A smart maintenance tool should provide the most complete possible vision of the asset's health status, and thus prevent the need for any maintenance operator intervention except basic servicing.

## 2.2. Big data: opportunities and challenges

Manufacturing companies collect daily large amounts of data about technological processes, consuming resources, failures and their identification, as well as repair work and maintenance actions. This creates the problem connected with the concept of big data which is popular in most disciplines, including, naturally, technical sciences. In effect, there exist many different definitions of big data. For example, the European Commission defines this term [39] as large amounts of different types of data produced from various types of sources such as people, machines or sensors. This data could be climate information, satellite imagery, digital pictures and videos, transition records or GPS signals. It may also involve personal data: that is, any information relating to an individual, and can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address.

As classified by Qi et al. [131], manufacturing data come from:

- Manufacturing resources, including equipment data, material and product data, as well as environmental data.
- Management data from manufacturing information systems and computer-aided systems.
- Internet data, including user data collected from e-commerce or public data.

In addition to that, the rapidly growing number of sensors and connected devices, as well as the increasing horizontal and vertical value chains, result in a large continuous data flow. As a consequence, the analysis and use of big data play a major role in Industry 4.0 applications. Moreover, a structured approach is the key enabler for the expected significant improvements in productivity and efficiency. Under the Industry 4.0 vision, data digitalisation and data processing are expected to bring major changes in manufacturing in general. Also, the spread of novel technologies will enable a stepwise increase in productivity of manufacturing companies. Nevertheless, even though companies invest in digital innovations, they usually must face different types of problems including in the research and development area.

Several barriers to successful digital transformation still exist:

- Information and communication technology (ICT) systems merely enable reporting. It is up to a user to evaluate the future behaviour of the system. Operators have to deal with a huge amount of data under uncertainty and the future values of the analysed signals can only be predicted by using their own skills and experience.
- Heterogeneous data streams cannot be well processed to realize the automated decision support due to the lack of powerful analytic capabilities. Sophisticated tools able to transfer data into information, extensive knowledge, and experience required pose a challenge to many companies.
- When information extracted from data is available, effective inference and executive actions are not implemented in real time. The development of predictive data analytics techniques in order to aggregate and process the sensor data to assist in maintenance operations or scheduling is, therefore, necessary.

One of the effects of utilizing big data is the high degree of complexity in the analysis of large and diverse data sets. A solution to this problem may be again the implementation of Industry 4.0 achievements such as the automation of data processing and expert system design, which, in turn, requires the use of effective analytical tools. The data collected with the use of maintenance or diagnostic measurement systems contains not only valuable information about manufacturing process that can be used for e.g. optimization purposes. It is also a source of valuable knowledge about e.g. the wear of production line key components, and, in turn, serves as a basis for determining the optimal time of repair or part replacement to prevent excess wear or catastrophic failure while controlling process cost or optimal performance. The transition from data to knowledge and knowledge-based executive actions without human action requires developing new analytical tools, due to several disadvantages of the known solutions; this will also solve a number of big data problems [9, 70, 116, 144]. Therefore, an increase in the number of scientific studies in this field can be observed recently. These studies focus on advanced processing of collected measurement data derived from diagnostic and monitoring systems of technological machines. On the one hand, this is an outcome of the rapid development of measurement and analytical techniques; on the other, it results from the growing importance of durability and reliability, as understood in broad terms.

Big data analytics is, therefore, considered as one of the Industry 4.0 pillars that could be used in various areas, including maintenance, particularly in fault prediction to reduce error probability or in big data driven predictive algorithms to reduce the harm before some damage may occur.

## 2.3. Time series data

As more automation and digitalization is used in manufacturing, the speed of response or executive action required in maintenance issues has inevitably become faster. As already mentioned, industrial enterprises store masses of maintenance data consisting of discrete events such as component failures, historical records of inspections and repairs, as well as online data streams from measurement devices. All of them being stored as time series data. According to [42], a time series is a collection of observations made chronologically. Time series data are characterized by large data size, high dimensionality, and the necessity for continuous updates. Moreover, due to their numerical and continuous nature, time series data must always be considered as a whole, and not as an individual numerical field problem.

Massive data in time series collected by industrial systems contain information about process performance, failure accidents or servicing actions and alarms during the production process. If properly analysed, the data can provide valuable information and knowledge. This means that when applying analytic data-based approaches, it is possible to find interpretive results for strategic decision-making, which, in turn, brings sustainable advantages such as maintenance cost reduction, machine fault reduction, increased remaining useful life of assets and their components, increased production, improved operator safety, overall profit, and many others [26, 32, 61, 70, 79–81].

Regarding time series representation, different data mining tasks are reported in the literature, and they can be roughly classified into four fields [42]: pattern discovery and clustering, classification, rule discovery, and summarization. Some of the studies concentrate on one of these fields, while others focus on more than one of the above processes. To analyse data and to draw some basic conclusions and generalizations about datasets, descriptive methods are employed. Descriptive statistics [155] are numbers that summarize the data with the purpose of describing what occurred in the sample. In contrast, inferential statistics are numbers that allow the investigator to determine whether there are differences between two or more samples and whether these differences are likely to be present in the population of interest. Descriptive statistics can also be used to compare samples from one study with another. A descriptive data analysis provides summaries and creates understanding of the content and relationships between variables. It may also reveal trends, clusters, and patterns. Statistical methods are generally applied to improve system reliability and performance both in design and in operation phases. Qualitative approaches such as cause-effect analysis and failure mode and effect analysis (FMECA) provide e.g. cause-consequence chains that connect failure indication with a chain of events, and link the emerging event with its expected consequences. The scenarios represent causal relationships rather than just co-occurrence that may appear in data-analytics. Risk analyses connected with the online monitoring of process parameters could help predict the emerging malfunctions or evaluate data reduction possibilities [128, 159, 170]. In terms of Industry 4.0 however, modelling requires the development of new and different capabilities, concepts, methods and algorithms, as well as models and tools. Prognostics and expert system tools are crucial challenges related to enabling predictive maintenance solutions. The existing solutions, such as advanced maintenance technique frameworks, or, alternatively, the widely applied maintenance techniques, such as experience-based methods and failure prediction methods based on the physics of failure, are inadequate to satisfy the requirements of practitioners in the field.

Maintenance data collected from different sensors and data sources come in the form of time series. As a result, in the last decades one could witness an explosion of interest in time series data modelling and analysis. There are two basic methods of maintenance time series analysis [78]. The first is related to the study of relationships between the series elements (correlation analysis). The target of this approach is the estimation of correlation functions using parametric methods. The other approach is related to the analysis of frequency characteristics of the tested series (spectral analysis of the time series). It uses a variety of spectral, asymptotic, and functional techniques. One can distinguish four basic factors in the time series: trend, seasonality, cyclicality, and randomness. Not all of these four factors are always involved in the time series. However, in all studies, the existence of random factors is assumed. Another important factor is the stochastic trend occurring in the time series as a result of the integration of previous disorders. Time series can be stationary or non-stationary.

Data processing is an important step of time series data analysis. A wide variety of models, algorithms, and tools for data analysis are reported in the literature, their aim being better understanding, interpretation, knowledge transfer of data. The selection of models, algorithms, and tools for data analysis primarily depends on the type of collected data. According to Jardine et al. [62], maintenance approaches to monitoring equipment condition for diagnostic and prognostic purposes can be grouped into: statistical approaches, artificial intelligence approaches, and model-based approaches. As a consequence, the widely used data analysis methods can be divided as follows:

- statistical methods, e.g. linear regression, multiple regression, variance analysis, contingency tables;
- methods based on artificial intelligence, machine learning and deep learning, e.g. classification trees, regression trees, reinforced trees, fuzzy trees, random forests, artificial neural networks, genetic algorithms, evolutionary algorithms, fuzzy sets, rough sets, association rules, support vector machines or Bayes classifiers.

Model-based approaches require advanced knowledge about the equipment to be monitored. On the other hand, statistical approaches require mathematical background. Still, smart approaches based on e.g. machine learning tools have been increasingly employed in PdM applications. The most effective for large data sets are usually the second group methods. For this reason, they are called modern or advanced methods for data analysis, owing to the fact that they support researchers in finding solutions for the most difficult tasks, ones that statistical methods cannot solve. In addition to that, they provide a powerful predictive tool for PdM applications; however, their efficiency and effectiveness depend on selecting the appropriate data processing technology .

According to Jardine et al. [62], advanced maintenance technologies have not been well implemented in industry because of either insufficient data due to an incorrect data collecting approach or no data collection and/or data storage at all, not to mention the lack of efficient communication between theory developers and practitioners in the area of reliability and maintenance. Another implementation barrier is a lack of efficient validation methods and difficulties with their utilisation due to frequent design modifications, together with frequent changes in business and management policies. All the above poses a great challenge for smart maintenance and time series data modelling. It is, therefore, no wonder that the literature (e.g. [154]) emphasizes that future research directions in production engineering will focus on the creation of intelligent sensors and their integration by means of digital systems and intelligent platforms. Hence, the key skill will be appropriate analysis and processing of signals recorded by these sensors, which may relate to the parameters of a manufacturing process. The objective of these activities will be, among others, to ensure reduced machine downtimes, improved product quality, and, generally, increased knowledge of manufacturing processes.

Moreover, many studies (e.g., [44, 53, 54]) stress that in order to cope with dynamic data growth, it is necessary to create methods and tools that not only collect data automatically, but that are also able to select the most relevant data and use appropriate analyses to extract knowledge therefrom. Manufacturing data sets usually include a vast number of input variables to be taken into account during a data mining process. Therefore, an important requirement for modern tools based on machine learning is primarily their ability to recognize significant data.

## 2.4. Predictive maintenance

Predictive maintenance requires different sources of data [29, 51]: process sensors, monitoring sensors, and test signals. The first category consists of maintenance methods which use data from the existing process sensors for measuring variables like temperature, pressure, vibration level, and flow. To give an example, the output of a pressure sensor in an operating plant can be used not only to indicate pressure. It can also be employed to verify the calibration and response time of the sensor itself and to identify process anomalies that may interfere with accurate measurements of the process parameters or disturb the plant's operation, safety or reliability. The second category of predictive maintenance methods uses data from sensors such as accelerometers and acoustic sensors.

At present, despite an easy access to several data sources, most decisions for maintenance and operation are made in a traditional manner, which is known as time-based preventive maintenance. To that end, suitable data analysis methods and techniques are employed, such as time series analysis and risk analysis. Recently, many studies have been conducted also in this area [8, 18, 37, 38, 46, 70, 79, 116, 159, 160]. Despite promising results, there still exist many factors which prevent effective implementation of expert systems for support maintenance services or system operators [20, 22, 70, 170], such as computerized maintenance management systems (CMMSs).

Production system elements, as well as the problems connected with their service life and reliability, pose significant research challenges, particularly with respect to predicting future states in order to enable inference and implementation of executive activities in terms of failure-preventing servicing. In addition, there arise challenges connected with sustainability aspects and requirements to be satisfied. Numerous studies have been conducted in this field. However, they are predominantly fragmentary analyses of specific areas. For instance, the authors of the work [128] used neural network modelling to predict surface roughness and tool flank wear over machining time for various cutting conditions in finish hard turning. The developed prediction system is capable of accurate surface roughness and tool wear prediction within the trained range. Although the best results were obtained in the tool wear prediction with chamfered tools, relatively fair results were achieved in the surface roughness prediction. In the study by Goyal et al. [49], the production process operation vibration analysis was proved to be an effective tool for determining the causes of inaccuracy in the manufacturing process and faulty components or in machine maintenance-related decisions. The non-contact measurement of vibration signal is considered essential for reliable structural health monitoring with respect to quality assurance, optimized product and service profitability, enhanced manufacturing productivity, and a reduced number of regular periodic inspections. Bearing this point of view in mind, the authors analysed the most popular signal processing techniques used for machine structural health monitoring, such as time series models, wavelet transform, and Hilbert–Huang transform, together with a combination of any two of them. The results led them to the conclusion that although there exist numerous studies investigating SHM or RUL in production engineering, they are usually limited to minor and academic problems.

In addition to the above, the monitoring of smart structures poses a big challenge in terms of fault or damage detection, due to a large amount of noisy data collected from many sensors on a periodic basis. According to Susto et al. [146], in prediction maintenance problems, regression-based formulations arise when predicting RUL of a process or equipment. On the other hand, classification-based maintenance formulations occur when seeking to discriminate between healthy and unhealthy conditions of the system being monitored. In effect, separate calculations are usually necessary for monitoring data (information used in preventive maintenance policies) and predicting RUL. The remaining useful life prediction is not only essential to verify whether the mission goals can be accomplished. It also provides a useful aid to online decision-making activities such as fault mitigation [136]. Calculations are usually based on the mean or median of maintenance cycle lengths, as well as on the definition of an optimal action threshold for distinguishing between faulty and non-faulty process iterations based on the observed data [16, 96, 120, 146–148]. While classification tools are a natural choice, they do not naturally map to the RUL factors that can be extrapolated from the process data for maintenance prediction and further decision making.

Merh [120] demonstrates the possibility of using data-driven models to predict the failure of machines used in continuous technological processes. The analysed example concerns the manufacturing of components for cars and tractors. The objective of the study was to predict the total downtime of machines involved in production. To achieve this goal, a multiple regression models, regression trees, and artificial neural networks were used. Furthermore, association rules were generated based on the manufacturing data set. The results have shown that the use of neural networks, regression trees, and association rules is adequate for that type of predictive task. Classification models, especially decision trees, random forests, and neural networks, were used by Accorsi et al. [1] for decision making in maintenance engineering. Based on the dataset derived from historical profiles of the energy variables of a high-speed packaging machine, they discovered strategies for the prediction of packaging machine failure. The experiment showed the best accuracy of the random forest compared to other methods. Despite the promising results of the study, the authors argue that there is a need for further experiments based on other data mining techniques.

As it was demonstrated in the state-of-the-art analysis above, a great focus is put on condition-based maintenance and, more recently, to its predictive option. Similar observations were made by Van Horenbeek et al. [57], who, additionally, indicated that although condition-based maintenance takes advantage of the known state of components, setting a degradation threshold beyond which preventive maintenance is carried out is not always an optimal solution compared to predictive maintenance, especially when considering interdependent multi-component systems. In addition to current degradation information, predictive maintenance also makes use of predictive information in the form of the RUL of components to optimally schedule maintenance actions, while condition-based maintenance only uses current component state information. Proactive maintenance decisions can be made based on the predictive information, which results in a dynamic maintenance schedule. Moreover, the predictive information makes it possible to include component interdependencies into the maintenance schedule. To do so, adequate time series modelling is necessary.

## 2.5. Cyber-physical systems in reliability analysis

Expectations related to the implementation of solutions based on the Industry 4.0 concept have led to the creation of machine-to-machine communication (M2M), industrial internet of things (IIoT), and smart technologies. Moreover, it may also become necessary to develop CPSs and their embedded system identification algorithms, taking into account the broad concept of sustainable development [66, 68, 94], which, consequently, poses a number of new research challenges. According to Lee at al. [90–92], recent advances in the manufacturing industry have paved the way for a systematic deployment of CPSs, within which the information from all related perspectives is closely monitored and synchronized between the physical factory floor and the cyber computational space. This requires advanced information analytics for networked machines to finally make them able to perform more efficiently and collaboratively. Although there exist numerous studies conducted in this area, they are mainly theoretical considerations in which new methods or mathematical models are usually verified only with the use of simulation data. Moreover, a review of recent studies reveals that most of them are devoted to the problem of CPS reliability, while hardly any tackles the reliability engineering based on CPSs.

Cyber-physical systems are defined [91, 102, 107, 130, 158, 168] as transformative technologies for managing interconnected systems between their physical assets and computational capabilities. With recent developments that have resulted in higher availability and cost-effectiveness of sensors, data acquisition systems, and computer networks, the competitive nature of today's industry forces more factories to move toward implementing high-tech methodologies. Consequently, the ever growing use of sensors and networked machines results in a continual generation of high volume data, which is known as big data. In such environment, CPSs can be further developed for managing big data and leveraging the interconnectivity of machines in order to reach the goal of creating intelligent, resilient, and self-adaptable machines. Furthermore, by integrating CPSs with production, logistics, and services in contemporary industrial practices, manufacturing companies could be transformed into Industry 4.0 factories with a significant economic potential [91].

The terms “CPS” and “reliability” are analysed bidirectionally. Reliability engineering for smart maintenance is based on CPSs. On the other hand [163], reliability is also considered as a critical requirement for cyber-physical systems.

When CPSs are taken as the key element of innovation under Industry 4.0, this presents more challenges for analysing failures and system reliability independently, related to, among others, the following aspects:

- The heterogeneity of interconnected elements. CPSs may include both physical and cyber components, but the conventional approaches to understanding physical dynamics do not model the behaviours of software and networks well, while cyber abstractions lack the considerations on temporal dynamics.
- The complexity of interfaces between elements. In a CPS, a computational element which receives data is not the one which controls the actuators, and networking and communications are existent in the cyber space.
- Unclear contributors to the system downtime. Since computation or information processing and physical processes are tightly integrated in some CPSs, it is difficult to identify whether the failures of a system are caused by computations, physical laws, or both working together.

In spite of the above, CPSs are not only the core element of Industry 4.0, but their implementation also creates smart maintenance, also known as Maintenance 4.0. As defined by Kans et al. [73], Maintenance

4.0 is a subset of the smart manufacturing system represented by self-learning and smart machines that predicts failure, makes diagnosis, and triggers maintenance actions. The smart equipment is in form of embedded or cyber-physical systems, i.e. it is the equipment where the physical and software components are intertwined. The CPS possesses computing resources for efficient data capture, processing, and communication in order to monitor and control the system.

Bokrantz et al. [14, 13, 108] distinguish four core dimensions of smart maintenance: data-driven decision making, human capital resource, internal integration, and external integration. Smart maintenance performance is, therefore, expected to have broader effects compared to the traditional view of maintenance.

## 2.6. Sustainability requirement

In recent years, one can witness the growing popularity of the concept of sustainability and of the idea of sustainable intelligent manufacturing in which it should be considered. One of its definitions [63] says that sustainability refers to the creation of a manufactured product with processes that have minimal negative impact on the environment, conserve energy and natural resources, are safe for employees and communities, and, finally, are economically sound. Accordingly, the primary objective of sustainable manufacturing is to create a balance between environmental, social, and economic dimensions. Stuchly et al. [65] observe that the creation of a sustainable production environment requires eliminating breakdowns and energy waste, and that the term “sustainability” is strongly related with sustainable maintenance. Sustainable maintenance can be defined as maintenance operations striving to maintain a balance between social, environmental, and financial dimensions. From a practical point of view, it requires changes in actions performed within the maintenance area. In general terms, this entails a combination of monitoring, control, and data processing capabilities to create smart technological machines that are able to learn, self-diagnose, and self-adapt. This kind of technological intelligence in maintenance can diminish the need for operator actions, as well as improve safety and reduce unnecessary costs.

He et al. [52] observe that, being the next generation manufacturing system, intelligent manufacturing ensures better quality, higher productivity, lower cost, and increased flexibility. This is why the concept of sustainable manufacturing is evolving due to a growing popularity of digital twin, and it contains sustainable manufacturing equipment, systems, and services that support each other. The idea of intelligent manufacturing has been transformed [52, 104] into digital twin-driven sustainable intelligent manufacturing, which has become a hot research trend owing to the fact that it produces good results in life cycle management, prediction, and sustainable manufacturing itself.

## 2.7. Smart maintenance challenges

The above state of the art analysis reveals that adequate and effective data modelling under Industry 4.0 requires the development of new and different capabilities and concepts. Novel methods, algorithms, models, and tools are highly needed, too. Prognostics and expert system tools, like predictive maintenance tools used as elements of CMMS, are the key challenges here. According to [62], the existing solutions, e.g. advanced maintenance technique frameworks or, alternatively, widely applied maintenance techniques such as experience-based methods and failure prediction methods based on the physics of failure, are inadequate to meet the requirements of practitioners in the field.

The currently used solutions for predicting failure are based on predictions that do not take account of qualitative features, sustainability or non-technical aspects. As a result, data-to-information and data-to-knowledge conversions–so important for executive actions–are still fragmented, incomplete, and far from being effective. Moreover, the existing prediction systems communicate the possibility of failure too late, with no support in deciding about a suggested moment of taking maintenance action based on the production schedule. Therefore, it is necessary to devise new solutions supporting fast and effective prediction and decision making.

On the one hand, to form a set of features describing the condition of an object, one must select an appropriate subset of these features providing important information about a device, system, plant, etc. On the other hand, the classifier of a system condition should be designed based on the data determined from the selected features. In addition, such classifier should specify the distance to the separation boundary. Remaining useful life can be predicted based on the time series of these distances. Hence, further basic studies on developing effective analytical methods and tools must be conducted (Fig. 1).

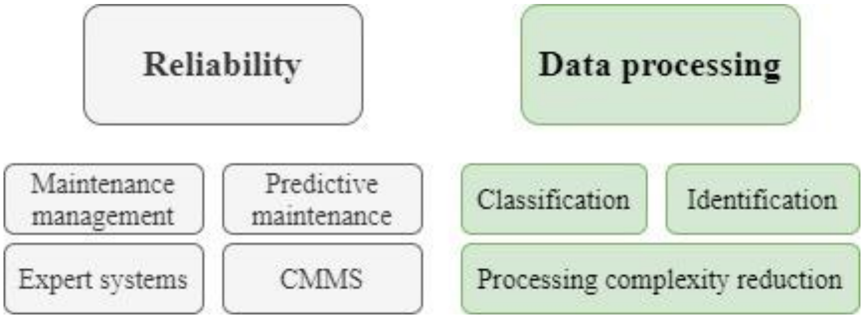


Fig. 1. R&D challenges in reliability and data processing

Their implementation at a later stage, e.g. as computerized maintenance management system (CMMS) modules, should be relatively easy on the operational level. This requires improved models and metrics for effective evaluation and optimization techniques on different levels, including product, process, and system. Consequently, one of the current research challenges is to develop new analytical methods for effective adoption of predictive tools, in order to refine data into business knowledge for a step-change in performance and cost-efficiency by novel capabilities. This, however, requires conducting basic research on the new solutions for technical infrastructure availability and reliability based on time series modelling.

PdM problems are generally tackled by means of machine learning (ML) techniques that are aimed at defining a health factor or, generally, PdM status of a process (e.g. the estimated number of runs before failure), which can then be employed either in an expert system (ES) or in a computerized maintenance management system (CMMS) to manage maintenance actions. As highlighted by Susto et al. [146], many PdM-related problems are still open from a modelling point of view, including high-dimensionality, data fragmentation, and the requirement for adaptive solutions and interpretable models. A new concept of the prediction and decision data analysis model will have to provide support for maintenance service activities. Such model should be first provided with values of the monitored parameters of a machine condition. Depending on their character, an intelligent system will, based on information criteria, automatically select a suitable model from a database with models or scenarios. The development of a system that enables the collection of different types of data, their aggregation, and selection of a suitable prediction model depending on the data type, as well as provides support for the decision-making process based on information about the production process, is an innovative approach, one that has not been yet reported in the literature.

Bumblauskas et al. [24] emphasize the significance of the above-mentioned aspects for Industry 4.0 applications in maintenance on account of the fact that smart maintenance, which is based on real-time transmission and analysis of data from across the factory, can have an extremely positive impact on all operational aspects. With the effective use of data collected through various sensors, it will be possible to predict the optimal time to perform maintenance maximizing uptime of the equipment while reducing costs. This shift from reactive and preventive maintenance to predictive maintenance is feasible only via the integration of large and diverse datasets by means of data-to-knowledge conversion methods and executive tools forming smart maintenance decision support systems (SMDSSs) operating in real-time.

# 3. Multiple sensor data modelling for smart maintenance

## 3.1. Sensors as a source of data

The demand for increased operational efficiency of technical infrastructure presents a challenge for enterprises, especially when the improvement in the quality of implemented activities is closely related to the reduction of various costs incurred by entrepreneurs. Companies strive for increased availability and maximum efficiency of the machine park while maintaining high quality of products. At the same time, however, they are looking for savings, usually expecting reduced expenses on renovation and repair. It is, therefore, no wonder that innovative production technologies and complex production processes pose stringent requirements for production equipment reliability. As a result, maintenance operations and activities become more and more significant from the operational point of view, but also because of economic reasons. Production equipment must be maintained in continuous operational efficiency, not to mention the fact that it is required that potential failures of production system components be predicted and full operational efficiency of the system be restored in the shortest time possible, without impairing the manufacturing process. These requirements combined with industrial needs form a basis of the prime objectives of maintenance activity, including [33, 77]:

- maintenance of specific quality of products/services,
- maximum prolongation of the working life of production equipment,
- assurance of conditions enabling safe operation of machines and devices,
- reduction of production costs to the minimum by decreasing the number of stoppages during production.

The application of adequate maintenance methods, tools, and techniques is crucial for efficient operation of any equipment or production process components. In addition to that, it exerts a considerable impact on the economy of the whole enterprise. Another significant aspect is the search for new more effective and innovative solutions for this particularly important field of production engineering. Consequently, novel yet viable maintenance solutions or strategies, as well as analytical support tools, should be produced to enhance operational efficiency. Obviously, effective maintenance actions require the use of adequate data and information sources that usually come in the form of monitoring or diagnostic systems. Monitoring is always a diagnostic tool. On the other hand, for diagnostics to be effective, it is essential that an additional element be implemented into the system– namely, knowledge extracted from data.

One of the afore-mentioned data and information sources could be condition monitoring, which is defined as [41] the process of determining the condition of machinery while in operation. This requires both defining machine parameters that provide information about the current condition of machines and evaluating contemporary data with prearranged nominal data for precise fault diagnostics. Condition-based maintenance utilizes measurements to appropriately schedule maintenance activities without interrupting normal machine operations. In this way, the basic requirements are met. Condition–based maintenance is a preventive maintenance approach supported by sensor measurements [105, 122]. Predictive maintenance should also be supported by sensor measurements; given the specifics and purpose of predictions, more sophisticated data processing methods are, however, required to that end. With sensor measurements and adequate data processing, not only is effective fault diagnosis feasible, but it is also possible to predict future states of an object or its remaining useful life.

In any complex technical system, a single sensor is incapable of collecting enough data for accurate fault diagnosis, prognosis or remaining useful life prediction, hence multiple sensors must be applied. With the rapid development of computer science and advanced sensor technology, there has been a growing tendency towards the use of multiple sensors for any modern maintenance strategy performance.

When multiple sensors are used, data collected from different sensors may contain different partial information about the same machine condition. As it was highlighted in [62], the problem is how to combine all partial information obtained from different sensors for more accurate machine diagnosis and prognosis. In addition to that, multiple sensor data means bigger and more complex databases to be analysed.

Furthermore, as discussed by Kene et al. [74], a novel approach to the problem should involve the combined use of some indirect sensors. This approach is generally referred to as sensor fusion or sensor data fusion. Data fusion consists of a fusion or combination of either the same family of entities or different families. In the sensor fusion approach, data from various sensors are combined by a common family or unified system of units representing the same environment for different sensor outputs. This approach has more advantages over the single-sensor use with respect to overcoming deficiencies, as well as providing accurate and precise results. The literature review [74] leads to the conclusion that the use of data fusion or sensor fusion has become indispensable in high precision manufacturing industries where product quality cannot be compromised over catastrophic tool failure and/or unexpected malfunctioning of any tool wear monitoring sensor.

All the above factors render data processing far more complicated, and, consequently, necessitate the use of more sophisticated and advanced analytical methods and models. In view of the multi-dimensionality and complex interactions of collected data, it is indispensable that multi-criteria decision analysis be implemented in predictive maintenance.

## 3.2. Limitations of prediction and decision support models

The immediate problem arising at the very beginning of prediction model design is how to reduce the dimension of an observation vector to a smaller subspace, so that it will not cause a loss of information necessary for correct definition of machinery condition [159]. Such reduction makes it possible to design simpler and more effective models for the above-described phenomena. Another problem connected with failure prediction pertains to random factors. The randomness of variables in formal models is one of the main difficulties in practical application of quantitative methods, hence it would seem justified that making a prediction based on randomness should also include random values from historic values of the observed signals [77].

In view of previous studies [76, 75, 77, 83, 159], it can be said that typical prediction models reported in the literature relate to predictions that are only based on technical factors. Such approach cannot be regarded as comprehensive, one that takes into consideration all factors affecting technical object reliability. Although the applied models use reliable data recorded automatically by real-time machine monitoring systems, they do not take non-technical aspects or qualitative features as factors affecting the reliability of a machine park. Non-technical aspects include, among other things, economic issues and seasonal character, while qualitative features are related to problems such as the operator's knowledge of a machine, reliability of performed repair work or audit results concerning the implementation of good practices in compliance with total productive maintenance (TPM). The idea behind the TPM concept [30] is to bring production and maintenance together by combining good working practices, teamworking, and continuous improvement. Another aspect affecting the accuracy and reliability of predictions is associated with object aging, where it is customary for a working time to have impact on failures and their frequency. To give an example [27], the understanding of aging mechanisms in electronic components is of crucial importance for the aerospace domain where these components make part of numerous critical subsystems, including avionics. To generate a prediction, appropriate input values of non-technical variables and qualitative features are necessary (Fig. 2).

Such prediction will provide information about the potential moment of machine failure. When scheduling the time of maintenance work, the decision process should include additional information about a production plan, a schedule of activities taken by maintenance services, as well as information about ordered spare parts [77]. Despite the difficulties related to the complexity of prediction and decision support model design (Fig. 2), it is additionally essential to adopt a multi-criteria decision approach supporting predictive maintenance.

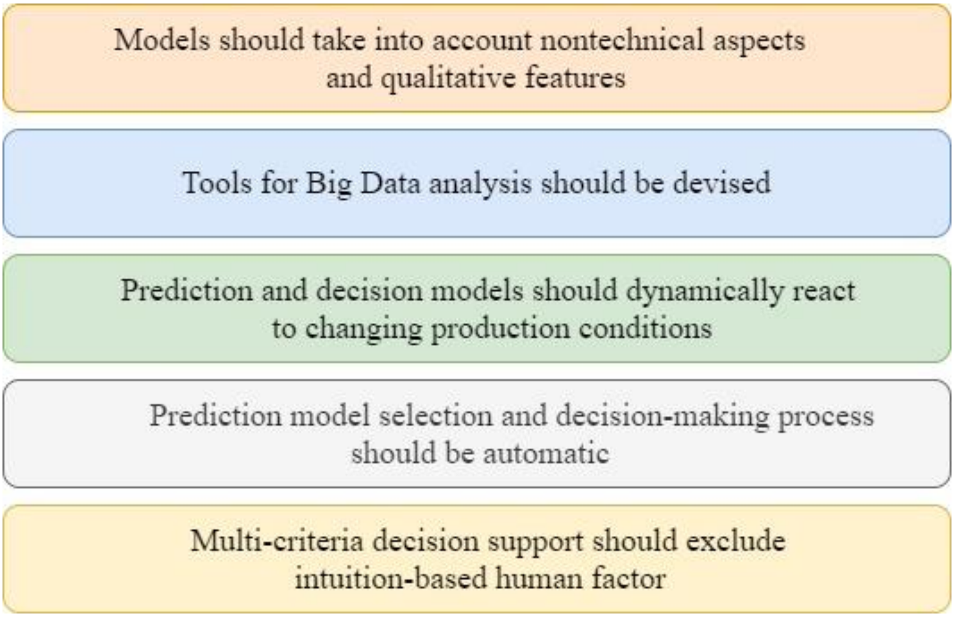


Fig. 2. Difficulties and limitations of prediction and decision support models

As a result, generated predictions will be more accurate, whereas decisions taken on their basis will undoubtedly bring real benefits for companies operating in the field of production engineering. The above advantages notwithstanding, the satisfaction of all these needs and requirements will render data processing and the model design process more complicated.

### 3.3. Multi-criteria decision support model

As previously mentioned, the currently used solutions for predicting failure are based on predictions that do not take into account qualitative features or non-technical aspects. Moreover, prediction systems communicate the possibility of failure after the expected time with no support in deciding about the suggested moment of taking maintenance or repair actions, relying merely on the production schedule and on the schedule of actions to be taken by maintenance services. For this reason, it is necessary to devise a solution supporting both prediction and decision. A schematic design of the proposed solution is shown in Fig. 3. This model [77] should be first provided with values of the machine condition parameters being monitored.

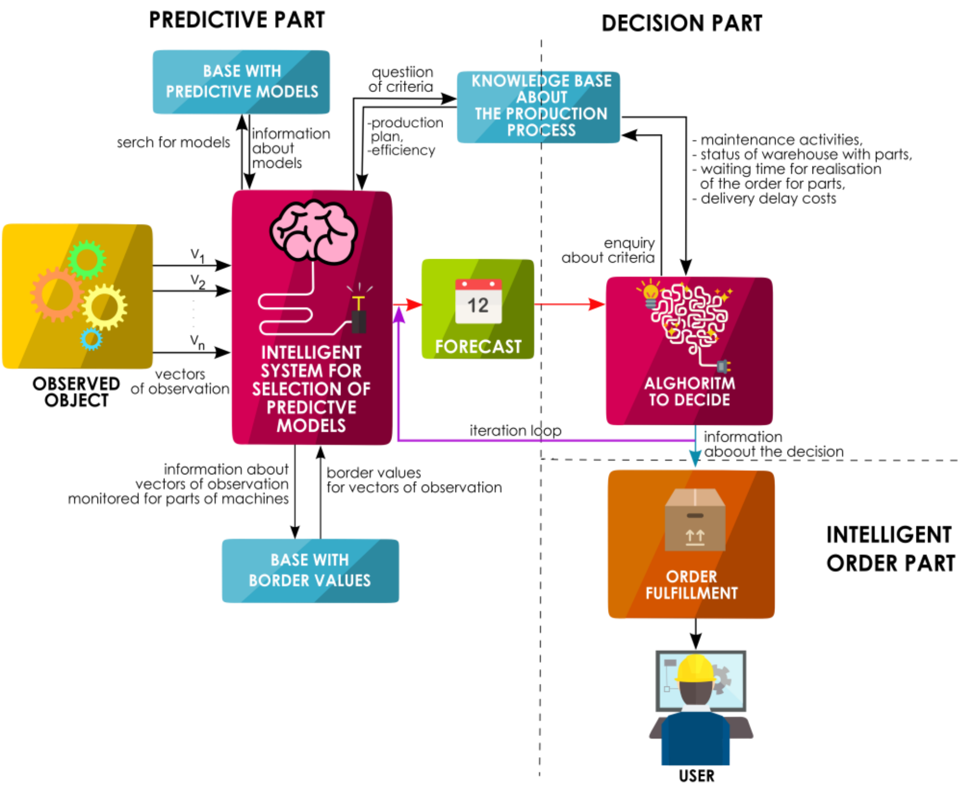


Fig. 3. Concept of multi-criteria decision support in maintenance of a machine park [77]

Depending on the character and valence of the data provided, an intelligent system will – based on information criteria – select a suitable data processing model from a database with models. This stage includes the selection of a specific method for searching patterns, which is typical of data mining. Having selected a model, the system will send a query to a database containing boundary values of specific monitored observation vectors for a machine or its parts. The return information, as well as the information from the production process about the production plan and production line efficiency, can be used to make a prediction, i.e. to determine the time after which the boundary value would be exceeded. The generated prediction should then be processed by a decision algorithm that – based on a query sent to the database about the production process regarding the schedule of maintenance and repair works – would indicate an optimal moment of repair or servicing, thus extending the time of failure-free operation [77].

Importantly, this model would use iterative techniques enabling prediction updates and generation of different status messages depending on the period of time preceding an undesired event. Obviously, the developed model for generating predictions and recommendations for preventive actions should include qualitative features and non-technical aspects, which means that it is necessary to devise standards describing these actions, as well as to determine their effect on the failure frequency of a machine park or its elements.

First of all, a module to select a mathematical model for observation vector data processing should be designed. The model selection will depend on the results of a deterministic part of the analysis; on the fact whether we are dealing with a sequence of homoscedastic or heteroscedastic variables; as well as on the results of information and/or prediction criteria [76–78].

Next, an algorithm of a decision support process based on current prediction results should be designed. Here, it is important to rank iteration-caused predictions. The model could still communicate failure risk either as a piece of information leading to no further action or as an alert that cannot be ignored, otherwise this would immediately cause a production line stoppage due to failure. The developed prediction model should also be verified on two different levels. On the first level, it should be verified based on numerical data, which would enable the consideration of potential improvements and the ex-post determination of prediction error. The second level of verification should involve testing solutions. This requires a two-way flow of data and their real-time analysis. The best solution for bidirectional data and information flow is digital twin (DT).

Based on a set of values relating to a given phenomenon (factor) and using proper information criteria, the intelligent system selects the appropriate mathematical model. At the same time, the prediction is done by models designed for time series analysis. Stationary process phenomena can be identified with the use of mathematical models for stationary series (i.e. autoregressive moving average models) and for non-stationary series (i.e. seasonal ARIMA, autoregressive conditional heteroscedasticity, generalized ARCH, etc.).

## 3.4. Overview of the state of the art

Predictive maintenance has nowadays become a new trend in prognostics and health management (PHM) for complex equipment. Predictive maintenance is invariably based on data, regardless of the implementation or execution method. Data collected from different sources and monitoring systems, having a different form or content, require the use of adequate processing methods. The more accurate the data are and the more effective the processing method is employed, the better prediction results might be achieved. Unfortunately, there exist a number of several identified limitations of this approach, such as non-technical or sustainable aspects which have to be taken into consideration when designing decision support models. In addition to that, the implementation of the proposed multi-criteria decision support concept requires an advanced knowledge base and a set of predictive models that must be compared for their effectiveness when used to analyse similar data.

It is, therefore, no wonder that over the past few years, studies were conducted to design smart maintenance solution elements using time series modelling and digital twin. To that end, a CNC machine tool (MT) was considered as a research object for maintenance data acquisition and modelling via machine learning and DT. Due to the fact that the MT is a complex mechanical, electrical, and hydraulic integrated system consisting of numerous different parts and components that are interdependent and interacting with each other, different kinds of faults may occur in different parts or subsystems. At the same time, the performance of MT components and parts undergoes constant degradation, which is a time-varying process. This will certainly give rise to the occurrence of unpredictable faults if troubleshooting is not timely. Therefore, predictive maintenance is indispensable to avoid any faults and to improve the reliability of this type of manufacturing machine. Simulations performed with the use of digital twins make it possible to obtain more accurate data about the condition of any part of the system from the equipment model without too many sensors installed, which - in turn - allows for more accurate and reliable smart predictive maintenance.

The analysis presented below is based on the results of a series of partial research activities. More recent of these results are systematized and discussed in a subsequent part of this monograph. They are based, among others, on:

- Mathematical modelling of real diagnostic data with the use of selected stochastic processes (Wiener and Ornstein-Uhlenbeck type processes) [159].
- Analysis and development of a method for forecasting the optimal moment of cutting tool replacement based on its wear degree identification [84].
- Development and verification of a mathematical model for effective real-time classification of cutting tool status [85].

- Development of a new classifier for condition assessment and remaining useful life prediction as an expert system tool for real-time monitoring of the manufacturing process [83].
- Maintenance strategies for system performance estimation, as well as cumulative maintenance cost and system lifecycle assessment [174].
- Development of a model for machine criticality assessment [67].
- Empirical study on the impact of maintenance function on more sustainable manufacturing processes [64].
- Investigation of the principal component analysis and logistic regression application for cutter condition identification [82].
- Analysis of decision trees as a decision support tool for machining process data analysis [6].

# 4. Machining process and smart maintenance

## 4.1. Machine tool reliability

Machining as a manufacturing technology has occupied a crucial role in manufacturing processes at many enterprises. This has not changed in modern industrial realities. It is estimated [8] that the expenditure on machining amounts to approximately 5% of the GDP in the developed countries. Therefore, the machining technology is constantly evolving. Numerous studies are conducted, investigating, among others, the accuracy of machined parts or the stability of high-speed machining processes [85, 83].

The machine tool is a standard manufacturing instrument. It is also regarded [111, 112] as the mother of all machines in the manufacturing industry. Due to their construction and equipment, modern machine tools can be included in the high-technology category. Machine tools are numerically controlled, they are equipped with sophisticated measurement tools or intelligent sensors. Computerized numerical control (CNC) machine tools also play a fundamental role in the development of smart manufacturing and all related new technological trends. Typical challenges for smart manufacturing, such as self-sensing, self-assessment, self-adjustment, self-maintenance and self-prediction, are also relevant in relation to CNC machine tools (CNCMTs). Smart manufacturing has to be based on the use of an intelligent CNCMT as a typical production line component, primarily owing to vast opportunities arising from information and communication technologies (ICT) and CPS development. As noted by Botkina et al. [19], the ever-increasing digitalization opens up multiple opportunities to significantly increase productivity and effectiveness of manufacturing processes and production machines; however, to ensure a proper behaviour of a complex production system, proper data analytics must be employed at all stages, including the production process itself. To predict, optimize, correct, and evaluate, engineers must be provided with instant access to data and well-structured, well-processed information extracted from the data collected throughout the whole production lifecycle.

The CNC machine tool is, of course, a complex system consisting of numerous structural elements. However, from the production process performance and the whole machine reliability it is possible to distinguish [67] its critical component-namely, a cutting tool. Tool failures [8, 49, 83, 95, 99] such as wear and breakage have impact on the machining process performance and can be harmful to a high-value machine. Therefore, a special focus is put on condition-based maintenance solutions and remaining useful life predictions for the cutting tool due to its importance for the machining process accuracy and machine tool reliability. Studies presenting different approaches to the CNCMT and its CBM or RUL prediction were analysed by Li et al. [99], who classified them depending on the proposed models into:

- Physics-based models describing tool wear by establishing mathematical models based on a failure mechanism.
- Reliability function models (statistical model-based approaches) assuming that the tool degradation process is subject to a certain distribution, and hence Gaussian process regression or hidden Markov models are proposed for tool RUL prediction.
- Data-driven models employing different kinds of sensors to monitor tool working condition.
- Hybrid models that are a combination of multiple models.

According to Li et al. [99], even though the physics-based approach can provide accurate RUL predictions, it may, in practice, be difficult, if not impossible, to make necessary extensive offline measurements, e.g. of tool wear width, which limits the scope of applications for this approach. In reliability function models, the most critical steps are the acquisition of failure or truncation data, as well as the assumption of a proper distribution. If failure data is unavailable, it is difficult to predict RUL using reliability functions, and if the assumed distribution is improper, the corresponding prediction error will be considerable. The data-driven approach requires failure data to establish models. If such failure data are unavailable or measurement results are inaccurate, it is difficult for these models to fully describe the tool wear process, which results in substantial RUL prediction errors. The combination of multiple models to create hybrid models attempts to capitalize on the advantages of different approaches by integrating them. Some studies combined the data-driven and physics-based methods to predict tool wear. Additionally, both the hidden Markov and polynomial regression models were employed to predict the RUL and health of tools. Nevertheless, these methods also require failure data to produce an accurate prediction model [47, 99]. In addition, they depend on the accuracy of data. To overcome the difficulties associated with limited or inaccurate data and to improve the accuracy of CBM and RUL predictions in terms of CNCMT smart maintenance application, a new data acquisition and modelling approach is needed.

## 4.2. Multiple-sensor system for tool condition assessment

In the first phase of the research described below, an advanced milling machine multi-sensor system was designed and constructed under laboratory conditions at the Rzeszów University of Technology [171]. It was assumed that the data would be collected from different signal sources in order to ensure health monitoring and then RUL prediction. Not only should the system be universal from the research perspective, but it should also be compliant with industrial conditions.

For analytical purposes, a typical industrial milling machine operating in real industrial conditions to that end, and it was equipped with sensors such as (Fig. 4): accelerometers, acoustic emission sensor, force and torque sensor, spindle velocity and spindle load sensor. In detail, the Haas VM-3 CNC machine was equipped with an inline direct-drive spindle and a set of sensors: seven accelerometers (sensitivity: 100 mV/g) integrated with temperature sensors, one acoustic emission sensor (sensitivity: 53 mV/Pa), and one force and torque sensor (3-axis, sensitivity: 1 mV/N, 10 mV/Nm). Two accelerometers were mounted on the lower bearing of the spindle and two on the higher bearing of the spindle, another two accelerometers were mounted along the Z-axis and one was mounted on the workpiece.

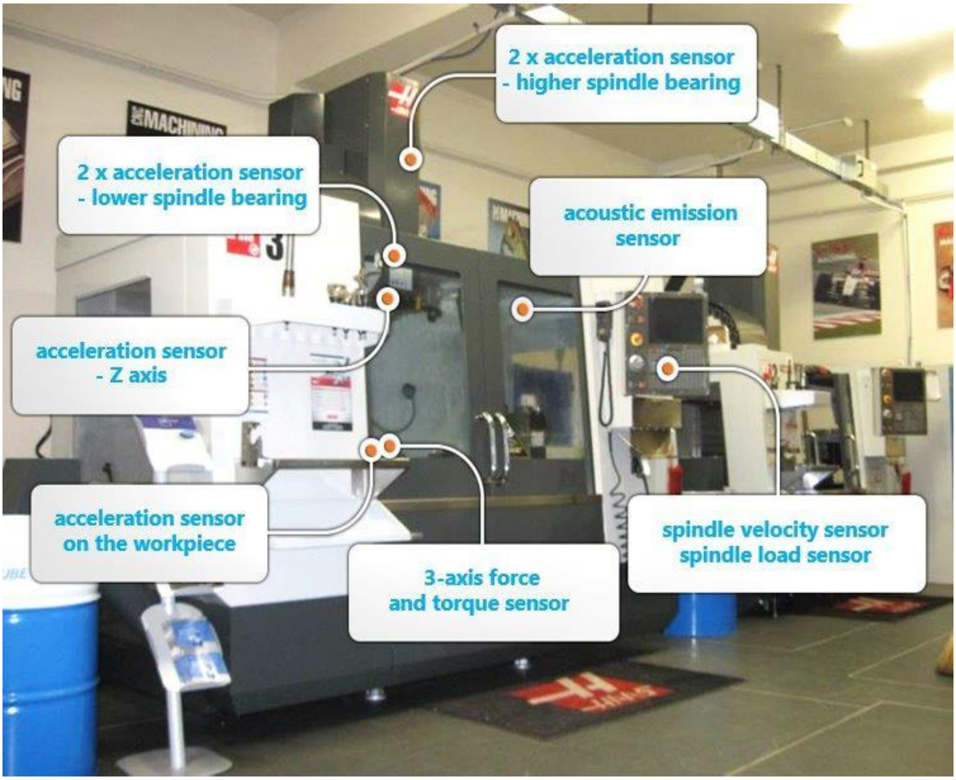


Fig. 4. Industrial CNC milling machine with a set of tool condition monitoring sensors [171]

The acoustic emission sensor was mounted in the machine cabin. Data collected with the use of this milling machine and multi-sensor tool condition monitoring system were used for different modelling purposes, their most important results being systematized and discussed in forthcoming chapters. The aim of the study was to apply both vibration and acoustic signal analysis for cutter health monitoring, condition classification, and its remaining useful life prediction. The primary goal of the experiment was to collect data about cutter condition during the milling process for further modelling. The cutter condition was divided into two classes: sharp and blunt. The experiment was performed using the industrial Haas VM-3 CNC machine. This machine is equipped with a 12,000 RPM direct drive spindle. The rotational speed of the spindle during machining was equal to 860 RPM.

A multi-component sensor, CL16 ZEPWN, was used in the tests. The sensor allows for force measurement in a range of 10 kN and torque measurement in a range of 1 kNm. The sensor has an accuracy class of 0.5, while its sensitivity is 1mV/V. The following signals were collected from the multi-component sensor: force sensor signals (*P1x, P2y, P3z*) and torque sensor signals (*M1x, M2y, M3z*). A platform for rapid prototyping of intelligent diagnostic systems was used to collect data during milling experiments [171]. The platform includes a Beckhoff industrial computer and an EtherCAT-based distributed I/O system. A hard disk of the engineering workstation was used to store data that were collected in the real time with a sampling interval of 2 ms. The duration of the signal buffer stored in one file was 640 ms.

## 4.3. Digital twin-oriented sensor data modelling

The production process management and control have some important limitations, as discussed in previous chapters. In addition, due to Industry 4.0 development challenges, there also exists [10] a lack of real-time interaction and closed-loop feedback mechanism between physical and virtual spaces. As already proved, this is the reason why the digital twin technology is a core and crucial tool enabling a close integration of manufacturing information and physical resources. Unfortunately, despite numerous analyses and discussions reported in the literature, the DT concept is still far from effective real-world implementation for more efficient production process control. To fill this gap, extensive research into digital twin-oriented sensor data modelling has been conducted in recent years, and its results are presented in this chapter.

During experiments data were collected from various real production tasks in a milling process performed using a machine. The data were used for:

- Identifying tool wear based on acoustic signal analysis [84].
- Developing a new mathematical model to effectively classify the cutting tool condition [85].
- Performing principal component analysis and logistic regression for tool condition identification [82].
- Analysing the potential use of decision trees as a decision support tool for machining process data analysis [6].
- Developing a new classifier for condition assessment and RUL prediction as an expert system tool for real-time monitoring [83].

Given the advantages of data-driven methods, including easy implementation and non-interruption of the manufacturing process, these methods have been a research hotspot for tool condition monitoring and RUL prediction. This situation stems from a lack of effective methods for developing product, process, and operation models based on virtual and physical convergence, which–at present–implies low real-time capability and predictability in production management, as well as necessitates future development.

### 4.3.1. Data reduction methods

As mentioned previously, information overload and difficulties with effective analysis and processing of collected data, due to the need of using sophisticated tools and possessing extensive knowledge and experience, pose a challenge to almost every company. One of information overload solutions used in data mining is data reduction [159]. Time series analysis, principal component analysis, and factor analysis methods are amongst the most popular techniques for reducing the number of variables or data in order to avoid the curse of dimensionality [78, 157]. Data reduction techniques are used as an effective information overload solution in several specific applications; however, proper data mining improvement techniques are still needed in the field of maintenance, where terabytes of different data are collected and processed online for failure indication. This is particularly true given a great need for appropriate methods and techniques ensuring the durability and reliability of production systems, which entails, among other things, the use of adequate predictive maintenance tools.

It was assumed [159] that many efforts were made to develop methods and tools for diagnosing failure for the purpose of predictive maintenance; however, several ground-breaking objectives have not been achieved yet. The positive effect of PM approaches in the improvement of operation and maintenance processes may be mitigated by different factors. A lack of adequate information concerning the maintenance process is one of them, which may result from, among other things, inadequate data acquisition and data analysis. This is clear for condition-based measures in particular, as better time prediction and better data mining would lead to a longer planning horizon. With more complex objects, raw time domain signals are too difficult to interpret directly. Collected data require a huge amount of storage space and seldom provide any insight into the condition of an element or object being monitored.

With complex machinery operating under a wide variety of conditions, however, more advanced approaches are needed. Recent advances in monitoring and real-time industrial data collection have also led to information overload due to an increased number of measurements, and thus a higher number of variables associated with every observation. In this way, high-dimensional datasets are obtained, not all of these measured variables and data being valid or important for understanding the analysed phenomena. In the case of predictive maintenance, those unimportant features may additionally lead to a significant decrease in prediction accuracy. Risk assessment tools can be considered [159] an equally effective solution, also in the case of data-driven predictive maintenance approach.

In the proposed solution [159], the behaviour of the observed diagnostic variable is modelled using two kinds of Levy-type diffusion processes. Levy processes are stochastic processes with independent increments that are also time homogeneous. The first one is the Wiener process with drift in a linear form $X(t)$, $t \geq 0$, expressed as a stochastic differential equation:

$$\mathrm{dX(t) = \theta dt + \sigma(t) dW(t)} \tag{1}$$

where the real constant $\theta$ and the non-negative time function $\sigma(t) \geq 0$ are non-random parameters. The solution of this stochastic differential equation with a firmly given boundary condition:

$$X(0) = x_0 > 0 \tag{2}$$

is a diffusion process:

$$X(t) = x_0 + \theta t + \int_0^t \sigma(s) dW(s), \ \ t \geq 0 \tag{3}$$

where the integral is considered as Itôo [127]. The expected mean value of this stochastic process is:

$$E(X(t)) = x_0 + bt, \quad t \geq 0 \tag{4}$$

as the mean value of the Itôo integral is equal to zero. The mean value of the $X(t)$ process is a linear function of time with an absolute component $x_0$ and a drift $\theta$. Using Itôo isometry, the variance of the $X(t)$ process has the following form:

$$VarX(t) = \int_0^t \sigma^2(s) ds, \ t \geq 0 \tag{5}$$

A variance of such process is a non-decreasing function of time.

The other process we apply is a different type of Levy stochastic diffusion processes, i.e. the Ornstein-Uhlenbeck process (OUP). We consider [159] the OUP with drift that is given by a stochastic differential equation according to Itôo:

$$dX(t) = \left[b + \frac{a+bt-X(t)}{\tau}\right] dt + \sigma dW(t), \quad t \geq 0, \tag{6}$$

with the deterministic boundary initial condition $X(0) = x_0$.

OUP is very often described as a mean reversion process. The parameters $a$ and $b$ stand for an absolute component and a gradient that is actually the asymptotic mean value of this process. Similarly to the Wiener process (WP) with drift (1) and (3), the parameter $b$ is called as a drift. The parameter $\sigma$>0 is a dispersion parameter. We can denote $\tau > 0$ as the time constant from a physical point of view. The mean reversion rate $1/\tau$ is actually the speed of an $X(t)$ process return to its mean value. To generate a random behaviour of the OUP, we further use the following form:

$$X(t) = a + bt + (x_0 - a)\, exp\left(-\frac{t}{\tau}\right) + U\sqrt{\frac{\tau\sigma^2}{2}}\sqrt{1 - exp\left(\frac{2t}{\tau}\right)}, \tag{7}$$

where $U \sim N(0;1)$.

The value $X(t)$ in time $t$ is, therefore, a random variable with normal distribution with its mean value:

$$EX(t) = a + bt + (x_0 - a)\, exp\left(-\frac{t}{\tau}\right) \tag{8}$$

and variance:

$$VarX(t) = \frac{\tau\sigma^2}{2}\left[1 - exp\left(-\frac{2t}{\tau}\right)\right]. \tag{9}$$

Real field data were used to estimate parameters for the WP with drift and the OUP with drift alike.

In most applications of the Wiener process, the maximum likelihood estimation (MLE) is used for parameter estimation. The application of MLE is relatively universal, robust, versatile and useful provided that the data form is normal and independent. However, there is a problem in terms of confidence/prediction interval calculations, especially in the edge parts of data cloud. Parameters related to confidence/prediction intervals are among the most important parameters later applied as diffusion parameters when constructing diffusion processes. Moreover, the analysed parameters, especially the diffusion parameter, are not constant but time dependent, which is more realistic.

For this purpose, therefore, based on the limits mentioned above, it was assumed that the more applicable approach for parameter estimation would be the least square method. The least square method was applied here both for estimating parameters of the Wiener process with drift and the OUP with drift. Based on the field data, we constructed a linear regression model (LRM) with 95% of prediction intervals [159].

The aim of the study, presented in detail in [159], was to prove that the number of datasets collected and processed in data acquisition systems can be limited by the use of periodic rather than continuous measurements. However, the periodicity of measurement must be precisely defined and regularly verified in compliance with changes in the technical condition of the analysed object. As regards the technical object [159], the measurement data collected and processed on a continuous basis, usually over a very long period of correct operation, do not generate information that is vital for the user.

The rejection of continuous measurement and replacing it with the diagnostics made at a precisely specified moment by a precisely specified time of measurement would lead to a more efficient use of the system for processing limited datasets. This can be achieved by adequate failure risk modelling, which results in its serving as a basis for drawing inferences about measuring frequency. Hence, the analysed process was categorized as a stochastic process with continuous time based on Levy-type diffusion processes.

The empirical distribution function was calculated together with the survival function and hazard function of the observed system, its behaviour simulated by the Wiener process with drift. From a practical point of view, it is relevant to think about soft failure occurrence, which is represented by the quantile 0.025, where – according to the modified Wiener process with drift [159] – failure may occur after 1082947 seconds, which means 300 hours of safe operational time. In such case, measurement should be made every 300 hours in order to guarantee adequate performance. As regards the OUP with drift, soft failure may occur after 295421 seconds (Table 1), which allows for measurement at latest every 82nd hour of operation.

Table 1. Moments of the distribution of $t_{FHT}$ in the observed system, as simulated by Wiener process with drift and OUP with drift [159]

| Moment | Wiener process with drift [s] | Modified Wiener process with drift [s] | OUP with drift [s] |
|---|---|---|---|
| Min | 142 | 886 200 | 420 |
| Median | 3 103 000 | 4 522 000 | 1 095 000 |
| Mean | 3 157 840 | 4 777 000 | 1 055 000 |
| Max | 9 279 000 | 15 880 000 | 1 858 000 |
| Var | 6.46E+12 | 5.21E+12 | 1.06E+11 |
| SD | 2 542 021 | 2 281 514 | 324 299 |
| 0.025 | 425 | 1 082 947 | 295 421 |
| 0.25 | 479 966 | 3 134 847 | 860 020 |
| 0.5 | 3 102 854 | 4 521 847 | 1 094 590 |
| 0.75 | 5 086 931 | 6 143 847 | 1 290 695 |
| 0.975 | 8 248 903 | 9 965 747 | 1 582 739 |

With more advanced system operation management, the potential system hard failure occurrence risk (represented by: 0.25 quantile for failure low risk level; 0.5 quantile for failure medium risk level; 0.75 quantile for failure high risk level; and 0.975 for almost sure situation of hard failure occurrence) occurs, respectively, after 9965747 seconds or 1582739 seconds (Table 1), which is equal to 115 days or 18 days of operation. Different modelling techniques bring different results, as shown in Table 1.

From a safety point of view, the lowest value of operational time before failure achieved in modelling should be taken into consideration, which means that measurements should be made every 82 hours. Consequently, this also implies a considerable reduction in the amount of collected data together with reduced acquisition and processing time and cost, thus ensuring high reliability and safety of the analysed asset.

Naturally, it is well known that ageing and degradation of any technical object proceed continuously [118]. This means that measurement intervals performed by the data acquisition system, identified according to the risk assessment modelling procedure described above, must be continually adjusted by the computerized maintenance management system [159].

### 4.3.2. Tool wear identification based on acoustic signal analysis

The identification of an optimal moment for cutting tool replacement is an important problem for every engineer who decides about machine tool condition. As tool-related costs make up a significant proportion of the total cost of production, the economical tool replacement policies are of great interest, especially for more expensive tools.

There are several studies on the problem of finding the time to tool replacement. One of them [149] presents a proportional hazard model for simulating tool reliability and hazard functions when the tool is used under different sequential values of cutting speed. Cutting speed is a covariate in this model. Experimental data are obtained and used to construct and validate the model, which is then used in decision making. Three criteria for tool replacement are proposed. The results show that the proposed solution provides a very good representation of the reliability and hazard functions at different speed values; nevertheless, the analysed object is s limited representation of the real machining process.

A different solution is proposed by Xu et al. [169], who have developed a quality failure rate model aimed at describing product quality deterioration during the cutting process. And then an approach for tool replacement decision-making with product quality deterioration and random tool failure is proposed. The optimal tool replacement time is obtained through balancing product quality loss, penalty cost for possible tool failure, production capacity loss and tool replacement cost. Unfortunately, the proposed solution is based on theoretical investigations, without any real production data analysis.

Another solution was presented by Zaretalab et al. [172], and it focuses on the stochastic approach for tool life modelling in the milling process. By considering the costs of tool condition monitoring methods, a hybrid policy was developed based on the reliability function for optimising the tool replacement time. The proposed policy covers seven functional modes, assuming discrete and continuous modes. The policy was applied in an experimental process using a CNC milling machine to assess its performance. Finally, the effect of different costs is demonstrated via sensitivity analysis, and the outcomes of the optimised policies under various conditions are compared. The results demonstrate that the proposed policy can optimise the tool replacement time due to its flexibility in covering the different functional modes efficiently; however, it does not precisely predict the remaining useful life of the cutting tool.

Traditional tool life models do not take into account the variation which is inherent in metal cutting processes. As a result, typical real-time diagnostics and data analysis seem to be more appropriate and more accurate with respect to final results. In terms of diagnosing machines and machining processes, acoustic emission is one of the diagnostic signals sources [133, 154, 173]. The acoustic signal can be used to detect and diagnose most phenomena relevant for the diagnosis of machines and processes; it, however, requires considerable computing power for signal processing and much higher sampling rates when compared to a vibration signal. For example, in the case of gearbox fault detection acoustic emission-based techniques require much higher sampling rates than vibration analysis-based techniques [133].

Therefore, it is questionable whether this technique would give a better or at least the same performance as the vibration analysis-based techniques using the same sampling rate. For this reason, the task of developing a solution for predicting the optimal moment of cutting tool replacement based on its wear degree identification by acoustic emission was analysed [84].

Correlation analysis, which is related to spectral analysis, was used to identify parametric properties of the signal [50, 78]. At the beginning [84], the standardization of time series relating to the acoustic signal was performed, so that the average values of each series $\{x_t\}_{1 \le t \le n}$ were equal to zero. The extended Dickey-Fuller test (augmented Dickey-Fuller test) was used to test the stationarity of the time series. The Ljung-Box test was used to test the significance of the correlation in the $\{x_t\}_{1 \le t \le n}$ series. The analyses [84] show that tool condition can be identified with the use of the calculated correlation values determined based on an acoustic signal sample. Therefore, additionally, the problem of cutter condition classification should be considered.

For this purpose, a support vector machine (SVM) was applied. SVM is an abstract concept of machine learning (ML), its main task being object classification or regression analysis. As far as object classification is concerned, the SVM algorithm [162] builds a model by determining a certain hyperplane that separates classes. Let:

$$S = \{(z_t, y_t): z_t \in R^m, y_t \in \{-1,1\}, 1 \le t \le T\} \tag{10}$$

denote the training dataset, where $z_t = \{r_\tau(t)\}_{1 \le \tau \le m}$ denotes the values of the autocorrelation function for the sample $t$, $y_t = -1$ for a sharp tool and $y_t = 1$ for a blunt tool. Observing the element $z_t \in R^m$ classification should be performed. Classes are nonlinearly separable if such hyperplane exists $f_h(z) = 0$, where:

$$f_h(z) = \sum_{j=1}^{l} \beta_j y_j k(z, z_j) + \beta_0 \tag{11}$$

and $f_h(z_t)y_t > 0$ for $1 \leq t \leq T$. The function $k: R^{m\times m} \rightarrow R$ is called as a kernel, while $l$ is the number of support vectors. Gaussian kernels were used for cutter condition identification [84], with the obtained tool condition classification error being below 2.6% (Table 2).

Table 2. Results of tool condition classification with the use of SVM [84]

| | Classification: sharp | Classification: blunt |
|---|---|---|
| Reference: sharp | 1209 | 27 |
| Reference: blunt | 29 | 908 |

In the analysed case, the number of support vectors was equal to $l = 502$. An analysis of the data in Table 2 reveals that out of 2,173 cases only 56 were misclassified (27 + 29). Thus, the cutting tool condition classification error is below 2.6%, as mentioned above.

### 4.3.3. Mathematical model for tool condition classification

The study [85] primarily aimed to determine tool condition. The determination whether the tool is sharp or not is essential to define its function, since the response variable is qualitative. Hence, values of this function should be included in a set of two elements describing the potential tool condition. The response variable has a categorical value in the case under consideration. The prediction of tool condition based on data obtained from sensors (accelerometers, microphones, etc.) can be referred to as a classification problem. Numerous classification techniques (classifiers) might be used to predict inclusion in the appropriate class [17, 34, 85, 97, 125, 142]. The use of logistic regression makes it possible to calculate the probability of the response variable's belonging to the appropriate category. Therefore, instead of tool condition determination, the probability of every possible condition was estimated. In other words, the application of logistic regression allows us to determine the distribution of the response variable based on the observation of input variables. Some observable input variables are strongly correlated, which was discussed in detail in [85]. The elastic-net method was used to minimize this problem.

Let us now consider a data set where the realization of the response variable belongs to a binary set. For any finite element, we analyse the training set $D = \{(x_i, y_i)\}_{1 \leqslant i \leqslant n}$, where $\{x_i\}_{1 \leqslant i \leqslant n}$ denotes the series of input variables, $\{y_i\}_{1 \leqslant i \leqslant n}$ is the series of the response variable, where $x_i \in R^m, y_i \in \{0,1\}$ for $1 \leqslant i \leqslant n$ denotes the number of samples, $m$ denotes the number of measurements obtained from the transducers (sensors). If the cutter is blunt, then we take $y_i = 1$ , otherwise we put $y_i = 0$. The training set can be presented as $D = \{Y, X\}$, where [85]:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = \begin{bmatrix} x_{(1)} \\ x_{(2)} \\ \vdots \\ x_{(n)} \end{bmatrix} \tag{12}$$

Observing the signal $x_i \in R^m$ obtained from the sensors, it is necessary to classify cutter condition. The task is to find a classifier such that $f: R^m \to \{0,1\}$, which would allow us to classify the cutter into categories y=1 or y=0, based on the observation $x \in R^m$. To solve an auxiliary task (and determine unknown parameters), the Newton-Raphson algorithm was applied. The application of this algorithm allows for unknown parameters to be estimated iteratively. Usually, measurements obtained from the sensors are correlated (which is referred to as a multicollinearity problem). Thus, the problem depends on the selection of appropriate predictors that should be included in the regression model. On the one hand, these predictors should influence the value of the response variable, while on the other they should not generate multicollinearity. The literature reports numerous techniques (e.g. singular value decomposition, regularization, least angle regression) for solving the problem of multicollinearity. Multicollinearity exists whenever an independent variable is highly correlated with one or more of other independent variables in a multiple regression equation. Multicollinearity is problematic because it undermines the statistical significance of an independent variable [178].

One of the possible ways to reduce multicollinearity between predictors is the application of the elastic-net method, which was successfully done in the presented investigation [85]. Acoustic signal properties were identified by correlation analysis that is related to spectral analysis. Therefore, the time series $\{x_t\}_{t \in N}$ was considered, which denotes the acoustic pressure and is (weakly) stationary in a broad sense, with realizations in the set of real numbers $R$.

The aim of research [85] was to investigate the possibility of designing a classifier that would identify tool condition. To develop the classifier, 2173 signals (series of acoustic pressure) obtained from a microphone were analysed, where 937 cases represent the blunt tool and 1236 the sharp tool. Each series $\{x_t^j\}_{1 \le t \le n}$ had 16000 measurements collected with a 25 kHz sampling frequency and was identified by the sequence of correlation values:

$$r^j = \{r_t^j\}_{1 \le t \le m} \in [-1,1]^m \tag{13}$$

for $0 \le j \le 2173$. Examples of signal realization and correlation sequence are given in Fig. 5.

For each analysed acoustic series $\{x_t^j\}_{1 \le t \le n}$, $0 \le j \le 2173$ an ADF test [106] was performed. Additionally, the Ljung-Box test was performed to examine the null hypothesis $H_0$ that elements of the acoustic series $\{x_t^j\}_{1 \le t \le n}$, $0 \le j \le 2173$ are independent.

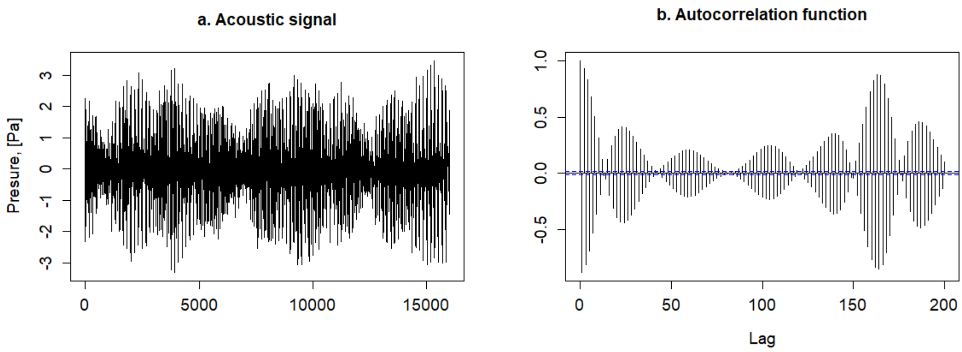


Fig. 5. Realization and autocorrelation for the acoustic signal $\{x_t^j\}_{1 \leq t \leq n}$ [85]

The achieved [85] relative error of classification was equal to (Table 3):

$$\frac{(77+106)}{2173} \approx 0.0842$$

The developed classification method enables effective identification of cutting tool condition, allowing further determination of the optimal moment of its replacement. Table 4 presents the identification ratios for the logistic regression model. The sensitivity (recall or probability of detection) was calculated as a proportion of the number of exact (relevant) identifications for blunt cutters to the number of samples that actually had a blunt cutter.

Table 3. Confusion matrix (identification results) for a classification level of 0.46 [85]

| | Classification: sharp | Classification: blunt |
|---|---|---|
| Reference: sharp | 1130 | 77 |
| Reference: blunt | 106 | 860 |

On the other hand, specificity was calculated as a proportion of the number of exact (relevant) identifications for sharp cutters to the number of samples that had a sharp cutter. The positive predicted value (precision) is a ratio of the number of exact (relevant) identifications for blunt cutters to the number of samples identified as blunt. Precision describes the purity of a classifier performed (purity in retrieval performance). Additionally, McNemar's chi-squared test was performed for symmetry of rows and columns for the confusion matrix presented in Table 4.

For the presented case, the McNemar chi-squared statistic was equal to 4.2842 and $p$-value was equal to 0.03847. This means that at a significance level of 0.05, there is no basis to reject the null hypothesis. Additionally, the $\kappa$-statistic was calculated. The $\kappa$-statistic is the proportion of disagreements expected by chance that did not occur. For identification based on the logistic regression model, the parameter $\kappa$ was equal to 0.829 [85].

Table 4. Basic ratios of quality for the logistic regression model [85]

| Parameter | Value |
|---|---|
| Accuracy | 0.9158 |
| Sensitivity | 0.9178 |
| Specificity | 0.9142 |
| Positive Predicted Value | 0.8903 |
| Negative Predicted Value | 0.9362 |
| Prevalence | 0.4312 |
| Detection Rate | 0.3958 |
| Detection Prevalence | 0.4445 |
| Balanced Accuracy | 0.9160 |
| False Alarm Rate | 0.0858 |

The developed solution enables effective identification of the cutting tool condition, which makes it possible to determine the optimal moment of its replacement. The above method is also a simple way to detect current tool condition.

### 4.3.4. Principal component analysis and logistic regression for tool condition identification

As mentioned previously, to cope with the dynamic data growth, it is necessary to create methods and tools that not only automatically collect data, but also select the most relevant data and use appropriate analyses to extract knowledge therefrom.

According to this identified research challenge, the possibility of applying major component analysis and logistic regression for machine element condition identification was analysed and verified based on industrial time series data [82]. For every observation $1 \leq i \leq n$ , we create elements of a learning dataset: for the blunt tool we assume $y_i = 1$ and for the sharp tool (no warnings) $y_i = 0$. This was done for 2173 samples that were attached to the data set. To that end, we analyzed signals coming from the experiments performed on a testbed comprising a CNCMT. The force and torque sensor was mounted in the chuck. In this study, data from the accelerometer mounted on the lower bearing of the CNC machine spindle and the $P2y$ force signal parallel to the Y-axis of the force and torque sensor were used.

The stationary property of data from the accelerometer mounted on the lower bearing of the CNC machine spindle was examined with the use of the augmented Dickey-Fuller (ADF) test. In addition, the significance of the correlation was examined via the Ljung-Box test. Using the Herglotz theorem for functional time series, data from the accelerometer mounted on the lower bearing of the CNC machine spindle should be identified by the autocorrelation function data [82]. Logistic regression was used to identify tool condition, with scores of principal components selected as explanatory variables.

The maximum likelihood method was applied to estimate structural parameters, and the tool condition probability was calculated for each case. For comparison, 15 and 30 components were used. Figure 6 shows the cutter probability distributions estimated with the use of 15 and 30 principal components, respectively. For 15 principal components the proportion of the explained variance is equal to 0.9793, while for 30 of them this value is 0.9977. The results, which are presented in detail in [82], prove that by applying logistic regression for tool condition identification and taking into account only 30 principal components (instead of 213 variables) we can achieve the expected result.

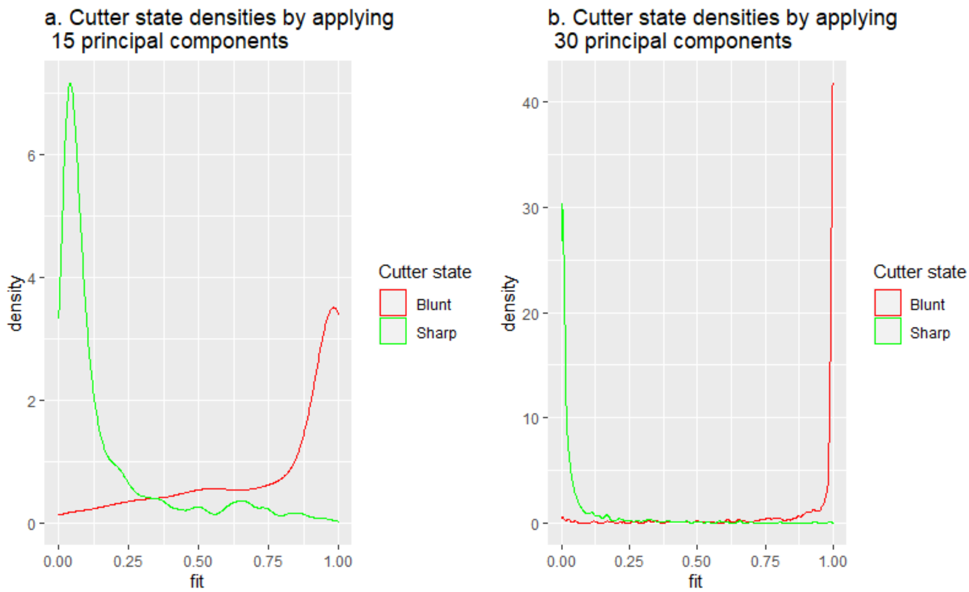


Fig. 6. Probability distribution [82]

The tool condition was predicted with a very small error (3.77%). Table 5 presents the classifier operating characteristics. The true positive rate (sensitivity, hit rate) is equal to 0.9349 and describes the proportion of real positive cases in which the blunt tool was correctly identified as blunt. The true negative rate (specificity) is equal to 0.983 and denotes the proportion of real negative cases in which the sharp tool was correctly identified as sharp.

The proposed application of principal component analysis and logistic regression enables the effective processing of sensor data related to tool condition identification based on signals generated during machining. As a result, the above methods may find practical application in condition monitoring systems. In particular, they may be helpful in the real-time identification of tool condition. In effect, the tool can be replaced the moment it loses its functional properties.

Table 5. Prediction model quality indicators and their values [82]

| Indictor | Value |
|---|---|
| Sensitivity | 0.9348986 |
| Specificity | 0.9830097 |
| Positive Predictive Value | 0.9765886 |
| Negative Predictive Value | 0.9521944 |
| Precision | 0.9765886 |
| Prevalence | 0.4312011 |
| Detection Rate | 0.4031293 |

In practice, this may mean that the replacement is made either too late or too soon. In industrial realities this may cause problems with ensuring the desired product quality and/or the cost effectiveness of manufacturing processes. The use of an effective condition monitoring system based on the analysed methods can prevent these adverse effects.

### 4.3.5. Decision trees as a decision support tool

Machining process data obtained from the multiple-sensor system for tool condition assessment described in Chapter 4.2 were used for different reliability purposes, as presented in Chapters 4.3.1-4.3.6. The research programme was continued in order to compare different modelling methods for cutting tool classification, wear identification, or remaining useful life prediction. The primary aim of the research described in this chapter was to identify tool condition based on data (instrument readings, signal readouts) obtained from accelerometers and microphones. Tool condition identification consists in determining whether the cutting tool is sharp or not. The outcome variable has a categorical value in the case considered. Thus, the presented task is related to a classification problem. In a study exhaustively described in [6], classification trees were employed for cutter condition identification.

A decision tree in data mining [56, 134] is a model which can be used to solve both classification and regression problems. With the growing importance of exploring large and complex data sets in time series analysis, knowledge discovery and data mining, the application of decision trees has become a powerful and popular approach. In operational research, decision trees refer to hierarchical models of decision and their consequences. When a decision tree is used for classification, it is referred to as a classification tree. When it is used for regression tasks – it is called a regression tree. This method allows to perform analyses that lead to finding a set of logical rule conditions, type, "if, then", helping to classify the examined objects clearly. In data mining and machine learning, decision trees are considered as predictive models. This is one of the most popular and effective methods of data mining which is also very often used for prediction.

Classification trees are created when a dependent variable is qualitative, and regression trees – with a continuous form of a dependent variable. Classification trees are used to determine the belonging of objects to classes, based on the measurements of one or more describing variables, determining their impact on a qualitative dependent variable – the forecast (predicted) variable. Prediction can be understood as a model that can be used to estimate (calculate) the value (or a range of values) of an attribute. This attribute value, in particular, may be a class label. As opposed to human abilities, classification trees help generate rules in very complex multidimensional cases. Obtained rules are usually presented in the form of a tree, and they are transparent even in the case of large trees [6].

To create a classification tree, the domain of input variables (a set of possible realizations of observations) should be divided into k separated regions. Each region corresponds to an appropriate categorical value (from a set of possible outcomes) of a response variable. To define a region, it is necessary to create conditions that can be satisfied by input variables, these conditions establishing a rule. A set of rules determines a classification tree.

The most advanced algorithm for building decision trees is that proposed by Breiman [23], which is known as the Classification and Regression Tree (CART) algorithm. The construction of a decision tree is carried out by conducting an in-depth search for all available variables and all possible splits in a data set for every decision node (t) by selecting an optimal split [11]. In the analysis and classification tree design, $\{(y_i, x_i)\}_{1 \le i \le n}$ denotes the analysed data set, where $y_i \in \{c_1, c_2, \dots, c_s\}$ and $x_i = (x_{i1}, x_{i2}, \dots, x_{ik}) \in R^k$. The values $c_1, c_2, \dots, c_s$ denote the possible classes for the feature $y$. The classification task involves dividing the space $R^k$ on $q$ separated regions, where each region corresponds to an appropriate class. Based on the observation feature $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$, we need to classify the analysed object. The entire space $R^k$ is divided into $q$ separated regions, $R_1 \cup R_2 \cup \dots \cup R_q = R^k$. For the node $m$, $1 \le m \le q$, representing the region $R_m$, the Gini index is determined as follows:

$$Q_G(m) = \sum_{j=1}^{s} p_{mi}(1 - p_{mi}) = 1 - \sum_{j=1}^{s} p_{mi}^2 \qquad (14)$$

where $p_{mi}$ is the conditional probability for the $j$ −th class in the node, s is the number of classes. In node $m$ with $n_m$ observations, the conditional probability for the $j$ −th class is equal to:

$$p_{mi} = \frac{\#\{y = c_i : x \in R_m\}}{n_m} \qquad (15)$$

The generated decision trees are qualitatively assessed via the use of k-fold cross-validation and confusion matrix. A confusion matrix shows classification errors by class. Based on the confusion matrix, numerical indicators can be calculated. They are determined as follows [40, 117, 143, 165]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (16)$$

which is determined as the sum of TP and TN, it indicates that the results are correctly classified to all the analysed data. This indicator evaluates the prediction ability of the model.

$$TruePositiveRate = Sensivity = \frac{TP}{TP + FN}, \quad (17)$$

is the rate that determines the fraudulent free transactions classified as fraudulent.

$$Specificity = 1 - FalsePositiveRate = \frac{TN}{TN + FP}, \quad (18)$$

is a measure of how well a test can identify true negatives.

$$PositivePredictiveValue = \frac{TP}{TP + FP}, \quad (19)$$

is an indicator that describes the relationship between the number of true positives and the total number of positives: true positives and false positives.

$$NegativePredictiveValue = \frac{TN}{TN + FN}, \quad (20)$$

is an indicator that describes the relationship between the number of true negatives and the total number of negatives: true negatives and false negatives.

$$Prevalence = \frac{TP + FN}{TP + TN + FP + FN}, \quad (21)$$

is an indicator that determines the frequency of occurrence of the distinguished class.

$$DetectionRate = \frac{TP}{TP + TN + FP + FN}, \quad (22)$$

is an index that measures the ratio of true positives to the total number of predictions.

$$DetectionPrevalence = \frac{TP + FP}{TP + TN + FP + FN}, \quad (23)$$

is an index defined as the number of predicted positive cases divided by the total number of predictions.

$$BalancedAccuracy = \frac{Sensivity + Specificity}{2}, \quad (24)$$

is a metric that one can use when evaluating how good is a binary classifier.

$$FalseAlarmRate = \frac{FP}{TP + FP}. \quad (25)$$

is the probability of falsely rejecting the null hypothesis for a particular test.

For the formulated research problem [6], a decision tree was created. The main goal was to obtain:

- a model that explains changes in the condition of a cutter blade depending on the measured parameters,
- a set of decision rules that can be used to predict cutter condition for given values of the measured parameters,
- a ranking of the importance of variables, and thus to provide information about parameters having the greatest impact on the cutter blade condition.

Figure 7 shows the most-developed decision tree consisting of 20 splits nodes and 22 terminal nodes. For each node, the probability value of the occurrence of a result in a given class and the percentage of the result in a data set are given.

The last element of the decision tree analysis was a confusion matrix. The generated confusion matrix is presented in Table 6, with the sharp cutter referred to as a negative case (N), while the blunt cutter is a positive case (P).

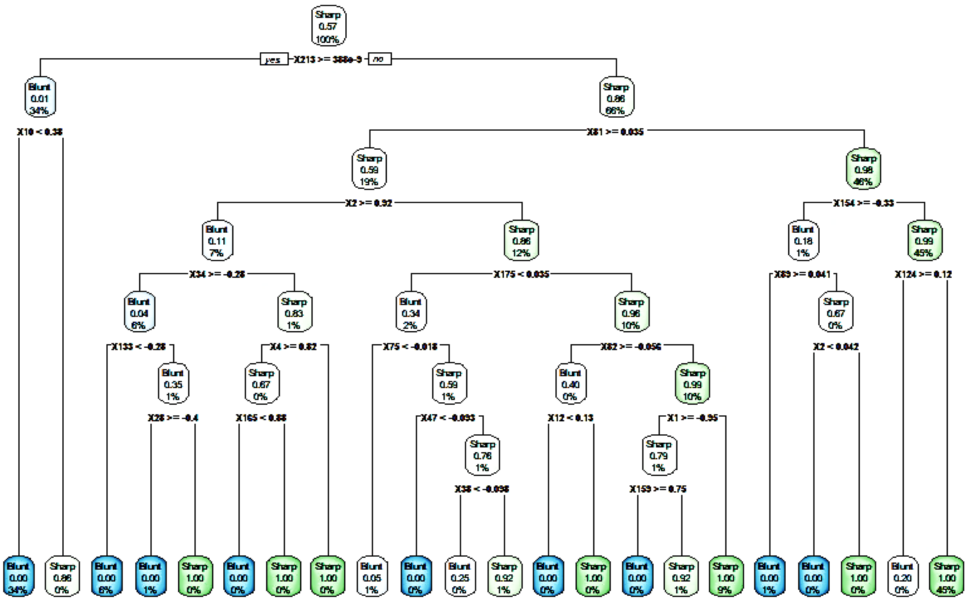


Fig. 7. Decision tree for cutter condition identification for cp = 0.001 [6]

The confusion matrix requires the following values to be determined: TP (True Positive) denotes the number of cases for which the blunt cutter was correctly identified, TN (True Negative) is the number of cases for which the sharp cutter was correctly identified, FP (False Positive) is the number of cases for which the blunt cutter was identified as sharp (false alarm), and FN (False Negative) denotes the number of cases for which the sharp cutter was identified as blunt [6].

Table 6. Confusion matrix (identification results) for *cp*=0.001 [6]

| | Prediction: sharp | Prediction: blunt |
|---|---|---|
| Reference: sharp | 1233 | 3 |
| Reference: blunt | 5 | 932 |

It means that the prediction error was (8/2173) ≈ 0.37%. The analyses of the indicator values of prediction model quality confirmed the prediction error value (Tab. 7).

Table 7. Prediction model quality indicators and their values [6]

| Indicator | Value |
|---|---|
| Accuracy | 0.9963 |
| Sensitivity | 0.9946638 |
| Specificity | 0.9975728 |
| Positive Predictive Value | 0.9967914 |
| Negative Predictive Value | 0.9959612 |
| Precision | 0.9967914 |
| Recall | 0.9946638 |
| Prevalence | 0.4312011 |
| Detection Rate | 0.4289001 |
| Detection Prevalence | 0.4302807 |
| Balanced Accuracy | 0.996 |

The accuracy value is 0.9963, which means that the prediction error is equal to 1 (Accuracy = 0.0037; 0.37%). This value indicates a very high predictive ability of the analysed tree.

Given both the preference for less complex classifiers (with a fewer decision and leaf nodes) and the established relationship between prediction error and *cp* value, the analysed decision tree was pruned to a *cp* value of 0.005 (Fig. 8).

The prediction error amounts to 1.93% and is equal to the accuracy value obtained, which indicates a decreased predictive ability of the analysed tree. The method presented in [6] offers a simple way for identifying cutter condition. It can also be considered as promising, since results obtained therewith have been confirmed by the prediction model quality indicators.

In this study, tool condition identification was based on the data obtained from accelerometers and microphones. Condition monitoring techniques, such as temperature, vibration, and acoustic signal analysis, play an important role as indicators of developmental failure and offer a wide range of different applications for fault diagnosis. A decision tree algorithm was also used by Vamsi et al. [161] to identify, among extracted features, the dominant feature (standard error) when assessing fault severity for a wind turbine gearbox under non-stationary loading by different condition monitoring techniques.

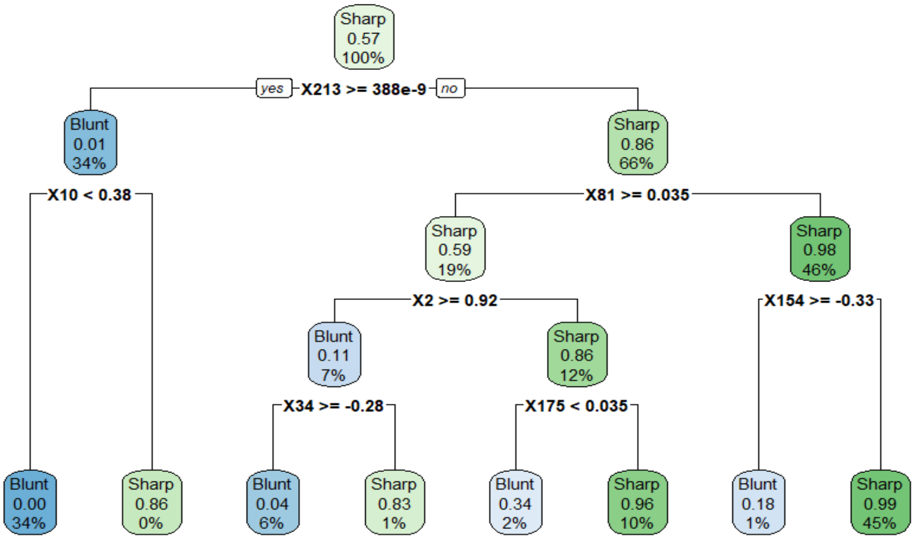


Fig. 8. Decision tree for cutter condition identification (*cp* = 0.005) [6]

SVM was used to classify features of different fault levels. The main purpose of this investigation was to verify the diagnostic capabilities of vibration signal analysis, compared to other techniques, in the fault detection of a gear tooth root crack and a gear tooth chip. The authors of [161] did not go beyond standard diagnostics. Unfortunately, however, neither RUL prediction based on collected data nor modelling techniques as a decision support tool were considered. Vibration generated during machining operations is directly linked to the problems of systems with rotating or reciprocating parts, such as bearings, engines, gear boxes, shafts, turbines, and motors. Vibration analysis has proved to be a measure for any cause of inaccuracy in manufacturing processes and manufactured components or any machine-related maintenance decision. It should, however, be remembered that vibration signal analysis most researchers focus on is not the only available condition monitoring technique. Technological machines and their typical example – machine tools, are equipped with several different sensors. Data gathered with these sensors could be alternatively used for effective structural health monitoring.

### 4.3.6. New classifier for tool condition assessment and prediction

Predicting the remaining useful life (RUL) of a tool, defined as the length of time from the current moment to the end of its useful life, is a key step in CBM implementation [93]. The primary objective of RUL is to predict the remaining useful time before a tool loses its capacity based on condition monitoring information [99].

RUL refers to the time left before observing a failure, given the current machine age and condition, as well as the past operation profile. It is defined as the conditional random variable [62]:

$$T - \frac{t}{T} > t, Z(t) \tag{26}$$

where $T$ denotes the random variable of time to failure, $t$ is the current age, and $Z(t)$ is the past condition profile up to the current time.

Jardine et al. [62] observe that many studies use the term "Remaining Useful Life Estimate" (RULE), which is used with a double meaning. In some cases, it refers to finding the distribution of RUL. In other cases, however, it simply denotes the expectation of RUL:

$$E[T - \frac{t}{T} > t, Z(t)] \tag{27}$$

As it was discussed in detail in [83, 147], in prediction maintenance problems, regression-based formulations arise when predicting the

RUL of a process or equipment. On the other hand, classification-based maintenance formulations occur when seeking to discriminate between healthy and unhealthy conditions of the system being monitored. Consequently, separate calculations are usually necessary for monitoring data (information used in PvM policies) and predicting RUL. Not only is the RUL prediction essential to verify whether the production process goals can be accomplished, but it also provides an important aid to online decision-making activities such as fault mitigation [136]. PvM calculations are usually based on the mean or median of maintenance cycle lengths, as well as on the definition of optimal action threshold for distinguishing between faulty and non-faulty process iterations based on observed data [147, 148].

While classification tools are a natural choice for PvM purposes, they do not map naturally to RUL factors that can be extrapolated from the process data for maintenance predictions and further decision making. Therefore, a new method was developed for proper estimation and representation of uncertainty in RUL prediction, based on the combined use of SVM as a classification tool and ARIMA-based identification to predict the remaining useful life [83]. A considerable reduction in calculation time and data processing complexity was achieved with the proposed method. From the entire learning dataset (time series data coming from a multiple-sensor measurement system of the machine tool in a production line), a suitable number of support vectors containing processed acoustic signals was selected, i.e. data essential for classifier design were extracted. In addition to that, a sequence of linear regression values spanned by support vector kernels was analysed to predict the number of cycles to object (cutting tool) failure.

All this was done in line with the expectation [59, 64, 67, 83, 85] regarding the development of technological machines in terms of service and repair optimization, status and process monitoring, intelligent maintenance, and the use of advanced diagnostic tools. Technological machines are, therefore, more and more advanced, and usually make part of an extensive stock of machine tools in enterprises operating in the area of the so-called high-tech industry. The high degree of technological advancement in CNC machine tools results, to a large extent, from the universality of various types of sensors, as well as measurement and diagnostic systems used in their design, which are among key integral elements of the automation system, allowing the control of many different parameters from the perspective of reliability and production process optimization alike. In modern manufacturing systems, monitoring based on the use of appropriate sensors is of key importance.

Solutions to the above problem can be found with machine learning (ML) methods. Within artificial intelligence [3, 26, 72, 129, 135, 137, 138], ML has emerged as a powerful tool for developing intelligent predictive algorithms in many applications. ML approaches have the ability to handle high-dimensional and multivariate data, as well as to extract hidden relationships within data in complex and dynamic environments such as industrial. Therefore, ML provides powerful predictive approaches for PdM applications, too. However, the performance of these applications depends on the appropriate choice of an ML technique.

The problem under consideration [83] boils down to effective identification of cutter condition via developing a suitable diagnostic and predictive algorithm. This constitutes a classical classification problem in the machine learning theory, the solution of which requires modelling an efficient kernel of the classifier (Fig. 9) used in SVM, which will then enable effective prediction and classification at the same time, based on an analysis of current measurements of acoustic signal and torques (Fig. 10).

Due to its high accuracy, SVM is another widely used and known ML method for performing classification and regression tasks [26, 58, 87, 113, 115]. SVM is characterized, among others, by high precision in the separation of different classes of data, which makes it capable of determining the best point for separating classes of data. SVM is a set of supervised learning methods that perform regression analysis and pattern recognition.

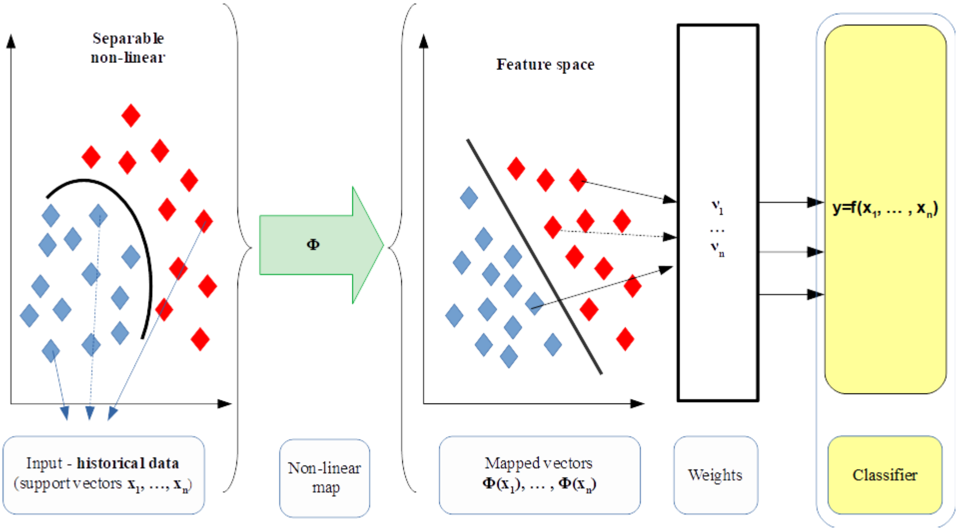


Fig. 9. SVM kernel model [83]

Therefore, in the analysed case, sensor data management involves the following tasks:

- using SVM to construct a classifier for cutter condition assessment,
- investigating the effect of acoustic signal correlation displacement length on the diagnostic error and the number of support vectors,
- developing a RUL prediction method.

By knowing the RUL value, engineers can schedule maintenance and thus avoid unplanned downtime. In addition, they can also optimize operational efficiency. As a result, a reliable RUL estimation is one of the priorities in predictive maintenance. In addition, predictive models with simulation [8] can be integrated into process realization systems in order to improve productivity and enhance product quality. Predictive performance models can also be effectively used in adaptive control for machining processes to reduce or eliminate error approaches.

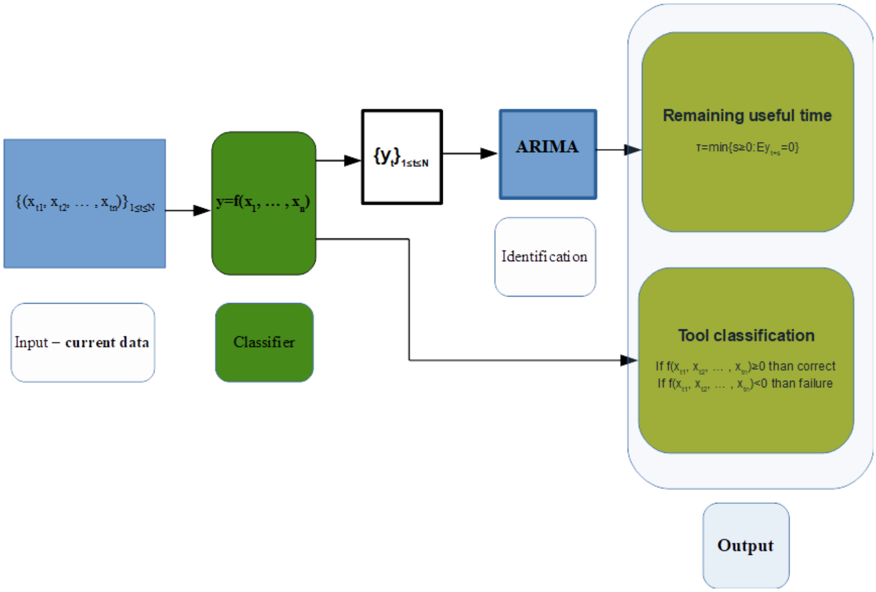


Fig. 10. RUL prediction and tool condition classification [83]

Time series describing the torque behaviour for the sharp and blunt tools, obtained with a three-axis sensor during the milling of thin-walled aircraft engine components, were modelled with the use of ARIMA models. Examples of torque signals are shown in Fig. 11. Significant differences between the correlation distributions of the sharp and blunt tools were verified by the Kolmogorov–Smirnov test.

Analyses [83] demonstrate that tool condition assessment can be made based on the calculated correlation values for acoustic signals and ARIMA model coefficients for torques. The task of tool wear assessment consists in devising a tool condition classifier and determining the expected number of cycles until failure. Such classifier was constructed with the use of SVM. A current measurement distance from the hyperplane was analysed for RUL estimation. The designed classifier for tool condition assessment uses a modification of the standard Gaussian kernel, based on acoustic signal autocorrelation.

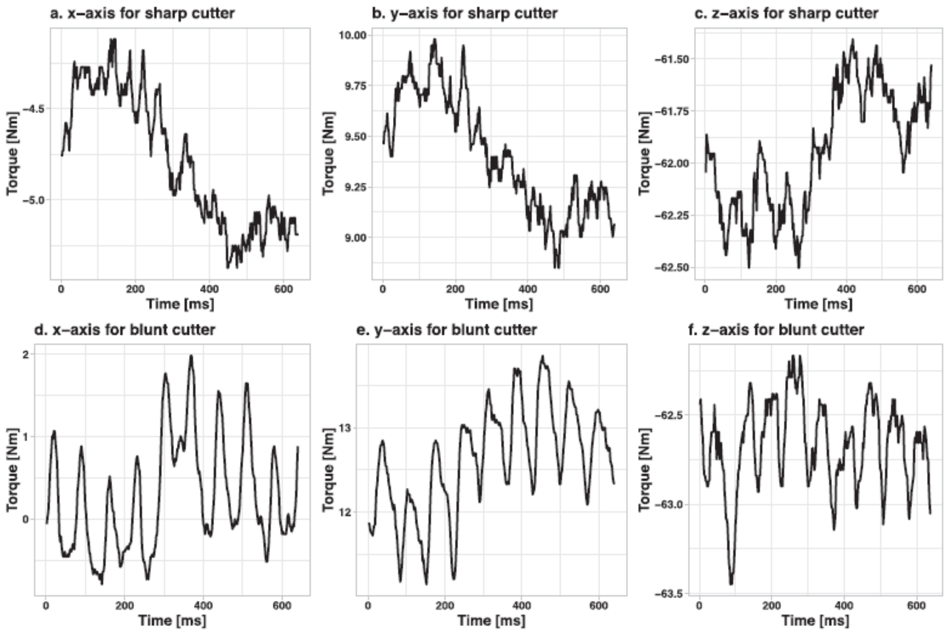


Fig. 11. Torques measured during milling of an aircraft engine thin-walled element [83]

A relationship between *p* (error probability), *n* (number of support vectors), and *m* (number of lags taken into account for the sequences of autocorrelation function values of acoustic signals) is presented in Fig. 12.

As it can be observed in Fig. 12, error probability decreases with increasing the number of lags m for the SVM with Gaussian and linear kernels. An increase in the number of m does not always result in a lower number n of support vectors. At a fixed lag number for Gaussian kernels, a lower probability of error in cutter condition assessment can be obtained.

Additionally, the classifier design based on Gaussian kernels requires the use of a smaller number of support vectors than is the case with linear kernels. To improve the quality of cutter condition assessment results, the classifier design also included the time series describing torque parameters, $M_x$, $M_y$ and $M_z$. Also, to reduce error probability, additional information from the torque data obtained with the chuck-mounted sensor was used and a modified kernel for classifier design was proposed.

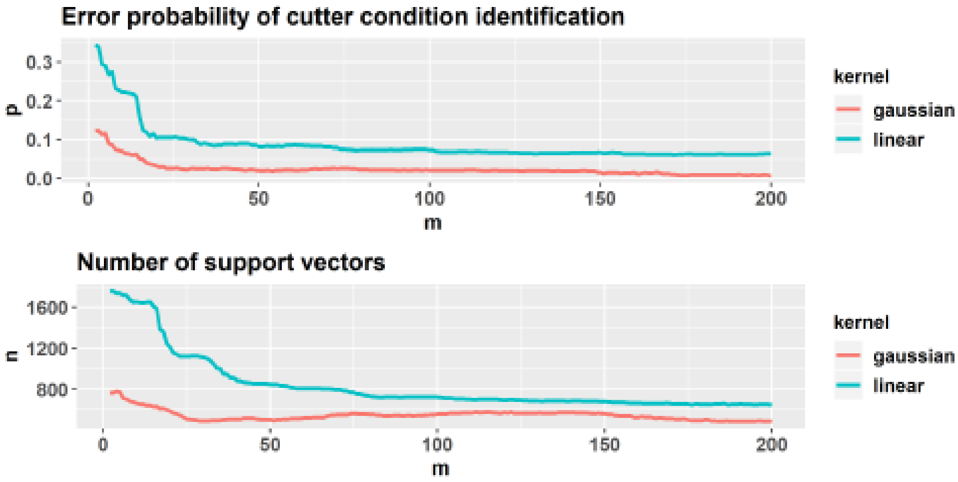


Fig. 12. Comparison of error probability and support vector number for kernels used in SVM [83]

A direct analysis of standard measurement data is not sufficient to assess tool condition. RUL prediction is unfeasible here, either. However, the results demonstrate that the use of historical data for the SVM classifier development followed by identification and prediction based on current monitoring data makes it possible to obtain the required information not only with lower error, but also at lower costs and in a shorter time than before. Prognostics deals with predicting future behaviour, and – as a result – there exist several sources of uncertainty affecting such prediction. Consequently, it is usually rarely feasible to obtain a precise RUL estimate. In fact, it makes no sense to make such predictions without first computing uncertainty associated with the RUL estimate. Researchers have developed different types of approach for quantifying uncertainty associated with RUL prediction and prognostics in general. The achieved prediction accuracy shows that this problem does not occur in the proposed method for online prognostics [83].

The proposed method was verified using data from the CNC machine monitoring system described in Chapter 4.2. The objective of RUL classification and prediction was to prevent the manufacturing of details that fall short of quality requirements when the process is performed with the use of a blunt tool. The proposed method is universal and can be effectively used as a mathematical model for processing CMMS-generated data in any manufacturing process. The presented kernel is a linear combination of Gaussian kernels. As a result of this modification, the cutter condition can be assessed with a relatively low error probability (Table 8). Additionally, Fig. 13 shows an example of cutter condition assessment and RUL prediction.

Table 8. Prediction error probability and the number of support vectors for the Gaussian and modified kernels [83]

| Kernel | | m=20 | m=30 | m=40 |
|---|---|---|---|---|
| Gaussian | prediction error<br>number of support vectors | 0.032<br>566 | 0.027<br>512 | 0.026<br>507 |
| modified $\lambda$=0.5, $\varphi$=0.3 | prediction error<br>number of support vectors | 0.025<br>552 | 0.023<br>513 | 0.021<br>509 |

As described in [99], most existing approaches for tool (which is a critical component in manufacturing processes such as grinding, milling, turning) RUL predictions are based on truncation and failure data. However, it is difficult to acquire these data for a new type of tool or for a similar tool that has just launched.

A novel tool RUL prediction method was therefore proposed in [99] based on an adaptive time window and deep bidirectional long short-term memory neural network. In this method, the wear factor is first determined by the current feature threshold ratio. Then, time windows are constructed to trace tool conditions well, and adjacent time windows are adaptively compressed while accounting for any outliers.

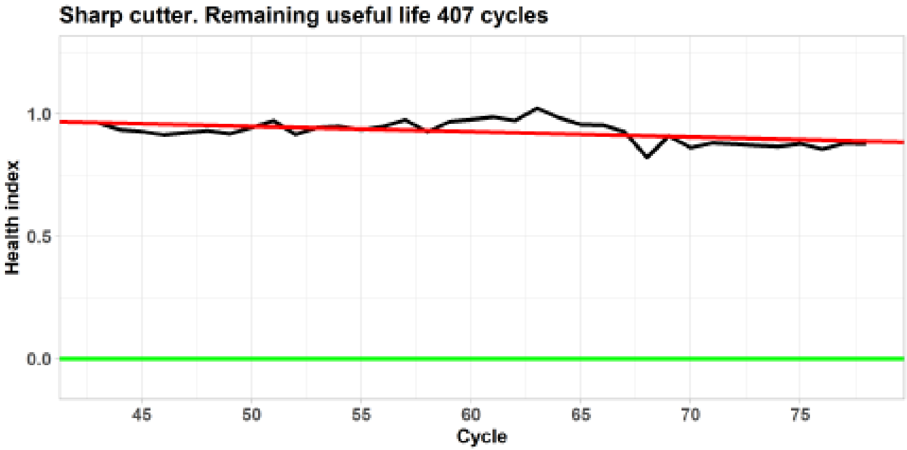


Fig. 13. Remaining useful life prediction for a sharp cutter [83]: red line – trend of health index, green line – demarcation between sharp and blunt cutter assessment

On this basis, a deep bidirectional long short-term memory is trained to capture and discover meaningful features using both forward and backward paths. Finally, multi-step ahead rolling prediction is performed to obtain RUL results. The presented results [99] show that with this method it is possible to predict the tool RUL. However, this algorithm is quite sensitive to changes in tool working conditions. The mean absolute error and root mean square error of the method proposed by [99] were 0.1130 and 0.1592. They are much higher than prediction errors achieved in this study.

# 5. Machine tool and its digital twin

## 5.1. Digital Twin: capabilities and applications

Advances in industrial digitization and ICT technologies have led to the creation of a new era known as Industry 4.0, in which Maintenance 4.0 is an emerging subject area. No less important are rapid advancements in cyber-physical systems, drawing attention to the digital twin concept and its great capability to realize Maintenance 4.0 by simulating real-time working conditions and performing intelligent decision-making. The DT concept is also an effective solution for the implementation of smart predictive maintenance approach as an enabling method of CPS, one that contains a physical model combined with real-time data filled with historical data about the process and wear. Digital twin can provide reliable data-to-knowledge transfer through intelligent data mining, which ensures high mapping faithfulness.

In its original form, the digital twin concept was described [86] as a digital informational construct about a physical system. It was created as an entity on its own and linked with a physical system. The digital representation should optimally include all information concerning the system asset that could be potentially obtained from its thorough inspection in the real world. A definition that better reflects the relationship between physical and digital objects was proposed in [48], according to which DT is an integrated multi-physics, multi-scale, probabilistic simulation of a complex product, one that uses the best available physical model, sensor updates, etc., to mirror the life of its corresponding twin. Regarding this definition, one should point out the role of measurements in real-time (sensors) and data processing with an adequate physical model enabling real-time simulation of the object or its constructional elements.

DT is also described in [101] as an effective tool to fulfil the requirements of smart manufacturing, designed as a digital duplication of entities with real-time two-way communication enabled between the cyber and physical space. The above clearly demonstrates the significance of data acquisition, data transmission, and data processing for DT, which is a carrier of model and data, one that can realize physical mapping in virtual space first and then bridge the physical and digital worlds. DT can integrate physical and virtual data throughout a product lifecycle, which results in a great volume of data that must be processed by advanced analytics. Results obtained from processed data can be used to improve the performance of a given product or process in the physical space.

The growing interest in digital twin capabilities and potential applications results from the fact that DT is known as a key enabler of digital transformation. Taking into consideration data collected from measurement systems as time series, as well as problems associated with the best possible data transformation model design, one can appreciate the importance of aspects explored in this monograph. Both collecting and analysing large amounts of manufacturing or maintenance data to find rules and knowledge have become the key to smart manufacturing and smart maintenance.

As mentioned by Lu et al. [107], data gathered from various sources to construct a digital twin will be regarded as big data. The efficient processing of big data collected from the physical space is one of the pillars of DT development. In addition, the data flow between an existing physical object and its digital twin has to be fully integrated in both directions, with DT playing the role of a control system for the physical object. In this way, virtual models could be capable of understanding the state of physical entities through sensing data, so as to predict, estimate, and analyse dynamic changes. At the same time, the physical objects would respond to the changes according to a simulation-optimized scheme. Through the cyber-physical closed loop [69, 131], digital twin could help optimize the entire manufacturing process.

The Digital Twin concept also offers several opportunities for manufacturing [7, 21, 28, 31, 88, 103, 109, 144], including maintenance, where it can be used for [86, 147]:

- inducing condition changes on upstream and downstream processes of production system identification,
- identifying and evaluating anticipatory maintenance measures,
- assessing machine condition based on descriptive methods and machine learning algorithms,
- integrating, managing and analysing machinery or process data during different stages of machine life cycle to handle data or information more efficiently, and thus to achieve better transparency of the machine's health condition.

It was pointed out [86] that the development of DT is still at its infancy, as the literature on the subject predominantly contains concept papers without concrete case studies. However, some applied case studies already exist, especially at lower levels of integration such as some types of computer-aided engineering (CAE) software. Among others, the DT-type software is included in the Siemens PLM/PDM portfolio. For example, Sinumerik One [179] is promoted as an advanced simulation tool enabling virtual analysis with advanced analytics and visualization of new machine tools during their design and optimization. Despite being an advanced CAD support tool, it does not meet the prime requirement imposed by the above-mentioned DT definition, namely that the flow of data between an existing physical object and its digital twin is fully integrated in both directions, with the DT playing the role of a system controlling the physical object. Another DT-type software offered by Siemens is Sinumerik Edge [180], which is described as a specification of the Siemens Industrial Edge platform. It provides access to all available machine data, with up to 1000 values per second per variable. The intelligent processing of these data enables optimization of process quality, increased machine availability, and higher machine productivity. The safe and scalable platform extends machine tools by additional functions and ensures the smooth operation of production. In this solution, the user himself decides which data are to be collected and when, i.e. for selected work steps or above certain thresholds. Recorded data can be imported into the shop floor management application Analyze MyWorkpiece/Toolpath for visualization and analysis. The need for user action and the use of relatively simple process improvement statistical methods make it an interesting solution for industrial practitioners. Nevertheless, at the current stage of its development, the solution cannot be classified as an advanced DT tool for the smart industry.

As mentioned in [119], DT has become one of the hottest topics in today's manufacturing because it promises improved innovation and design, visually enhanced collaboration, and ongoing operation of connected products and assets. It provides live, or near real-time, information and insights for manufacturers and asset operators to proactively improve, optimise, and transform businesses using emerging technologies such as IoT, big data, edge computing, machine learning, and predictive analytics. A DT does not have to be an accurate copy or realistic representation of the product or physical asset, it is rather an applicable abstraction reflecting valuable data; hence, it should be developed based on its being fit for the purpose. Unlike a traditional simulation, DT has the capability for determining a preventive maintenance schedule, understanding how the physical twin performs, observing system performance, promoting traceability, facilitating refinement of assumptions, enabling maintainers to troubleshoot, and combining IoT data with the physical system. The value of DT is highly related to its capability to utilise data generated during different stages of the product life cycle, from its design to the disposal phase [119].

Recent studies on the application of DT in manufacturing focus on production planning and control, as it is the main data-sink within a production system that ties everything together. As summarized in [12, 68, 119, 139, 153], the implementation of DT still poses numerous challenges, including a lack of detailed methodology and standards, as well as difficulties with collecting and storing large amounts of data, developing data acquisition systems, synchronisation, and modelling complex systems, not to mention a lack of awareness and corporate resistance to adopting the DT technology, together with difficulties in constructing, understanding, controlling, and simulating real-time changes in the system. High-fidelity models are required to simulate and test a given product or process in a virtual environment by reducing development time and cost.

Several gaps regarding the use of DT in three phases of industrial operation were identified in [119], the filling of which could benefit further research in the field:

- More methods should be developed to implement DT based on current knowledge.
- Benchmark studies between models developed for the same sector are needed to compare and identify best practices.
- The focus on after-sales remains limited, and more research is needed to support this operational phase. Therefore, there is a demand for more studies investigating the use of the DT technology in the area of after-sales services.
- The literature review reveals that the potential use of DT to support the production and supply chain sector has not yet been studied exhaustively. Therefore, future research is necessary on the methodology and implementation of DT in this area, with special emphasis on the food supply chain.

## 5.2. Smart maintenance with digital twin

With the rapid advancement in cyber-physical systems, the Digital Twin concept constitutes an effective solution for implementing the predictive maintenance approach as an enabling method for CPS [60, 73, 110, 153], one that contains a physical model, real-time sensing, and historical running data.

Predictive maintenance is considered as the most crucial potential application for digital twin. Among others, it can be used for [103]:

- Monitoring condition to analyse the deviations between collected data and expected values, since digital twin can interpret collected

data from different perspectives. Comparisons between digital twin simulated data and collected data can help determine the failure mode. Digital twin provides a high-fidelity accurate model and keeps updating through the product lifecycle, thus it can reproduce the current state of a physical object in virtual space with less data.
- Predicting performance, as digital twin can help optimize and predict product performance.

Barricelli et al. [12] also point out that DT can be successfully applied for predictive analytics to predict future statuses and changes such as failures in the product lifecycle, including its reliability aspects. DT uses results of descriptive and predictive techniques as an input for prescriptive analytics. Apart from applying predictive and prescriptive algorithms, the DT codes compute prescriptions and optimization schema by exploiting proper ontologies and high-dimensional data-coding techniques. As a result, feedback is sent both to the physical twin and to other DTs in the whole environment. Finally, DT provides modelling and simulation applications for representing – in a realistic and natural way – both the current status of the physical twin and different what-if scenarios [12]. In a detailed state-of-the-art analysis of DT applications in the production industry, Tao et al. [150, 152, 153] conclude that the DT-driven PHM shows great advantages over the traditional PHM methods in terms of four aspects: model, data, interaction, and decision making, which stems from the fact that DT:

- holistically merges physical data and virtual data, real-time data and historical data, as well as ensures data fusion;
- connects the physical and virtual spaces; as a result, not only can the physical entity be better controlled, but the virtual model can also be progressively optimized and upgraded.

Several authors [15, 123, 126, 151, 153, 164, 166, 167, 175, 176] stress some limitations of the current research on PHM. For instance, present applications mainly focus on the high-value equipment, which limits the broader applicability of DTs. Furthermore, not only are DTs useful for fault diagnosis and lifetime prediction, but they are also applicable for equipment maintenance and repair. One of the challenges related to DT implementation concerns effective realisation of cyber–physical fusion. Cyber–physical fusion involves the use of many technologies such as data acquisition, data transmission, data mining, and collaborative control. Many issues must be addressed in order to realize the cyber–physical fusion for DTs.

Lim et al. [101] additionally suggest that the incorporation of big data analytics into DT will provide more insight, resulting in better decision-making support, while improvements in DT simulations will enable better monitoring and enhanced transparency during processes. Unfortunately, as pointed out by Liu et al. [103], previous studies offer hardly any comprehensive and in-depth analyses of digital twin from the perspective of concepts, technologies, and industrial applications. There remain numerous technical challenges to respond to in order to implement digital twin applications. Even in the case of high-value equipment, data is still insufficient, because some data cannot be measured in the actual situation.

Therefore, further research combining data processing, DT, CPS, and smart maintenance is necessary. According to He et al. [52], manufacturing systems that can monitor physical processes can also create digital twins in the physical world, receive real-time information from the physical world for simulation analysis, and make informed decisions through real-time communication and collaboration with humans. The combination of digital twin and intelligent manufacturing will make manufacturing smarter and more efficient, especially when creating smart maintenance.

By reading real-time parameters from the sensors or control systems, a digital monitoring model is built to analyse the state of products by an adequate data processing method and issue an early warning in case of any potential failure. This is one potential DT scenario, but not the only one, of a digital twin applied to online real-time product inspection, equipment fault diagnosis or RUL prediction. Given its potential to generate accurate data from physical and virtual spaces and from real-time data and condition-based monitoring, the digital twin concept is the future technology for predicting failure and remaining useful life (Fig. 14).

Shao et al. [139] suggest that digital twins should only collect data that are relevant for the use case of interest rather than all available data from the physical system. Similarly, there should be not just one digital twin of a physical system but a variety of digital twins depending on their use.

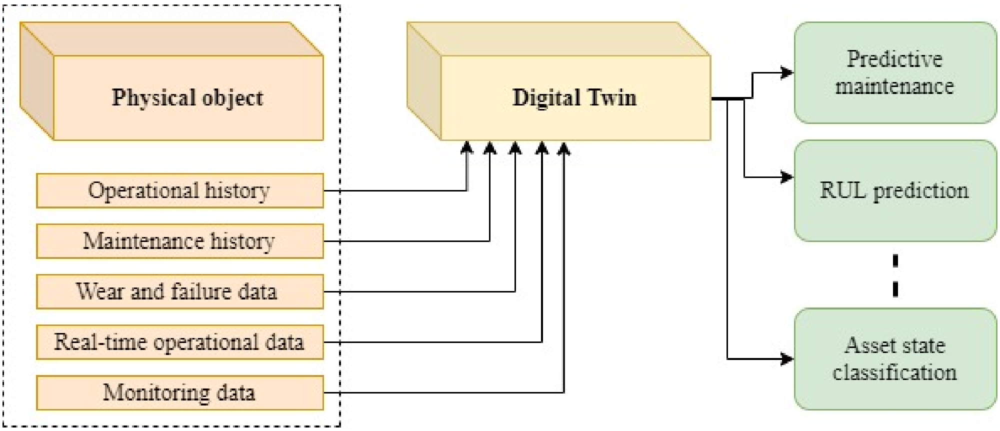


Fig. 14. Relationship between digital twin, physical object, and reliability aspects

The creation of a digital twin requires good understanding of its scope and constraints in its specific applications, among which the machine tool is considered to be one of the most challenging.

## 5.3. Data-driven digital twin of a cutting tool

The problem of sustainability has become an important topic, hence combining sustainable concepts with intelligent manufacturing to achieve sustainable intelligent manufacturing is a future research direction of paramount significance [52, 121]. Sustainable intelligent manufacturing technologies could reduce emissions in the life cycle of products, thereby satisfying the requirements of both intelligent manufacturing and comprehensive sustainability from the perspective of environmental, economic, and social aspects. The same also refers to CNCMT and a cutting tool as its key element. According to [102], a cyber-physical machine tool represents a new generation of complete cyber-physical system-based machine tools that thoroughly integrate machine tool and machining process with computation and networking. Compared to other existing systems, CPMTs offer a higher level of connectivity, intelligence, and autonomy. The core of any cyber-physical machine tool lies in the digital twin of a physical object.

In light of the above, as a natural consequence and continuation of previous studies, the concept of a data-driven DT model for a cutting tool is proposed, reflecting actual working conditions. An effective DT of a cutting tool should use data from different sensors (multi-sensor system) for tool wear identification, condition classification, and RUL prediction. To that end, adequate models must be selected with different data processing methodology and efficiency (as presented in Chapters 4.3.2-4.3.6), which is the fundamental element of the proposed multi-criteria decision support model (see Chapter 3.3). A database included in the multi-criteria decision support model serves for storing real-time and historical data gathered from the CNCMT multi-sensor system. The knowledge base contains a set of knowledge learned through relevant data processing algorithms. Rules for the knowledge base are refined from historical maintenance data and are provided for inference and executive actions. Expertise and domain knowledge are expressed by rules in the knowledge base. Through data processing model selection, further data analysis, and decision-making driven by machine learning algorithms, the digital twin model of a machine tool will provide several intelligent services such as design improvement, running process optimization, fault prediction, and accurate maintenance scheduling. Problems relating to big data and data reduction method selection must also be taken into consideration (see the discussion in Chapters 2.2 and 4.3.1). The proposed approach meets several of the current research challenges that were also discussed in Chapter 1 and Chapter 2. As it was mentioned in [102], due to the complex structure and diverse functions of CNC machine tools, this modelling approach poses a great challenge.

Since the machine tool is a complex electro-mechanical and hydraulic integrated system [110] consisting of various parts and components that are interdependent and interacting with each other, various kinds of faults may occur concurrently at different parts or subsystems. The performance of machine tool components and parts undergoes implicit degradation, which means that it is a time-varying process, too. This will undoubtedly give rise to unpredictable faults if troubleshooting is not timely. Therefore, predictive maintenance is necessary to avoid faults and improve reliability of this type of manufacturing machine. With the use of DT, a one-to-one real-time mapping model of these physical devices based on physical degradation has to be prepared and analysed. Following the multi-physics model of MT design, real-time data acquisition, and machine tool implementation, a DT for MT will have to come to reality (Fig. 15).

Through digital twin simulation, more accurate data about the condition of any part of a system or machine can be obtained from the equipment model. This offers a possibility of more faithful modelling, object reproduction, and, finally, reliable predictive maintenance. The new knowledge about data processing can help reduce the number and type of physical sensors required for monitoring or diagnostic system construction.

Data-driven and model-based methods should be combined to produce an accurate prediction result. Different types of sensors on the physical CNCMT are required to provide data support for the data-driven modelling and data processing methods, and thus to predict the remaining useful life of a cutting tool. In data-driven methods, the historical sensing data must go through several steps, including noise reduction, data pre-processing, feature extraction, and condition recognition, to finally become applicable for prediction (Fig. 16).

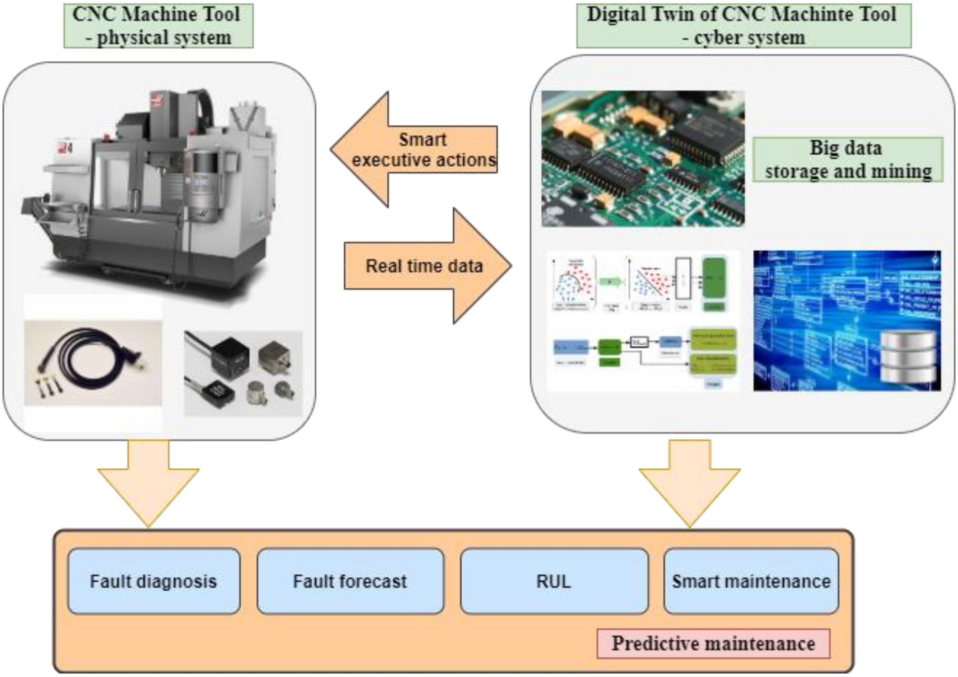


Fig. 15. Machine tool and its digital twin for predictive maintenance

The RUL predicted by data-driven methods will be used as system observation values of MT. The system's state-space model and the simulated system's inner values will make it possible to predict the condition of MT via prior knowledge. Then, the system observation values will be used to revise the values of the predicted condition. The physical object transfers manufacturing information and fault data from the physical space to DT. The database stores the real-time and historical data gathered from MT. The knowledge base contains both knowledge and a set of rules learned through relevant machine learning methods.

Rules for the knowledge base are refined from history maintenance data and are provided for inference. Through analysis and decision-making driven by e.g. machine learning algorithms, DT is able to provide several intelligent services such as design improvement, technological process optimization, fault prediction, or accurate scheduled maintenance.

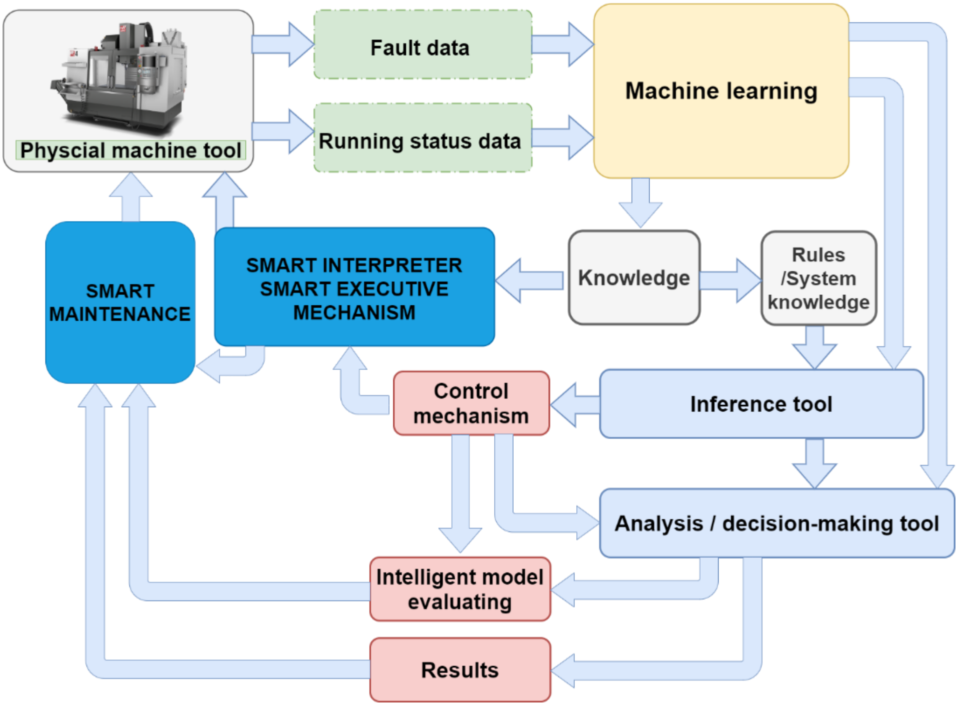


Fig. 16. Machine learning and digital twin base algorithm for smart maintenance

Therefore, DT will achieve goals such as precise simulation, self-sensing, self-adjustment, self-prediction, and self-assessment. In effect, when a new fault occurs, the running system data and fault data are mapped to a DT model. Through the analysis and inference of the algorithm model with relevant machine learning methods, smart fault reasoning and smart decisions will be provided.

To verify the accuracy of the developed machine learning and DT base algorithm for smart technological infrastructure in a sustainable environment, as well as its execution action correctness and effectiveness, adequate final tests must be performed. Obtained results will allow for the final tuning of the algorithm. All this will be incorporated into the simulation models (Fig. 17).

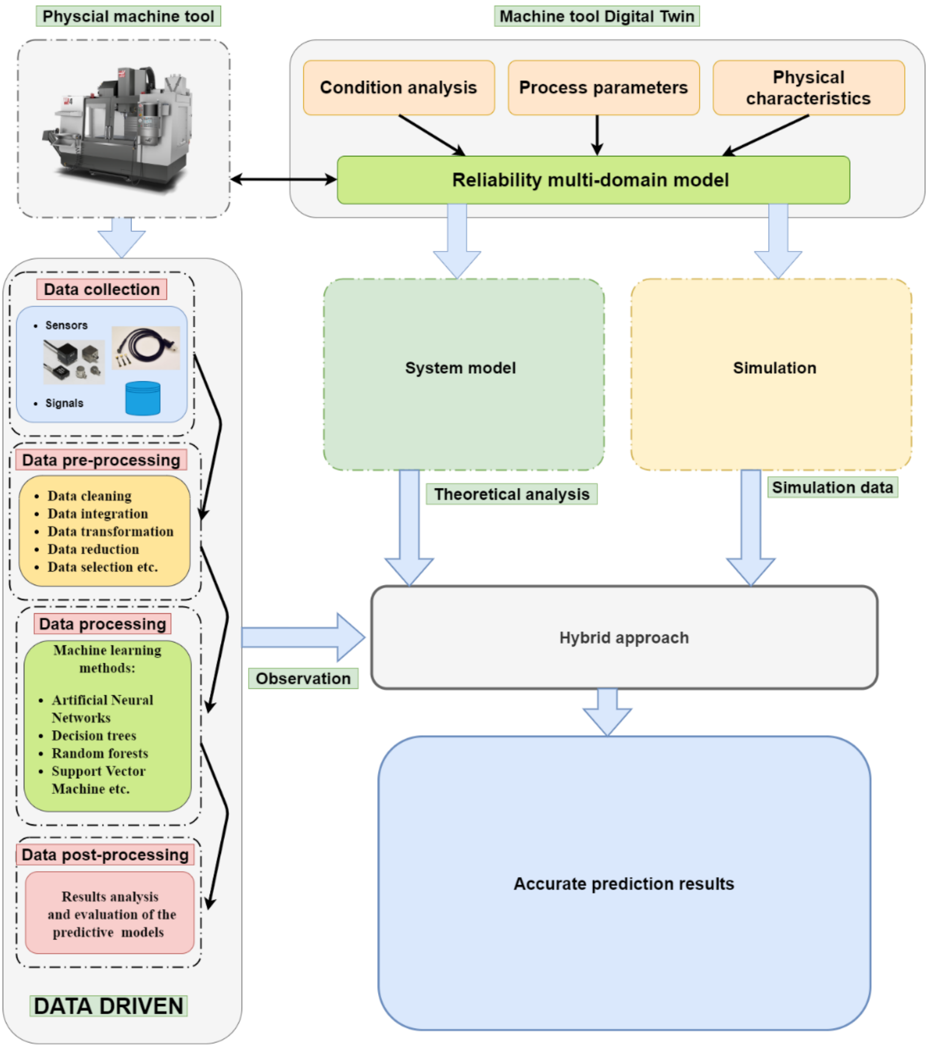


Fig. 17. Predictive maintenance framework based on machine tool and its digital twin

Predictive maintenance will be considered as a machine learning and DT-driven decision-making strategy enabling the real-time diagnosis of impending failures and the prognosis of future equipment condition, where the decision to perform maintenance is reached by observing the state of a system and its components. According to [3-5], previous studies offer no comprehensive and in-depth analysis of DT in terms of concepts, technologies, and industrial applications. In addition to that, the sustainability of production assets can be improved by the development of new data analytic methodologies and tools for production and maintenance schedule optimisation, based on the current operating condition and residual life estimation.

One can pinpoint several research needs (lacks) concerning successful DT implementation in smart maintenance:

- Data processing automation and expert system design require the use of effective analytical tools. Moreover, the transition from data to knowledge, as well as knowledge-based executive actions without human action, requires developing new analytical tools and solving a number of big data problems.
- Systematic deployment of CPS systems, wherein information from several related perspectives is exchanged, requires advanced information analytics. This means that for effective data modelling under Industry 4.0 a number of new capabilities, concepts, methods and algorithms, models and tools must be developed.
- Current studies on technical infrastructure reliability and time series modelling feasibility predominantly include fragmentary analyses of specific areas and are usually limited to minor and academic problems. As a result, they are inadequate to the needs of practitioners in the field.
- In addition to the above, there is a need to create methods and tools that will not only collect data automatically but will also select the most relevant data and use appropriate analyses to extract knowledge therefrom.

- At present, technical infrastructure reliability assessment is usually based on predictions that do not take qualitative features, sustainability or non-technical aspects into account.

In light of the above, research problems that require solving in the nearest future can be summarized as the following set of questions to be answered:

- How can time series data collected during operation be used with the support of DT to improve the availability and sustainability of production assets?
- How can useful information be extracted with high certainty from the amount of time series data gathered from different sensors, measuring instruments, and other sources such as digital twins?
- How can heterogeneous formats of data be processed in potential degradation and failure identification?
- How can real-time accuracy of remaining useful life (RUL) prediction be improved and verified?
- How can smart maintenance be scheduled and adjusted instantly to optimize the lifecycle availability and sustainability of production systems?
- How can new knowledge acquired from improved time-series modelling be used to create a smart maintenance decision process?

## References


1. Accorsi R, Manzini R, Pascarella P et al. Data Mining and Machine Learning for Condition-based Maintenance. Procedia Manufacturing 2017; 11: 1153–1161.
2. Ahmad R, Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering 2012; 63(1): 135–149.
3. Alzubi J, Nayyar A, Kumar A. Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series 2018; 1142(1): 012012.
4. Amruthnath N, Gupta T. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. 2018 5th International Conference on Industrial Engineering and Applications, ICIEA 2018, Institute of Electrical and Electronics Engineers Inc.: 2018: 355–361.
5. Amruthnath N, Gupta T. Fault class prediction in unsupervised learning using model-based clustering approach. 2018 International Conference on Information and Computer Technologies (ICICT), IEEE: 2018: 5–12.
6. Antosz K, Mazurkiewicz D, Kozłowski E, Sęp J, Żabiński T. Machining Process Time Series Data Analysis with a Decision Support Tool. In: Machado J., Soares F., Trojanowska J., Ottaviano E. (eds) Innovations in Mechanical Engineering. icieng 2021. Lecture Notes in Mechanical Engineering. Springer, Cham 2022, https://doi.org/10.1007/978-3-030-79165-0_2.
7. Armendia M, Alzaga A, Peysson F, Euhus D. Twin-Control Approach. Twin-Control, Cham, Springer International Publishing: 2019: 23–38.

8. Arrazola P J, Özel T, Umbrello D et al. Recent advances in modelling of metal machining processes. CIRP Annals 2013; 62(2): 695–718.
9. Baglee D, Marttonen-Arola S, Galar D. The need for Big Data collection and analyses to support the development of an advanced maintenance strategy Examination of big data analytics for manufacturing View project DIMECC S4 Fleet P3, View project. 2015.
10. Bao J, Guo D, Li J, Zhang J. The modelling and operations for the digital twin in the context of manufacturing. Enterprise Information Systems 2019; 13(4): 534–556.
11. Baraldi P, Di Maio F, Al-Dahidi S et al. Prediction of industrial equipment Remaining Useful Life by fuzzy similarity and belief function theory. Expert Systems with Applications 2017; 83: 226–241.
12. Barricelli B R, Casiraghi E, Fogli D. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. IEEE Access 2019; 7: 167653–167671.
13. Bokrantz J, Skoogh A, Berlin C et al. Smart Maintenance: an empirically grounded conceptualization. International Journal of Production Economics 2020, https://doi.org/10.1016/j.ijpe.2019.107534.
14. Bokrantz J, Skoogh A, Berlin C, Stahre J. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. International Journal of Production Economics 2017; 191: 154–169.
15. Booyse W, Wilke D N, Heyns S. Deep digital twins for detection, diagnostics and prognostics. Mechanical Systems and Signal Processing 2020; 140: 106612.
16. Borgi T, Hidri A, Neef B, Naceur M S. Data analytics for predictive maintenance of industrial robots. 2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET), IEEE: 2017: 412–417.

17. Borucka A, Kozlowski E, Oleszczuk P, Mazurkiewicz D. The Use of Mathematical Models Describing the Spread of Covid-19 in Strategic State Security Management. European Research Studies Journal 2020; XXIII (Special Issue 3): 82–98.
18. Borucka A, Wiśniowski P, Mazurkiewicz D, Świderski A. Laboratory measurements of vehicle exhaust emissions in conditions reproducing real traffic. Measurement 2021; 174: 108998.
19. Botkina D, Hedlind M, Olsson B et al. Digital Twin of a Cutting Tool. Procedia CIRP, Elsevier B.V.: 2018; 72: 215–218.
20. Bousdekis A, Lepenioti K, Apostolou D, Mentzas G. Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. IFAC-PapersOnLine 2019; 52(13): 607–612.
21. Boyes H, Hallaq B, Cunningham J, Watson T. The industrial internet of things (IIoT): An analysis framework. Computers in Industry 2018; 101: 1–12.
22. Braaksma J, Tiddens W. Selecting Suitable Candidates for Predictive Maintenance. International Journal of Prognostics and Health Management 2018.
23. Breiman L, Friedman J H, Olshen R A, Stone C J. Classification and Regression Trees. New York, Chapman and Hall: 1984.
24. Bumblauskas D, Gemmill D, Igou A, Anzengruber J. Smart Maintenance Decision Support Systems (SMDSS) based on corporate big data analytics. Expert Systems with Applications 2017; 90: 303–317.

25. Cachada A, Barbosa J, Leitno P et al. Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture. 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE: 2018: 139–146.
26. Carvalho T P, Soares F A A M N, Vita R et al. A systematic literature review of machine learning methods applied to predictive maintenance. Computers and Industrial Engineering 2019; 137: 106024.
27. Celaya J R, Patil N, Saha S et al. Towards accelerated aging methodologies and health management of Power MOSFETs (Technical Brief). Annual Conference of the Prognostics and Health Management Society, PHM 2009, 2009.
28. Cheng J, Zhang H, Tao F, Juang C F. DT-II:Digital twin enhanced Industrial Internet reference framework towards smart manufacturing. Robotics and Computer-Integrated Manufacturing 2020; 62: 101881.
29. Civerchia F, Bocchino S, Salvadori C et al. Industrial Internet of Things monitoring solution for advanced predictive maintenance applications. Journal of Industrial Information Integration 2017; 7: 4–12.
30. Cooke F L. Implementing TPM in plant maintenance: Some organisational barriers. International Journal of Quality and Reliability Management 2000; 17(9): 1003–1016.
31. Culot G, Nassimbeni G, Orzes G, Sartor M. Behind the definition of Industry 4.0: Analysis and open questions. International Journal of Production Economics 2020; 226: 107617.
32. Dalenogare L S, Benitez G B, Ayala N F, Frank A G. The expected contribution of Industry 4.0 technologies for industrial performance. International Journal of Production Economics 2018; 204: 383–394.

33. Daniewski K, Kosicka E, Mazurkiewicz D. Analysis of the correctness of determination of the effectiveness of maintenance service actions. Management and Production Engineering Review 2018; 9(2): 20–25.
34. Dhillon B. Engineering Maintainability. Elsevier 1999, https://doi.org/10.1016/B978-0-88415-257-6.X5000-5.
35. Do P, Voisin A, Levrat E, Iung B. A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. Reliability Engineering and System Safety 2015; 133: 22–32.
36. Duffey R B, Saull J W. Managing and predicting technological risk and human reliability: A new learning curve theory. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2008; 222(2): 245–254.
37. Efthymiou K, Papakostas N, Mourtzis D, Chryssolouris G. On a predictive maintenance platform for production systems. Procedia CIRP 2012; 3(1): 221–226.
38. Endrenyi J, Aboresheid S, Allan R N et al. The present status of maintenance strategies and the impact of maintenance on reliability. IEEE Transactions on Power Systems 2001; 16(4): 638–646.
39. EU. The EU data protection reform and big data. Publications Office of the EU, 2016.
40. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006; 27(8): 861–874.
41. Fleischmann H, Kohl J, Franke J et al. Improving maintenance processes with distributed monitoring systems. 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), IEEE: 2016; 0: 377–382.
42. Fu T. A review on time series data mining. Engineering Applications of Artificial Intelligence 2011; 24(1): 164–181.

43. Fumagalli L, Macchi M, Colace C et al. A Smart Maintenance tool for a safe Electric Arc Furnace. IFAC-PapersOnLine 2016; 49(31): 19–24.
44. Gao R, Wang L, Teti R et al. Cloud-enabled prognosis for manufacturing. CIRP Annals - Manufacturing Technology 2015; 64(2): 749–772.
45. Gavrilov L A, Gavrilova N S. The Reliability Theory of Aging and Longevity. Journal of Theoretical Biology 2001; 213(4): 527–545.
6. Gilabert E, Fernandez S, Arnaiz A, Konde E. Simulation of predictive maintenance strategies for cost-effectiveness analysis. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 2017; 231(13): 2242–2250.
47. Girdhar P, Scheffer C. Practical Machinery Vibration Analysis and Predictive Maintenance. 2004, https://doi.org/10.1016/0301-679X(78)90097-X.
48. Glaessgen E H, Stargel D S. The digital twin paradigm for future NASA and U.S. Air force vehicles. Collection of Technical Papers AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2012, https://doi.org/10.2514/6.2012-1818.
49. Goyal D, Pabla B S. The Vibration Monitoring Methods and Signal Processing Techniques for Structural Health Monitoring: A Review. Archives of Computational Methods in Engineering 2016; 23(4): 585–594.
50. Hamilton J D. Time Series Analysis. Princeton, Princeton University Press: 1994.
51. Hashemian H M. State-of-the-art predictive maintenance techniques. IEEE Transactions on Instrumentation and Measurement, 2011; 60(1): 226–236.

52. He B, Bai K J. Digital twin-based sustainable intelligent manufacturing: a review. Advances in Manufacturing 2021; 9(1): 1–21.
53. Hedberg T D, Hartman N W, Rosche P, Fischer K. Identified research directions for using manufacturing knowledge earlier in the product life cycle. International Journal of Production Research 2017; 55(3): 819–827.
54. Helu M, Hedberg T, Barnard Feeney A. Reference architecture to integrate heterogeneous manufacturing systems for the digital thread. CIRP Journal of Manufacturing Science and Technology 2017; 19: 191–195.
55. Holgado M, Macchi M. Exploring the role of E-maintenance for value creation in service provision. 2014 International Conference on Engineering, Technology and Innovation: Engineering Responsible Innovation in Products and Services, ICE 2014, IEEE Computer Society: 2014. doi:10.1109/ICE.2014.6871586.
56. Holzinger A. Data Mining with Decision Trees: Theory and Applications. Online Information Review 2015; 39(3): 437–438.
57. Van Horenbeek A, Pintelon L. A dynamic predictive maintenance policy for complex multi-component systems. Reliability Engineering & System Safety 2013; 120: 39–50.
58. Hsu, Chih-Wei; and Lin C-J, Hsu C-W and L. A Comparison of Methods for Multiclass Support Vector Machines. IEEE Transactions on Neural Networks 2002: 1–26.
59. Hwang G, Lee J, Park J, Chang T W. Developing performance measurement system for Internet of Things and smart factory environment. International Journal of Production Research 2017; 55(9): 2590–2602.

60. Ibrahim M S, Fan J, Yung W K C et al. Machine Learning and Digital Twin Driven Diagnostics and Prognostics of Light-Emitting Diodes. Laser and Photonics Reviews 2020; 14(12): 2000254.
61. Ikeda Y, Mazurkiewicz D. Application of Neural Network Technique to Combustion Spray Dynamics Analysis. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2002; 2281: 408–425.
62. Jardine A K S, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing 2006; 20(7): 1483–1510.
63. Jasiulewicz-Kaczmarek M. Sustainable maintenance assessment model of enterprise technical infrastructure. Poznań, Wydawnictwo Politechniki Poznańskiej: 2019.
64. Jasiulewicz-Kaczmarek M, Antosz K, Wyczółkowski R et al. Application of MICMAC, Fuzzy AHP, and Fuzzy TOPSIS for Evaluation of the Maintenance Factors Affecting Sustainable Manufacturing. Energies 2021; 14(5): 1436.
65. Jasiulewicz-Kaczmarek M, Stuchly V. Maintenance in sustainable manufacturing. LogForum 2014; 10(3): 273–284.
66. Jasiulewicz-Kaczmarek M, Żywica P. The concept of maintenance sustainability performance assessment by integrating balanced scorecard with non-additive fuzzy integral. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20(4): 650–661.
67. Jasiulewicz–Kaczmarek M, Antosz K, Żywica P et al. Framework of machine criticality assessment with criteria interactions. Eksploatacja i Niezawodnosc 2021; 23(2): 207–220.

68. Jayal A D, Badurdeen F, Dillon O W, Jawahir I S. Sustainable manufacturing: Modeling and optimization challenges at the product, process and system levels. CIRP Journal of Manufacturing Science and Technology 2010; 2(3): 144–152.
69. Jiang H, Qin S, Fu J et al. How to model and implement connections between physical and virtual models for digital twin application. Journal of Manufacturing Systems 2021; 58: 36–51.
70. de Jonge B. Maintenance Optimization based on Mathematical Modeling. 2017.
71. de Jonge B, Teunter R, Tinga T. The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance. Reliability Engineering and System Safety 2017, https://doi.org/10.1016/j.ress.2016.10.002.
72. Kanawaday A, Sane A. Machine learning for predictive maintenance of industrial machines using IoT sensor data. 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), IEEE: 2017: 87–90.
73. Kans M, D. G. The Impact of Maintenance 4.0 and Big Data Analytics within Strategic Asset Management. Engineering 2017.
74. Kene A P, Choudhury S K. Analytical modeling of tool health monitoring system using multiple sensor data fusion approach in hard machining. Measurement 2019; 145: 118–129.
75. Kosicka E, Kozłowski E, Mazurkiewicz D. The use of stationary tests for analysis of monitored residual processes. Ekspolatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(4): 604–609.

76. Kosicka E, Kozłowski E, Mazurkiewicz D. Intelligent Systems of Forecasting the Failure of Machinery Park and Supporting Fulfilment of Orders of Spare Parts. Advances in Intelligent Systems and Computing, Springer Verlag: 2018; 637: 54–63.
77. Kosicka E, Mazurkiewicz D, Gola Arkadiusz. Multi-criteria decision support in maintenance of machine part. Innowacje w Zarządzaniu i Inżynierii Produkcji, monografia pod red. Ryszarda Knosali. Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, T II., 2016: 584–593.
78. Kozłowski E. Analiza i identyfikacja szeregów czasowych. Lublin, Wydawnictwo Politechniki Lubelskiej: 2015.
79. Kozłowski E, Kowalska B, Kowalski D, Mazurkiewicz D. Survival Function in the Analysis of the Factors Influencing the Reliability of Water Wells Operation. Water Resources Management 2019; 33(14): 4909–4921.
80. Kozłowski E, Kowalska B, Kowalski D, Mazurkiewicz D. Water demand forecasting by trend and harmonic analysis. Archives of Civil and Mechanical Engineering 2018; 18(1): 140–148.
81. Kozłowski E, Mazurkiewicz D, Kowalska B, Kowalski D. Application of a Multidimensional Scaling Method to Identify the Factors Influencing on Reliability of Deep Wells. Advances in Intelligent Systems and Computing, 2019; 835: 56–65.
82. Kozłowski E, Mazurkiewicz D, Sęp J, Żabiński T. The use of principal component analysis and logistic regression for cutter state identification. In: Machado J., Soares F., Trojanowska J., Ottaviano E. (eds) Innovations in Mechanical Engineering. icieng 2021. Lecture Notes in Mechanical Engineering. Springer, Cham 2022, https://doi.org/10.1007/978-3-030-78170-5_34.

83. Kozłowski E, Mazurkiewicz D, Żabiński T et al. Machining sensor data management for operation-level predictive model. Expert Systems with Applications 2020; 159: 113600.
84. Kozłowski E, Mazurkiewicz D, Żabiński T. Identyfikacja stopnia zuzycia frezu na podstawie analizy sygnału akustycznego. Inżynieria Zarządzania – Cyfryzacja Produkcji – Aktualności badawcze 1, pod red. R. Knosali, 2019: 561–571.
85. Kozłowski E, Mazurkiewicz D, Żabiński T et al. Assessment model of cutting tool condition for real-time supervision system. Eksploatacja i Niezawodnosc 2019; 21(4): 679–685.
86. Kritzinger W, Karner M, Traar G et al. Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine 2018; 51(11): 1016–1022.
87. Kroll B, Schaffranek D, Schriegel S, Niggemann O. System modeling based on machine learning for anomaly detection and predictive maintenance in industrial plants. 19th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2014, https://doi.org/10.1109/ETFA.2014.7005202.
88. Kumar A, Shankar R, Thakur L S. A big data driven sustainable manufacturing framework for condition-based maintenance prediction. Journal of Computational Science 2018; 27: 428–439.
89. Lee J, Bagheri B. Cyber-Physical Systems in Future Maintenance. Lecture Notes in Mechanical Engineering, Springer Heidelberg: 2015; 20: 299–305.
90. Lee J, Bagheri B, Jin C. Introduction to cyber manufacturing. Manufacturing Letters 2016; 8: 11–15.

91. Lee J, Bagheri B, Kao H A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters 2015; 3: 18–23.
92. Lee J, Lapira E, Bagheri B, Kao H. Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters 2013; 1(1): 38–41.
93. Lee W J, Wu H, Yun H et al. Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data. Procedia CIRP 2019; 80: 506–511.
94. Lele A. Industry 4.0. Smart Innovation, Systems and Technologies, 2019; 132: 205–215.
95. Leng J, Liu Q, Ye S et al. Digital twin-driven rapid reconfiguration of the automated manufacturing system via an open architecture model. Robotics and Computer-Integrated Manufacturing 2020; 63: 101895.
96. Leturiondo U, Salgado O, Ciani L et al. Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data. Measurement 2017; 108: 152–162.
97. Levitt J. The Handbook of Maintenance Management. 2009: 455.
98. Lewis A. Benchmarking Best Practice in Maintenance Management. 2012, https://doi.org/10.1108/14635771211218407.
99. Li H, Wang W, Li Z et al. A novel approach for predicting tool remaining useful life using limited data. Mechanical Systems and Signal Processing 2020; 143: 106832.
100. Li Z, Wang Y, Wang K S. Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario. Advances in Manufacturing 2017; 5(4): 377–387.

101. Lim K Y H, Zheng P, Chen C-H. A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspectives. Journal of Intelligent Manufacturing 2020; 31(6): 1313–1337.
102. Liu C, Hong X, Zhu Z, Xu X. Machine Tool Digital Twin: Modelling methodology and applications. Proceedings of International Conference on Computers and Industrial Engineering, CIE, 2018.
103. Liu M, Fang S, Dong H, Xu C. Review of digital twin about concepts, technologies, and industrial applications. Journal of Manufacturing Systems 2021; 58: 346–361.
104. Liu Q, Leng J, Yan D et al. Digital twin-based designing of the configuration, motion, control, and optimization model of a flow-type smart manufacturing system. Journal of Manufacturing Systems 2021; 58: 52–64.
105. Liu R, Yang B, Zio E, Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing 2018; 108: 33–47.
106. Lopez J H. The power of the ADF test. Economics Letters 1997; 57(1): 5–10.
107. Lu Y, Liu C, Wang K I K et al. Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. Robotics and Computer-Integrated Manufacturing 2020; 61: 101837.
108. Lundgren C, Bokrantz J, Skoogh A. Performance indicators for measuring the effects of Smart Maintenance. International Journal of Productivity and Performance Management 2020, https://doi.org/10.1108/IJPPM-03-2019-0129.

109. Lundgren C, Skoogh A, Bokrantz J. Quantifying the Effects of Maintenance - A Literature Review of Maintenance Models. Procedia CIRP, Elsevier: 2018; 72: 1305–1310.
110. Luo W, Hu T, Ye Y et al. A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin. Robotics and Computer-Integrated Manufacturing 2020; 65: 101974.
111. Luo W, Hu T, Zhang C, Wei Y. Digital twin for CNC machine tool: modeling and using strategy. Journal of Ambient Intelligence and Humanized Computing 2019; 10(3): 1129–1140.
112. Luo W, Hu T, Zhu W, Tao F. Digital twin modeling method for CNC machine tool. 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), IEEE: 2018: 1–4.
113. Ma Y, Guo G. Support vector machines applications. Support Vector Machines Applications 2014; 9783319023: 1–302.
114. Macchi M, Roda I, Fumagalli L. On the Advancement of Maintenance Management Towards Smart Maintenance in Manufacturing. IFIP Advances in Information and Communication Technology, Springer New York LLC: 2017; 513: 383–390.
115. Mahadevan S, Shah S L. Fault detection and diagnosis in process data using one-class support vector machines. Journal of Process Control 2009; 19(10): 1627–1639.
116. Manco G, Ritacco E, Rullo P et al. Fault detection and explanation through big data analysis on sensor streams. Expert Systems with Applications 2017; 87: 141–156.
117. Matthews B W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure 1975; 405(2): 442–451.

118. Mazurkiewicz D. Analysis of the ageing impact on the strength of the adhesive sealed joints of conveyor belts. Journal of Materials Processing Technology 2008; 208(1–3): 477–485.
119. Melesse T Y, Di Pasquale V, Riemma S. Digital Twin models in industrial operations: State-of-the-art and future research directions. IET Collaborative Intelligent Manufacturing 2021; 3(1): 37–47.
120. Merh N. Applying Predictive Analytics in a Continuous Process Industry 2019: 105–115.
121. Monostori L, Kádár B, Bauernhansl T et al. Cyber-physical systems in manufacturing. CIRP Annals 2016; 65(2): 621–641.
122. Mourtzis D, Vlachou E, Milas N, Xanthopoulos N. A Cloud-based Approach for Maintenance of Machine Tools and Equipment Based on Shop-floor Monitoring. Procedia CIRP 2016; 41: 655–660.
123. Mustakerov I, Borissova D. An intelligent approach to optimal predictive maintenance strategy defining. 2013 IEEE International Symposium on Innovations in Intelligent Systems and Applications, IEEE INISTA 2013, https://doi.org/10.1109/INISTA.2013.6577666.
124. Nakagawa T. Maintenance Theory of Reliability. London, Springer Verlag: 2005.
125. Neugebauer R, Denkena B, Wegener K. Mechatronic Systems for Machine Tools. CIRP Annals 2007; 56(2): 657–686.
126. Nguyen K T P, Medjaher K. A new dynamic predictive maintenance framework using deep learning for failure prognostics. Reliability Engineering and System Safety 2019; 188: 251–262.
127. Oksendal B. Stochastic differential Equations – An introduction with applications. London, Springer: 2000.

128. Özel T, Karpat Y. Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks. International Journal of Machine Tools and Manufacture 2005; 45(4–5): 467–479.
129. Paolanti M, Romeo L, Felicetti A et al. Machine Learning approach for Predictive Maintenance in Industry 4.0. 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), IEEE: 2018: 1–6.
130. Peres R S, Dionisio Rocha A, Leitao P, Barata J. IDARTS – Towards intelligent data analysis and real-time supervision for industry 4.0. Computers in Industry 2018; 101: 138–146.
131. Qi Q, Tao F. Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. IEEE Access 2018; 6: 3585–3593.
132. Qin J, Liu Y, Grosvenor R. A Categorical Framework of Manufacturing for Industry 4.0 and beyond. Procedia CIRP 2016; 52: 173–178.
133. Qu Y, He D, Yoon J et al. Gearbox Tooth Cut Fault Diagnostics Using Acoustic Emission and Vibration Sensors - A Comparative Study. Sensors 2014; 14(1): 1372–1393.
134. Rokach L, Maimon O. Data Mining with Decision Trees. World Scientific: 2014, https://doi.org/10.1142/9097.
135. Sakib N, Wuest T. Challenges and Opportunities of Condition-based Predictive Maintenance: A Review. Procedia CIRP 2018; 78: 267–272.
136. Sankararaman S, Goebel K. Why is the Remaining Useful Life Prediction Uncertain . Annual Conference of the Prognostics and Health Management Society 2013 2012: 1–13.
137. Schmidt B, Wang L. Predictive Maintenance: Literature Review and Future Trends. Proceedings of the 25th International Conference on Flexible Automation and Intelligent Manufacturing 2015, 1.

138. Scuro C, Sciammarella P F, Lamonaca F et al. IoT for structural health monitoring. IEEE Instrumentation and Measurement Magazine 2018; 21(6): 4–14.
139. Shao G, Helu M. Framework for a digital twin in manufacturing: Scope and requirements. Manufacturing Letters 2020; 24: 105–107.
140. Shi J, Li Y, Wang G, Li X. Health index synthetization and remaining useful life estimation for turbofan engines based on run-to-failure datasets. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 621–631.
141. Sikorska J Z, Hodkiewicz M, Ma L. Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing 2011; 25(5): 1803–1836.
142. Skoczyński W, Stembalski M, Roszkowski A et al. Sensory we współczesnych obrabiarkach sterowanych numerycznie. Mechanik 2016; 11(11): 1740–1744.
143. Sokolova M, Japkowicz N, Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. AAAI Workshop - Technical Report, Springer, Berlin, Heidelberg: 2006; WS-06-06: 1015–1021.
144. Soldatos J, Lazaro O, Cavadini F. The Digital Shopfloor: Industrial Automation in the Industry 4.0 Era. 2019, https://doi.org/10.13052/rp-9788770220408.
145. Spendla L, Kebisek M, Tanuska P, Hrcka L. Concept of predictive maintenance of production systems in accordance with industry 4.0. 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE: 2017: 000405–000410.

146. Susto G A, Beghi A. Dealing with time-series data in Predictive Maintenance problems. IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2016: 1–4.
147. Susto G A, Schirru A, Pampuri S et al. Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. IEEE Transactions on Industrial Informatics 2015; 11(3): 812–820.
148. Susto G A, Schirru A, Pampuri S et al. A predictive maintenance system for integral type faults based on support vector machines: An application to ion implantation. IEEE International Conference on Automation Science and Engineering 2013: 195–200.
149. Tail M, Yacout S, Balazinski M. Replacement time of a cutting tool subject to variable speed. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 2010; 224(3): 373–383.
150. Tao F, Anwer N, Liu A et al. Digital twin towards smart manufacturing and industry 4.0. Journal of Manufacturing Systems 2021; 58: 1–2.
151. Tao F, Cheng J, Qi Q et al. Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology 2018; 94(9–12): 3563–3576.
152. Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. Journal of Manufacturing Systems 2018; 48: 157–169.
153. Tao F, Zhang H, Liu A, Nee A Y C. Digital Twin in Industry: State-of-the-Art. IEEE Transactions on Industrial Informatics 2019; 15(4): 2405–2415.
154. Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Annals 2010; 59(2): 717–739.
155. Thompson C B. Descriptive Data Analysis. Air Medical Journal 2009; 28(2): 56–59.

156. Tsui K L, Chen N, Zhou Q et al. Prognostics and Health Management: A Review on Data Driven Approaches. Mathematical Problems in Engineering 2015; 2015: 1–17.
157. Uhm D, Jun S-H, Lee S-J. A Classification Method Using Data Reduction. International Journal of Fuzzy Logic and Intelligent Systems 2012; 12(1): 1–5.
158. Vaidya S, Ambad P, Bhosle S. Industry 4.0 – A Glimpse. Procedia Manufacturing 2018; 20: 233–238.
159. Vališ D, Mazurkiewicz D. Application of selected Levy processes for degradation modelling of long range mine belt using real-time data. Archives of Civil and Mechanical Engineering 2018; 18(4): 1430–1440.
160. Valis D, Mazurkiewicz D, Forbelska M. Modelling of a transport belt degradation using state space model. 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE 2017: 949–953.
161. Vamsi I, Sabareesh G R, Penumakala P K. Comparison of condition monitoring techniques in assessing fault severity for a wind turbine gearbox under non-stationary loading. Mechanical Systems and Signal Processing 2019; 124: 1–20.
162. Vapnik V. Statistical Learning Theory. John Wiley & Sons: 1998.
163. Wang S, Wang D, Su L et al. Towards Cyber-Physical Systems in Social Spaces: The Data Reliability Challenge. 2014 IEEE Real-Time Systems Symposium, IEEE 2014: 74–85.
164. Wu D, Jennings C, Terpenny J et al. A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. Journal of Manufacturing Science and Engineering 2017; 139(7): 1–9.

165. Wu X, Kumar V, Ross Quinlan J et al. Top 10 algorithms in data mining. Knowledge and Information Systems 2008; 14(1): 1–37.
166. Wuest T, Weimer D, Irgens C, Thoben K-D. Machine learning in manufacturing: advantages, challenges, and applications. Production and Manufacturing Research 2016; 4(1): 23–45.
167. Xu L Da, He W, Li S. Internet of things in industries: A survey. IEEE Transactions on Industrial Informatics 2014; 10(4): 2233–2243.
168. Xu L Da, Xu E L, Li L. Industry 4.0: state of the art and future trends. International Journal of Production Research 2018; 56(8): 2941–2962.
169. Xu W, Cao L. Optimal tool replacement with product quality deterioration and random tool failure. International Journal of Production Research 2015; 53(6): 1736–1745.
170. Yan J, Meng Y, Lu L, Li L. Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes, and Applications for Predictive Maintenance. IEEE Access 2017; 5: 23484–23491.
171. Żabiński T, Mączka T, Kluska J. Industrial platform for rapid prototyping of intelligent diagnostic systems. Advances in Intelligent Systems and Computing, Springer Verlag: 2017; 577: 712–721.
172. Zaretalab A, Haghighi H S, Mansour S, Sajadieh M S. Optimisation of tool replacement time in the machining process based on tool condition monitoring using the stochastic approach. International Journal of Computer Integrated Manufacturing 2019; 32(2): 159–173.
173. Zenisek J, Wolfartsberger J, Sievi C, Affenzeller M. Modeling Sensor Networks for Predictive Maintenance. International Journal of Prognostics and Health Management, 2019: 184–188.

174. Zhang A, Srivastav H, Barros A et al. Performance Analysis of Redundant Safety-instrumented Systems Considering the Imprecision of Information in Proof Tests. Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference, Singapore 2020: 3561–3568.
175. Zhao R, Yan R, Chen Z et al. Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing 2019; 115: 213–237.
176. Zio E. Reliability engineering: Old problems and new challenges. Reliability Engineering and System Safety 2009; 94(2): 125–141.
177. EN 13306:2017 Maintenance - Maintenance terminology. 2017.
178. The problem of multicollinearity. Understanding Regression Analysis, Boston, MA, Springer US: 2007: 176–180.
179. Sinumerik One. [https://new.siemens.com/global/en/products/ automation/systems/sinumerik-one.html].
180. Sinumerik Edge. [https://new.siemens.com/global/en/products/ automation/topic-areas/industrial-edge/sinumerik-edge.html].