Springer Series in Reliability Engineering

Kim Phuc Tran *Editor*

Artificial Intelligence for Smart Manufacturing

Methods, Applications, and Challenges

Springer

Springer Series in Reliability Engineering

Series Editor

Hoang Pham, Industrial and Systems Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

Today's modern systems have become increasingly complex to design and build, while the demand for reliability and cost effective development continues. Reliability is one of the most important attributes in all these systems, including aerospace applications, real-time control, medical applications, defense systems, human decision-making, and home-security products. Growing international competition has increased the need for all designers, managers, practitioners, scientists and engineers to ensure a level of reliability of their product before release at the lowest cost. The interest in reliability has been growing in recent years and this trend will continue during the next decade and beyond.

The Springer Series in Reliability Engineering publishes books, monographs and edited volumes in important areas of current theoretical research development in reliability and in areas that attempt to bridge the gap between theory and application in areas of interest to practitioners in industry, laboratories, business, and government.

Now with 100 volumes!

Indexed in Scopus and EI Compendex

Interested authors should contact the series editor, Hoang Pham, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854, USA. Email: **hopham@soe.rutgers.edu, or Anthony Doyle, Executive Editor, Springer, London. Email:anthony.doyle@springer.com.**

Kim Phuc Tran
Editor

Artificial Intelligence for Smart Manufacturing

Methods, Applications, and Challenges

Springer

Editor
Kim Phuc Tran
ENSAIT and GEMTEX
University of Lille
Roubaix, France

ISSN 1614-7839 ISSN 2196-999X (electronic)
Springer Series in Reliability Engineering
ISBN 978-3-031-30509-2 ISBN 978-3-031-30510-8 (eBook)
https://doi.org/10.1007/978-3-031-30510-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction to Smart Manufacturing with Artificial Intelligence

Kim Phuc Tran

Abstract This chapter provides an overview of Artificial Intelligence-based methods, applications, and challenges for Smart Manufacturing in Industry 5.0. We elaborate on the essential issues related to the applications and the potential of Artificial Intelligence algorithms in Smart Manufacturing. We will introduce crucial topics that will be discussed in the following chapters of the book.

Keywords Artificial intelligence · Smart manufacturing · Industry 5.0

1 Introduction

As the fourth industrial revolution has passed an early stage of development, many companies are developing intelligent systems and cutting-edge innovations of Industry 4.0 to improve productivity and quality Nguyen et al. [1], Truong et al. [2], Huong et al. [3]. Artificial intelligence (AI) emerges as a powerful technology to deploy Smart Manufacturing with performance enhancement, cost control, optimization of processes, etc. Wang et al. [4], Tran [5]. Meanwhile, the next phase of industrialization has been started to introduce, known as Industry 5.0. One of the most prominent features of Industry 5.0 is that it places humans at the center of the manufacturing process. It is now at the center of Industry 5.0 and attracts a lot of interest from governments, enterprises, and researchers for implementing Smart Manufacturing. According to a discussion in Breque et al. [6], Industry 5.0 places the well-being of the worker at the center of the production process and uses new technologies to provide prosperity beyond jobs and growth while respecting the production limits of the planet. Based on this vision, people call Industry 5.0 a human-centric solution. In the next phase of transformation, Industry 5.0 is expected to bring many more technologies as well as many challenges for Smart Manufacturing Xu et al. [7].

K. P. Tran (✉)
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France
e-mail: kim-phuc.tran@ensait.fr

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_1

This book will serve as a reference on the latest advancement of 4.0 technologies in smart manufacturing, as well as, discuss difficulties, challenges, and perspectives on important topics in Smart Manufacturing in the context of Industry 5.0 such as Product Design, Predictive Maintenance, Quality control, Digital Twin, Wearable technology, Quantum Machine Learning, etc.

2 Main Features of This Book

The key features of this book are given as follows:

1. A survey on AI-based methods, applications, and challenges for smart manufacturing in Industry 5.0 to clarify some important issues related to the applications and the potential of AI algorithms in Smart Manufacturing.
2. Quality control for Smart Manufacturing in Industry 5.0.
3. Dynamic Process Monitoring Using Machine Learning Control Charts in Smart Manufacturing.
4. Benefits of using Digital Twin for online fault diagnosis in Smart Manufacturing.
5. Fault Prediction of Papermaking Process Based on Gaussian Mixture Model and Mahalanobis Distance.
6. Multi-objective optimization of flexible flow-shop intelligent scheduling based on a hybrid intelligent algorithm.
7. Personalized pattern recommendation system for fashion product design.
8. Perspective of Efficient and Trustworthy Federated Learning-based Explainable Anomaly Detection for Smart Manufacturing.
9. Multimodal machine learning in prognostics and health management of manufacturing systems.
10. Explainable Artificial Intelligence for Cybersecurity in Smart Manufacturing.
11. Wearable technology for Smart Manufacturing.
12. The case studies presented and analyzed in several chapters, as well as, some source codes of algorithms, are also shared with readers.

3 Structure of the Book

This book uncovers fundamental principles and recent developments in Artificial Intelligence for Smart Manufacturing. The book contains 12 chapters.

In the Introductory chapter "Introduction to Artificial Intelligence for Smart Manufacturing," the book editor, Kim Phuc Tran, elaborates on the overview of Artificial Intelligence-based methods, applications, and challenges for Smart Manufacturing in Industry 5.0.

Huu Du Nguyen and Kim Phuc Tran investigate in their chapter, "Artificial Intelligence for Smart Manufacturing in Industry 5.0: Methods, Applications, and Challenges”, a survey on AI-based methods, applications, and challenges for smart manufacturing in Industry 5.0 to clarify some important issues related to the applications and the potential of AI algorithms in smart manufacturing.

Huu Du Nguyen, Phuong Hanh Tran, Thu Ha Do, and Kim Phuc Tran develop in their chapter, “Quality control for Smart Manufacturing in Industry 5.0,” a comprehensive background review of important notions and advanced techniques related to quality control for smart manufacturing such as Machine Learning, computer vision, the Internet of Things, and Artificial Intelligence.

Xiulin Xie and Peihua Qiu develop in their chapter, “Dynamic Process Monitoring Using Machine Learning Control Charts,” modified machine learning control charts using nonparametric longitudinal modeling and sequential data decorrelation algorithms.

Guojian Chen, Zhenglei He, YiMan, Jigeng Li, Mengna Hong, and Kim Phuc Tran develop in their chapter, “Fault Prediction of Papermaking Process Based on Gaussian Mixture Model and Mahalanobis Distance” a data-driven approach to predict fault in the papermaking process on the basis of correlation analysis and clustering algorithms.

Huanhuan Zhang, Zhenglei He, Yi Man, Jigeng Li, Mengna Hong, and Kim Phuc Tran develop in their chapter, “Multi-objective optimization of flexible flow-shop intelligent scheduling based on a hybrid intelligent algorithm,” the intelligent scheduling problem of a flexible flow shop and establishes a two-stage flexible flow-shop scheduling mode.

Guillaume Tartare, Cheng Chi, and Pascal Bruniaux develop in their chapter, “Personalized pattern recommendation system of men’s shirts”, a personalized garment pattern recommendation system to fit by integrating the designer’s knowledge and 3D measurement.

Thu Ha Do, Phuong Bac Ta, Kim Duc Tran, and Kim Phuc Tran develop in their chapter, “Efficient and Trustworthy Federated Learning-based Explainable Anomaly Detection,” a comprehensive perspective on federated learning-based anomaly detection as well as, future research direction.

Sagar Jose, Khanh T. P. Nguyen, and Kamal Medjaher investigate in their chapter, “Multimodal machine learning in prognostics and health management of manufacturing systems”, the possibility of exploiting additional data sources in manufacturing besides monitoring sensors, e.g. production line cameras or maintenance reports with multimodal learning for industrial prognostics and health management.

Phuong Bac Ta, Thu Ha Do, Kim Duc Tran, and Kim Phuc Tran, investigate in their chapter, “Explainable Artificial Intelligence for Cybersecurity in Smart Manufacturing”, a potential approach of Explainable Artificial Intelligence to enable Smart manufacturing in the Industrial Revolution 5.0.

Tho Nguyen, Kim Duc Tran, Ali Raza, Quoc Thong Nguyen, Huong Mai Bui, and Kim Phuc Tran investigate in their chapter, “Wearable technology for Smart Manufacturing in Industry 5.0” a survey on the applications and challenges of wearable Internet of Things devices in industrial sectors, particularly in manufacturing.

4 Conclusion

This book will introduce readers to the latest results of theoretical and applications of AI approaches with respect to the current challenges and opportunities in Smart Manufacturing in the context of Industry 5.0. The book will also provide ready-to-use algorithms for readers and practitioners to deploy Smart Manufacturing in practice. To do so, case studies will also be introduced to readers and practitioners in each chapter to readers.

References

1. Nguyen HD, Tran KP, Thomassey S, Hamad M (2021) Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. Int J Inf Manag 57:102282
2. Truong HT, Ta BP, Le QA, Nguyen DM, Le CT, Nguyen HX, Do HT, Nguyen HT, Tran KP (2022) Light-weight federated learning-based anomaly detection for time-series data in industrial control systems. Comput Ind 140:103692
3. Huong TT, Bac TP, Ha KN, Hoang NV, Hoang NX, Hung NT, Tran KP (2022) Federated learning-based explainable anomaly detection for industrial control systems. IEEE Access 10:53854–53872
4. Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: methods and applications. J Manuf Syst 48:144–156
5. Tran KP (2021) Artificial intelligence for smart manufacturing: methods and applications
6. Breque M, De Nul L, Petridis A (2021) Industry 5.0: towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Luxembourg, LU
7. Xu X, Lu Y, Vogel-Heuser B, Wang L (2021) Industry 4.0 and industry 5.0-inception, conception and perception. J Manuf Syst 61:530–535

Artificial Intelligence for Smart Manufacturing in Industry 5.0: Methods, Applications, and Challenges

Huu Du Nguyen and Kim Phuc Tran

Abstract As the fourth industrial revolution has passed an early stage of development, many companies are developing intelligent systems and cutting-edge innovations of Industry 4.0 to improve productivity and quality. Meanwhile, the next phase of industrialization has been started to introduce, known as Industry 5.0. One of the most prominent features of Industry 5.0 is that it places the well-being of humans at the center of the manufacturing process. Advanced technologies are also created to keep up with the trend of the fifth industrial revolution. Artificial intelligence (AI) algorithms have proven to play a key role in Industry 4.0. Moving to Industry 5.0, with the human-centric orientation, AI was developed in combination with human intelligence (HI), leading to the new concept of Augmented Intelligence (AuI). AI and AuI algorithms are expected to bring significant benefits for enabling smart manufacturing in Industry 5.0. In this study, we provide a survey on AI-based methods, applications, and challenges for smart manufacturing in Industry 5.0. The discussions will help to clarify some important issues related to the applications and the potential of AI algorithms in smart manufacturing.

Keywords Artificial intelligence · Smart manufacturing · Industry 5.0 · Augmented intelligence · Cyber-physical systems · Human-centric manufacturing

1 Introduction

Since the beginning of Industry 4.0, humanity has witnessed significant changes in all aspects of life, especially the manufacturing sector. By integrating many advanced technologies such as Artificial Intelligence (AI), Industrial Internet of

H. D. Nguyen
International Research Institute for Artificial Intelligence and Data Science, Dong A University, Danang, Vietnam

K. P. Tran (✉)
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France
e-mail: kim-phuc.tran@ensait.fr

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_2

Things (IIoT), Cyber-Physical Systems (CPS), Big Data, and Cloud computing, Industry 4.0 has substantially revolutionized industrial manufacturing. Automated production is raised to a new level: "smart". Instead of simply executing pre-programmed commands in automation, AI algorithms with the ability to imitate some human thought have made some stages of the production process more intelligent. Along with the interconnection of CPS and the power of cloud computing, they enable manufacturers to improve significantly productivity and product quality, maintain effective production systems, and optimize the production workflow. As a result, production efficiency will also be enhanced. A few years ago, Rojko [1] cited information on some of Industry 4.0's quantified contributions in terms of efficiency and costs in production, including a reduction of 10–30% of production costs, 10–30% of logistics costs, and 10–20% of quality management costs. With the rapid and continuous development of new techniques and the improvement of existing algorithms, it can be trusted that the productivity and efficiency of Industry today are even higher.

Despite such great contributions, there are still limitations in Industry 4.0. Several existing problems have been discussed in Nahavandi [2]. For instance, due to the aim of making manufacturing more productive, Industry 4.0 focuses purely on improving mass production and reducing the intervention of humans in the manufacturing process. This inadvertently omits or lowers the human cost resulting from the optimization of processes. Automated machines and robots can substitute for humans in many important stages of the production process. However, it also raised concerns about the mass loss of jobs among workers. Another concern is related to environmental issues and resources. Experiencing the industrial revolutions with the large-scale exploitation and use of fossil fuel resources such as gasoline, oil, coal, and gas, the earth is being seriously polluted, while the resources are gradually exhausted. Although smart manufacturing in Industry has paid more attention to reducing waste generation and the negative influences on the environment from its operation, its focus is not strong enough. There is still a need in Industry 4.0 for better technological solutions that can save the environment and increase sustainability. These problems of Industry 4.0 are among the driving forces for the introduction of Industry 5.0.

Industry 5.0 refers to the fifth industrial revolution. For some, it may still be too early to talk about Industry 5.0 because many businesses are still in the early stages of getting started with new technologies in Industry 4.0. However, the next phase of industrialization is already started. While Industry 4.0 interconnects AI, machines, processes, and systems to optimize performance, the vision of Industry 5.0 is beyond efficiency and productivity: it aims at refining the collaborative interactions between humans and machines Breque et al. [3]. Similar to Industry 4.0, Industry 5.0 is also based on advanced technologies to build a smart manufacturing (SM) platform. However, these technologies will play a different role, supporting and collaborating with people rather than superseding them. The concept of a robot is extended to a cobot. While robots operate independently and with no human involvement, cobots are designed to physically interact with people on shared factory floors. In the same way, another extension of AI algorithms is also introduced, i.e.

the Augmented Intelligence (AuI) algorithm. AuI refers to an integration of artificial intelligence (AI) and human intelligence (HI) to enhance both human and machine capabilities. As discussed in Yau et al. [4], AuI is more likely a human-centered intelligent system that exploits effectively existing experiences or creates new skills and capabilities for humans, allowing them to even exceed their potential. From this point of view, AuI is an appropriate solution that aligns with the strategic objectives of Industry 5.0.

Smart and human-centric manufacturing is a core of Industry 5.0. Among many advanced technologies applied to enable the future of manufacturing, AI and AuI algorithms play the role of a brain that controls, monitors, and optimizes the whole manufacturing process. They are a strong driving force to achieve Industry 5.0's objectives. This study provides a survey to clarify the potential applications of AI and AuI algorithms contributing to SM in Industry 5.0. Several challenges and perspectives for these applications will be discussed and suggested. We also consider many AI-based methods for the transition from Industry 4.0 to Industry 5.0.

The rest of the study is organized as follows: In Sect. 2, we provide an overview of AI and AuI algorithms. Section 3 presents the transition from Industry 4.0 to Industry 5.0. The applications of AI and AuI algorithms in the field of smart manufacturing are discussed in Sect. 4. Section 5 focuses on the difficulties, challenges, and perspectives for further applications of these algorithms for smart manufacturing in Industry 5.0. Finally, some concluding remarks are given in Sect. 6.

2 Overview of Artificial Intelligence and Augmented Intelligence Techniques

2.1 The AI Algorithms

AI, or simply intelligence demonstrated by machines, is a scientific term that describes computer-based algorithms to make computers capable of mimicking and displaying some human thinking skills like *learning* and *making decisions*. In the literature, there are several points of view regarding the definition of AI. Some of them have been summarized in Mikalef and Gupta [5]. The author also suggested understanding AI as "the ability of a system to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals". In fact, the concept of AI has been was initiated decades ago but its strong development and great applications in many aspects of life have only been witnessed in the past 10 years thanks to the breakthrough achievements in cloud computing and IoT. Applications of AI can be seen in all aspects of life today. In agriculture, they are used to predict product yield, build smart agricultural production systems, detect pests, and adjust the amount of irrigation water and fertilizers. In the medical field, AI algorithms support doctors in diagnosing and detecting diseases early based on X-ray images or other biochemical data. In industry, they are the key to building

Table 1 Overviews of the applications of AI in several fields in real life

Field	References
Medicine	?, Yu et al. [6]
Healthcare	Kaul et al. [7], [8], [9]
COVID-19	Kumar et al. [10], Elsheikh et al. [11]
Transport	Abduljabbar et al. [12]
Law	Surden [13]
Education	Chassignol et al. [14], Zhai et al. [15]
Cyber sercurity	Leenen and Meyer [16]
Power electronics	Zhao et al. [17]
rock mechanics	Lawal and Kwon [18]
Automobile Industry	Ajitha and Nagra [19]
Smart manufacturing	Wang et al. [20], Kotsiopoulos et al. [21, 22]

smart factories with the functions of optimizing the production process, performing predictive maintenance, detecting abnormalities in the production line, and others. One can mention many more applications of AI in other fields such as classifying customers and ensuring system security in banking and finance, developing recommendation systems in business and marketing, and so on. Several references that summarize the applications of AI in real life are presented in Table 1.

In general, AI consists of machine learning and deep learning algorithms. A fundamental way to classify the AI algorithms is to split them into conventional techniques and advanced techniques [23]. Based on the kinds of data available for training the model, the conventional techniques are further subdivided into groups of supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning algorithms. Supervised learning uses labeled data to train the model. It is suitable for tasks of classification and regression relying on historical data. Many classical machine learning algorithms like decision trees, random forest, Naive Bayes, k-nearest neighbor, and support vector machine belongs to this group. Unsupervised learning can find undefined patterns in the dataset without labels, widely used for clustering and reducing the dimension of data tasks. Typical unsupervised learning algorithms include hierarchical clustering, k-means clustering, principal component analysis (PCA), etc. A systematic review of supervised and unsupervised machine learning algorithms can be seen in Alloghani et al. [24]. A hybrid method between supervised and unsupervised learning is semi-supervised learning where the model training relies on both labeled and unlabeled data. In a semi-supervised model, the unsupervised learning algorithm is first used to cluster similar data, and then the labeled data is used to label the remaining unlabeled data [25]. The last group in the conventional techniques is about the algorithms for making a sequential decision, reinforcement learning. An agent learns optimal behaviors through trial-and-error interactions with a dynamic environment. In the process of performing a specific task, an RL model tries random trials to maximize the reward and minimize the penalty, learning intelligent strategies and virtuoso skills. Li [26] provided a comprehensive overview of recent exciting achievements of reinforcement learning.

For the advanced techniques, the authors meant deep learning algorithms, the multi-layered artificial neural networks that can learn automatically key features of unstructured data without previous knowledge. In the literature, deep learning is considered as a subset of machine learning, and then AI. It provides powerful analytic tools for processing and analyzing big data in the era of Industry 4.0. The most well-known deep learning architectures include recurrent neural network (RNN), convolution neural network (CNN), Restricted Boltzmann machine (RBM), and autoencoders. Each type of architecture contains different variants. For example, a variant of RNN is long short-term memory (LSTM) networks that can capture long-term dependencies of different time scales. A comprehensive overview of techniques, taxonomy, and applications of deep learning algorithms is presented in Sarker [27].

2.2 The AuI Techniques

The AI techniques allow machines to operate autonomously and carry out specific functions, but they cannot think in the same way that humans do. Therefore, human intervention should not be ignored by intelligent systems to avoid the potential risks or even harms caused by the limitations of AI in some important applications. Moreover, human supervision, interaction, and participation will enhance the confidence level in these systems while human knowledge will be optimally used. The concept of Augmented Intelligence (AuI), also known as intelligence amplification, or intelligence augmentation, is then introduced.

AuI focuses on the assistive role of AI in collaboration with human intelligence (HI). It is defined as "halfway between the entirely human and entirely automated capabilities" through the formula [4]:

$$\mathrm{AuI} = \mathrm{HI}(50\%) + \mathrm{AI}(50\%).$$

That is, AuI combines both AI and HI to perform a task, aiming to strengthen the abilities of humans with the help of AI. The role of HI and AI in AuI is classified into three schemes, including the HI-AI approach, the AI-HI approach, and the combined HI and AI decisions approach. In the first scheme, the HI is utilized to input more useful information from the raw data to the AI system, helping the AI system to raise awareness of the context. The final decision is drawn from the AI system by analyzing the raw data and taking these inputs into consideration. The second scheme means that the AI is first applied to extract important information from raw materials. This information plays a guidance role that helps humans to have a deeper insight into the problem and then make a better decision. That is, in this scheme, the final decisions are made by humans, and the AI algorithms help to improve the correctness of these decisions. The method in the third scheme belongs to parallel decision-making models, where both HI and AI inputs contribute to making the final decision process simultaneously. There is no common rule to choose an appropriate scheme for any situation in the literature. The AuI approaches should be designed depending on each situation and for different purposes. No matter which scheme is applied, by combining both HI and AI in an algorithm, the AuI techniques can improve both human and machine capabilities, leading to better performance compared to using HI or AI separately. As an example, a study related to detecting lymph node cancer cells released from an IBM report showed that an AI system had a 7.5% error rate while human pathologists had a 3.5% error rate. This error rate, however, was reduced to 0.5% by combining input from both the AI system and pathologists [28].

The difference in each AuI model is due to the use of different AI algorithms and different combinations of AI and HI, in which AI plays a core advanced technology role. In the literature, several machine learning and deep learning algorithms have been applied to develop AuI models. Wang et al. [29] used the CNN algorithm to develop a computer vision-based AuI model for designing ergonomic wearable devices to control smart home objects remotely. The AuI architecture for designing a speech recognition system in Hebbar [30] was based on the blend of RNN and DNN to power a digital assistant interface. El Koujok et al. [31] built a decision support tool for real-time fault diagnosis in industrial processes relying on the PCA method to isolate faults and the random forest algorithm to classify faults. The digital assistance system for training employees in industrial manufacturing combining both AI with reinforcement learning strategy and HI with knowledge of experienced employees has been suggested in Maettig and Foot [32].

In several AuI models, AI techniques are not used for taking part in any decision-making process. Instead, by analyzing data and exploring patterns, then it reports those patterns to users, allowing HI to take over. As a result, the AuI approach can overcome many obstacles and technological challenges standing facing when using the AI method only. Although it is still open and inclusive debate, many researchers have considered the AuI the next generation for the future of AI [33, 34]. However, one can be sure of one thing the development and applications of AuI will be a powerful driver of significant change in many areas in real life.

3 Transition from Industry 4.0 to Industry 5.0

It can be said that "Industry 5.0" is a quite distant concept to more than just a few people, and there are reasons for this fact. In the past, each industrial revolution needed to go through a long enough period of time to develop and accumulate a substantial amount of new techniques and technologies to be able to advance to the next evolution. Indeed, since the first industrial revolution began in 1760, it took mankind more than 100 years to begin the second in 1870, and another 100 years to progress to the third in 1969. The fourth industrial revolution, or Industry 4.0, was after Industry 3.0 in a shorter time, but also took up to 40 years originated from a project in the high-tech strategy of the German government in 2011. In

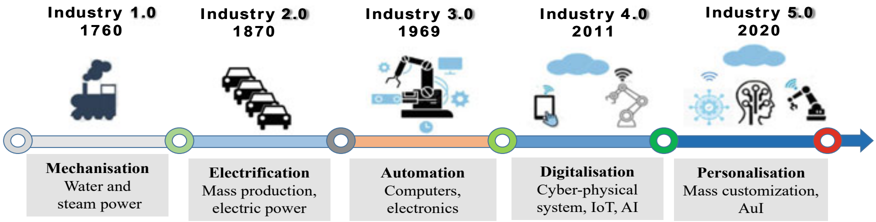


Fig. 1 The flow of industrial revolutions

many developing countries, implementing Industry 4.0 is still a challenge; applying advanced technologies to the fourth industrial revolution is even a goal or vision for a number of years to come. This is also true for many companies in developed countries which are considering Industry 4.0 as a trend they are currently adopting their strategy to. However, while industrial companies are still on the path of this fourth revolution, Industry 5.0 is already well on its way. Just a few years later the formal beginning of Industry 4.0, the term Industry 5.0 has been introduced in [35]. In 2020, the European Commission formally called for the fifth industrial revolution in 2020 by releasing the document titled "Industry 5.0: Towards a Sustainable, Human-centric, and Resilient European Industry" [3]. That is to say, humanity takes only about 10 years to transform from Industry 4.0 to Industry 5.0, a very short time compared to previous revolutions. Figure 1 shows the flow of industrial revolutions from the first to the fifth.

Traditionally, each industrial revolution was often initiated by the use of a new kind of energy and technological breakthroughs in production. From Industry 1.0 to Industry 2.0, the conversion is from steam power and mechanization of production to electric power and assembly line production. From Industry 2.0 to Industry 3.0, is the shift from manual production to automation using computers and electronic devices. The automated production in Industry 3.0 is raised to a new level, smart manufacturing, and digitization, in the process of moving to Industry 4.0 based on the application of cyber-physical systems (CPS) using AI algorithms and IoT technology. However, this is not the case for Industry 5.0. When it comes to technology, the recent strong development and applications of advanced and intelligent technologies such as AI algorithms, Big Data, IoT, and CPS are not enough to take Industry 4.0 to a higher level of evolution. Instead, these advanced technologies only help accelerate smart manufacturing processes in it. What makes the difference that leads to the introduction of Industry 5.0 is the human factor.

As it is just the dawn of evolution, there are various ways to define this concept and different points of view to explain its objectives in the literature. Most of the definitions consent that Industry 5.0 is a human-centric solution in which humans collaborate with autonomous machines to enable customization manufacturing. Industry 5.0 could be considered the first industrial revolution led by humans. It brings back and highlights the role of humans in the smart manufacturing process by further utilizing human brainpower and creativity. Designing synergy between humans and machines, Industry 5.0 improves the human touch with manufacturing and raises the ability of robotics to effectively collaborate with humans. This increases the process efficiency and accelerates the speed and accuracy of industrial automation. In Industry 4.0, the application of advancements like IoT, Big Data, or AI is for the goal of minimizing human involvement and prioritizing process automation. However, this trend is reversed in Industry 5.0: the goal of Industry 5.0 is to strike a balance whereby machine-human interaction empowers humans to express themselves in the form of personalized products and services. Moreover, the vision of Industry 5.0 is even beyond efficiency and productivity as the sole goals. In Breque et al. [3], the authors stated that Industry 5.0 reinforces the contribution of industry to society, places the well-being of humans at the center of the production process and uses new technologies to provide prosperity beyond jobs and growth while respecting the production limits of the planet. They also mentioned three interconnected core values of the new revolution: human centricity, sustainability, and resilience. The human-centric value places humans at the heart of the manufacturing process, converting technology-driven progress in previous industrial industries to a human-centric and society-centric solution. The sustainable value respects the planetary boundaries by developing solutions to recycle natural resources and reduce waste and environmental impact, aiming for a viable economy. Finally, resilient value considers a higher degree of robustness in industrial production to provide critical infrastructure and better cope with disruptions in case of crisis.

Regardless of the difference between Industry 4.0 and Industry 5.0, there is an important issue that needs to be addressed when talking about these industrial revolutions. That is, they are mutually complementary, not mutually exclusive. Unlike some earlier cases where the introduction of each industrial revolution led to the oblivion of the previous one, Industry 4.0 has not become obsolete due to the emergence of Industry 5.0. The progress from Industry 4.0 to 5.0 is not like climbing up a ladder and it is not necessary to go through Industry 4.0 to start Industry 5.0. Instead, the two revolutions can be parallel as Industry 5.0 is to take advantage of achievements developed in Industry 4.0 and bring their benefits to humans. It is a fact that so far many companies and businesses are still trying to keep up with trends, applying or adapting advanced technologies of Industry 4.0. However, the coexistence of these two industrial revolutions allows them to think about implementing Industry 5.0 even if it is only with certain stages of the production process.

In the literature, several academic efforts have been published to discuss important issues related to the fifth industrial revolution. Aslam et al. [36] presented a framework of absolute innovation management in the era of IoT and Industry 5.0. Xu et al. [37] ascertained the context of both Industry 4.0 and Industry 5.0 and clarified some questions related to the coexistence of the two revolutions. A number of critical components of Industry 5.0 towards a successful adoption in the field of manufacturing have been presented in Javaid and Haleem [38]. Nahavandi [2] outlined many key features and concerns manufacturers could face in Industry 5.0. The author also provided an example in which robots work with humans to increase production efficiency. A comprehensive survey on technologies and their potential applications in Industry 5.0 can be seen in Maddikunta et al. [39]. In the next sequel, we will discuss the key technologies, namely AI and AuI, for smart manufacturing in Industry 5.0 in more detail.

4 Artificial Intelligence and Augmented Intelligence Techniques for Smart Manufacturing in Industry 5.0

At the heart of any industrial revolution is production. Since mankind engages in Industry 4.0, industrial production has changed to a new nature, smart manufacturing (SM). The goal of SM is to use advanced technologies to monitor the manufacturing process and adapt to changing requirements and conditions of production reality to increase its performance. SM does not simply refer to a manufacturing process in the factory to produce a specific product, it also includes all the steps involved to operate that process, from managing, planning, preparing raw materials, forecasting, performing system maintenance, transporting, distributing, and storing, etc. In Industry 5.0, the production shift from mass production to mass customization requires companies and businesses to expand production in the sense of providing new products more often and increasing the size of product assortments. Higher and more individual demands from the market pose more tasks for SM and make it an essential objective for Industry 5.0.

Among many cutting-edge techniques applied in SM, AI is considered to play a very important role: it is a "brain" that makes manufacturing "smart" [40]. In fact, AI has impacts on all productive, social and economic sectors. This is evidenced by the growing influence of AI: they make progress in the adaptability and flexibility of businesses. In a report on the state of AI in 2021, Michael et al. [41] stated that AI adoption is rising steadily, accounting for 56% of all respondents in at least one function, up from 50% in 2020. Due to its great contributions, AI is today regarded as one of the most prominent innovations in Industry 4.0 that enable SM. The AI algorithms provide efficient solutions for a wide range of manufacturing stages. For example, He et al. [42] constructed a decision support system used for textile manufacturing process optimization based on a deep reinforcement learning scheme. Romeo et al. [43] applied different machine learning (ML) approaches like decision Trees, k-Nearest Neighbors, and neighborhood component features selection to develop an innovative design support system for optimization problems in engineering design processes. A systematic literature study that showed the applications of deep learning algorithms for multi-criteria making-decision approaches has been presented in Yang et al. [44]. As discussed in Wang et al. [20], deep learning algorithms have been investigated for a wide range of manufacturing systems recently. The authors also highlighted two important applications of deep learning for predictive maintenance in SM, which are diagnostic analytics for fault assessment and predictive analytics for defect prognostic. In manufacturing processes, failures due to machinery degradation after certain periods of working or abnormal working conditions can result in unexpected and serious consequences. Monitoring machine conditions and recognizing the incipient defects is then vital for manufacturers. The efficiency of AI algorithms in diagnosing the root cause of failures in a manufacturing process makes them a powerful tool to perform the task. The CNN, DBN, and the Auto Encoder algorithms have been applied to a large number of specific situations related to fault diagnostics like aircraft engines [45], rolling-element bearing [46, 47], reciprocating compressor [48], gearbox [49], a wind generator [50], and wind turbine [51]. Another important problem in SM that can be solved effectively by AI algorithms is scheduling. Scheduling is a method regularly used to allocate resources and tasks in industrial companies to optimize one or more objectives during a certain period of time. Automatic and autonomous scheduling management can lead to a zero-defect manufacturing Serrano-Ruiz et al. [52]. Among many AI algorithms, reinforcement learning algorithms contribute significantly to scheduling solutions. The use of reinforcement learning for specific schemes of scheduling in industrial manufacturing can be seen in a number of studies, see, for example, Shiue et al. [53], Waschneck et al. [54]. Moreover, the application of AI algorithms in SM smart manufacturing is not limited to the topics mentioned above. Kotsiopoulos et al. [21] presented a smart grid paradigm of machine learning and deep learning algorithms enabling SM.

In Industry 5.0, industrial production shifts to mass customization and penalization. The role of humans in the advanced technologies used in SM is highlighted, leading to the concept of human-centric solutions. The development and application of AI in SM also become more flexible and customizable with the introduction of AuI.

As discussed above, AuI is a combination of AI and HI to solve a specific problem. It can both strengthen the power of AI and exploit the intelligence of humans simultaneously. One of the most visible contributions of AuI to SM in Industry 5.0 is perhaps analyzing Big data. We all know that the cutting-edge technologies in Industry 4.0 like the Internet of Things (IoT), Cloud computing, 5G, and CPS have led to the emergence of a new era, the era of Big Data. In Industry 5.0, as production shifts to personalization and customization, the products tend to be personalized with many different requirements. These changes will also contribute to a significant increase in the V's of the Bid data, i.e. Volume, Velocity, Variety, Value, and many others. According to an estimation from the International Data Corporation, the "digital universe" could contain over 5,200 gigabytes of data for each person and 40 trillion gigabytes of data points for a total by 2020 [55]. Big Data contains a huge and useful source of information covering all aspects of smart manufacturing. Exploiting this huge source of data allows manufacturers to have an insight into the issue of concern, and then make accurate decisions to optimize all the stages of the production process. AuI is an effective tool to accomplish this task. While the use of AI only tends to make a system operate autonomously without human involvement, AuI methods use AI algorithms to provide engineers and managers with actionable data. Decisions will be made in a more flexible and system-appropriate manner. For this reason, AuI would be at the center of the blossoming fields of data analytics in Industry 5.0.

The applications of AuI can be seen in several phases contributing to SM. For example, in Agrawal et al. [56], the AuI has been utilized to solve the 2D packing problem having irregular shapes and sizes without overlapping in a rectangular 2D container. Yu et al. [57] described the way to combine HI, i.e. knowledge, experience, and inspiration of experts and designers, and AI, i.e. the genetic algorithm and the case-based reasoning, in designing new products. An adaptive driver support system, called Driver Advocate, that merges various AI techniques like agents, ontology, production systems, and machine learning algorithms to help drivers have a safer, more enjoyable, and more productive driving experience was reported in Hwang et al. [58]. Yau et al. [4] provided a summary of applications of AuI in many sectors in real life, including industry and manufacturing. There is a perspective concept arising in Industry 5.0 that can be considered as a visible representation of AuI, the next generation of robots, called collaborative robots (or Cobots). Instead of a programmable machine that can perform sequential tasks, these robots that can work side by side with a human. Designing based on AI algorithms, Cobots can interact with humans, learn, be flexible, and have the possibility of wide use and quick adjustment. The integration of Cobots in production makes them possible to advance from coexistence to collaboration between workers and robots. A literature review of the application, development and future opportunities of Cobots in SM can be seen in Liu et al. [59].

5 Dificulties, Challenges, and Perspectives for Application of Artificial Intelligence Techniques for Smart Manufacturing in Industry 5.0

In this section, we discuss some difficulties and opportunities for the application of AI and AuI to SM in Industry 5.0.

5.1 5G, 6G, and Smart Manufacturing

McCann et al. [60] has considered Industry 4.0 as the way how manufacturing industry exploits the innovations of 5G wireless communications by automating industrial technologies and utilizing AI technologies. This view shows the importance of 5G for advanced industrial manufacturing. 5G refers to the fifth generation of wireless technology that enables faster connectivity, and higher capacity but also lower latency than its previous generations. For example, in the right conditions, the download speeds in 5G are up to 100 times faster than in 4G, reaching 10 gigabits per second. The 5G latency rate is also significantly lower compared to the 4G latency, i.e. 1 ms versus the average of 200 ms. The better connections in 5G enable machines to take advantage of the massive computing power and cloud storage options without being tethered by physical wires. It is also a core technology that enhances the effectiveness of the Industrial Internet of Things (IIoT), a network of physical devices embedded with sensors, actuators, electronics, software, and network connectivity. In SM, sensors are installed in every important part of the machine and every step of the production line to collect data from the whole manufacturing process. The use of 5G in IIoT networks provides a powerful solution for connecting, interacting, and exchanging data between these objectives. Those are the prerequisites for many vital stages in SM like real-time production monitoring or predictive maintenance. Due to its tremendous advantages, the 5G networks can bring many benefits to companies and businesses. The technology is estimated to increase the global manufacturing gross domestic product by 4%, or approximately $740 billion by 2030 [61]. While the 5G technology is commercially being rolled out in Industry 4.0, research on 6G systems has also begun. Although its experience can only be expected in a near future, the 6G wireless technology is expected to outperform 5G in multiple specifications. In particular, the next generation has stricter requirements in higher frequency and data transmission rate, faster mobility, lower latency, and better spectral efficiency than 5G. The evolution of wireless communication characteristics from 5G and envisioned for 6G as well as the compassion of 6G features with respect to 5G can be seen in De Alwis et al. [62] and Zong et al. [63].

In Industry 4.0 and then Industry 5.0, the connection role of 5G and 6G systems is becoming more and more important. While AI algorithms are likened to the brain that controls the smart manufacturing process, 5G and 6G can be considered the lifeblood of the system, ensuring data is updated and shared in real-time, leading to effective analysis, monitoring, and even multi-access edge computing. The availability of 6G will dramatically change the performance of the AI-based technologies applied in SM like virtual reality, augmented reality, and federated learning.

Regardless of the great potential of 5G and 6G, the technology will not be ready for widespread use by manufacturers. Several hurdles for industry adoption of 5G and then 6G were discussed in Mourtzis et al. [61], involving the cost of procuring and installing the required equipment, safety, and deployment knowledge. Other infrastructure requirements for the application of new wireless systems have been mentioned in O'Connell et al. [64].

5.2 Product Design

The transition of the industry from mass production to mass customization and personalization will lead to a huge change in product design. In the early days of the second industrial revolution starting mass production, a famous quote was "A customer can have a car painted any color he wants, so long as it is black". This has changed rapidly as in the competitive economy after that the manufacturing industry is oriented towards customer needs, taking the criteria of meeting customer needs as the goal. Along with the creative ideas of the designers, the product design in the Industry 4.0 is significantly based on data analysis from users. The era of Big data opened up providing manufacturers with a tremendous amount of valuable information about customer needs and preferences. AI algorithms are developed to analyze these data, allowing manufacturers to better understand consumer desires and support their ideas and suggestions in designing new products. However, Industry 5.0 is taking the task of product design to the next level: It provides customers the ability to personalize an item from start to finish. In particular, customers are empowered to directly involve in the product design process and production. They can provide ideas, require specific features from the product, or even design the product according to their own style. This co-design process is enabled by advanced technologies like 3D or 4D printing, open product architectures, on-demand manufacturing systems, and responsive cyber-physical systems [65]. As a result, consumers are provided with a product or a service specifically tailored to them. A straightforward example is the fashion industry where clients can choose precisely the kind of shoes or clothing with their very own colors, style, or material. The task of the design team and the manufacturer will now include consulting and finalizing products based on the customer's own opinion. The AuI models can be applied to analyze the efficiency and features of new products. Yu et al. [57] presented an example of AuI based design system that integrates several bits of intelligence to design new products. A system that can take into account user preferences and decide which product configuration is the best compromise has been designed in Mladineo et al. [66]. Yang et al. [67] proposed an integration approach to synthesize similar consumer preferences in product design. Zhan et al. [68] presented a data mining approach for bridging customer knowledge to innovative product development.

The personalization of commercial products also brings certain difficulties for manufacturers. When a product is highly personalized, the number of products produced for that requirement is not much, leading to increased production costs. The unique features in each product will also lead to significant changes in the production process as the manufacturers might have to update new features in the production line. Several steps for smart design processes have been discussed in Zhan et al. [69].

5.3 Predictive Maintenance

Machinery maintenance is a crucial mission for any industrial manufacturing company. The purpose of the task is nothing but to ensure the ongoing operation of the whole system, maximizing the life cycle of the machine, minimizing machine downtime and associated costs, and improving the quality of production. It can benefit significantly businesses by reducing 10–20% cost in related labor and materials, accounting for 15% of the total costs of a typical manufacturing company Jamie et al. [70]. Typically, maintenance involves regular servicing of equipment, routine checks, repair work, and replacement of worn or nonfunctional parts. Over the course of industrial revolutions, maintenance strategies have gone through an evolution from reactive maintenance, and preventive maintenance to predictive maintenance. Nowadays, reactive maintenance is no longer appropriate as in this strategy technicians perform the maintenance only after failures are detected. These failures could cause remarkable damage to businesses such as disrupting production, reducing product quality, and creating defective products. To overcome this disadvantage of the reactive method, the preventive or condition-based strategy aims to rigidly schedule maintenance that can prevent faults from occurring. In smart manufacturing, the maintenance strategy is transformed to the next level, i.e. predictive maintenance. Based on the idea of the preventive method, the predictive strategy uses the advanced technologies of Industry 4.0 like IIoT and Big Data, cyber-physical systems, and AI algorithms to monitor the machines in real-time, evaluate their future health or remaining useful life, and formulate proactive maintenance plans. An efficient predictive maintenance solution can significantly enhance the quality, productivity, efficiency safety, and profitability of the manufacturing process. In the literature, a large number of studies have been conducted to propose effective methods for predictive maintenance. A comprehensive review of the advancements of ML techniques widely applied to PdM for smart manufacturing in I4.0 has been conducted in Çınar et al. [71]. Another overview of methods and applied tools for intelligent predictive maintenance models in Industry 4.0 can be seen in [72].

While AI-based algorithms have shown considerable performance in designing predictive maintenance models, the use of hybrid-augmented intelligence algorithms has been also conducted. By involving humans in the predictive maintenance process, an AuI-based system can encode knowledge continuously in a self-learning loop that optimizes the system's performance and provides more benefits to maintenance engineers. However, there is also a challenge to optimizing human-machine collaboration in a predictive maintenance system. Two specific reasons for this fact have been discussed in Wellsandt et al. [73], including the constraints to encode maintenance knowledge and experience into the system, and the lack of appropriate approaches to include human maintenance actors in the decision-making process in current systems. Based on this discussion, the authors also suggested a novel approach for maintenance experts and operators to interact with a predictive maintenance system through a digital intelligent assistant.

5.4 Quality Control

Quality control refers to processes by which products are measured in order to maintain their quality. In other words, it is the process to make certain that all the products are of the predefined standards and expectations. A right quality control process aims at monitoring different stages of manufacturing and brings manufacturers numerous advantages. It helps manufacturers not only prevent defects but also identify problems and correct them as soon as they occur, rather than at the end of the manufacturing process. It also facilitates consistently high-quality products, resulting in the prestige of manufacturers and the satisfaction of customers.

Quality control methods have a long history in industrial manufacturing, traced back to the times of Industry 1.0 when the modes of production were still in their infant stage, and does not have a uniform process, namely, its type depends on the industry. For example, in the food and pharmaceutical industry where defective products can have severe negative effects on human health, chemical and microbiological testing of samples from the production line should be performed. Automobile manufacturing focuses on parts meeting specifications and tolerances, ensuring engines, drive trains, and other mechanical parts operate efficiently and safely as designed. Even so, along with the development through the various stages of the industrial revolution, there are still general methods used in quality control. Statistical process control (SPC) is an effective method for quality control that can be applied to any process where the conforming product output can be measured. A large number of control charts, a key tool used in SPC, have been designed in the literature for the purpose of controlling processes and maintaining product quality, see, for example, Nguyen et al. [74], Nguyen and Tran [75]. Recently, traditional control charts have been combined with machine learning architectures to create higher-performance models. For example, Maboudou-Tchao et al. [76] designed control charts based on Support Vector Data Description (SVDD) to release a restrictive assumption of multivariate normal distribution of quantities of interest. A modified multivariate cumulative sum control chart based on SVDD for multivariate statistical process control was proposed in Xia et al. [77].

Since Industry 4.0, the manufacturing industry has shifted to smart manufacturing, and advanced technologies have been applied to make production more efficient. Technological products also have higher quality standards, requiring higher precision on a larger production scale. In that context, the use of computer vision and AI algorithms have become widely used in quality control. Villalba-Diez et al. [78] showed an application deep learning soft sensor combined with a high-resolution optical quality control camera to increase the accuracy and reduce the cost of an industrial visual inspection process in the Printing Industry 4.0. Several applications of computer vision in manufacturing, including production process control, were discussed in Zhou et al. [79].

Under the Industry 5.0 circumstance where the products tend to be personalized, producing a single product should remain profitable for manufacturers. Revolutionary changes in smart manufacturing should be conveyed by smart quality control to ensure the delivery of the best quality products to customers [69]. Therefore, the use of AI-AuI-based advanced technologies is necessary and also a challenge for businesses to speed up quality control with high accuracy.

5.5 Digital Twin

One of the advanced technologies that can advance remarkably toward smart manufacturing in Industry 5.0 is the digital twin. A digital twin model refers to a digital replica of a physical a system, aiming to understand and control and analyze the physical system based on analyzing the cyber system. The digital twin technology creates digitized virtual models for real objects by simulating as closely as possible the physical components, structures, states, and operations of the objects. The optimal parameters obtained by analyzing and predicting the dynamic changes of the virtual models will be transferred to the physical objects to optimize the entire manufacturing process. A digital twin can be applied to many stages in smart manufacturing, involving product design, product development, predictive maintenance, monitoring facilities and processes, decision-making support, etc. The application of digital twins brings great benefits to businesses. For example, it is expected that the global market of digital twins will reach $26.07 billion by 2025 with a compound annual growth rate of 38.2% [80].

As discussed in Grieves and Vickers [81], a digital twin consists of three fundamental components: a physical space containing physical objects, a cyberspace containing cyber objects, and a connection mechanism flowing data from the physical space to the cyberspace to synchronize the physical and cyber systems. The physical space is connected to a virtual part during its whole life-cycle by integrating and combining data from multiple sources. The accurate and comprehensive information achieved from the real world is the key material to perform analysis in the virtual version. New advanced technologies in Industry 4.0 and Industry 5.0 such as 5G and 6G, IIoT and cyber-physical systems, AI algorithms, virtual reality, and augmented reality can facilitate the realization of the digital twins in a smart manufacturing system. Lee et al. [80] presented a reference architecture for the integration of AI and digital twins in a structure of cyber-physical systems toward smart manufacturing. A comprehensive review of the concept, methods, applications, and function of the digital twins in smart manufacturing was conducted in Son et al. [82].

There are several issues rising from the application of digital twins in Industry 5.0. Ramu et al. [83] mentioned the lack of trust and privacy issues in sharing sensitive data that could lead to a lag in the wide adoption of digital twins in manufacturing and industrial sectors. Federated learning was then suggested by combining in digital twin model to solve the problem. The human-centric manufacturing requirement is also a challenge for applying digital twins. Four main directions related to the difficulty when deploying digital twins into current production processes are raised in Lattanzi et al. [84].

5.6 Cybersecurity in Smart Manufacturing

The application of digital technologies in smart manufacturing brings not only great benefits but also new risks and challenges to manufacturers. The use of IIoT and cloud computing technologies means that all the components involved in the production process such as equipment, machines, electronic devices, software, and systems of monitoring, controlling, and analyzing are connected via wireless network or wired Ethernet. The possibility of Internet connectivity and delivering data through it leaves potential vulnerability for attackers to exploit and take control of the system. As a consequence, manufacturing systems are becoming more accessible than ever before and cyber-attacks are posing a significant threat to the field. In fact, critical IIoT devices are now vulnerable to a large number of cyber-attacks. As discussed in Lezzi et al. [85], 31% of businesses have experienced cyber-attacks on operational technology while 38% expect attacks to extend from information technology to operational technology. Despite perceiving the enormous damage that cyber-attacks can bring, only 16% consider their company well prepared to face cybersecurity challenges. Krundyshev and Kalinin [86] reported that cyber attacks in the energy sector alone cost $13.2 million annually. Due to exposing people, intellectual property, and the whole process to the risks, cyber-attacks are now considered a complex challenge and the most prevalent threat to manufacturing in Industry 5.0.

In the literature, numerous study efforts have addressed the problem of cybersecurity for some aspects of smart manufacturing. A review of machine learning techniques applied to cybersecurity has been presented in Martínez Torres et al. [87]. The deep learning methods are also proposed to solve several problems in the area of cyber-security, involving intrusion detection, malware detection, phishing/spam detection, and website defacement detection. A Survey of the application of the deep learning methods can be seen in Mahdavifar and Ghorbani [88] and Ahmad and Rahimi [89].

Among many industrial assets involved in cyber-attacks, Industrial Control Systems (ICS) are considered to be the vulnerable system that needs to be protected. ICS refers to an automation system that controls and manages industrial technical facilities to ensure they run automatically. The ICS is a general term that encompasses several types of control systems like Supervisory Control and Data Acquisition (SCADA) systems and Distributed Control Systems (DCS). An IIoT-based ICS is now one of the top industries attacked by various threats [90]. Recently, federated learning (FL) has emerged as a a promising solution to deal with these threats of an ICS. FL is a popular distributed learning framework developed for edge devices. This collaborative AI method enables training AI models by averaging local updates aggregated from multiple learning IIoT devices without data exchange. It mitigates privacy leakage risks by allowing the private data to stay locally while leveraging large-scale computation from edge devices [91]. Huong et al. [92] presented an FL-based AI model for anomaly detection in ICS. A hybrid architecture combining FL, Autoencoder, Transformer, and Fourier mixing sublayer designing for anomaly detection within the ICS contexts has been provided in Truong et al. [93]. In general, there are promising perspectives for the idea of combining different advanced deep learning algorithms with FL to solve cyber-security problems in smart manufacturing.

5.7 Wearable Technology for Smart Manufacturing in the Industry 5.0

Wearable devices refer to devices equipped with microchips, sensors, and wireless communications capabilities that can be wearable. Various types of wearable devices are available in the market such as smart watches, wristbands, smart glasses, electronic garments, skin patches, and so on. When it comes to these devices, one could immediately think of applications in medical and healthcare. Indeed, it is quite popular that they are utilized to sense and collect physiological data and detect the posture and motion of the wearer's body. This information is useful for the diagnosis and treatment of patients or monitoring of individuals in the home and community settings. However, their applications are not limited to this field. In fact, wearable technology has the potential to be useful in manufacturing, especially in Industry 5.0. Several applications of the technology in smart manufacturing can be named such as production, warehousing, logistics, maintenance, safety, and security. For instance, in severe working environments like in the oil, gas, and automotive industries, wearable technology can create a safer shop floor by monitoring surrounding conditions and alarming potential incidents. They can be integrated with voice functions to become hands-free instructions and communication devices, allowing workers to stay focused on their tasks, obtain additional information, or deliver remote commands [40]. Hao and Helo [94] described a specific maintenance service scenario in which the smart glasses are utilized to perform maintenance and remote assistance. The eye-wearable devices for improving the performance in machine maintenance have also been interested in [95]. Nguyen et al. [96] suggested an ensemble machine learning algorithm for dealing with wearable sensor data-based human activity recognition. Kucukoglu et al. [97] introduced a digital assembly glove to measure vibration and force values on the fingers to classify proper and defective operations in connector assembly tasks from an automotive company. From an economic perspective, wearable technology can increase employees' productivity by 8.5% and life and job satisfaction by 3.5% [98].

With wearable technology, the structure and operation of each device are not simply a combination of electronic-based sensors and other parts to be wearable. It is the integration of a variety of advanced technologies, ranging from hardware and IIoT to cloud computing, Big Data analytics, and AI algorithms. A detailed study about the explanation of wearable technology from A to Z has been presented in Godfrey et al. [99]. Kong et al. [100] proposed an industrial wearable system of human-centric empowering technology to establish a human cyber-physical symbiosis to support real-time, trusting, and dynamic interaction among operators, machines, and production systems. In Industry 5.0, human resources are considered an important manufacturing resource and assets that could be connected to the Internet. The IIoT application in cloud manufacturing was even proposed to a deeper level of Internet of Users [94]. As a result, advanced technologies like AI algorithms, Cloud computing, cyber-physical systems, and 5G and 6G are powerful solutions to answer the questions of how to develop wearable products and how to use these devices in industrial manufacturing.

5.8 Human-Centric and Sustainable Manufacturing

Unlike other industrial revolutions, the emergence of Industry 5.0 is not the result of a technological breakthrough or the discovery of new energy. It is the introduction of the human element at the heart of the revolution. While Industry 4.0 focuses on smart systems, Industry 5.0 is considered a synergy between humans and autonomous machines. While Industry 4.0 and previous ones focus on improving the efficiency and productivity of the manufacturing process, leading to significant oblivion of the human cost, in Industry 5.0, the human element is the goal, leading to the concept of human-centric manufacturing.

The human element in human-centric manufacturing can be separated into two objectives, customers and workers. Customers are not simply recipients of products and bring economic benefits to businesses, they are able to join in the manufacturing process to a certain extent. As an example, they can come up with ideas and participate in the product design process. The new products aim to meet the requirements and styles of each individual. This is the meaning of mass customization and personalization. The second objectives include all the people involved directly in the manufacturing process. The human-centric approaches focus on this object. Through these approaches, humans, and machines are paired to further utilize human intelligence and creativity to increase process efficiency by combining workflows with intelligent systems. The human-machine relationship in human-centric manufacturing is encapsulated in a 5C journey: Coexistence, Cooperation, Collaboration, Compassion, and Coevolution [101]. Moreover, a new role of workers is developed, shifting from "cost" to "investment". Advanced technologies are adaptive to serve people in safer, more comfortable, and more motivating working environments. Researchers have also come up with many solutions to this problem. For instance, Peruzzini et al. [102] conducted an industrial case study for using virtual manufacturing to design human-centric factories. Lu et al. [101] provided an outlook on human-centric manufacturing towards Industry 5.0, answering many key questions related to human needs, the role of humans in future manufacturing, models and technologies enabling human-centric manufacturing, and system framework of human-centric manufacturing.

An important property of manufacturing towards human-centric manufacturing is sustainability. Nowadays, humanity is facing many severe problems such as climate change (global warming, melting ice, rising sea level, drought, floods, etc.), environmental pollution, and depletion of traditional natural resources. The issue of sustainable manufacturing becomes more and more important and urgent. In addition, the world's population is still growing steadily, leading to higher demand and consumption. Sustainability strategies enable manufacturers to meet growing demands with minimal impacts on the environment and society. Jawahir and Bradley [103] introduced the 6R concepts for sustainable practices, including Reduce, Reuse, Recycle, Redesign, Recover, and Remanufacture. Sustainability in manufacturing is also meeting the needs of those directly involved in the production process. The human needs are mentioned in Lu et al. [101] with five levels, i.e. Safety, Health, Belonging, Esteem, and Self-actualization. A systematic literature review on Industry 4.0 applications for sustainable manufacturing has been carried out in Ching et al. [104].

The advanced technologies and approaches in Industry 5.0 play a decisive role in the long-term and challenging process toward human-centric and sustainable manufacturing. Although it has been adopted to some extent, the application of these technologies is still in its infancy. This remains one of the biggest challenges for manufacturers in Industry 5.0 with “the power of industry to achieve societal goals beyond jobs and growth, to become a resilient provider of prosperity, by making production respect the boundaries of our planet and placing the well-being of the industry worker at the center of the production process” [3].

5.9 Quantum Machine Learning

Quantum machine learning can be simply understood as the implementation of machine learning algorithms based on quantum computing. That is, in quantum machine learning, quantum algorithms are developed to solve typical problems of machine learning using the efficiency of quantum computing. In the era of Big Data, quantum machine learning have distinct benefits, having a decisive influence on whether to effectively exploit and use this Big data source or not. Previously, AI algorithms were only implemented with classical computation, when combined with quantum computing technology, the power of these algorithms is multiplied to a new level. It has been mentioned in the literature that the new generation of computing technology involves a type of computer hundreds of millions of times faster than the world’s most powerful computer that humans created today. For example, a Google quantum computer could do calculations in less than four minutes while a traditional supercomputer would need about 10,000 years to complete.

The power of quantum computing makes it outperform classical computing in finding new correlations in data, recognizing patterns, and advancing classification or regression. Huang et al. [105] considered two paths that quantum technology can enhance machine learning. The first path is related to the power of quantum computing that can speed up the training process or enhance inference in graphical models by finding better optima or finding optima with fewer queries. As it can simulate samples from probability distributions that are exponentially difficult to sample from classically, the second one is about the possibility of quantum computing to generate correlations between variables that are inefficient to represent using classical computation. According to the discussion in Malina and Woerner [106], quantum machine learning is expected to have a significant influence on manufacturing in several areas. For instance, it can analyze additional interactive factors and processes to increase production yield in semiconductor chip fabrication. In production flows and robotics scheduling for complex products where the simulation and optimization require intensive computation, it can also enable faster optimization runs and perform optimizations more dynamically. Another example is in quality control for software development with progressively sophisticated software validation, verification, and fault analysis, quantum computers can have the capability to analyze software systems that are much more complex than classical ones.

There are still many open questions related to the quantum learning procedure. For example, how can we efficiently implement an optimization problem (that is usually solved by iterative and dissipative methods such as gradient descent) on a coherent and thus reversible quantum computer? And the overall question is whether there is a general way how quantum physics can in principle speed up certain problems of machine learning. Although the research on and the manufacturing of quantum computers are still in their infancy, it is believed that once quantum technology is widely deployed, it will create a new revolution not only in computing but also in industrial production as well as many other areas of real life.

5.10 Human-Centered Design of AI and Explainable AI for Industry 5.0

The widespread use of AI in SM as well as in many other fields of real life brings not only benefits but also potential risks. As more and more enterprise and government policies are relying on AI algorithms, legal, ethical and practical concerns related to this technology have started to appear. The autonomous AI's thinking could be rational but that also might be unethical or unjustified. The inference of AI models based on incomplete or distorted data may result in inappropriate "thinking", exaggerating prejudice, and even causing physical harm. Other worries are related to the problem of losing human control over external AI agents and replacing/rejecting human beings with future AI systems. The concept of human-centered design (HCD) was then introduced to provide a solution in response to these fears.

HCD refers to the design methods that aim to center humans in the development of a design. In this context, "humans" has a broad meaning, including experiences, desires, motivations, emotions, behavior, and perspective of people whose involve in the process. A core of HCD is that it shifts the previous view of humans as a part of the system like "human factors," or "human resources" into a central view in all the aspects of the design.

In the context of AI, HCD leads to a buzzword, namely Human-Centered AI (HAI) or Human-in-the-loop AI. This means placing humans in the loop through their active involvement in the phases of preparing, learning, and making decisions about AI-based systems. HAI is crucial to ensure that the end-user and human values are prioritized by AI solutions [107]. As humans are increasingly interacting with AI-based intelligent systems, there are several questions arise related to the design of these systems. Some have been mentioned in Riedl [108] like: How to design smart systems that can help people to understand their decisions, should these systems have the problem-solving abilities more like humans, how people can trust the systems' decisions or be comfortable working with them, how to design the intelligent systems that can bring users who may not be experts or well versed in may not information to its users in a meaningful and understandable manner. After assigning two critical capacities for human-centered AI, namely "understanding humans, and being able to help humans understand the AI systems", the authors have come to the conclusion that human-centered AI does not mean that an AI-based agent must think like humans; instead, it is able to understand the expectations and needs of humans and to help humans understand them in return. A similar point of view for human-centered AI could also see in Xu [109], where the authors provided a working HAI framework with three major components: an ethically aligned design for avoiding discrimination, maintaining fairness and not replacing humans; technology enhancement for reflecting the depth characterized by HI; and human factors design for making AI solutions explainable, comprehensible, useful, and usable. An example of an HAI approach in developing an AI-based clinical decision support system has been presented in Bond et al. [110].

One of the challenges facing HAI is the opacity of machine learning and deep learning algorithms, the core technology of AI. Many machine learning-based models are considered black boxes and lack transparency, which is not comfortable for end-users. How exactly do these models make decisions and interpretations? How to identify, quantify or troubleshoot the errors? As humans often seek rationale and reassurance for their decisions, this black-box phenomenon may cause skepticism and distrust. Moreover, in critical situations like autonomous driving and remote surgery, any mistakes can result in dire consequences and non-explainable output from AI models is most unacceptable. As a consequence, it may affect the adoption of AI-based intelligent systems. Therefore, it is necessary to make AI explainable and the concept of explainable AI (XAI) is introduced in the literature. The aim of XAI is to handle the black box problem of AI models. By allowing users to understand the algorithms and parameters used, XAI makes the outputs more reliable, improving the users' decision-making efficiency. An insight into XAI from the social science point of view has been shared in Miller [111]. Zhu et al. [112] proposed a human-centered approach using XAI for designers, specifically for game designers. By the year 2015, a program called DARPA has been organized with the cooperation of 13 universities and research institutes to approach XAI [113]. In the field of smart manufacturing, Huong et al. [92] proposed an XAI-based solution for detecting anomalies in industrial control systems. An XAI approach to enhance the perspective and reliable results of an LSTM-based Autoencoder-OCSVM learning model for anomaly detection in industrial manufacturing has been suggested in Ha et al. [90].

As a conclusion in Gunning et al. [113] after the four-year DARPA program, there is no universal solution to XAI. Different user types require different types of explanations. Especially, as new AI techniques are developed, they will continue to need explanation. However, XAI should be considered essential for HAI and then HCD, the important solutions aiming at the central role of humans in Industry 5.0.

5.11 Manufacturing in the Metaverse

When it comes to Metaverse, we are talking about the most trending topic today in the next-generation digital technologies. Semantically, the term Metaverse comes from a combination of the prefix "meta", meaning transcendence, and the suffix "verse" taken from "universe". It refers to a computer-generated world that is parallel to, linked to, and reflected in the physical world. To some extent, Metaverse can be considered an expansion of digital twins into the fields of people and society [114].

Integrating many advanced emerging technologies, it provides a fully immersive and self-sustaining virtual shared space for humans with various interactive and collaborative activities. For example, virtual reality and augmented reality technologies are applied to produce an immersive experience, the digital twin technology is for building a replica of physical reality (including structure and functionality), the wearable sensors and brain-computer interface technology allow the interaction in the Metaverse, and large-scale metaverse creation and rendering are developed based on AI algorithms [115]. Metaverse is turning the idea of a virtual world in science fiction into an interactive virtual world in reality.

Although Metaverse is just coming out of its infancy to become a reality in the near future, the great benefits it brings are visible, especially in the field of manufacturing. Many stages in smart manufacturing benefit from Metaverse. A virtual shared space enables design engineers to exchange ideas and collaborate remotely from around the world to create 3D virtual designs for a specific product. It also provides opportunities for stakeholders to join in the product design process, and share their ideas with manufacturers, thereby shortening the product life cycle. Metaverse can also serve as a virtual space for testing new products. For example, it allows robot manufacturers to examine the user acceptability of novel robot agents. From this digital environment, humans can also understand robot operations, building trust and confidence when working with robots, and shifting towards human-robot collaboration [116]. This supports the aim of human-centric manufacturing in Industry 5.0. Moreover, as the virtual world generated in Metaverse reflects approximately the structure and functionality of reality, it can simulate possible scenarios in the physical world. Information obtained from these hypothetical situations provides a useful source of data from the physical asset that allows manufacturers to control the manufacturing process more efficiently. It could be planning a factory floor and setting up the right production line to maximize resources. It could be monitoring processes in real-time and performing predictive maintenance to detect abnormal conditions, remove defective products, save cost and minimize disruptions. Another advantage of Metaverse in smart manufacturing is in training workers. It is not easy for new employees to be familiar with operating a certain machine in complex systems or getting used to the production lines. Thanks to the immersive environment in Metaverse, workers can practice operating systems and production lines instead of having them use physical equipment that can be dangerous or difficult in a real training situation.

Realizing the Metaverse applied to industrial manufacturing requires an integration of many advanced technologies. The advent of 5G and 6G networks, AI and AuI algorithms Augmented Reality, Virtual Reality, and Mixed Reality technologies, Blockchain, digital twins, edge computing, and other advanced technologies should be all taken into consideration to serve as the tools and building blocks of the Metaverse. Wang et al. [117] provided a comprehensive survey on the fundamentals of Metaverse, including the general architecture, key characteristics, and enabling technologies. Another survey on technological singularity, virtual ecosystem, and research agenda of Metaverse has been presented in Lee et al. [116]. After reviewing the state-of-the-art technologies as enablers to the development of the Metaverse, the authors discussed the gap between the latest technology and the requirements of reaching the Metaverse. Among several layers of a Metaverse platform, the presence of AI algorithms inside each layer is noticeable. The role of AI algorithms in the foundation and development of the Metaverse has been explored in Huynh-The et al. [23].

6 Concluding Remarks

In this chapter, we have surveyed AI's methods and applications in smart manufacturing in Industry 5.0. We have also discussed several challenges and perspectives on the potential applications of AI techniques for smart manufacturing. It can be said that the concept of Industry 5.0 has not achieved a lot of attraction yet so far. However, its progress is inevitable. The kick-start of the European Commission to explore the shaping of Industry 5.0 has provided a vision of this fifth industrial revolution. This vision will help manufacturers to answer both the questions of how they will benefit, and how they can best leverage new technologies to obtain optimal performance from human-machine interactions in smart manufacturing. Like in Industry 4.0, AI is the key to the success of smart manufacturing in Industry 5.0. It provides an effective solution for many problems from training new workers, designing new products, maintaining machine systems, scheduling, and storing, to optimizing the whole manufacturing process. As Industry 5.0 places humans at its center, the role of AI is not to eliminate the employees but facilitate and collaborate with them in the manufacturing process. The collaboration between humans and machines leads to the introduction of new concepts like AuI and cobots. These will be the core factors of smart manufacturing to drive a more performing and human-centric goal in Industry 5.0. Due to several challenges, opportunities, and also great benefits of AI for manufacturing, more studies need to be conducted to gain insights from Industry 5.0, enabling businesses to be ready for the transition to this new industrial revolution.

References

1. Rojko A (2017) Industry 4.0 concept: background and overview. Int J Interact Mob Technol 11
2. Nahavandi S (2019) Industry 5.0-A human-centric solution. Sustainability 11:4371
3. Breque M, De Nul L, Petridis A (2021) Industry 5.0: towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Luxembourg, LU
4. Yau K-LA, Lee HJ, Chong Y-W, Ling MH, Syed AR, Wu C, Goh HG (2021) Augmented intelligence: surveys of literature and expert opinion to understand relations between human intelligence and artificial intelligence. IEEE Access 9:136744–136761
5. Mikalef P, Gupta M (2021) Artificial intelligence capability: conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Inf Manag 58:103434

6. Yu K-H, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. Nat Biomed Eng 2:719–731
7. Kaul V, Enslin S, Gross SA (2020) History of artificial intelligence in medicine. Gastrointest Endosc 92:807–812
8. Lorkowski J, Grzegorowska O, Pokorski M (2021) Artificial intelligence in the healthcare system: an overview. In: Best practice in Health Care, pp 1–10
9. Farina E, Nabhen JJ, Dacoregio MI, Batalini F, Moraes FY (2022) An overview of artificial intelligence in oncology. Fut Sci OA 8:FSO787
10. Kumar V, Singh D, Kaur M, Damaševičius R (2021) Overview of current state of research on the application of artificial intelligence techniques for Covid-19. PeerJ Comput Sci 7:e564
11. Elsheikh AH, Saba AI, Panchal H, Shanmugan S, Alsaleh NA, Ahmadein M. Artificial intelligence for forecasting the prevalence of Covid-19 pandemic: an overview. In: Healthcare, vol 9, MDPI, p 1614
12. Abduljabbar R, Dia H, Liyanage S, Bagloee SA (2019) Applications of artificial intelligence in transport: an overview. Sustainability 11:189
13. Surden H (2018) Artificial intelligence and law: an overview. Ga St UL Rev 35:1305
14. Chassignol M, Khoroshavin A, Klimova A, Bilyatdinova A (2018) Artificial intelligence trends in education: a narrative overview. Procedia Comput Sci 136:16–24
15. Zhai X, Chu X, Chai CS, Jong MSY, Istenic A, Spector M, Liu J-B, Yuan J, Li Y (2021) A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity 2021
16. Leenen L, Meyer T (2021) Artificial intelligence and big data analytics in support of cyber defense. In: Research anthology on artificial intelligence applications in security, IGI Global, pp 1738–1753
17. Zhao S, Blaabjerg F, Wang H (2020) An overview of artificial intelligence applications for power electronics. IEEE Trans Power Electron 36:4633–4658
18. Lawal AI, Kwon S (2021) Application of artificial intelligence to rock mechanics: an overview. J Rock Mech Geotech. Eng 13:248–266
19. Ajitha P, Nagra A (2021) An overview of artificial intelligence in automobile industry-A case study on tesla cars. Solid State Technol 64:503–512
20. Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: methods and applications. J Manuf Syst 48:144–156
21. Kotsiopoulos T, Sarigiannidis P, Ioannidis D, Tzovaras D (2021) Machine learning and deep learning in smart manufacturing: the smart grid paradigm. Comput Sci Rev 40:100341
22. Rai R, Tiwari MK, Ivanov D, Dolgui A (2021) Machine learning in manufacturing and industry 4.0 applications
23. Huynh-The T, Pham Q-V, Pham X-Q, Nguyen TT, Han Z, Kim D-S (2022) Artificial intelligence for the metaverse: a survey. arXiv:2202.10336
24. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ (2020) A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Supervised and unsupervised learning for data science, pp 3–21
25. Sheikhpour R, Sarram MA, Gharaghani S, Chahooki MAZ (2017) A survey on semi-supervised feature selection methods. Pattern Recognit 64:141–158
26. Li Y (2017) Deep reinforcement learning: an overview. arXiv:1701.07274
27. Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2:1–20
28. Broschert S, Coughlin T, Ferraris M, Flammini F, Florido JG, Gonzalez AC, Henz P, de Kerckhove D, Rosen R, Saracco R et al (2019) Symbiotic autonomous systems: white paper iii
29. Wang K-J, Zheng CY, Mao ZH (2019) Human-centered, ergonomic wearable device with computer vision augmented intelligence for VR multimodal human-smart home object interaction. In: 2019 14th ACM/IEEE international conference on human-robot interaction (HRI). IEEE, pp 767–768
30. Hebbar A (2017) Augmented intelligence: Enhancing human capabilities. In: 2017 third international conference on research in computational intelligence and communication networks (ICRCICN), IEEE, pp 251–254

31. El Koujok M, Ragab A, Ghezzaz H, Amazouz M (2020) A multiagent-based methodology for known and novel faults diagnosis in industrial processes. IEEE Trans Ind Inform 17:3358–3366
32. Maettig B, Foot H (2020) Approach to improving training of human workers in industrial applications through the use of intelligence augmentation and human-in-the-loop. In: 2020 15th international conference on computer science & education (ICCSE). IEEE, pp 283–288
33. Hassani H, Silva ES, Unger S, TajMazinani M, Mac Feely S (2020) Artificial intelligence (AI) or intelligence augmentation (IA): what is the future? AI 1:8
34. De Felice F, Petrillo A, De Luca C, Baffo I (2022) Artificial intelligence or augmented intelligence? impact on our lives, rights and ethics. Procedia Comput Sci 200:1846–1856
35. Rada M (2020) Industry 5.0 definition. https://michael-rada.medium.com/industry-5-0-definition-6a2f9922dc48
36. Aslam F, Aimin W, Li M, Ur Rehman K (2020) Innovation in the era of IoT and industry 5.0: absolute innovation management (AIM) framework. Information 11:124
37. Xu X, Lu Y, Vogel-Heuser B, Wang L (2021) Industry 4.0 and industry 5.0-inception, conception and perception. J Manuf Syst 61:530–535
38. Javaid M, Haleem A (2020) Critical components of industry 5.0 towards a successful adoption in the field of manufacturing. J Ind Integr Manag 5:327–348
39. Maddikunta PKR, Pham Q-V, Prabadevi B, Deepa N, Dev K, Gadekallu TR, Ruby R, Liyanage M (2022) Industry 5.0: a survey on enabling technologies and potential applications. J Ind Inf Integr 26:100257
40. Nguyen HD, Tran KP, Castagliola P, Megahed FM (2022) Enabling smart manufacturing with artificial intelligence and big data: a survey and perspective. In: Advanced manufacturing methods. CRC Press, pp 1–26
41. Michael C, Bryce H, Alex S, Alex S (2021) The state of AI in 2021. https://www.mckinsey.com/business-functions/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021
42. He Z, Tran K-P, Thomassey S, Zeng X, Xu J, Yi C (2021) A deep reinforcement learning based multi-criteria decision support system for optimizing textile chemical process. Comput Ind 125:103373
43. Romeo L, Loncarski J, Paolanti M, Bocchini G, Mancini A, Frontoni E (2020) Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. Expert Syst Appl 140:112869
44. Yang M, Nazir S, Xu Q, Ali S (2020) Deep learning algorithms and multicriteria decision-making used in big data: a systematic literature review. In: Complexity 2020
45. Tamilselvan P, Wang P (2013) Failure diagnosis using deep belief learning based health state classification. Reliab Eng Syst Safety 115:124–135
46. Lu C, Wang Z, Zhou B (2017) Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv Eng Inform 32:139–151
47. Chen Z, Deng S, Chen X, Li C, Sanchez R-V, Qin H (2017) Deep neural networks-based rolling bearing fault diagnosis. Microelectron Reliab 75:327–333
48. AlThobiani F, Ball A et al (2014) An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks. Expert Syst Appl 41:4113–4122
49. Wang P, Yan R, Gao RX et al (2017) Virtualization and deep recognition for system fault classification. J Manuf Syst 44:310–316
50. Dong H, Yang L, Li H (2016) Small fault diagnosis of front-end speed controlled wind generator based on deep learning. WSEAS Trans Circ Syst 15:64–72
51. Yang Z-X, Wang X-B, Zhong J-H (2016) Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach. Energies 9:379
52. Serrano-Ruiz JC, Mula J, Poler R (2021) Smart manufacturing scheduling: a literature review. J Manuf Syst 61:265–287
53. Shiue Y-R, Lee K-C, Su C-T (2018) Real-time scheduling for a smart factory using a reinforcement learning approach. Comput Ind Eng 125:604–614

54. Waschneck B, Reichstaller A, Belzner L, Altenmüller T, Bauernhansl T, Knapp A, Kyek A (2018) Optimization of global production scheduling with deep reinforcement learning. Procedia Cirp 72:1264–1269
55. Gantz J, Reinsel D (2007) The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Anal Fut 2012:1–16
56. Agrawal A, Yadav P, Upadhyay C, Corney JR, Vasantha GA, Jagadeesan A, Lynn A (2015) A novel hybrid intelligence approach for 2d packing through internet crowdsourcing. In: 2015 IEEE technological innovation in ICT for agriculture and rural development (TIAR). IEEE, pp 33–39
57. Yu G, He Z, Lai C, Sun Y. An optimization design system with hybrid intelligence. In: Fifth World Congress on intelligent control and automation (IEEE Cat. No. 04EX788), vol 3. IEEE, pp 2790–2794
58. Hwang CH, Massey N, Miller BW, Torkkola K. Hybrid intelligence for driver assistance. In: FLAIRS conference, pp 281–285
59. Liu L, Guo F, Zou Z, Duffy VG (2022) Application, development and future opportunities of collaborative robots (cobots) in manufacturing: a literature review. Int J Hum Comput Interact 1–18
60. McCann J, Quinn L, McGrath S, O'Connell E (2018) Towards the distributed edge—An IoT review. In: 2018 12th international conference on sensing technology (ICST). IEEE, pp 263–268
61. Mourtzis D, Angelopoulos J, Panopoulos N (2021) Smart manufacturing and tactile internet based on 5G in industry 4.0: challenges, applications and new trends. Electronics 10:3175
62. De Alwis C, Kalla A, Pham Q-V, Kumar P, Dev K, Hwang W-J, Liyanage M (2021) Survey on 6G frontiers: trends, applications, requirements, technologies and future research. IEEE Open J Commun Soc 2:836–886
63. Zong B, Fan C, Wang X, Duan X, Wang B, Wang J (2019) 6G technologies: key drivers, core requirements, system architectures, and enabling technologies. IEEE Veh Technol Mag 14:18–27
64. O'Connell E, Moore D, Newe T. Challenges associated with implementing 5G in manufacturing. In: Telecom, vol 1, MDPI, p 5
65. Hu SJ (2013) Evolving paradigms of manufacturing: from mass production to mass customization and personalization. Procedia Cirp 7:3–8
66. Mladineo M, Zizic MC, Aljinovic A, Gjeldum N (2022) Towards a knowledge-based cognitive system for industrial application: case of personalized products. J Ind Inf Integr 27:100284
67. Yang B, Liu Y, Liang Y, Tang M (2019) Exploiting user experience from online customer reviews for product design. Int J Inf Manag 46:173–186
68. Zhan Y, Tan KH, Huo B (2019) Bridging customer knowledge to innovative product development: a data mining approach. Int J Prod Res 57:6335–6350
69. Zheng P, Sang Z, Zhong RY, Liu Y, Liu C, Mubarok K, Yu S, Xu X et al (2018) Smart manufacturing systems for industry 4.0: conceptual framework, scenarios, and future perspectives. Front Mech Eng 13:137–150
70. Jamie M, Michael Z, Daniel S (2013) The maintenance advantage in manufacturing. https://www.bcg.com/publications/2013/lean-maintenance-advantage-manufacturing-achieving
71. Çınar ZM, Abdussalam Nuhu A, Zeeshan Q, Korhan O, Asmael M, Safaei B (2020) Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability 12:8211
72. Achouch M, Dimitrova M, Ziane K, Sattarpanah Karganroudi S, Dhouib R, Ibrahim H, Adda M (2022) On predictive maintenance in industry 4.0: overview, models, and challenges. Appl Sci 12:8081
73. Wellsandt S, Klein K, Hribernik K, Lewandowski M, Bousdekis A, Mentzas G, Thoben K-D (2022) Hybrid-augmented intelligence in predictive maintenance with digital intelligent assistants. Ann Rev Control
74. Nguyen HD, Nadi AA, Tran KD, Castagliola P, Celano G, Tran KP (2022) The shewhart-type RZ control chart for monitoring the ratio of autocorrelated variables. Int J Prod Res 1–26

75. Nguyen HD, Tran KP (2020) Effect of the measurement errors on two one-sided Shewhart control charts for monitoring the ratio of two normal variables. Qual Reliab Eng Int 36:1731–1750
76. Maboudou-Tchao EM, Silva IR, Diawara N (2018) Monitoring the mean vector with Mahalanobis kernels. Qual Technol Quant Manag 15:459–474
77. Xia B, Jian Z, Liu L, Li L (2018) An effective multivariate control chart for detecting small mean shifts using support vector data description. Adv Mech Eng 10:1687814018810625
78. Villalba-Diez J, Schmidt D, Gevers R, Ordieres-Meré J, Buchwitz M, Wellbrock W (2019) Deep learning for industrial computer vision quality control in the printing industry 4.0. Sensors 19:3987
79. Zhou L, Zhang L, Konz N (2022) Computer vision techniques in manufacturing. IEEE Trans Syst Man Cybern: Syst
80. Lee J, Azamfar M, Singh J, Siahpour S (2020) Integration of digital twin and deep learning in cyber-physical systems: towards smart manufacturing. IET Collab Intell Manuf 2:34–36
81. Grieves M, Vickers J (2017) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Transdisciplinary perspectives on complex systems. Springer, pp 85–113
82. Son YH, Kim G-Y, Kim HC, Jun C, Noh SD (2022) Past, present, and future research of digital twin for smart manufacturing. J Comput Des Eng 9:1–23
83. Ramu SP, Boopalan P, Pham Q-V, Maddikunta PKR, Huynh-The T, Alazab M, Nguyen TT, Gadekallu TR (2022) Federated learning enabled digital twins for smart cities: concepts, recent advances, and future directions. Sustain Cities Soc 79:103663
84. Lattanzi L, Raffaeli R, Peruzzini M, Pellicciari M (2021) Digital twin for smart manufacturing: a review of concepts towards a practical industrial implementation. Int J Comput Integr Manuf 34:567–597
85. Lezzi M, Lazoi M, Corallo A (2018) Cybersecurity for industry 4.0 in the current literature: a reference framework. Comput Ind 103:97–110
86. Krundyshev V, Kalinin M. Prevention of cyber attacks in smart manufacturing applying modern neural network methods. In: IOP conference series: materials science and engineering, vol 940. IOP Publishing, p 012011
87. Martínez Torres J, Iglesias Comesaña C, García-Nieto PJ (2019) Machine learning techniques applied to cybersecurity. Int J Mach Learn Cybern 10:2823–2836
88. Mahdavifar S, Ghorbani AA (2019) Application of deep learning to cybersecurity: a survey. Neurocomputing 347:149–176
89. Ahmad HM, Rahimi A (2022) Deep learning methods for object detection in smart manufacturing: a survey. J Manuf Syst 64:181–196
90. Ha DT, Hoang NX, Hoang NV, Du NH, Huong TT, Tran KP (2022) Explainable anomaly detection for industrial control system cybersecurity. arXiv:2205.01930
91. Ding J, Tramel E, Sahu AK, Wu S, Avestimehr S, Zhang T (2022) Federated learning challenges and opportunities: an outlook. In: ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 8752–8756
92. Huong TT, Bac TP, Ha KN, Hoang NV, Hoang NX, Hung NT, Tran KP (2022) Federated learning-based explainable anomaly detection for industrial control systems. IEEE Access 10:53854–53872
93. Truong HT, Ta BP, Le QA, Nguyen DM, Le CT, Nguyen HX, Do HT, Nguyen HT, Tran KP (2022) Light-weight federated learning-based anomaly detection for time-series data in industrial control systems. Comput Ind 140:103692
94. Hao Y, Helo P (2017) The role of wearable devices in meeting the needs of cloud manufacturing: a case study. Robot Comput-Integr Manuf 45:168–179
95. Zheng XS, Foucault C, Matos da Silva P, Dasari S, Yang T, Goose S (2015) Eye-wearable technology for machine maintenance: effects of display position and hands-free operation. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp 2125–2134

96. Nguyen HD, Tran KP, Zeng X, Koehl L, Tartare G (2020) An improved ensemble machine learning algorithm for wearable sensor data based human activity recognition. In: Reliability and statistical computing. Springer, pp 207–228
97. Kucukoglu I, Atici-Ulusu H, Gunduz T, Tokcalar O (2018) Application of the artificial neural network method to detect defective assembling processes by using a wearable technology. J Manuf Syst 49:163–171
98. Barth J (2013) The human cloud: wearable technology from novelty to production. The open cloud company, Rackspace
99. Godfrey A, Hetherington V, Shum H, Bonato P, Lovell N, Stuart S (2018) From a to z: wearable technology explained. Maturitas 113:40–47
100. Kong XT, Luo H, Huang GQ, Yang X (2019) Industrial wearable system: the human-centric empowering technology in industry 4.0. J Intell Manuf 30:2853–2869
101. Lu Y, Zheng H, Chand S, Xia W, Liu Z, Xu X, Wang L, Qin Z, Bao J (2022) Outlook on human-centric manufacturing towards industry 5.0. J Manuf Syst 62:612–627
102. Peruzzini M, Grandi F, Cavallaro S, Pellicciari M (2021) Using virtual manufacturing to design human-centric factories: an industrial case. Int J Adv Manuf Technol 115:873–887
103. Jawahir I, Bradley R (2016) Technological elements of circular economy and the principles of 6R-based closed-loop material flow in sustainable manufacturing. Procedia Cirp 40:103–108
104. Ching NT, Ghobakhloo M, Iranmanesh M, Maroufkhani P, Asadi S (2021) Industry 4.0 applications for sustainable manufacturing: a systematic literature review and a roadmap to sustainable development. J Clean Prod 130133
105. Huang H-Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H, McClean JR (2021) Power of data in quantum machine learning. Nat Commun 12:1–9
106. Malina R, Woerner S (2019) Exploring quantum computing use cases for manufacturing. IBM Corporation
107. Margetis G, Ntoa S, Antona M, Stephanidis C (2021) Human-centered design of artificial intelligence. In: Handbook of human factors and ergonomics, pp 1085–1106
108. Riedl MO (2019) Human-centered artificial intelligence and machine learning. Hum Behav Emerg Technol 1:33–36
109. Xu W (2019) Toward human-centered AI: a perspective from human-computer interaction. Interactions 26:42–46
110. Bond RR, Mulvenna MD, Wan H, Finlay DD, Wong A, Koene A, Brisk R, Boger J, Adel T. Human centered artificial intelligence: weaving UX into algorithmic decision making. In: RoCHI, pp 2–9
111. Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. Artif Intell 267:1–38
112. Zhu J, Liapis A, Risi S, Bidarra R, Youngblood GM. Explainable AI for designers: a human-centered perspective on mixed-initiative co-creation. In: 2018 IEEE conference on computational intelligence and games (CIG). IEEE, pp 1–8
113. Gunning D, Vorm E, Wang JY, Turek M (2021) Darpa's explainable AI (XAI) program: a retrospective
114. Lv Z, Qiao L, Li Y, Yuan Y, Wang F-Y (2022) Blocknet: beyond reliable spatial digital twins to parallel metaverse. Patterns 3:100468
115. Ning H, Wang H, Lin Y, Wang W, Dhelim S, Farha F, Ding J, Daneshmand M (2021) A survey on metaverse: the state-of-the-art, technologies, applications, and challenges. arXiv:2111.09673
116. Lee L-H, Braud T, Zhou P, Wang L, Xu D, Lin Z, Kumar A, Bermejo C, Hui P (2021) All one needs to know about metaverse: a complete survey on technological singularity, virtual ecosystem, and research agenda. arXiv:2110.05352
117. Wang Y, Su Z, Zhang N, Xing R, Liu D, Luan TH, Shen X (2022) A survey on metaverse: fundamentals, security, and privacy. IEEE Commun Survey Tutor

Quality Control for Smart Manufacturing in Industry 5.0

Huu Du Nguyen, Phuong Hanh Tran, Thu Ha Do, and Kim Phuc Tran

Abstract Smart manufacturing is widely accepted as the new emerging transformation of the manufacturing industry today. In addition, quality control, an important aspect that contributes to the successful process of smart manufacturing attracts attention from the community. However, there are certain challenges in implementing quality control methods in Industry 5.0. Thus, this chapter aims to provide a comprehensive background review of important notions and advanced techniques related to quality control for smart manufacturing such as Machine Learning, computer vision, the Internet of Things, and Artificial Intelligence. Then, several difficulties and opportunities in the implementation of these techniques for quality control in Industry 5.0 are discussed. Finally, a case study on monitoring wine production in the food industry is also considered to show the performance of Machine Learning-based techniques for quality control.

Keywords Human-centric quality control · Explainable artificial intelligence · Industry 5.0 · Computer vision · Anomaly detection

1 Introduction

It has been more than a decade since human being pursued the fourth industrial revolution, or Industry 4.0. Since that time, industrial production has undergone tremendous changes. Automated production is transformed into smart manufacturing thanks to the application of artificial intelligence (AI) algorithms and other advanced technolo-

H. D. Nguyen
International Research Institute for Artificial Intelligence and Data Science, Dong A University, Danang, Vietnam

P. H. Tran (✉)
HEC Liège - Management School of the University of Liège, Liège, Belgium
e-mail: phtran@uliege.be

T. H. Do · K. P. Tran
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_3

gies such as Autonomous Robots, the Internet of Things, 3D Printing, Cloud Computing, Virtual Reality, Augmented Reality, Digital Twins, and Cyber-Physical systems. The aim of Industry 4.0 is to achieve automated decision-making and machinery interconnection, increase productivity across the value chain and enable the efficient production of goods.

Along with improving productivity and quality in industrial production, many cutting-edge technologies and advanced techniques have been invented and introduced in Industry 4.0. However, the application of these achievements to production is not straightforward, especially for small and medium enterprises. Even in many countries around the world, the new-age technologies of the 4.0 revolution are still just a destination or "the next" that has never been experienced or applied. While many companies, businesses, and industries are still in the development phase of the fourth revolution, the next revolution, or Industry 5.0 is already well on its way. According to Breque et al. [1], Industry 5.0 will "provide a vision of an industry that aims beyond efficiency and productivity as the sole goals" and "place the well-being of the worker at the center of the production process and uses new technologies to provide prosperity beyond jobs and growth while respecting the production limits of the planet". That is, rather than focusing on developing advanced technologies that can enhance the efficiency and productivity of industrial manufacturing, Industry 5.0 brings a new paradigm, emphasizing the collaboration between humans and machines and placing human factors at the heart of the revolution. By considering human well-being and the limited resources on the planet, it aims for a manufacturing industry with three crucial keywords: sustainable, human-centric, and resilient.

There is one important thing that needs to be clarified from the introduction of Industry 5.0. Unlike the previous industrial revolutions, it does not replace the implementation of Industry 4.0. Instead, its main idea is to promote the advanced technologies of Industry 4.0 and bring benefits to human beings. The common point between the two revolutions is smart manufacturing with high productivity and perfect products. This is especially important for businesses, especially in today's competitive global economy. In this study, we would like to pay attention to the second factor, which is to ensure product quality through quality control.

In industrial manufacturing, quality control (QC) has been considered a key to the success of any business nowadays. It is an important task to improve the quality of the production process and final products. The primary goal of quality control is to make sure that the products produced are up to standards and to eliminate unsatisfactory products during the production process. Obviously, manufacturers do not want to send defective products out for purchase, they do not want to waste time and money due to faulty products. In such a mission, the main objectives of quality control include enhancing product quality and reducing risks, gaining production efficiencies, and then garnering customer loyalty. It is thus desirable that all manufactured products should be put under strict quality control processes.

There are certain challenges in the quality control methods in the new context of Industry 5.0. In this chapter, we aim to provide a survey of important problems related to quality control for smart manufacturing, including machine learning, computer vision, the Internet of Things, and AI. We discuss several difficulties and opportunities for these techniques for quality control in Industry 5.0. A case study is also given to demonstrate the ML-based method for quality control.

The rest of the chapter is organized as follows. In Sect. 2, we summarize briefly the concept of quality control in industrial manufacturing. Section 3 presents the application of machine learning and computer vision techniques for quality control. Section 4 is to discuss some difficulties and opportunities for AI-based quality control in Industry 5.0. A case study is given in Sect. 5 where we present a machine learning-based method to monitor the wine quality in the food industry. Finally, Sect. 6 is for some concluding remarks.

2 Quality Control in Industrial Manufacturing

In the past decade, the manufacturing industry (MI) has made great strides in the context of Industry 4.0 which is a term denoted for The Fourth Industrial Revolution. The first appearance of this term at Hanover Fair, Germany in 2011 [2] has been seen as a sign of the beginning of a new era in which the industry transitioned to the new evolutionary step of high technologies and robotization of the production process. From the innovations in Industry 4.0, the MI has opportunities to gain significant benefits from various emerging technologies such as additive manufacturing [3], digital twin [4], virtual reality [5], augmented reality [6], artificial intelligence (AI) [7], Internet of Things (IoTs) [8], computer vision [9], etc. Together with these developments, quality control (QC) has been seen as a key issue contributing to the success of the implementation of these technologies in the MI by reducing cost [10], improving product quality [11], and increasing productivity [12]. From the point of view of Total Quality Management (TQM), QC is a significantly important factor in process control. In this regard, without QC in manufacturing should lead to unfathomable consequences for the processes such as (1) There are neither standards nor tools to monitor product quality, (2) The appearance of more and more defective products as a result of the out-of-control state of quality, (3) Opportunity for early warning of signs of instability in the production process related to quality deterioration is at risk of being missed, (4) The increase in the maintenance, operating, disposal, and rework costs regarding defective products, (5) Effects of poor quality products on customer satisfaction and loyalty. Furthermore, it is clear that there is a gradual shift from industry 4.0 to 5.0 today to be suitable for the innovations of advanced techniques as well as the demands of society in the new era. This move not only evolves to a new step of MI, but QC also has to evolve accordingly. Therefore, it would be helpful for both researchers and practitioners to explore more about the development history of QC in the manufacturing process as well as related issues such as notions and classification. A brief review of QC in the MI will be presented in these subsections as follows.

2.1 The History of Quality Control

The timeline for the development of QC can be divided into two big periods (1) From the Middle Ages to the late 1970s, (2) From the early 1980s till current. Two stages are described as follows.

2.1.1 From the Middle Ages to the Late 1970s

QC has been an issue that has attracted attention from leaders since the Middle Ages. That is the commitment to quality shown in great historical works like the pyramids or exquisite handcrafted products through activities such as construction and crafting [13]. According to [14], from the Middle Ages to the late 1970s, QC has gone through six stages including the operator quality control period, the foreman quality control period, the inspection quality control period, the statistical quality control, the total quality control, and the total quality control organization-wide phase. Each period is described in more detail in the following Fig. 1, we just discuss here prominent events during these ones (Table 1).

It is important to note that the foundations of statistical aspects of QC were developed during the 1920s. A variety of scientific results in this period plays a fundamental role in statistical process control. One of the notable leading authors is Shewhart who is the father of statistical charts to control the variables of a product [15]. These tools are also named control charts or Shewhart control charts. The others are H. F. Dodge and H. G. Romig with the idea of acceptance sampling plans. The appearance of these plans is an alternative to 100% inspection. Although three of them work at Bell Telephone Laboratories in the U.S., their research has not received the attention of the mother country.

In the early 1950s, products made in the West countries, especially from the United States (US) are famous for their top quality while being considered low quality if made in Japan [16]. However, then the world had to change its prejudices about

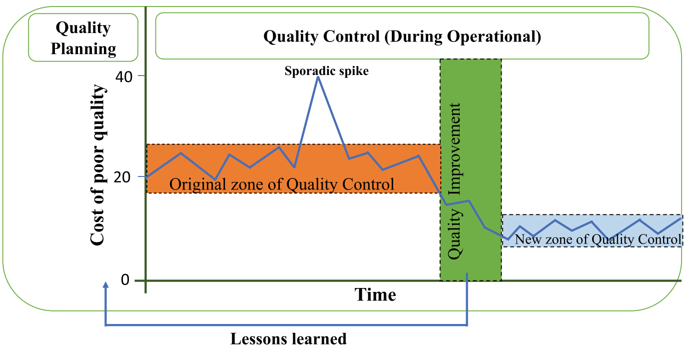


Fig. 1 Juran Trilogy diagram. Adapted from the [29]

Table 1 The evolution of quality control from the middle ages to the late 1970s

N.	Milestone	Period name	Characteristics of production process	Quality control is responsible by
1	From the middle ages to 1900	Operator quality control	– The products were created directly by a person or a small group. – Limited quantity of certain products in production.	Operator
2	From the beginning of the 20th century to 1920	Foreman quality control	– Emergence of industrial revolution. – A worker was responsible for only a part of the production line instead of an entire product. – Mass production-based specialization of labor.	Foremen or supervisor
3	From 1920 to 1940	Inspection quality control	– More complicated products and processes. – Increased production volume. – Quality standards were set. – Substandard products are discarded and reworked.	Inspectors
4	From 1940 to 1960	Statistical quality control	– Product demand increased during World War 2. – Recognized sampling principle instead of 100% inspection. – The American Society for Quality Control (ASQC) foundation and then renamed to the American Society for Quality (ASQ). – A set of sampling inspection plans for attributes called MIL-STD-105A was developed by the military in 1950. Modified versions are MIL-STD-105B, MIL-STD-105C, MIL-STD-105D, MIL-STD-105E, and MIL-STD-414. – Japan was deeply influenced by the philosophies of Statistical Quality Control after the teaching visits of Deming (1950) and Juran (1954). They conducted quick training and education program about QC. – Quality Control and Reliability Handbook H-107 (1958) developed by the Department of Defense (DOD), which dealt with single-level continuous sampling procedures and tables for inspection by attributes. Revised Quality Control and Reliability Handbook H-108 (1959) covered multilevel continuous sampling procedures as well as topics in life testing and reliability.	Worker or Production foreman, or people from the inspection and quality control department
5	During the 1960s	Total quality control	– Involvement of various departments and managers in the quality control. – Emergence of zero defects concept which centered around achieving productivity through worker involvement. – Increase in using Quality circles in Japan based on the participative style of management with an uplift of morale and motivation achieved through consultation and discussion in informal subgroups.	Several departments and management personnel

(continued)

Table 1 (continued)

N.	Milestone	Period name	Characteristics of production process	Quality control is responsible by
6	During the 1970s	Total quality control organization-wide	– A quality system is the agreed-on companywide and plantwide operating work structure, documented in effective, integrated technical and managerial procedures, for guiding the coordinated actions of the people, the machines, and the information of the company and plant in the best and most practical ways to assure customer quality satisfaction and economical costs of quality. – The expanded use of a graphical tool known as the cause-and-effect Diagram (i.e. fishbone diagram, Ishikawa diagram) in Japan developed by K. Ishikawa	Everyone in the company, from the operator to the first-line supervisor, manager, vice president, and even the chief executive officer

Japanese products with the appearance of outstanding quality products from Japanese brands such as Xerox, Motorola, etc. The secret to this magical transformation is that Japan has made great strides and changes in QC to have high-quality products. It is important to note that the activities of Deming and Juran contributed significantly to the success of QC implementation in Japan. Deming introduced his lectures on statistical quality control to Japanese engineers and managers. According to Deming, statistical quality control is the application of statistical principles and techniques in all stages of production directed toward the most economic manufacture of a product that is maximally useful and has a market. He also emphasized that QC involves uniform production of a product that meets the needs of the consumer [17]. Furthermore, Juran highlighted to management the benefits of making QC part of the overall management philosophy and strategy [16]. With the increasing product demand, the U.S. quickly realized the new requirements in sampling plans. In order to respond to these demands, a Quality Control and Reliability Handbook H-107 is introduced by the Department of Defense (DOD) in 1958. It guides practitioners dealt with single-level continuous sampling procedures and tables for inspection by attributes. Then, a revised version namely Quality Control and Reliability Handbook H-108 in 1959 covered multilevel continuous sampling procedures as well as topics in life testing and reliability.

The 1960s marked the widespread implementation of total quality management and the involvement of various departments and managers in the QC process instead of only the inspection and QC department being responsible for it as before. Meanwhile, the appearance of the zero defects concept centered around achieving productivity through worker involvement is implemented very successfully, especially for critical products and assemblies. Furthermore, a quality circle is a concept that started to be widely applied in Japan. This notation indicates that the open discussion between members of the organization such as operators, supervisors, managers, and other members, together with a participative style of management, will help the quality of the product or service be improved. Feigenbaum [14] cited that the 1970s is the period of total quality control organization-wide in which the quality of a product or service is produced in relation to the entire staff inside the organization. Moreover, a tool found by Ishikawa in 1943 [18], namely a fishbone diagram or a cause and effect diagram or Ishikawa diagram, is popularized in this period in Japan. This graph helps the practitioners to find possible causes that lead to out-of-control state and their effects on the procedure.

2.1.2 From the Early 1980s Till Current

During the 1980s, QC is divided into areas such as vendor quality control, product design assurance, product and process quality audit, and related areas. Quality control software gradually became widely used. Furthermore, big companies began to find ways to implement statistical quality control in their factories. Thus, it is no surprise that there is an increase in the need to establish official guidelines or standard settings for quality regulation in companies in the U.S. The appearance of quality systems standards such as ISO 9000, QS-9000, and the Baldrige National Quality Award standards fill this lack. One of the quality systems standards widely known as ISO was developed by the International Organization for Standardization with five versions from ISO 9000 to ISO 9004 in the late 1980s. They were revised in 1994, 2000, and 2008 [17], respectively.

From the last decade of the 20th century to now, the blooming of information and computational technology, IoTs, and Big Data affected the development of QC. There is an urgent need in analyzing the huge data collected from the internet or equipment (IoTs) in real-time. Thus, the current century will persistently witness a drive in the expansion of QC and improvement techniques that can seamlessly integrate data using technology and perform real-time analysis with zero tolerance for errors. Currently, the industry has begun to enter a new phase called the industrial revolution 5.0. During this period, there is various changes in the characteristic of the manufacturing industry that is different from the previous decades. Industry 5.0 concentrated on aspects such as human-centricity, sustainability, and resiliency [19] while Industry 4.0 focused on aspects such as interoperability, decentralization, virtualization, real-time capability, modularity, and service orientation [20]. Together with the development from Industry 4.0 to Industry 5.0, MI is currently gradually consolidating into a new form known as Smart Manufacturing (SM) [21]. This complex concept refers to the manufacturing of the future which consists of both object and process integration, incorporating present and future manufacturing resources equipped with sensors, computing platforms, communication technology, data-intensive modeling, control, simulation, and predictive engineering, see Kusiak [22]. Generally, SM relates to terms such as cyber-physical systems, IoT, cloud computing, service-oriented computing, AI, and DS that aims to optimize and improve the efficiency, flexibility, and competitiveness of the MI. As a result, innovation changes in SM lead to adaptive improvement in QC in various areas such as Additive Manufacturing, DNA plasmid Manufacturing, Food Industry, etc. Therefore, QC has made adjustments including new models, concepts, and methodologies to adapt to the SM in a new era.

2.2 Quality Control Definition

Activities related to QC can be seen from time immemorial. However, this term has only been discussed in the scientific community since the 1920s. Much of the work to clarify QC notation has been performed, which has shown it is a broad concept.

From the beginning of the 1920s, the term control of quality is first introduced in the publication of Radford [23]. In the early years of this decade, the concept had a rather broad connotation including quality planning and inspection at a particular stage. Later, it was narrowed to the concept of QC. Both Shewhart [24] and Deming [25] considered QC as a prevention aspect. Early detection of defective products in the process is not real control. Taking corrective action to eliminate the causes of the problem is a necessary part of the control. These actions may involve aspects well beyond the immediate manufacturing stage such as a change in managerial approach, better vendor relationships, design improvement, clearer specifications, etc. This is a major aspect that is discussed in the concept of Japanese company-wide QC. Besides, these authors also mentioned the statistical quality control notion involves statistical methods used in QC. It is important to note that the reliability concept refers to quality at the time of test than throughout life in QC implementation. According to ANSI\ ASQC A3 (American National Standards Institute\American Society for Quality Control) standard in 1978 [26], QC is identified as the operational techniques and the activities which sustain a quality of product or service that will satisfy given needs; also the use of such techniques and activities. Then, Alford and Beatty [27] defined QC as an industrial management technique or group of techniques by means of which products of uniformly acceptable quality are manufactured. It is indeed the mechanism by which products are made to measure up to specifications determined by customer's demand and transformed into sales, engineering, and manufacturing requirements. It is concerned with making things right rather than discovering and rejecting those made wrong. Freund [28] noted that in some industries and countries, QC is used in a total or company-wide sense of guiding, directing, or managing all aspects affecting quality while in others, QC is thought of as being equivalent to an inspection operation. According to Juran [29], QC is a regulatory process that includes a variety of activities that are closely related to each other. In particular, the actual quality performance of the process is measured and compared with standards and then appropriate response actions will be conducted. In addition, based on the total quality management (TQM) approach, he also proposed a diagram related to this notation namely Juran Trilogy Diagram illustrated in Fig. 1. It is used to describe the interrelation among quality planning, quality improvement, and QC and the fundamental managerial processes. Jura considered QC as one of the stages in the quality management process for conducting operations to prevent adverse change and maintain a steady state. Figure 1 has shown that a number of actions need to be taken to improve the performance of the process to avoid waste even though the process is in control. After numerous quality improvement actions were taken, the process was controlled at a better quality performance level. Kumar and Suresh [30] defined QC as a system that is used to maintain a desired level of quality in a product or service. It is a systematic control of various factors that affect the quality of the product. It depends on materials, tools, machines, type of labor, working conditions, etc. According to Eldin [31], QC is a universal managerial process for conducting operations so as to provide stability to prevent adverse change and maintain a stable status. To maintain stability, the QC process evaluates actual performance, compares actual performance to goals, and takes action on the difference. QC is one of the three basic managerial processes through which quality can be managed. The others are quality planning and quality improvement. Following the idea of Kumar and Suresh [30], Mitra [13] made additional contributions to the QC concept is defined as a system that maintains a desired level of quality, through feedback on product/service characteristics and implementation of remedial actions, in case of a deviation of such characteristics from a specified standard. It is divided into three main sub-areas: offline QC, statistical process control, and acceptance sampling plans. Aft [17] highlighted that quality future production is the goal of QC activity. QC aims at the prevention of defects at the source and relies on an effective feedback system and corrective action procedure. QC uses inspection as a valuable tool.

In Europe's countries as well as in the United States, the term QC is now used in a narrower sense than previous definitions. A European interdisciplinary organization has recently a new name as European Organization for Quality (EOQ) instead of the European Organization for Quality Control to European Organization for Quality (EOQC) founded in 1954. The term total quality control (TQC) was replaced by total quality management (TQM) by The Union of Japanese Scientists and Engineers (JUSE) in 1997.

2.3 Quality Control Methods Classification

Depending on the category of product or type of testing, there are various types of QC methods used in these processes. For example, QC testing methods developed and validated for different areas of plasmid-based pharmaceuticals are different. In order to test raw materials, compendial methods obtained from the U.S. Pharmacopoeia viewed as an offline quality control method, are used to verify the identity and quality of each raw material must be in place or a thorough vendor qualification program established by the end of Phase III clinical trials. Meanwhile, QC testing of pure bulk plasmid must be capable of determining plasmid identity, purity, sterility, and potency for Pure bulk plasmid [32]. Based on our current knowledge, QC methods can be divided into three main sub-areas, consisting of (1) Online quality control, (2) Offline quality control, and (3) Acceptance sampling plans.

- **Online quality control methods**: These QC activities are conducted at the manufacturing stage to reduce defects in the manufacturing process and keep it in statistical control. One of the most common online quality control methods is in-process inspection which involves the inspection of the product during the production process to identify potential issues. In-process inspection can be performed by trained personnel or by automated systems, such as cameras or other imaging equipment. This type of inspection is particularly useful for detecting deviations from the specified requirements and ensuring that the production process is under control. Another common online quality control method is process monitoring. Specialized equipment are used to continuously monitor the production process and identify potential issues. For example, temperature sensors can be used to monitor the temperature of a production process, while vibration sensors can be used to monitor the stability of the equipment. The data collected by the process monitoring equipment can be analyzed in real-time to detect deviations from the specified requirements and identify potential problems with the production process. A third common online quality control method is statistical process control (SPC) with tools such as control charts and Ishikawa diagrams. SPC is a data-driven approach to quality control that involves the use of statistical methods to monitor and control the production process. SPC is based on the principle that a process will produce predictable results if it is under control. The data collected during the production process is analyzed using statistical methods to identify potential sources of error and determine if the production process is under control. SPC can be used to identify trends in the production process and make data-driven decisions to improve product quality. Another important online quality control method is process control. Process control involves the use of automated systems to control the production process and ensure that it is operating within specified parameters. For example, a temperature control system can be used to maintain the temperature of a production process within a specified range. Process control systems can be programmed to respond to changes in the production process, making adjustments as needed to maintain the desired process conditions.
  Online quality control methods play the role of identifying and resolving issues with the production process before the final product is produced. These actions help to eliminate the amount of waste generated by the production process and improve product quality as well as improve the efficiency of the production process by reducing the need for manual inspections and shortening the time required to resolve quality issues. These methods are also considered the useful tools to identify trends in the production process and make data-driven decisions to improve product quality.
- **Offline quality control methods**: The offline quality control methods are defined as systematic methods of optimizing production processes and product designs. This approach is tended to select a controllable product and process parameters in such a way that the deviation between the product or process output and the standard will be minimized. They aim to improve product manufacturability and reliability in order to reduce product development and lifetime costs. Kackar [33] stressed that quality and cost control activities are conducted at the product and the

process design stages in the product development cycle based on offline quality control methods. Some of the offline quality control methods such as Taguchi method, design reviews, sensitivity analyses, prototype tests, accelerated life tests, and reliability studies are used by leading manufacturers of high-quality products. However, these methods are neither as thoroughly developed nor as widely applied as online quality control methods. It is widely used in Japan to produce high-quality products at low cost. Phadke et al. [34] proposed four keys steps of off-line quality control as: (1) Identify important process factors that can be manipulated and their potential working levels; (2) perform fractional factorial experiments on the process using orthogonal array designs; (3) analyze the resulting data to determine the optimum operating levels of the factors; (4) conduct an additional experiment to verify that the new factor levels indeed improve the quality control.

One of the benefits of offline quality control methods is that they help to ensure that the final product meets the specified requirements and is of high quality. This can improve customer satisfaction and reduce the risk of product returns or complaints. Offline quality control methods can also help identify and resolve any quality issues with the production process, improving product quality and reducing waste. Like online quality control methods, these offline quality control ones are also helpful in improvement the efficiency of the production process by reducing the need for manual inspections and reducing the time required to resolve quality issues. Automated systems, such as cameras and test equipment, can be used to perform offline quality control methods, reducing the need for manual labor and increasing the speed and accuracy of the inspection process.

- **Acceptance sampling plans**: A statement regarding the required sample size for product inspection and the acceptance or rejection criteria of the lot based on meeting certain stipulated conditions for sentencing lot. Conducting 100% inspections is not always necessary and can be costly in various manufacturing cases, such as the production of weapons, forged products, food, etc. Acceptance sampling plans are used in a variety of industries, including not only manufacturing but also construction and services. This technique is first applied by the U.S. military for testing bullets during World War II. It is one of the practical tools of SPC used in classical quality control applications today. Acceptance sampling plans are based on statistical methods and involve the use of probability. The sample size is selected based on the desired level of confidence and the desired level of risk of accepting a non-compliant batch. The level of confidence is the degree of certainty that the sample accurately represents the batch. The level of risk is the probability that a non-compliant batch will be accepted. There are two main types of acceptance sampling plans: single sampling and double sampling. Single sampling is a method where only one sample is taken from the batch and tested for compliance with the quality standards. The decision to accept or reject the batch is based on the results of this single sample. Single sampling is used when the cost of testing each item in the batch is high and when the risk of accepting a non-compliant batch is low. Double sampling is a method where two samples are taken from the batch and tested for compliance with the quality standards. The first sample is used to determine the acceptability of the batch. If the first sample is found to be in compliance with the quality standards, the batch is assumed to be in compliance and no further action is taken. However, if the first sample is found to be non-compliant, a second sample is taken and tested. The decision to accept or reject the batch is based on the results of the second sample. Double sampling is used when the cost of testing each item in the batch is low and when the risk of accepting a non-compliant batch is high. Acceptance sampling plans provide several benefits, including (1) Improved Quality, (2) Reduced Testing Costs, (3) Improved Efficiency, and (4) Increased Confidence

It important to note that QC methods can be classified based on the type of testing performed or the level of product evaluation. These classifications are useful for understanding the different approaches to quality control and selecting the most appropriate method for a particular production process. Understanding the different classifications of QC methods is essential for ensuring consistent product quality and improving the production process.

3 Machine Learning and Computer Vision for the Quality Control

As discussed in Villalba-Diez et al. [35], the task of controlling the quality of products in modern industrial manufacturing should be performed automatically. It would be very difficult to achieve efficiency if it relied heavily on humans. Workers often tend to find them difficult to concentrate and are tired of a task that is repeated for a long time continuously. In such situations, even simple visual defects may not be recognized for correct classification. In the traditional inspection processes, the first step is to determine features of products relevant to the inspection like edges, curves, corners, color patches, etc. When a new type of defect occurs, its features are often described manually by engineers. Based on these features, a rule-based system is created to classify an object. The resulting system is then applied to automatically identify if a product is up to quality standards. Although these methods have been effective in some situations, there are cases in which they are ineffective. In fact, there are always unknown faults that cannot be determined in advance and it is impossible to cover all kinds of errors as they can come from any type of size, shape, or orientation. The task of inspection then requires applying advanced methods to solve the problem.

Since the beginning of Industry 4.0, industrial manufacturing has transferred to digitalization. Many companies and businesses have started the process by visualizing the data they stored. For the image data captured from cameras attached in the stages of the manufacturing process, computer vision technology is a powerful tool to handle these data. The use of computer vision allows manufacturers to achieve improved quality inspection, an important part of quality control, at speeds, latency, and costs beyond the capabilities of human inspectors.

Computer vision refers to the computer-based approaches to gaining insight into visual data like digital images or videos. It includes many steps to extract key features of visual data from collecting and processing to analyzing digital images. Conceptually, this technology tries to mimic parts of the human visual system to allow computers to identify and process images or videos in the same way that humans do. In recent years, there have been significant advances in computer vision techniques thanks to the application of machine learning algorithms. Computer vision methods now are able to surpass humans in some difficult tasks related to detecting and labeling objects. Rather than having a system relying on the rules created by the experts and engineering, the machine learning-based computer vision system can learn automatically key features from data and create rules to define quality products. With machine learning methods, it is not necessary to cover manually a machine vision model for all the possible production situations. Numerous problems can be solved by using machine learning-based computer vision such as segmentation, feature extraction, refining visual models, pattern matching, shape representation, and surface reconstruction. These problems are reflected in many practical problems of different fields such as engineering, medicine, agriculture, astronomy, sports, and education [36].

In fact, manufacturers have been using computer vision for years, but only when applying machine learning algorithms, especially deep learning algorithms, can computer vision technology's efficiency be significantly increased. Previously, due to the limitations of computing technology as well as computer memory, it is a challenge to apply the deep learning method to develop computer vision systems. Researchers focus mainly on traditional machine learning algorithms like K-means, Naive Bayes classifier, Decision Tree, Boosting, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). However, as the computing power of the computer is increased, deep learning algorithms have revolutionized computer vision technology. In particular, the use of deep learning algorithms like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNNs), Deep Belief Networks (DBNs), Restricted Boltzmann Machines (RBMs), and Autoencoders can fuel great applications in many different computer vision problems, such as object detection, motion tracking, action recognition, and human pose estimation. A brief review of deep learning for computer vision can be seen in Voulodimos et al. [37]. Another critical review of emerging deep learning techniques and application scenarios has been presented in Chai et al. [38] where the authors focused on four key tasks of computer vision, including recognition, visual tracking, semantic segmentation, and image restoration.

It is now widely applied to quality control in different manufacturing industries. Zhou et al. [39] discussed several applications of computer vision for inspection in mechanics, automotive, and 3-D printing. Banús et al. [40] proposed a deep-learning-based solution to automatically perform the quality control of the sealing of matrix-shaped thermoforming food packages while satisfying production cadence. Three deep-learning architectures which are commonly used for image classification have been considered, namely ResNet, VGG, and DenseNet. A survey on fabric defect detection techniques based on computer vision for the inspection of real fabric defects was conducted in Kumar [41]. Rajan et al. [42] built a self-learning CNN architecture to develop a vision system that uses cameras attached to the raspberry pi model to capture the image, analyze and identify the defects of bolts.

In Industry 5.0, the human factor is placed at the center of manufacturing, leading to the concept of human-centric manufacturing. However, in computer vision, the term “human-centric” is being considered in a different sense. In general, human-centric computer vision refers to techniques designed to interpret visual data about humans for various tasks. Simply, this is the adoption of computer vision to detect, identify and understand humans from camera imagery. This understanding can be found in several studies, for example, Ebadi et al. [43, 44]. Obviously, this interpretation needs to be clarified to avoid confusion with a more general term of human-centric manufacturing. In Industry 5.0, it could be understood as the association with humans, serving humans to enhance the performance of production process monitoring and quality control. Humans can label a certain number of data based on their experience to feed the ML-based computer vision models. In turn, these models can support humans in automatically monitoring the production process and detecting defects or anomalies, supporting humans to make the final decisions. In the next sections, we will discuss the human role in quality control in Industry 5.0.

4 Dificulties, Challenges, and Perspectives for AI-Based Quality Control for Smart Manufacturing in the Industry 5.0

In Industry 5.0, as smart manufacturing is associated with the human factor; aiming at human-centric production, the requirements for quality control are becoming more and more important. The application of advanced technologies in the quality control process is a necessity. This not only brings advantages through improving controlling efficiency but also leads to difficulties and challenges for enterprises. In this section, we discuss some difficulties and opportunities for AI-based quality control for Smart Manufacturing in Industry 5.0.

4.1 The Use of Anomaly Detection Techniques for Quality Control

One of the most important problems in quality control is to analyze directly the products along the production chain. Smart manufacturing integrates hardware, software, and communication technologies to optimize the production process which allows for increasing productivity and reducing the cost. In essence, quality control is the supervision and monitoring to promptly detect abnormalities in the production process. A good product is one that is manufactured under standard conditions with no defects, meeting the strict requirements of customers. Timely detection of abnormalities is a prerequisite to ensure product quality. By detecting abnormalities that arise in the manufacturing process, manufacturers can take appropriate actions to find out and correct the causes of these abnormalities as well as limit the number of defects. That is to say, the contribution of anomaly detection techniques to quality control goes beyond detecting defective products. It also allows monitoring of the whole manufacturing system and operations, preventing the causes that can adversely affect the production process.

In the quality control literature, most anomaly detection techniques' applications focus on quality inspection, an integral part that inspects the quality of the actual manufactured products. This technique has been applied in many industries. For example, in the polymer production industry, Peng et al. [45] built a multi-class classifier method based on a CNN with InceptionV3, VGG16, and ResNet50 for pellet quality inspection. In the pharmaceutical industry, Mac and Hung [46] proposed an automated pill quality inspection method that fully automates the image analysis of internal crack/contamination detection using deep learning algorithms and computer-vision-based processing. In the automobile industry, Yang et al. [47] presented a lightweight deep learning model for the inspection of laser welding defects on the safety vent of the power battery. The authors combined CNN with the technique of transfer learning, i.e. a pre-trained SqueezeNet, to create a model with a small model size and low computation complexity. The same idea of using CNN and transfer learning (but with different architectures) was also applied to quality inspection of different sugarcane varieties in the agricultural industry in Alencastre-Miranda et al. [48]. Recently, ML algorithms are becoming more and more popular to solve anomaly detection problems in visual inspection. As pointed out in Rožanec et al. [49], the use of machine learning to detect defects of industrial products could be seen in a number of specific applications such as controlling the quality of printing, inspecting printed circuit board production, detecting defects during metallic powder bed fusion in additive manufacturing, determining the quality of bottles, manufactured vehicle parts, and aerospace components. In other situations where quantitative variables are the characteristics of interest, ML algorithms were combined with traditional methods like control charts in statistical process control (SPC) to enhance their effectiveness. A typical architecture of this solution has been applied in Xia et al. [50]. In general, the combination of ML with SPC methods can provide a promising solution for the quality control of quantitative characteristics in smart manufacturing

Despite the effectiveness of the existing anomaly detection methods, there are still challenging problems in quality control. For example, Napoletano et al. [51] discussed some difficulties regarding an automated visual inspection. Recent AI-based models tend to apply hybrid methods that combine several complex machine-learning structures. The complexity of these models can be a major obstacle in applying them in industrial production due to the requirements of real-time performance. Therefore, future researchers should pay more attention to keeping a balance between algorithm complexity, detection accuracy, and operating time. The authors also suggested the directions of using 3D imaging and hyperspectral imaging to increase the accuracy of defect detection and characterization in image-based quality control. Other difficulties related to the distribution of defect-free samples in some complicated industrial scenarios were mentioned in Liu et al. [52]. Although a number of advanced deep learning-based models have been proposed to overcome these difficulties like dual prototypes autoencoder [52], CNN-autoencoder [51], and restricted Boltzmann machine [53], the improvement of the existing methods to get better performance of the inspection should be considered for further researches in the field.

4.2 AI and IIoT-Based Solutions for Quality Control

In Industry 5.0, quality suffers when products tend to be personalized and customized products are usually manufactured in small batch sizes. This is a significant challenge for manufacturers to utilize the same common methods for controlling quality. Products made in small batches means more likely to lack sample data and erroneous measurements, leading to flawed assessments and false alarms. Thus, the state-of-the-art technologies of Industry 5.0 should be integrated and applied to quality control in order to increase the precision of the measurements. This also facilitates the real-time monitoring process which is important in smart manufacturing. Among these technologies, the AI and Internet of Things (IoT) methods are indispensable.

In smart manufacturing, sensors are embedded in all the components of machinery systems and production lines to monitor the whole manufacturing process. The aim of these electronic devices is to collect data related to the current status of the manufacturing process, involving the status of the products in production, the operating condition of the machines, and related operating procedures. The collected data from a single sensing device are then integrated with the ones from other devices by the connective technique of IoT. The IoT technology enables objects in a network of physical devices embedded with sensors, actuators, electronics, software, and network connectivity to connect, interact and exchange data. After that, these data are processed and analyzed by AI algorithms to draw conclusions about the manufacturing process or the products. That is to say, IoT technologies are to collect data and AI algorithms are to process these data, leading to automatic real-time monitoring of the production process in general as well as automatic controlling of product quality in particular. This is the core idea for the application of AI and IoT solutions to quality control in the literature. For example, Ha and Jeong [54] proposed an AI-based defect inspection method for injection molding using industrial IoT systems. A pilot project in China, called AIoT (Internet of Agricultural Things), which aimed to leverage the technologies of the IoT to ensure food safety was introduced in Liu et al. [55]. An IoT-based framework for quality inspection in gearboxes production was presented in Cicconi and Raffaeli [56] where the authors stated that using their proposed approach could result in about 60% time reduction for quality control operations. Eichelberger et al. [57] introduced an AI-enabled IoT platform for intelligent industrial production and an early validation in terms of a demonstrator use case on AI-enabled visual quality inspection

Besides the outstanding achievements of AI and IoT technologies that enhance the performance of product quality control in smart manufacturing, there are certain difficulties in applying these technologies in Industry 5.0. Manufacturers must integrate new technologies into existing systems, leading to significant investments in new machines and devices. When a new technology is updated, it usually requires highly trained workers. However, these are worthwhile investments for any business in the context of Industry 5.0. The effective collaboration between humans and machines in Industry 5.0 will fully exploit human intelligence and flexibility, and the durable and accurate working ability of machines and autonomous robots to speed up the efficiency of controlling product quality and achieve their desired products.

4.3 Human-Centric Quality Control

When it comes to a smart manufacturing system, it is usually thought of as a combination of many advanced technologies like the Internet of Things, AI, Digital Twin, Cyber-Physical Systems, Cloud Computing, and Big Data analysis. That means the human factor is sometimes underestimated despite the fact that it is one of the most important aspects of manufacturing. However, humans have a significant influence on the performance of any manufacturing process in practice. Research in Pacaux-Lemoine et al. [58] indicated that including a human component can increase substantially productivity and improve product quality. The human-centric is emphasized as a key pillar in building a more resilient industry, and therefore it is essential to integrate the human component into smart manufacturing, according to a report from the European Commission on Industry 5.0. An outlook on human-centric manufacturing towards Industry 5.0 was provided in Lu et al. [59].

The role of humans was also the main topic of interest in many references related to zero defect manufacturing (ZDM), the aim of quality control. Powell et al. [60] argued that the human factor should not be neglected. With the use of technologies and human-robot collaboration for operator assistance, the capabilities will be extremely flexible when the human factor is kept in the loop. suggested exploring further the role of people on overall manufacturing effectiveness and quality. In an in-depth and comprehensive study on human-centric ZDM, Wan and Leirmo [61] separated the role of humans into managers, engineers, and operators. In particular, the managers play a core role in advancing manufacturing although they do not directly involve in the production process. They are people who are in charge of defining a clear vision of the businesses and setting the standard performance in manufacturing based on the customers' desires. Meanwhile, engineers are responsible for three main missions in a manufacturing system: reducing systemic errors and identifying the root causes of defects; incorporating advanced technologies to improve the efficiency of quality control, and involving in daily operators' work and supporting them with clear and concise instructions. Agreeing with this discussion, we would like to add one more element related to the human factor in quality control in particular as well as smart manufacturing in general, namely customers. In the

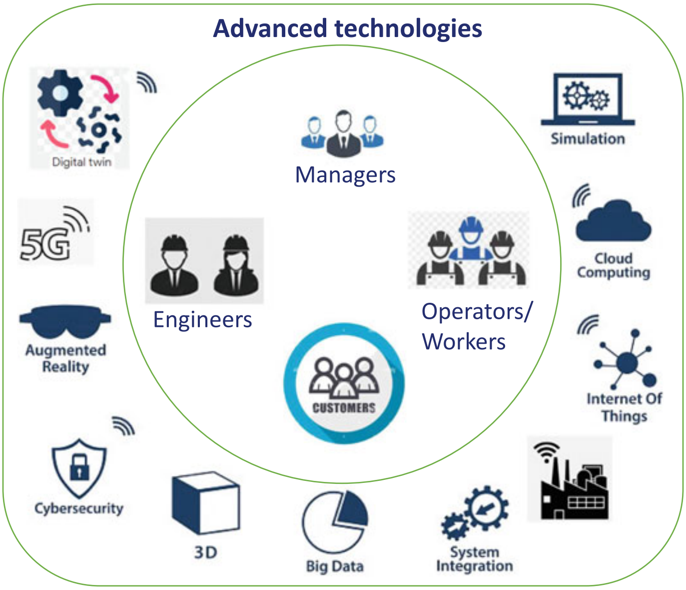


Fig. 2 Human-centric quality control

context of Industry 5.0, as manufacturing transfers from mass production to mass personalization, the customer's role is not only as a consumer but also as an active participant in the product design and quality assurance process. With the wish for personalized products with emotion and identity attachment, customers may also desire to have access to a real-time quality process to ensure that the final products satisfy their requests. The involvement of customers in the quality control process only brings benefits to the business. On the one hand, businesses can take advantage of the knowledge and experience of customers to prevent product defects. On the other hand, it also increases the customers' satisfaction and trust in the business. Figure 2 illustrates the human-centric paradigm in quality control.

Inheriting advanced inventions and technologies from Industry 4.0, the introduction and application of disruptive technologies to smart manufacturing in Industry 5.0 is an inevitable trend. This not only brings advantages to help improve the effectiveness of the human-centric approach in quality control but brings certain difficulties at the same time. In many situations, the reluctance to change was the critical barrier for organizations to adopt new technologies. Toward human-centric quality control requires the willingness to cooperate, the equipment of necessary skills and knowledge, and the integration of all human roles. Businesses should prioritize the cultivation of human resources and developing human capital and the technological adaptations must start from the human perspective, based on its fundamental view on human resources [61].

5 A Case Study

In this case study, we present a machine-learning pipeline to monitor the wine quality in the food industry. We apply seven machine learning algorithms (i.e. Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Support Vector Machine (SVM), K-Nearest Neighbors, Gaussian Naive Bayes, XGBOOST (XGB)) to classify between good and bad quality wine. Besides, we analyze the feature importance by using the Random Forest algorithm and explainable AI—SHapley Addictive exPlanations (SHAP). In the context of the wine industry, knowing important factors could aid human control and improve product quality. First, we provide information about the wine database (containing red and white wine datasets), data preprocessing, and data analysis step. Then, seven machine learning models and results are presented (Table 2).

5.1 Dataset

The wine industry is invested in both wine production and marketing purposes. Wine certification and quality assessment are critical components in this context. Certification prevents the illegal adulteration of wines and ensures quality for the wine market. Wine quality evaluation refers to a part of the certification process that examines the manufactured wine. It has the potential to be used to improve winemaking (by identifying the most influential factors) and to categorize wines such as premium brands (useful for setting prices). The wine dataset (presented in [62]) contains red and white variants of "Vinho Verde" wine for research on wine quality (from Portugal) including a total of 6495 samples (i.e 4897 entries (white wine) and 1598 entries (red wine)) with 12 features described in Table 3. The quality of wine is graded by experts ranging from 0 (very bad) to 10 (very excellent).

5.1.1 Data Preprocessing

At first, the wine dataset is cleaned by checking to handle the missing data and duplicated values. Depending on the wine quality score, wine is determined as bad

Table 2 The statistic of wine dataset

Dataset	Wine dataset
Number of observations	6497
Number of features	12
Number of red wine samples	1599
Number of white wine samples	4898

Table 3 Feature descriptions in the wine quality dataset [62]

Attribute	Meaning
Fixed activity	Fixed acids (i.e. tartaric, malic, citric, and succinic acids) are found in grapes (except succinic)
Volatile acidity	These acids could be extracted from the wine before the manufacturing process can be completed
Citric acid	A kind of fixed acid that contributes to the freshness of wine
PH	Potential of hydrogen (PH) is a numerical scale used to determine the acidity or basicity of wine. Most wines have a pH of 2.9–3.9 because of acidity
Residual sugar	Refers to the natural sugar found in grapes that remains after the fermentation process has ended
Chlorides	The concentration of chloride in wine is usually the main contributor to saltiness in wine
Free sulfur dioxide	Known as sulfites, and too many of them are undesirable because they emit a strong odor
Total sulfur dioxide	This is the sum total of the bound and the free sulfur dioxide. This is primarily used to kill harmful bacteria while also preserving quality and freshness
Sulfates	Are mineral salts containing sulfur. They are involved in the fermentation process and have an impact on the aroma and flavor of the wine
Density	Is commonly used to calculate the rate of sugar-to-alcohol conversion
Quality	“Good” and “bad” quality of wine, range of score between 0 (very bad) and 10 (excellent)

(score <6) and good (score ≥ 6), labeled as 0 and 1, respectively. The data set is divided into training and testing sets before being fed into the model with a ratio of (8:2) after being normalized. According to the learning data, different machine learning algorithms learn after making predictions according to test data. The data processing is as followed:

1. Cleaning data
2. Labeling data
3. Normalize value
4. Data split.

5.1.2 Data Analysis

In this section, the wine dataset is analyzed to provide intuitively the characteristics and behavior of data. In Fig. 3, the quality attribute illustrates the quality score of white wine on a scale of 10 with a dominant score of 5–7. We suppose that the score of good wine is greater than or equal to 6 (labeled as 1) while the score of low-quality

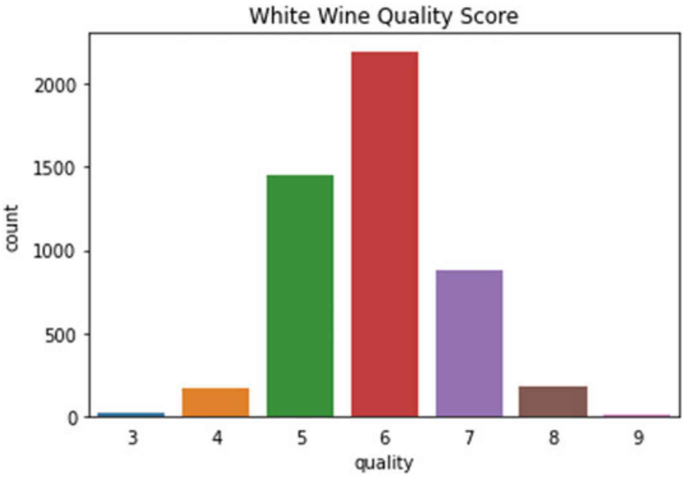


Fig. 3 Quality score of white wine before label

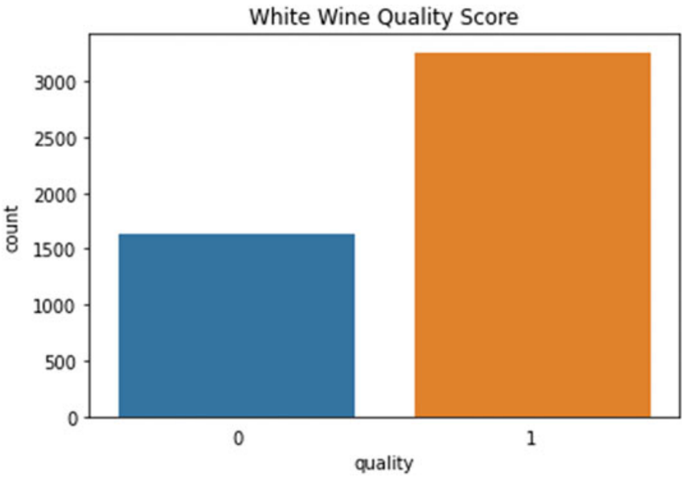


Fig. 4 Quality score of white wine after label

wine is lower than 6 (labeled as 0). After labeling, the quality attribute is classified as binary (0 or 1) in which high-quality wine doubles low-quality wine about the number of samples, Fig. 4.

Next, the correlation matrix is applied to explore the relationship between categorical variables. The correlation value could imply two variables can be stated as strong correlation or no correlation, detailed in Table 4. Heatmap and correlation matrix are visualized in Fig. 5 in which **density** shows a strong positive correlation with **residual sugar** while implying a strong negative correlation with **alcohol**. In addition, in Fig. 6 we can see most variables do not have linear relation, except **den-**

Table 4 Correlation level explanation

Correlation value	Interpretation
+/– 0.9 to +/– 1.0	Strong positive/negative correlation
+/– 0.7 to +/– 0.9	High positive/negative correlation
+/– 0.5 to +/– 0.7	Moderate
+/– 0.3 to +/– 0.5	Low positive/negative correlation
+/– 0.0 to +/– 0.3	Negligible correlation

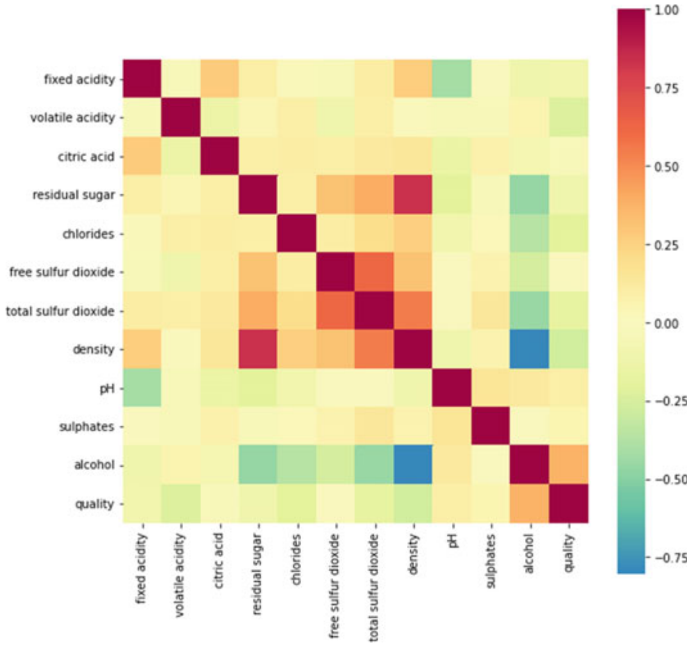


Fig. 5 Heatmap correlation between variables

sity attribute tends to be linear with **residual sugar** attribute (upward trend) and **alcohol** (downward trend).

In the context of food manufacturing, evaluating and understanding the primary factors that influence product quality is crucial since we want to handle and improve the quality. Besides, feature importance could bring profound benefits to data and model understanding, and model improvement. There are several techniques for extracting feature importance in the literature. In this study, we apply two methods which are feature importance based on the Random Forest algorithm and SHapley Addictive exPlanations (SHAP) [63]—an Explainable AI technique to get a comprehensive insight into data.

Random forest (RF) composes many individual decision trees (constructed by internal nodes and leaves) in parallel. Each tree is constructed using a random extrac-

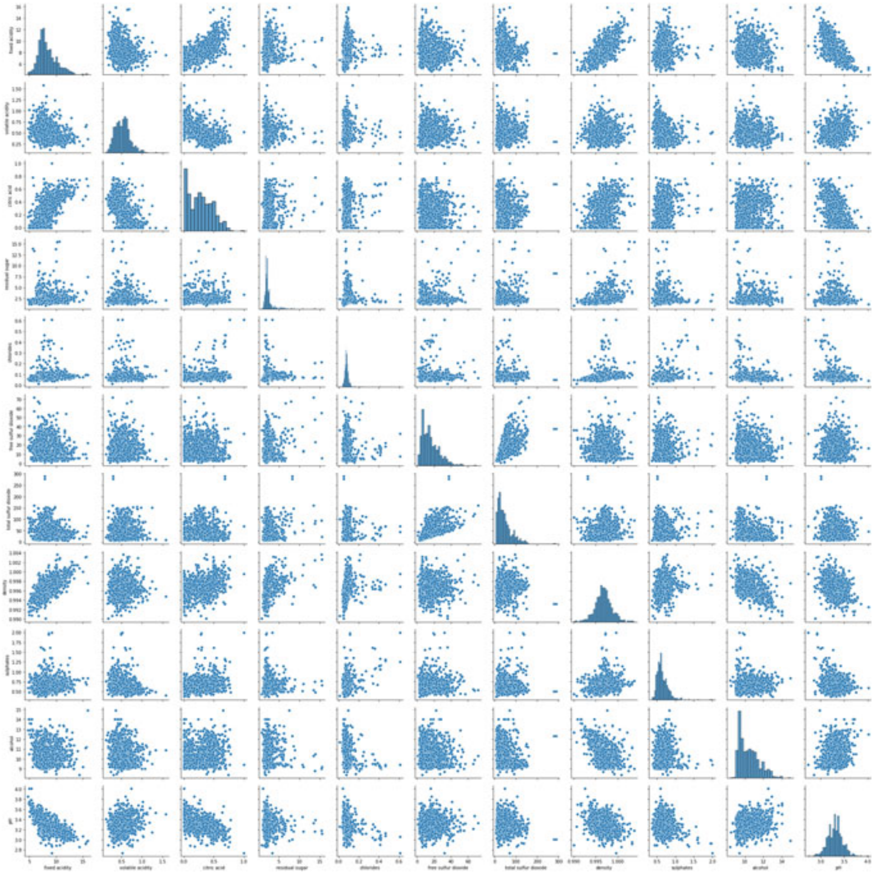


Fig. 6 Correlation between variables

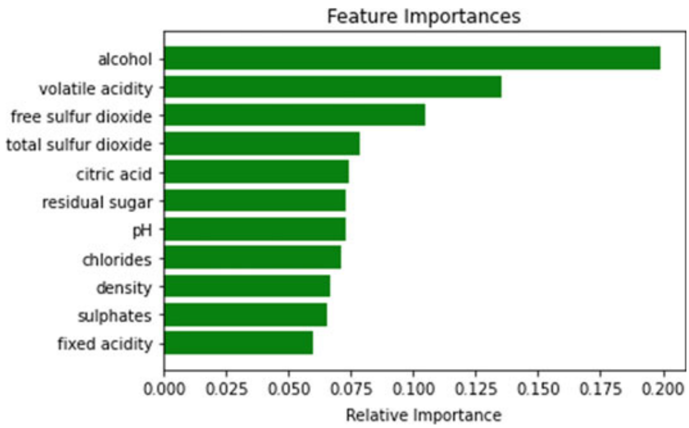


Fig. 7 Random forest feature importance

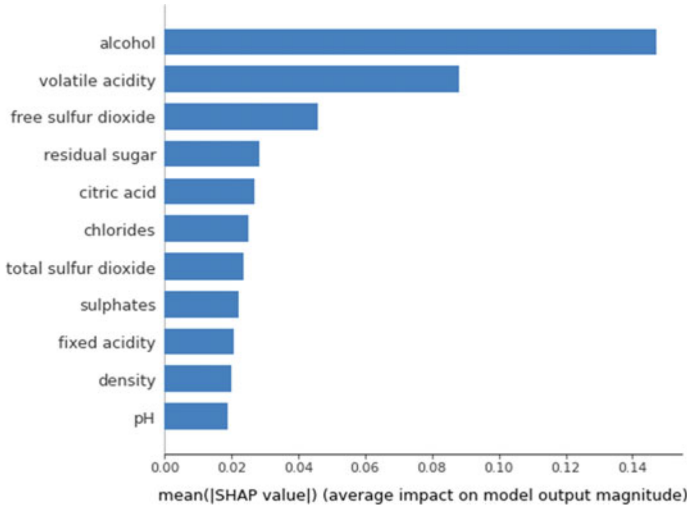


Fig. 8 SHAP summary plot—bar chart

tion of observations from the sub-dataset and features. The maximum voting of all decision tree outputs indicates the feature importance. SHAP is a game-theory-based approach for globally or locally interpreting the output of any ML or DL model introduced in [63]. This technique estimates the contribution level of each feature to the predicted results by using Shapley values, i.e., the average expected marginal contribution of a feature over all the possible combination sets or coalitions.

The results of feature importance based on the Random Forest algorithm are illustrated in Fig. 7, we can see obviously that three factors (i.e. **alcohol**, **volatile acidity**, and **free sulfur dioxide**) have the most effect on white wine quality. The features are organized in order of relevance, with the top feature being the most essential and the bottom one being the least. A similar result could be implied from the SHAP method in Fig. 8. The feature importance can be plotted with more details, showing the feature value in Fig. 9. The SHAP summary plot provides an overall description of a model that combines feature importance with feature effect. On a summary plot, the shapely value for a feature and specific sample is shown as a point. Shapely values are on the X-axis, and features are on the Y-axis. Colors represent low/high values.

5.2 *Machine Learning Model*

In this study, we implement seven popular machine learning algorithms which are Logistic Regression, Decision Tree, Random Forests, Support Vector Machine (SVM), K-Nearest Neighbors, Gaussian Naive Bayes, XGBOOST (XGB) for the

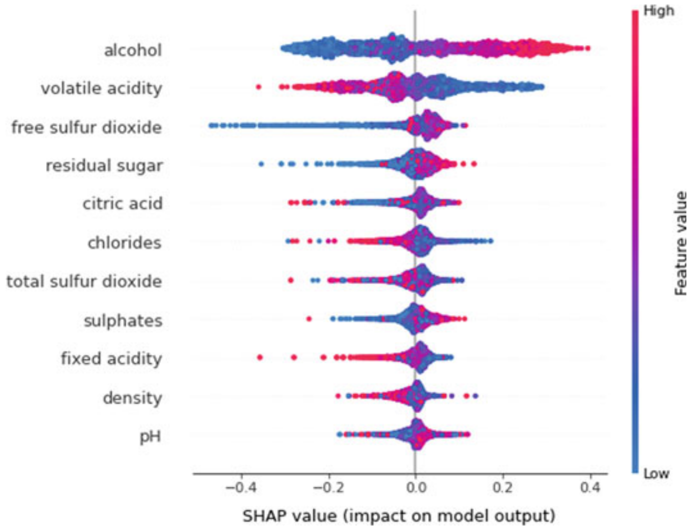


Fig. 9 SHAP summary plot in detail

classification task. The detail of each algorithm has been explained in detail in previous work, therefore we only provide the overview of these algorithms.

- **Logistic Regression**: a supervised learning algorithm frequently used in binary or multiclass classification problems (note that logistic regression is not a regression algorithm). Logistic regression comes from statistics used to find the best fit logistic function to describe the relationship between one dependent variable and one or more independent variables. The value of the logistic function ranges between zero and one [64].
- **Decision Tree algorithm**: a supervised learning algorithm capable of solving both regression and classification problems. A decision tree is constructed by nodes (which represent features), branches (which represent rules), and leaves (which represent an outcome). The advantages are high interpretability and less time execution. However, it is considered unstable because a small change in data can lead to a vastly different decision tree and can suffer from over-fitting (that can be addressed in the random forests algorithm)
- **Random Forest algorithm**: a supervised learning algorithm that can be used for classification and regression by ensembling multiple decision trees. Random Forest randomly selects data and specific features to build multiple decision trees. The result is obtained by averaging the decisions of multiple decision trees. Random Forests have the potential to reduce variance and eliminate overfitting. Furthermore, the random forests algorithm is more robust and accurate than a single decision tree.
- **Support Vector Machine (SVM)**: a supervised machine learning algorithm primarily used for classification and regression analysis. Intuitively, SVM attempts

to find an optimal hyperplane to separate data by maximizing the margin or by minimizing the Euclidean distance (L2-norm) of training data. Testing data could be classified based on the hyperplane.

- **K-Nearest Neighbors (K-NN)**: an unsupervised learning algorithm that divides n observations into chosen k clusters with each observation belonging to the cluster mean that is closest to it. K-NN retains all training examples in memory. In particular, when a new, previously unseen example x is received, the k-NN algorithm finds k training examples that are nearest to x. The output is derived based on the majority label (classification task) or the average label (regression task) [64].
- **Gaussian Naive Bayes**: a supervised learning algorithm for classification problems based on probabilistic (i.e. Bayes' Rule). The term "naive" refers to the assumption that input features are independent of each another. In addition, "Gaussian" implies that the continuous value distribution associated with each feature follows the Gaussian distribution (also called Normal distribution). The algorithm learns the likelihood of an object depending on Bayes' Theorem with certain characteristics belonging to a specific group or class.
- **Extreme Gradient Boosting (XGBoost)** [65]: a scalable machine learning system for tree boosting used for supervised learning problems. XGBoost builds upon decision trees, ensemble learning, and gradient boosting, providing parallel tree boosting for regression, classification, and ranking problems.

5.3 Results

The performance of the machine learning algorithms approach for quality classification is evaluated in terms of Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), and F1-Score (Eq. 4). In more detail, True Positive (TP) is an outcome where the model correctly predicts the good wine, TN (True Negative) stands for the number of samples correctly predicted as bad wine, FP (False Positive) is the number of good wine samples incorrectly classified, and FN (False Negative) stands for the number of false bad wine prediction.

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{1}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

$$\mathrm{F1Score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{4}$$

Table 5 The Performance of seven machine learning methods evaluating on wine dataset

Method	White wine dataset				Red wine dataset			
	Precision	Accuracy	Recall	F1 Score	Precision	Accuracy	Recall	F1 Score
Gaussian Naive Bayes	0.7371	0.6826	0.7808	0.7583	0.7529	0.7312	0.7441	0.7485
Logistic regression	0.7391	0.6826	0.8976	0.8107	0.7683	0.7312	0.7906	0.7793
K-nearest neighbors	0.7410	0.7459	0.9248	0.8227	0.7185	0.7343	**0.8313**	0.7708
Decision tree	0.8068	0.7673	0.8352	0.8207	0.8062	0.7687	0.7500	0.7771
Support vector	0.6390	0.6397	**1.0000**	0.7797	0.6024	0.6187	0.8546	0.7067
Random forest	**0.8187**	**0.8142**	0.9104	**0.8621**	**0.8452**	**0.8250**	0.8255	**0.8352**
XGBoost	0.7779	0.7632	0.8800	0.8258	0.7771	0.7500	0.7500	0.7633

Accuracy refers to the number of correctly classified data samples over the total number of data samples. If the dataset is unbalanced, accuracy may not be a good metric. As a result, other metrics are needed to evaluate the model. Precision (positive predictive value) is a metric that measures how many of the positive predictions made are correct. The true positive rate, also known as recall, is a measure of how many positive cases the classifier correctly predicted out of all positive cases in the data. The F1 score is a measure that combines precision and recall and is used to achieve a balance between precision and recall.

Table 5 shows the predicted result of seven machine learning models on the Wine dataset (red and white wine). Overall, seven machine learning models could interpret the characteristic of the Wine dataset in order to classify the quality of wine in which evaluation parameters ≥ 0.6 and random forest achieves high performance over the others.

6 Concluding Remarks

In this chapter, we discussed Quality control for Smart Manufacturing, especially the application of machine learning, AI, and computer vision for quality control in Industry 5.0. Machine learning and computer vision have been demonstrated to be effective methods for detecting faults, anomalies, and monitoring quality. However, as smart manufacturing is associated with the human factor in the context of Industry 5.0, there are more apparent difficulties and challenges along with the emergence of opportunities and promised solutions. With the evolution of machine learning-based techniques, modern AI-based Anomaly detection could be more robust, trustworthy, and reliable to be deployed and integrated human-centric in the manufacturing system. Finally, we presented a case study on monitoring products in the wine industry. In particular, we demonstrated an example of classification between good and bad wine by using seven machine-learning techniques. The results are evaluated based on accuracy, recall, precision, and F1-score parameters.

References

1. Breque M, De Nul L, Petridis A (2021) Industry 5.0: towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Luxembourg, LU
2. Mosconi F (2015) The new European industrial policy: global competitiveness and the manufacturing renaissance. Routledge
3. Bottani E, Vignali G (2019) Augmented reality technology in the manufacturing industry: a review of the last decade. IISE Trans 51:284–310
4. Lopes MR, Costigliola A, Pinto R, Vieira S, Sousa JM (2020) Pharmaceutical quality control laboratory digital twin-a novel governance model for resource planning and scheduling. Int J Prod Res 58:6553–6567
5. Sahija D (2021) Critical review of machine learning integration with augmented reality for discrete manufacturing. Independent Researcher and Enterprise Solution Manager in Leading Digital Transformation Agency, Plano, USA
6. Ho PT, Albajez JA, Santolaria J, Yagüe-Fabra JA (2022) Study of augmented reality based manufacturing for further integration of quality control 4.0: A systematic literature review. Appl Sci 12:1961
7. Fahle S, Prinz C, Kuhlenkötter B (2020) Systematic review on machine learning (ML) methods for manufacturing processes-identifying artificial intelligence (AI) methods for field application. Procedia CIRP 93:413–418
8. Li J, Maiti A, Springer M, Gray T (2020) Blockchain for supply chain quality management: challenges and opportunities in context of open manufacturing and industrial internet of things. Int J Comput Integr Manuf 33:1321–1355
9. Aghbashlo M, Hosseinpour S, Ghasemi-Varnamkhasti M (2014) Computer vision technology for real-time food quality assurance during drying process. Trends Food Sci Technol 39:76–84
10. Shivajee V, Singh RK, Rastogi S (2019) Manufacturing conversion cost reduction using quality control tools and digitization of real-time data. J Clean Prod 237:117678
11. Gunasekaran S (1996) Computer vision technology for food quality assurance. Trends Food Sci Technol 7:245–256
12. Mazumder J (2015) Design for metallic additive manufacturing machine with capability for “certify as you build.” Procedia CIRP 36:187–192
13. Mitra A (2016) Fundamentals of quality control and improvement. Wiley
14. Feigenbaum AV (1991) Total quality control, New York
15. Shewhart WA (1924) Some applications of statistical methods to the analysis of physical and engineering data. Bell Syst Tech J 3:43–87
16. Juran JM (1981) Product quality-a prescription for the west, part i. Manag Rev 70:8–14
17. Aft LS (2018) Fundamentals of industrial quality control. CRC Press
18. Ishikawa K (1960) Ishikawa diagram
19. Leng J, Sha W, Wang B, Zheng P, Zhuang C, Liu Q, Wuest T, Mourtzis D, Wang L (2022) Industry 5.0: prospect and retrospect. J Manuf Syst 65:279–295
20. Carvalho N, Chaim O, Cazarini E, Gerolamo M (2018) Manufacturing in the fourth industrial revolution: a positive prospect in sustainable manufacturing. Procedia Manuf 21:671–678
21. Tran PH, Heuchenne C, Thomassey S (2022) Enhanced CUSUM control charts for monitoring coefficient of variation: a case study in textile industry. IFAC-PapersOnLine 55:1195–1200
22. Kusiak, A. (2018). Smart manufacturing. Int J Prod Res 56(1–2):508–517
23. Radford GS (1922) The control of quality in manufacturing. Ronald Press Company

24. Shewhart WA (1931) Economic control of quality of manufactured product. Macmillan Co Ltd, London
25. Deming WE (1982) Quality, productivity, and competitive position. Massachusetts Inst Technology
26. A3 A (1978) Quality systems terminology. American Society for Quality Control
27. Alford LP, Beatty HR (1951) Principles of industrial management: revised and rewritten. Ronald Press Co., New York
28. Freund RA (1985) Definitions and basic quality concepts. J Qual Technol 17:50–56
29. Juran JM (1986) The quality trilogy: a universal approach to managing for quality. Qual Prog 19:19–24
30. Kumar SA, Suresh N (2006) Production and operations management. New Age International
31. Eldin AB (2011) modern approaches to quality control. BoD–Books on Demand
32. Durland RH, Eastman EM (1998) Manufacturing and quality control of plasmid-based gene expression systems. Adv Drug Deliv Rev 30:33–48
33. Kackar RN (1985) Off-line quality control, parameter design, and the Taguchi method. J Qual Technol 17:176–188
34. Phadke MS, Kackar RN, Speeney DV, Grieco MJ (1983) Off-line quality control in integrated circuit fabrication using experimental design. Bell Syst Tech J 62:1273–1309
35. Villalba-Diez J, Schmidt D, Gevers R, Ordieres-Meré J, Buchwitz M, Wellbrock W (2019) Deep learning for industrial computer vision quality control in the printing industry 4.0. Sensors 19:3987
36. Khan AI, Al-Habsi S (2020) Machine learning in computer vision. Procedia Comput Sci 167:1444–1451
37. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. In: Computational intelligence and neuroscience 2018
38. Chai J, Zeng H, Li A, Ngai EW (2021) Deep learning in computer vision: a critical review of emerging techniques and application scenarios. Mach Learn Appl 6:100134
39. Zhou L, Zhang L, Konz N (2022) Computer vision techniques in manufacturing. IEEE Trans Syst Man Cybern: Syst
40. Banús N, Boada I, Xiberta P, Toldrà P, Bustins N (2021) Deep learning for the quality control of thermoforming food packages. Sci Rep 11:1–15
41. Kumar A (2008) Computer-vision-based fabric defect detection: a survey. IEEE Trans Ind Electron 55:348–363
42. Rajan AJ, Jayakrishna K, Vignesh T, Chandradass J, Kannan T (2021) Development of computer vision for inspection of bolt using convolutional neural network. Mater Today: Proc 45:6931–6935
43. Ebadi SE, Dhakad S, Vishwakarma S, Wang C, Jhang Y-C, Chociej M, Crespi A, Thaman A, Ganguly S (2022) Psp-hdri +: a synthetic dataset generator for pre-training of human-centric computer vision models. arXiv:2207.05025
44. Ebadi SE, Jhang Y-C, Zook A, Dhakad S, Crespi A, Parisi P, Borkman S, Hogins J, Ganguly S (2021) Peoplesanspeople: a synthetic data generator for human-centric computer vision. arXiv:2112.09290
45. Peng Y, Braun B, McAlpin C, Broadway M, Colegrove B, Chiang L (2022) Contamination classification for pellet quality inspection using deep learning. Comput Chem Eng 107836
46. Mac TT, Hung NT (2021) Automated pill quality inspection using deep learning. Int J Modern Phys B 35:2140050
47. Yang Y, Yang R, Pan L, Ma J, Zhu Y, Diao T, Zhang L (2020) A lightweight deep learning algorithm for inspection of laser welding defects on safety vent of power battery. Comput Ind 123:103306
48. Alencastre-Miranda M, Johnson RM, Krebs HI (2020) Convolutional neural networks and transfer learning for quality inspection of different sugarcane varieties. IEEE Trans Ind Inform 17:787–794
49. Rožanec JM, Trajkova E, Dam P, Fortuna B, Mladenić D (2022) Streaming machine learning and online active learning for automated visual inspection. IFAC-PapersOnLine 55:277–282

50. Xia B, Jian Z, Liu L, Li L (2018) An effective multivariate control chart for detecting small mean shifts using support vector data description. Adv Mech Eng 10:1687814018810625
51. Napoletano P, Piccoli F, Schettini R (2021) Semi-supervised anomaly detection for visual quality inspection. Expert Syst Appl 183:115275
52. Liu J, Song K, Feng M, Yan Y, Tu Z, Zhu L (2021) Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection. Opt Laser Eng 136:106324
53. Zhang Y, Peng P, Liu C, Zhang H. Anomaly detection for industry product quality inspection based on gaussian restricted Holtzmann machine. In: 2019 IEEE international conference on systems, man and cybernetics (SMC). IEEE, pp 1–6
54. Ha H, Jeong J (2021) Cnn-based defect inspection for injection molding using edge computing and industrial IoT systems. Appl Sci 11:6378
55. Liu Y, Han W, Zhang Y, Li L, Wang J, Zheng L (2016) An internet-of-things solution for food safety and quality control: a pilot project in china. J Ind Inf Integr 3:1–7
56. Cicconi P, Raffaeli R (2020) An industry 4.0 framework for the quality inspection in gearboxes production. Comput-Aided Des Appl 17:813–824
57. Eichelberger H, Palmer G, Reimer S, Vu TT, Do H, Laridi S, Weber A, Niederée C, Hildebrandt T (2022) Developing an AI-enabled IIoT platform—Lessons learned from early use case validation. arXiv:2207.04515
58. Pacaux-Lemoine M-P, Trentesaux D, Rey GZ, Millot P (2017) Designing intelligent manufacturing systems through human-machine cooperation principles: a human-centered approach. Comput Ind Eng 111:581–595
59. Lu Y, Zheng H, Chand S, Xia W, Liu Z, Xu X, Wang L, Qin Z, Bao J (2022) Outlook on human-centric manufacturing towards industry 5.0. J Manuf Syst 62:612–627
60. Powell D, Magnanini MC, Colledani M, Myklebust O (2022) Advancing zero defect manufacturing: a state-of-the-art perspective and future research directions. Comput Ind 136:103596
61. Wan PK, Leirmo TL (2023) Human-centric zero-defect manufacturing: state-of-the-art review, perspectives, and challenges. Comput Ind 144:103792
62. Cortez P, Cerdeira A, Almeida F, Matos T, Reis J (2009) Modeling wine preferences by data mining from physicochemical properties. Decis Support syst 47:547–553
63. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems, vol 30
64. Burkov A (2019) The hundred-page machine learning book, vol 1. Andriy Burkov Quebec City, QC, Canada
65. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 785–794

Dynamic Process Monitoring Using Machine Learning Control Charts

Xiulin Xie and Peihua Qiu

Abstract Machine learning methods have been widely used in different applications, including process control and monitoring. For handling statistical process control (SPC) problems, the existing machine learning approaches have some limitations. For instance, most of them are designed for cases in which in-control (IC) process observations at different time points are assumed to be independent and identically distributed. In practice, however, serial correlation almost always exists in the observed sequential data, and the longitudinal pattern of the process to monitor could be dynamic in the sense that its IC distribution would change over time (e.g., seasonality). It has been well demonstrated in the literature that control charts could be unreliable to use when their model assumptions are invalid. In this chapter, we modified some representative existing machine learning control charts using nonparametric longitudinal modeling and sequential data decorrelation algorithms. The modified machine learning control charts can well accommodate time-varying IC process distribution and serial data correlation. Numerical studies show that their performance are improved substantially for monitoring different dynamic processes.

Keywords Control chart · Data correlation · Dynamic processes · Machine learning · Seasonality · Statistical process control

1 Introduction

Statistical process control (SPC) provides a major tool for online monitoring of sequential processes [16, 22, 26]. Most conventional SPC charts are designed for detecting process distributional shifts under the assumptions that process observations at different time points are independent and identically distributed (i.i.d.) with a parametric (e.g., Normal) distribution when the process under monitoring is in-control (IC). In practice, however, observed data of a sequential process are often

X. Xie · P. Qiu (✉)
Department of Biostatistics, University of Florida, 2004 Mowry Road, Gainesville, FL 32610, USA
e-mail: pqiu@ufl.edu

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_4

serially correlated, and dynamic in the sense that their IC distribution varies over time. This chapter focuses on online monitoring of dynamic processes with serially correlated data.

In the SPC literature, many control charts have been developed, which can be roughly classified into the following four categories: Shewhart, cumulative sum (CUSUM), exponentially weighted moving average (EWMA), and change-point detection (CPD) charts (cf., [17, 24, 30, 31]). As mentioned above, early control charts are designed mainly for cases when the observed IC data are i.i.d. and parametrically distributed. After SPC finds more and more applications for disease surveillance, environmental monitoring, business management, and many others, the conventional model assumptions mentioned above are rarely valid in these applications. It has been well demonstrated in the literature that control charts would be unreliable to use in cases when one or more of their model assumptions are invalid (e.g., [4, 20, 27]). So, some recent SPC research has considered cases when the IC process distribution does not have a parametric form (e.g., [10, 27]), process observations are serially correlated (e.g., [5, 29, 38]), or the IC process distribution is time-varying (e.g., [28, 36]).

In recent years, machine learning methods have been under rapid development (e.g., [1, 8, 14]). Since an SPC problem can be regarded as a binary class classification problem, in which each process observation needs to be classified into either the IC or the out-of-control (OC) status during sequential process monitoring, some machine leaning methods using both the IC and OC historical data have been used for process monitoring in the SPC literature. For instance, support vector machine (SVM), linear discriminant analysis (LDA), and k-nearest neighbors (KNN) methods have been employed for various process monitoring problems (i.e., [21, 39]). However, unlike the conventional classification problem, most SPC applications only involve IC training data before online process monitoring. To overcome this difficulty, some machine leaning algorithms, such as KNN, SVM, and random forest (RF), have been adapted to develop control charts using the one-class classification, artificial contrast, real-time contrast, and some other novel ideas (e.g., [12, 21, 34]). An attractive feature of these control charts based on machine learning algorithms is that they usually do not impose restrictive model assumptions explicitly. However, most of them require the implicit assumptions that process observations at different observation times are independent and identically distributed in order to define their decision rules properly. Therefore, such machine learning approaches have much room for improvement.

In Xie and Qiu [37], we modified some representative existing machine learning control charts so that the modified charts can properly accommodate serial correlation in process observations. However, these modified charts still assume the IC process distribution to be time-independent. Thus, they cannot be used in cases when the IC distribution is actually time-varying. In this chapter, we further modify the representative existing machine learning control charts considered in Xie and Qiu [37] so that the modified charts can accommodate both serial data correlation and time-varying IC process distribution, by using nonparametric longitudinal modeling and sequential data decorrelation algorithms. More specifically, in a modified control chart, an IC dataset is required to obtain an initial estimate of the IC longitudinal pattern of the dynamic process under monitoring using a nonparametric longitudinal modeling approach. Then, at the current time point during online process monitoring, the observed data are first standardized using the estimated IC longitudinal pattern and then decorrelated with all historical data using a sequential data decorrelation algorithm. Next, a machine learning control chart is applied to the standardized and decorrelated data for making a decision whether the process has a distributional shift at the current time point or not. Numerical studies show that the modified machine learning control charts are substantially improved for monitoring different dynamic processes after such a modification.

The remaining parts of the chapter are organized as follows. In Sect. 2, some representative existing machine learning control charts are briefly described. In Sect. 3, the proposed modification for certain machine learning control charts are described in detail. Some simulation studies are presented in Sect. 4 to evaluate their numerical performance. A real-data example to demonstrate the application of the modified machine learning control charts is discussed in Sect. 5. Finally, some remarks conclude the chapter in Sect. 6.

2 Some Representative Machine Learning Control Charts

In this section, we introduce some representative recent machine learning control charts. Assume that $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ is a vector of $p \geq 1$ numerical quality characteristics to monitor about a sequential process, and its observation at time n is $\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{np})'$. To online monitor the sequential process $\{\mathbf{X}_n, n \geq 1\}$, an initial IC dataset $\mathcal{X}_{IC} = \{\mathbf{X}_{-m_0+1}, \mathbf{X}_{-m_0+2}, \ldots, \mathbf{X}_0\}$ of size m_0 is assumed to be available in advance for all methods.

2.1 Control Chart Based on Artificial Contrasts

To solve a classification problem by a supervised machine learning method, a training dataset containing observations of both classes (e.g., IC and OC) is required. However, in many SPC applications, we only have an IC dataset before online process monitoring, and no OC process observations would be available in advance. To overcome this difficulty, Tuv and Runger [34] proposed the idea of artificial contrast. By this idea, artificial data are generated from a given distribution (e.g., Uniform) to represent the off-target data from the process. More specifically, for individual variables X_l, their artificial contrasts are generated independently from uniform distributions whose ranges are the same as those of X_l values in the IC dataset, for $l = 1, 2, \ldots, p$. Then, these artificial observations can be used as OC observations. By generating an artificial OC dataset, it converts the process monitoring problem to a supervised learning problem so that any machine learning classifiers, such as SVM and RF, can be used. However, this type of control charts are basically Shewhart charts, since the decision at a given time point during online process monitoring only relied on the observation at that time point. To overcome these limitations, Hu and Runger [19] suggested a modification by using the ideas of generalized likelihood ratio test and EWMA. To make a decision about the status of the process under monitoring at the current time point n, the log likelihood ratio of observed data $\mathbf{X}_n$ is first calculated as $l_n = \log\left[\hat{p}_1(\mathbf{X}_n)\right] - \log\left[\hat{p}_0(\mathbf{X}_n)\right]$, for $n \geq 1$, where $\hat{p}_1(\mathbf{X}_n)$ and $\hat{p}_0(\mathbf{X}_n)$ are the estimated probabilities of $\mathbf{X}_n$ in each class obtained by the RF classifier. Then, they considered the following univariate EWMA charting statistic (cf., [30]):

$$E_n = \lambda l_n + (1-\lambda)E_{n-1}, \text{ for } n \geq 1, \tag{1}$$

where $\lambda \in (0, 1]$ is a weighting parameter. The chart gives a signal of process mean shift at time n if

$$E_n > h_{AC}, \tag{2}$$

where h_{AC} is a control limit. The chart (1)–(2) is called AC chart hereafter to represent "artificial contrast."

For the AC chart (1)–(2), its control limit h_{AC} can be determined by a 10-fold cross-validation (CV) procedure to achieve a given value of the IC average run length (ARL), denoted as ARL_0. More specifically, 90% of the IC dataset $\mathcal{X}_{IC}$ and the artificial contrast dataset is first used to train the RF classifier. Then, a bootstrap sample can be drawn with replacement from the remaining 10% of the IC dataset, and the chart (1)–(2) with a given h_{AC} can be applied to the bootstrap sample. The run length (RL) value, defined to be the number of observation times from the beginning of process monitoring to the signal time, can then be recorded. Finally, the above procedure can be repeated for V times, and the average of the corresponding V values of RL can be used as the estimate of the ARL_0. Then, h_{AC} can be searched so that a given level of ARL_0 is reached. In this searching process, the bisection algorithm (Qiu [26], this chapter) or its modifications [9] can be used.

2.2 Control Chart Based on Real Time Contrasts

The classifier in the method AC is trained only one time using the IC dataset $\mathcal{X}_{IC}$ and the artificial OC dataset, which may not represent the actual off-target process observations well in a given application. To overcome this limitation, Deng et al. [12] suggested the so-called real-time contrast (RTC) method. The RTC method treats the process monitoring problem as a real-time classification problem, in which process observations in the IC dataset and those within a moving window of the current time point form a training dataset, with the former as IC observations and the latter as OC observations. More specifically, a dataset with N_0 observations, which is denoted as $\boldsymbol{S}_0$, is first randomly selected from the IC dataset $\mathcal{X}_{IC}$. Then, during online process monitoring, process observations in a window of the current observation time point

n are treated as OC data and denoted as $\boldsymbol{S}_n = \{\mathbf{X}_{n-w+1}, \mathbf{X}_{n-w+2}, \ldots, \mathbf{X}_n\}$, where w is the window size. Then, the RF classifier can be retrained sequentially overtime using the training dataset that combines $\boldsymbol{S}_0$ and $\boldsymbol{S}_n$. As discussed in Deng et al. [12], there could be several possible charting statistics based on the RF algorithm. As in their simulation studies, the average estimated classification rate in the dataset $\boldsymbol{S}_0$ can be used as the charting statistic, which is defined to be

$$R_n = \sum_{i=-m_0+1}^{0} \hat{p}_0^{(n)}(\mathbf{X}_i) I(\mathbf{X}_i \in \boldsymbol{S}_0)/N_0, \text{ for } n \geq 1, \tag{3}$$

where $\hat{p}_0^{(n)}(\mathbf{X}_i)$ are the estimated probabilities of $\mathbf{X}_i$ in the IC class obtained by the RF classifier trained at time n, and $I(u)$ is the indicator function that equals 1 when u is "true" and 0 otherwise. The chart gives a signal at time n if

$$R_n > h_{RTC}, \tag{4}$$

where h_{RTC} is a control limit of the RTC chart.

The control limit of the RTC chart (3)–(4) can be determined by the following bootstrap procedure suggested by Deng et al. [12]. First, we draw with replacement a sample from the IC dataset after the observations in $\boldsymbol{S}_0$ are excluded. Then, the chart (3)–(4) with the control limit h_{RTC} is applied to the bootstrap sample to obtain a RL value. This bootstrap re-sampling procedure is repeated for B = 1,000 times, and the average of the B values of RL is used to approximate the ARL_0 value for the given h_{RTC}. Finally, h_{RTC} can be searched by a numerical algorithm so that the assumed ARL_0 value is reached.

2.3 Control Chart Based on Support Vector Machine

Even though the RTC chart based on the RF classifier is useful and can be applied to a variety of monitoring problems, its charting statistic takes discrete values, which makes it less effective in some cases. As an alternative, He et al. [18] proposed a distance-based control chart. It uses the SVM framework to measure the distance between the support vectors and real time observations in $\boldsymbol{S}_n$. As discussed in He et al. [18], the distance from a sample of process observations to the boundary surface defined by the support vectors can be either positive or negative. They suggested transforming the distance using the following standard logistic function:

$$g(d) = \frac{1}{1 + \exp(-d)}.$$

Then, the following average value of the transformed distances from individual observations in $\boldsymbol{S}_n$ to the boundary surface can be defined to be the charting statistic:

$$M_n = \sum_{j=n-w+1}^{n} g(d(\mathbf{X}_i))/w, \text{ for } n \geq 1, \tag{5}$$

where $d(\mathbf{X}_i)$ is the distance from the observation $\mathbf{X}_i$ to decision boundary determined by the SVM classifier obtained at time n. The chart gives a signal at time n if

$$M_n > h_{SVM}, \tag{6}$$

where h_{SVM} is the control limit of the chart. The chart (5)–(6) is denoted as DSVM hereafter, to reflect the fact that it is a Distance-based control chart using SVM. The control limit of DSVM can be determined by a bootstrap procedure, similar to the one described above for the RTC chart.

In the above DSVM chart (5)–(6), the SVM algorithm needs to be implemented, and there are several qualititess involved that need to be selected in advance, including the kernel function and the penalty parameter [11]. In SVM, one of the most commonly used kernel functions is the Gaussian radial basis function (RBF), which is defined as: for any two observations $\mathbf{X}_i, \mathbf{X}_j$,

$$G(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{\sigma^2}\right),$$

where σ^2 is the spread parameter. He et al. [18] suggested using the above RBF as the kernel function with $\sigma^2 > 2.8$. They also suggested choosing the penalty parameter to be 1 for training SVM.

2.4 Control Chart Based on the KNN Classification

Another machine learning control chart, proposed by Sukchotrat et al. [33], is based on the KNN data description procedure. This chart is denoted as KNN hereafter. The charting statistic of KNN is defined as the average distance between a given observation X_n and its k nearest observations in the IC dataset $\mathcal{X}_{IC}$, and it is defined as follows:

$$C_n^2 = \sum_{j=1}^{k} \|\mathbf{X}_n - N_j(\mathbf{X}_n)\|/k, \text{ for } n \geq 1, \tag{7}$$

where $N_j(\mathbf{X}_n)$ is the jth nearest neighboring observation of $\mathbf{X}_n$ in the IC dataset $\mathcal{X}_{IC}$, and $\|\cdot\|$ is the Euclidean distance. Then, for online process monitoring, the process is declared to be OC at a given time n if the charting statistic C_n^2 of the related process observation exceeds the control limit h_{KNN}.

In the above KNN chart, the control limit h_{KNN} can be determined by the following bootstrap procedure: (i) a total of $B = 1,000$ bootstrap samples are obtained from the training dataset by the random sampling procedure with replacement and each bootstrap sample has the same size as the training dataset, (ii) the C_n^2 values defined in (7) of the individual observations in the bootstrap sample can be computed, (iii) the $(1-\alpha)$th percentile of all C_n^2 values can be computed from each bootstrap sample, and (iv) h_{KNN} is chosen to be the mean of the B such percentiles.

3 Suggested Modified Machine Learning Control Charts for Dynamic Process Monitoring

For many longitudinal processes, their distributions could change over time, even when their performance is considered to be IC. One example is about sequential monitoring of environmental variables, such as air temperature and various pollutant levels. These variables usually have seasonal variation. To monitor such dynamic processes, the machine learning control charts introduced in the previous section are obviously inappropriate to use because they require the IC process distribution to be unchanged over time. Recently, Xie and Qiu [36] suggested a new method for dynamic process monitoring. The basic idea of that method is to specify a time period as a baseline time period, estimate the regular longitudinal pattern of the quality variables in that period, and then compare the future performance of the process under monitoring with its performance in the baseline time period. In this section, a procedure for estimating the regular longitudinal pattern is first discussed in detail. Then, the suggested modification of some representative machine learning control charts for monitoring dynamic processes using is discussed.

3.1 Estimation of the Regular Multivariate Longitudinal Pattern

The time period of the initial IC dataset $\mathcal{X}_{IC}$ is set as a baseline time interval, and the IC dataset is assumed to follow the nonparametric longitudinal model:

$$\mathbf{X}_j = \boldsymbol{\mu}_j + \boldsymbol{\epsilon}_j, \qquad \text{for } j = -m_0+1, -m_0+2, \ldots, 0, \tag{8}$$

where $\boldsymbol{\mu}_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jp})'$ is the mean of $\mathbf{X}_j$, and $\boldsymbol{\epsilon}_j$ is the p-dimensional zero-mean error term. In Model (8), the covariance structure is described by $\mathrm{Cov}(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_{j^*})$, for any $j, j^* \in [-m_0+1, 0]$. Furthermore, it is assumed that the serial correlation among the IC process observations is stationary, and the serial correlation exists only when two observations are within $b_{\max} > 0$ in their observation indices. More specifically, it is assumed that $\boldsymbol{\gamma}(s) = \mathrm{Cov}(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_{j+s})$ only depends on s when j changes, and $\boldsymbol{\gamma}(s) = \mathbf{0}$ when $s > b_{\max}$. The above assumptions should be reasonable in many applications.

To obtain an initial estimate of $\boldsymbol{\mu}_j$, we can compute the local linear kernel (LLK) smoothing estimates of all components of $\boldsymbol{\mu}_j$ (cf., [35]). In matrix notation, let $\mathbf{W} = (X_{-m_0+1,1}, \ldots, X_{0,1}, \ldots, X_{-m_0+1,p}, \ldots, X_{0,p})'$, $\mathbf{Z}_j = \left[(1, -m_0+1-j)', \ldots, (1, -j)'\right]'$, and $\mathbf{K}_j = \mathrm{diag}\{K(\frac{i-j}{h_l}), i = -m_0+1, -m_0+2, \ldots, 0, l = 1, 2, \ldots, p\}$, where $K(\cdot)$ is a kernel function and $\{h_l, l = 1, 2, \ldots, p\}$ are bandwidths. Then, the initial estimate of $\boldsymbol{\mu}_j$, for $j = -m_0+1, -m_0+2, \ldots, 0$, can be obtained by the following LLK smoothing procedure:

$$\min_{\boldsymbol{\beta}\in R^{2p}} \left[\mathbf{W} - (I_{p\times p} \otimes \mathbf{Z})\boldsymbol{\beta}\right]' \mathbf{K}_j \left[\mathbf{W} - (I_{p\times p} \otimes \mathbf{Z})\boldsymbol{\beta}\right], \tag{9}$$

where $\otimes$ denotes the Kronecker product, $I_{p\times p}$ is the $p \times p$ identity matrix, and $\boldsymbol{\beta} = (\beta_{01}, \beta_{11}, \ldots, \beta_{0p}, \beta_{1p})'$ are coefficients. The solution of (9) has the expression

$$\widehat{\boldsymbol{\beta}} = [(I_{p\times p} \otimes \mathbf{Z}_j)'\mathbf{K}_j(I_{p\times p} \otimes \mathbf{Z}_j)]^{-1}(I_{p\times p} \otimes \mathbf{Z}_j)'\mathbf{K}_j\mathbf{W}.$$

Then, the initial estimate of $\boldsymbol{\mu}_j$, for $j = -m_0+1, -m_0+2, \ldots, 0$, is given by:

$$\widehat{\boldsymbol{\mu}}_j = \widehat{\boldsymbol{\beta}}'(I_{p\times p} \otimes \boldsymbol{\xi}_1), \tag{10}$$

where $\boldsymbol{\xi}_1 = (1, 0)'$. In the above LLK procedure, the kernel function $K(\cdot)$ is usually chosen to be the Epanechnikov kernel function, i.e., $K(u) = \frac{3}{4}(1-u^2)I(|u| \le 1)$, because of its good properties [13]. For the bandwidths $\{h_l, l = 1, 2, \ldots, p\}$, it has been well discussed in the literature that the conventional cross-validation (CV) procedure would not perform well when process observations at different time points are serially correlated, since the CV procedure cannot properly distinguish the data correlation structure from the data mean function (e.g., [2, 23]). Thus, we suggest choosing them using the following modified cross-validation (MCV) procedure that was originally suggested by Brabanter et al. [7] for handling bandwidth selection in a univariate regression setup with correlated data. By this approach, the bandwidths $\{h_l, l = 1, \ldots, p\}$ can be chosen by minimizing the following MCV score:

$$\mathrm{MCV}(h_1, h_2, \ldots, h_p) = \frac{1}{m_0} \sum_{j=-m_0+1}^{0} \left(\mathbf{X}_j - \widehat{\boldsymbol{\mu}}_{-j}\right)' \left(\mathbf{X}_j - \widehat{\boldsymbol{\mu}}_{-j}\right),$$

where $\widehat{\boldsymbol{\mu}}_{-j}$ is the leave-one-out estimate of $\boldsymbol{\mu}_j$ by (10) when the observation $\mathbf{X}_j$ is excluded in the computation and when the kernel function $K(\cdot)$ is modified to be

$$K_\varepsilon(u) = \frac{4}{4 - 3\varepsilon - \varepsilon^3} \begin{cases} \frac{3}{4}(1-u^2)I(|u| \le 1), & \text{when } |u| \ge \varepsilon, \\ \frac{3(1-\varepsilon^2)}{4\varepsilon}|u|, & \text{when } |u| < \varepsilon, \end{cases}$$

where $\varepsilon \in (0, 1)$ is a small constant. The modified kernel function $K_\varepsilon(u)$ equals 0 at $u = 0$ and is small around $u = 0$, to diminish the impact of data autocorrelation on bandwidth selection.

As mentioned above, two original process observations are allowed to be correlated if their observation times are within $b_{\max}$ apart and the serial correlation is assumed to be stationary. Then, the covariance matrices $\boldsymbol{\gamma}(s)$, for $0 \le s \le b_{\max}$, can be estimated by the following moment estimates:

$$\widehat{\boldsymbol{\gamma}}(s) = \frac{1}{m_0 - s} \sum_{j=-m_0+1}^{-s} \left(\mathbf{X}_{j+s} - \widehat{\boldsymbol{\mu}}_{j+s}\right)\left(\mathbf{X}_j - \widehat{\boldsymbol{\mu}}_j\right)', \qquad \text{for } 0 \le s \le b_{\max}.$$

3.2 *Dynamic Process Monitoring*

Next, we discuss online monitoring of the p-dimensional dynamic process with the observations $\{\mathbf{X}_n, n \ge 1\}$. When the process is IC, it is assumed that it follows the regular longitudinal pattern described by Model (8) in the sense that

$$\mathbf{X}_n = \boldsymbol{\mu}_n + \boldsymbol{\epsilon}_n, \qquad \text{for } n \ge 1, \tag{11}$$

where $\boldsymbol{\mu}_n = \boldsymbol{\mu}_{n^*}$, n^* is an integer in $[-m_0 + 1, 0]$, $n = n^* + T m_0$, $T \ge 1$ is an integer, and the error term $\boldsymbol{\epsilon}_n$ has the same covariance structure as that in Model (8).

Then, for monitoring dynamic processes using machine learning control charts, we suggest first standardizing the observed data at the current time point using the estimated IC longitudinal pattern in (10), and then decorrelating the observed data with historical data. After a proper data standardization and decorrelation of the observed data, a machine learning control chart can be used to make a decision whether the process is IC or not at the current time point. The modified machine learning control charts for monitoring dynamic processes with serial data correlation can then be summarized below.

Proposed Dynamic Process Monitoring Scheme using Machine Learning Control Charts

Step 1 **Initial Estimation**: Obtain the initial estimates $\{\widehat{\boldsymbol{\mu}}_j, -m_0 + 1 \le j \le 0\}$ and $\{\widehat{\boldsymbol{\gamma}}(s), 0 \le s \le b_{\max}\}$ from the initial IC data $\mathcal{X}_{IC}$, as discussed in Sect. 3.1.

Step 2 **Data Standardization and Decorrelation**: At the current time point n, if $n = 1$, then define the standardized observation to be

$$\mathbf{e}_1^* = [\widehat{\boldsymbol{\gamma}}(0)]^{-1/2} (\mathbf{X}_1 - \widehat{\boldsymbol{\mu}}_1).$$

Otherwise, the estimated covariance matrix of $(\mathbf{X}'_{n-b}, \mathbf{X}'_{n-b+1}, \ldots, \mathbf{X}'_n)'$ is defined to be

$$\widehat{\boldsymbol{\Sigma}}_{n,n} = \begin{pmatrix} \widehat{\boldsymbol{\gamma}}(0) & \cdots & \widehat{\boldsymbol{\gamma}}(b) \\ \vdots & \ddots & \vdots \\ \left[\widehat{\boldsymbol{\gamma}}(b)\right]' & \ldots & \widehat{\boldsymbol{\gamma}}(0) \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{n-1,n-1} & \widehat{\boldsymbol{\Sigma}}_{n-1,n} \\ \widehat{\boldsymbol{\Sigma}}'_{n-1,n} & \widehat{\boldsymbol{\gamma}}(0) \end{pmatrix},$$

where $b = \min(n-1, b_{\max})$. Then, the decorrelated and standardized observation at time n is defined to be

$$\mathbf{e}_n^* = \widehat{\mathbf{D}}_n^{-1/2}\left[-\widehat{\boldsymbol{\Sigma}}'_{n-1,n}\widehat{\boldsymbol{\Sigma}}^{-1}_{n-1,n-1}\widehat{\mathbf{e}}_{n-1} + (\mathbf{X}_n - \widehat{\boldsymbol{\mu}}_n)\right],$$

where $\widehat{\mathbf{D}}_n = \widehat{\boldsymbol{\gamma}}(0) - \widehat{\boldsymbol{\Sigma}}'_{n-1,n}\widehat{\boldsymbol{\Sigma}}^{-1}_{n-1,n-1}\widehat{\boldsymbol{\Sigma}}_{n-1,n}$, and $\widehat{\mathbf{e}}_{n-1} = [(\mathbf{X}_{n-b} - \widehat{\boldsymbol{\mu}}_{n-b})', (\mathbf{X}_{n-b+1} - \widehat{\boldsymbol{\mu}}_{n-b+1})', \ldots, (\mathbf{X}_{n-1} - \widehat{\boldsymbol{\mu}}_{n-1})']'$.

Step 3 **Decision-Making**: Apply a machine learning control chart to the decorrelated and standardized data $\{\mathbf{e}_n^*, n \geq 1\}$ to see whether a signal is triggered.

4 Simulation Studies

In this section, we investigate the numerical performance of the four existing machine learning control charts AC, RTC, DSVM and KNN described in Sect. 2, in comparison with their two modified versions. The first modified version of these four control charts are denoted as AC-D-WOC, RTC-D-WOC, DSVM-D-WOC and KNN-D-WOC, where "D" indicates that the Dynamic nature of the process under monitoring is considered in the chart, and "WOC" represents 'WithOut considering serial Correlation. The second modified version of these four control charts are denoted as AC-D-C, RTC-D-C, DSVM-D-C and KNN-D-C, where the last letter "C" denotes serial Correlation has been considered. This modified version is discussed in Sect. 3.2. In all simulation examples, the nominal ARL_0 values of all charts are fixed at 200. The number of quality characteristics is set to be $p = 5$, and the parameter $b_{\max}$ is chosen to be 15. Regarding the IC process distribution and the IC serial data correlation, the following four cases are considered:

Case I: IC process observations $\{\mathbf{X}_n, n \geq 1\}$ are i.i.d. with the IC distribution $N_p(\mathbf{0}, \boldsymbol{I}_p)$.

Case II: IC process observations $\{\mathbf{X}_n, n \geq 1\}$ are generated from Model (11). Their mean and correlation structures are specified in Model (8), where the means are defined to be

$$\boldsymbol{\mu}_j = \left[sin(2\pi t_j), cos(2\pi t_j), sin^2(2\pi t_j), cos^2(2\pi t_j), sin(2\pi t_j) + cos(2\pi t_j)\right],$$

$t_j = (j + m_0)/m_0$ for $j = -m_0 + 1, -m_0 + 2, \ldots, 0$, each component of the error term $\boldsymbol{\epsilon}_j$ has the standardized χ_3^2 distribution, and the covariance matrix of $\boldsymbol{\epsilon}_j$ is $\boldsymbol{I}_p$.

Case III: Same as Case II, except that the error terms $\{\boldsymbol{\epsilon}_j\}$ are assumed to follow the vector-AR(1) model $\boldsymbol{\epsilon}_j = 0.2\boldsymbol{\epsilon}_{j-1} + \boldsymbol{\eta}_j$, where $\boldsymbol{\epsilon}_0 = \mathbf{0}$, each component of $\boldsymbol{\eta}_j$ has the standardized χ_3^2 distribution, and the covariance matrix of $\boldsymbol{\eta}_j$ is $\boldsymbol{I}_p$.

Case IV: Same as Case III, except that the covariance matrix of $\boldsymbol{\eta}_j$ is $\boldsymbol{\Sigma} = (\sigma_{l_1 l_2})_{p\times p}$ with $\sigma_{l_1 l_2} = 0.5^{|l_1 - l_2|}$, for $l_1, l_2 = 1, 2, \ldots, p$.

For the four cases described above, Case I is the conventional case with i.i.d. mean $\mathbf{0}$ IC process observations and the normal IC process distribution. Cases II–IV consider three different dynamic processes. The dynamic process in Case II still has independent observations at different observation times, and the p quality variables are independent with each other as well. The dynamic process in Case III is the same as that in Case II, except that process observations are serially correlated. Dynamic process observations in Case IV are serially correlated and the p quality variables are mutually associated as well.

4.1 Evaluation of the IC Performance

We first evaluate the IC performance of the related control charts. In the simulation study, the IC sample size m_0 is fixed at 2,000. The weighting parameter λ in the chart AC and its two modified versions are chosen to be 0.2, as suggested in Hu and Runger [19], the moving window size w in the charts RTC and DSVM and their modified versions are chosen to be 10, as suggested in Deng et al. [12] and He et al. [18], and the number of nearest observations k in the chart KNN and its modified versions is chosen to be 30, as suggested in Sukchotrat et al. [33]. For each method, its actual ARL_0 value is computed as follows. First, an IC dataset of size m_0 is generated from the IC model, and the IC parameters are estimated from the IC data. Second, the conditional ARL_0 value of the chart given the IC dataset is calculated based on 1,000 replicated simulations of online process monitoring of 2,000 sequential process observations. Third, the previous two steps are repeated for 100 times, and the sample average of the 100 conditional ARL_0 values is used as the estimated actual ARL_0 value of the chart. The standard error of the estimated actual ARL_0 value can also be computed as the standard deviation of the 100 conditional ARL_0 values divided by $\sqrt{100}$. The estimated ARL_0 values in different cases considered are shown in Table 1.

From Table 1, we can have the following conclusions. (i) The four original machine learning control charts all have a reasonable performance in Case I when the process observations are i.i.d. with a normal distribution, but they are unreliable to use in all other cases when some or all of these assumptions are violated because their estimated actual ARL_0 values are substantially different from the nominal ARL_0 level of 200 in these cases. (ii) The first modified version of the four machine learning control charts AC-D-WOC, RTC-D-WOC, DSVM-D-WOC, and KNN-D-WOC perform well in Cases I and II when the independence assumption is valid, but their performance is quite poor in Cases III and IV when this assumption is violated. (iii) As a comparison, the second modified version of the four machine learning control charts AC-D-C, RTC-D-C, DSVM-D-C, and KNN-D-C have a reasonably good performance in all cases considered, since its estimated actual ARL_0 values are always within 10% of the nominal ARL_0 level. Therefore, this example confirms that the IC performance of the machine learning control charts can be improved in a substantial way by using the suggested modification discussed in Sect. 3.2.

Table 1 Actual ARL_0 values and their standard errors (in parentheses) of four machine learning control charts and their modified versions when their nominal ARL_0 values are fixed at 200, $p = 5$, and $m_0 = 2,000$

Methods	Case I	Case II	Case III	Case IV
AC	207(4.22)	53.7(1.40)	44.7(1.44)	42.8(1.37)
AC-D-WOC	195(4.7)	187(3.77)	107(3.38)	110(3.44)
AC-D-C	203(3.96)	193(3.65)	189(3.42)	191(3.74)
RTC	191(3.88)	103(2.33)	89.6(1.74)	90.4(1.83)
RTC-D-WOC	187(3.76)	194(4.27)	142(2.95)	139(2.87)
RTC-D-C	189(3.69)	192(4.08)	206(4.93)	209(5.03)
DSVM	210(4.19)	125(2.67)	113(2.05)	110(2.56)
DSVM-D-WOC	193(4.30)	196(4.22)	147(3.03)	148(2.99)
DSVM-D-C	191(4.04)	202(3.99)	194(4.12)	195(4.15)
KNN	208(4.36)	141(2.45)	143(2.56)	139(2.41)
KNN-D-WOC	205(4.27)	190(3.15)	170(3.27)	167(2.95)
KNN-D-C	189(3.96)	193(4.20)	202(4.32)	204(4.41)

4.2 Evaluation of the OC Performance

Next, we evaluate the OC performance of the related charts in case when $m_0 = 2,000$. In order to make the comparison more meaningful, we intentionally adjust the control limits of different control charts so that their actual ARL_0 values equal the nominal ARL_0 value of 200 in all cases considered. In the next simulation example, it is assumed that all quality variables have a same shift at the beginning of online process monitoring with the shift size δ changing from 0 to 1 with a step of 0.25. Because different control charts have different procedure parameters (e.g., the moving window sizes of RTC and DSVM) and their performance may not be comparable if their parameters are set to be the same, here we compare their optimal OC performance to make the comparison fair. Namely, to detect a given shift by a chart, the related procedure parameter is chosen by minimizing the OC average run length, denoted as ARL_1, of the chart while maintaining its ARL_0 value at 200. The resulting ARL_1 value is called optimal ARL_1 value hereafter. The results of the optimal ARL_1 values of these machine learning control charts and their modified versions in Cases I–IV are presented in Fig. 1.

From the figure, we can have the following conclusions. First, all four machine learning control charts and their two modified versions perform reasonably well in Case I when the process observations are i.i.d. with a normal distribution, since their model assumptions are all satisfied. Second, The first modified version of the related control charts AC-D-WOC, RTC-D-WOC, DSVM-D-WOC, and KNN-D-WOC are the most effective one among the three version of all charts in Case II when the independence assumption is valid, but are less effective in Cases III and IV when this assumption is invalid. Third, the second modified version of the four machine learning control charts AC-D-C, RTC-D-C, DSVM-D-C, and KNN-D-C have the

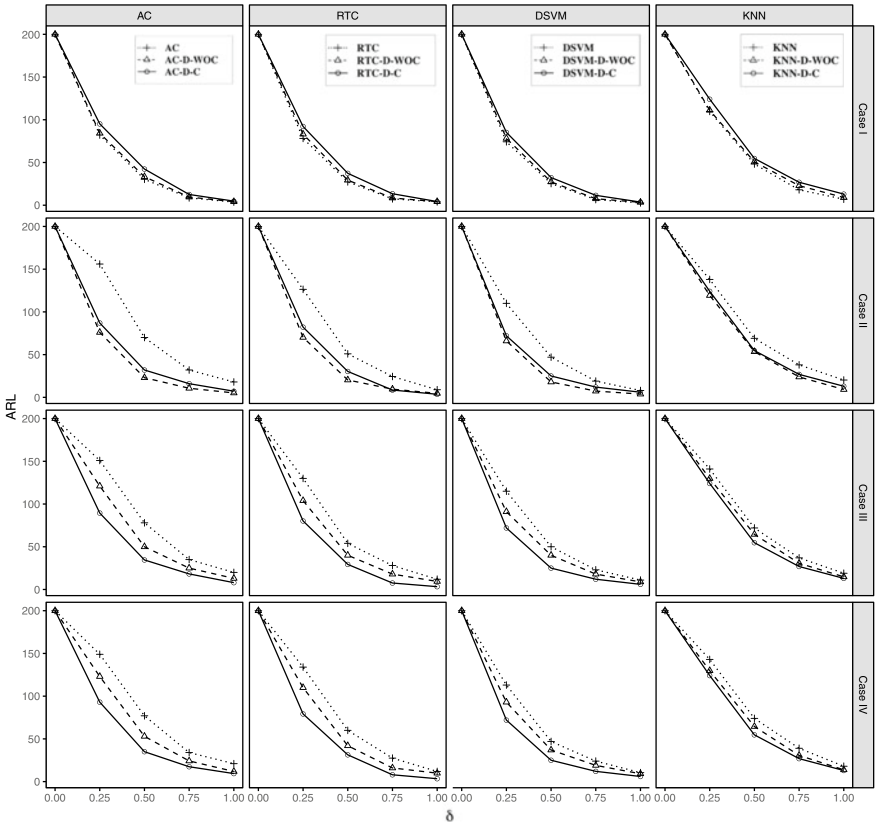


Fig. 1 Optimal ARL_1 values of the four control charts and their modified versions when their nominal ARL_0 values are fixed at 200, $p = 5$, m_0 = 2,000, and all quality characteristics have the same shift with the shift size δ changing among 0.25, 0.5, 0.75, and 1

best performance among the three versions of all charts in Cases III and VI when the process under monitoring is dynamic with serial data correlation.

5 An Application

In this section, we demonstrate the application of the modified machine learning control charts discussed in Sects. 3 and 4 using a real dataset, which contains electricity generation and weather data in Spain. This dataset can be downloaded from the web page of Kaggle with the link https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather. Electricity is generated using a variety of resources, including coal, natural gas, nuclear energy, and solar energy, and its usage usually depends on the meteorological conditions [32]. For examples, the colder months often bring more electricity usage as more electricity is spent on the heating system. However, when excess electricity is generated, much time and resources would be wasted because electricity cannot be stored in large quantities efficiently [25]. Therefore, it is important to online monitor the electricity generation and demand in the electric industry. If something unusual happens (e.g., unseasonable cold weather), the electric utility companies can take actions quickly to adjust the amount of electricity generated to meet demand. In this analysis, the amount of electricity generated by three most common energy sources, including gas, coal and oil, and two important environmental variables, i.e., temperature and humidity are considered. The dataset used here contains observations of the five variables during a time period from January 1, 2015 to December 31, 2016. The original data of these five variables are shown in Fig. 2. From the figure, it can be seen that there is a quite

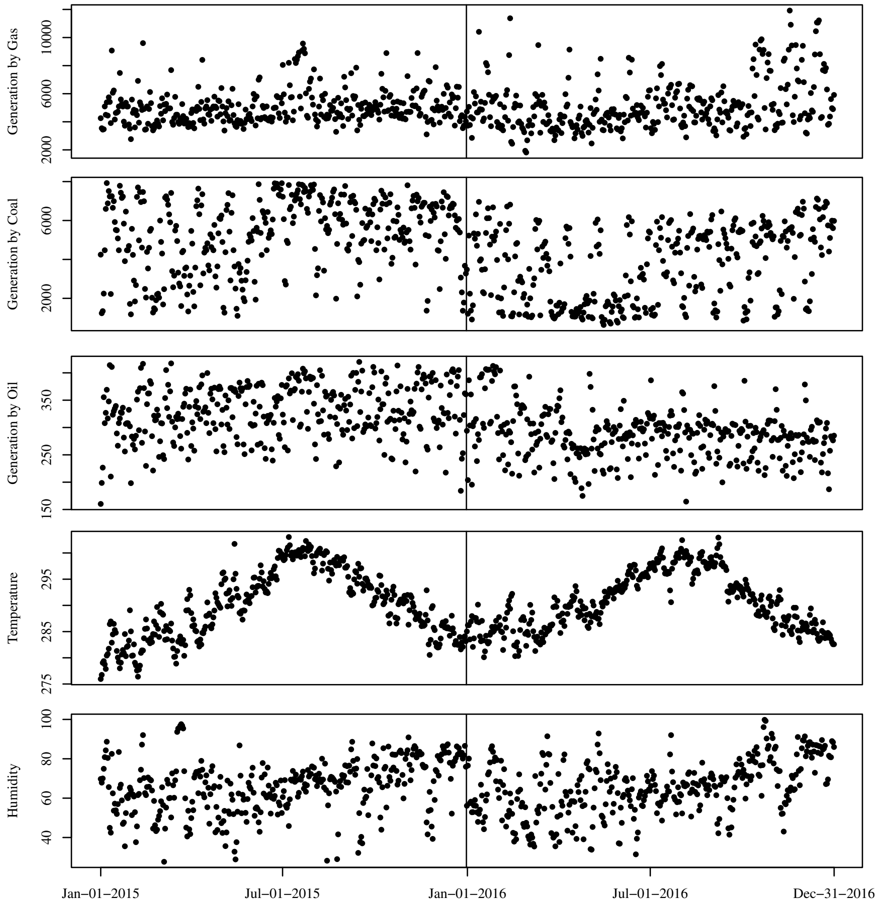


Fig. 2 Original observations of five variables considered in the electricity example. The solid vertical line in each plot separates the initial IC data from the data for online process monitoring

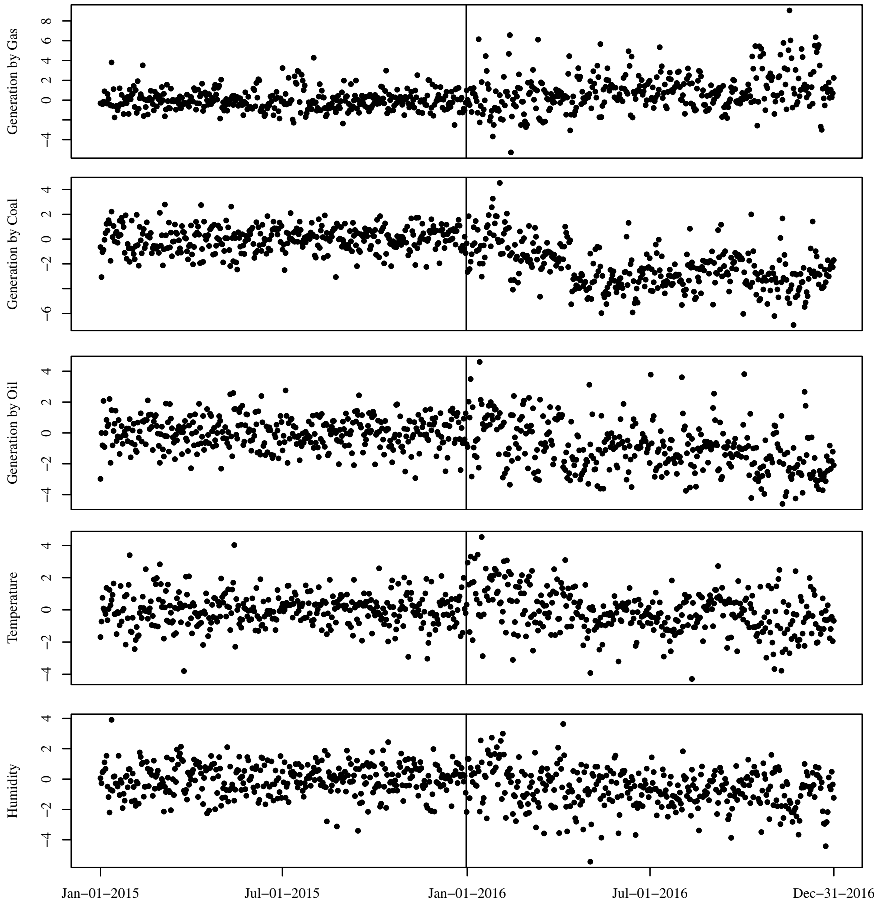


Fig. 3 Standardized and decorrelated observations of five variables considered in Fig. 2. The solid vertical line in each plot separates the initial IC data from the data for online process monitoring

obvious yearly seasonality in the observed data, the temperature is higher during summer times, and the amount of electricity generated by coal seems higher in the last six months of each year. In our analysis, the data in the first year are used as the IC data for estimating the regular longitudinal pattern of the five variables, and the data in the second year are used for online process monitoring.

For the IC data, we first compute the initial LLK estimates $\widehat{\boldsymbol{\mu}}_j$ by (10), and then obtain the residuals $\mathbf{X}_j - \widehat{\boldsymbol{\mu}}_j$, for all j. Next, we use the Ljung-Box test for checking serial data correlation in the residuals of each variable. The p-values of this test are all $< 2.3 \times 10^{-9}$ for the five variable. Thus, there is a significant autocorrelation in the IC data. The Augmented Dickey-Fuller (ADF) test for stationarity of the autocorrelation gives p-values of <0.01 for all five variables, which implies that the stationarity assumption is valid in this case. To check the normality assumption for

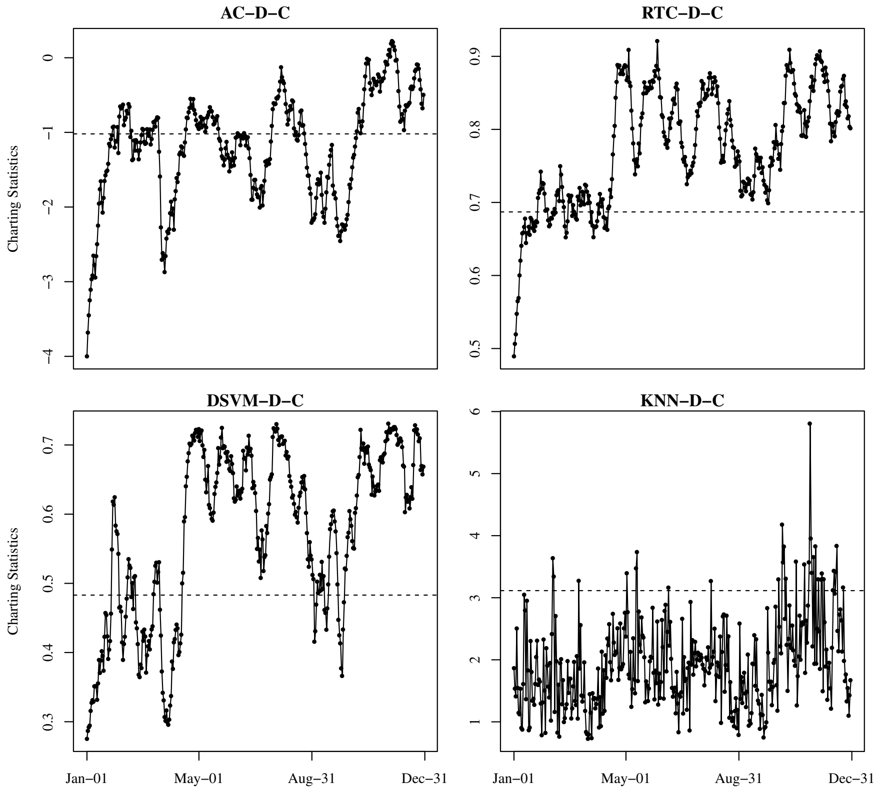


Fig. 4 Control charts for online monitoring of the data during January 1 and December 31, 2016. In each plot, the horizontal dashed line denotes the control limit of the related control chart

the data, the Shapiro test is performed, and it gives a p-values of 2.9×10^{-7}, which implies that the distribution of the standardized IC data is significantly different from a normal distribution. Therefore, the four modified control charts AC-D-C, RTC-D-C, DSVM-D-C and KNN-D-C should be appropriate to use in this example, because the IC data have a dynamic pattern, significant stationary serial data correlation, and a non-normal distribution. The standardized and decorrelated data of the five variables by the procedure discussed in Sect. 3.2 are shown in Fig. 3, from which it can be seen that the standardized and decorrelated IC data are indeed quite stable, and the the standardized and decorrelated data in the second year seem to be quite different from the IC data starting from the very beginning of the second year.

Next, we apply the four charts AC-D-C, RTC-D-C, DSVM-D-C and KNN-D-C to this dataset for online process monitoring starting from January 1, 2016. In all control charts, the nominal ARL_0 values are fixed at 200, and their control limits are computed in the same way as that in the simulation study of Sect. 4. All four control charts are shown in Fig. 4. From the plots in the figure, the charts AC-D-C, RTC-D-C, DSVM-D-C and KNN-D-C give their first signals on the Jan 30th, Jan 28th, Jan 27th, and Feb 13th, respectively. By checking the standardized and decorrelated process observations shown in Fig. 3, it seems that all fours chart can detect a systematic change in the process well and the chart DSVM-D-C gives the earliest signal among them.

6 Concluding Remarks

Some control charts based on machine learning approaches have been developed recently in the SPC literature. However, most existing machine learning control charts are based on the assumptions that the process observations at different time points are independent and identically distributed. So, they would be unreliable to use in case when the IC process distribution changes over time. In this chapter, we have suggested a modification procedure for some representative existing machine learning control charts using the nonparametric longitudinal modeling and sequential decorrelation algorithms. Numerical studies show that the performance of these modified control charts is substantially better than their original versions in cases when the IC process distribution is time-varying.

There are still some issues about the modified machine learning control charts that need to be addressed in our future research. For example, these machine learning methods require a relatively large IC dataset. But, in some applications, a relatively large IC dataset may not be available. In such cases, self-starting control charts might be helpful (cf., [15]). In addition, the current proposed methods assume that the serial correlation in process observations is short-ranged and stationary. Even through these assumptions should be reasonable in many applications, the serial data correlation could be long-range and non-stationary in some other applications (cf., [3, 6]).

References

1. Aggarwal CC (2018) Neural networks and deep learning. Springer, New Yorker
2. Altman NS (1990) Kernel smoothing of data with correlated errors. J Am Stat Assoc 85:749–758
3. Altmann EG, Cristadoro G, Esposti MD (2012) On the origin of long-range correlations in texts. Proc Natl Acad Sci USA 109:11582–11587
4. Apley DW, Lee HC (2008) Robustness comparison of exponentially weighted moving-average charts on autocorrelated data and on residuals. J Qual Technol 40:428–447
5. Apley DW, Tsung F (2002) The autoregressive T^2 chart for monitoring univariate autocorrelated processes. J Qual Technol 34:80–96
6. Beran J (1992) Statistical methods for data with long-range dependence. Stat Sci 4:404–416
7. Brabanter KD, Brabanter JD, Suykens JAK, De Moor B (2001) Kernel regression in the presence of correlated errors. J Mach Learn Res 12:1955–1976
8. Breiman L (2001) Random forests. Mach Learn 45:5–32
9. Capizzi G, Masarotto G (2016) Efficient control chart calibration by simulated stochastic approximation. IIE Trans 48:57–65

10. Chakraborti S, Graham MA (2019) Nonparametric (distribution-free) control charts: an updated overview and some results. Qual Eng 31:523–544
11. Cortes C, Vapnik VN (1995) Support-vector networks. Mach Learn 20:273–297
12. Deng H, Runger G, Tuv E (2012) System monitoring with real-time contrasts. J Qual Technol 44:9–27
13. Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. Theory Probab Appl 14:153–158
14. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning—Data mining, inference, and prediction. Springer, Berlin
15. Hawkins DM (1987) Self-starting CUSUMS for location and scale. Statistician 36:299–315
16. Hawkins DM, Olwell DH (1998) Cumulative sum charts and charting for quality improvement. Springer, New York
17. Hawkins DM, Qiu P, Kang CW (2003) The changepoint model for statistical process control. J Qual Technol 35:355–366
18. He S, Jiang W, Deng H (2018) A distance-based control chart for monitoring multivariate processes using support vector machines. Ann Oper Res 263:191–207
19. Hu J, Runger G (2010) Time-based detection of changes to multivariate patterns. Ann Oper Res 174:67–81
20. Li J (2021) Nonparametric adaptive CUSUM chart for detecting arbitrary distributional changes. J Qual Technol 53:154–172
21. Li W, Zhang C, Tsung F, Mei Y (2021) Nonparametric monitoring of multivariate data via KNN learning. Int J Prod Res 59:6311–6326
22. Montgomery DC (2012) Introduction to statistical quality control. Wiley, New York
23. Opsomer J, Wang Y, Yang Y (2001) Nonparametric regression with correlated errors. Stat Sci 16:134–153
24. Page ES (1954) Continuous inspection scheme. Biometrika 41:100–115
25. Poonpun P, Jewell W (2014) Analysis of the cost per kilowatt hour to store electricity. IEEE Trans Energy Convers 23:529–534
26. Qiu P (2014) Introduction to statistical process control. Chapman Hall/CRC, Boca Raton, FL
27. Qiu P (2018) Some perspectives on nonparametric statistical process control. J Qual Technol 50:49–65
28. Qiu P, Xiang D (2014) Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior. Technometrics 56:248–260
29. Qiu P, Xie X (2022) Transparent sequential learning for statistical process control of serially correlated data. Technometrics 64:487–501
30. Roberts SV (1959) Control chart tests based on geometric moving averages. Technometrics 1:239–250
31. Shewhart WA (1931) Economic control of quality of manufactured product. D. Van Nostrand Company, New York
32. Staffell I, Pfenninger S (2018) The increasing impact of weather on electricity supply and demand. Energy 145:65–78
33. Sukchotrat T, Kim SB, Tsung F (2009) One-class classification-based control charts for multivariate process monitoring. IIE Trans 42:107–120
34. Tuv E, Runger G (2003) Learning patterns through artificial contrasts with application to process control. Trans Inf Commun Technol 29:63–72
35. Xiang D, Qiu P, Pu X (2013) Nonparametric regression analysis of multivariate longitudinal data. Stat Sinica 23:769–789
36. Xie X, Qiu P (2023) Control chart for dynamic process monitoring with an application to air pollution surveillance. Ann Appl Stat 17:47–66
37. Xie X, Qiu P (2022) Machine learning control charts for monitoring serially correlated data. In: Tran KP (ed) Control charts and machine learning for anomaly detection in manufacturing. Springer, pp 131–147
38. Xue L, Qiu P (2021) A nonparametric CUSUM chart for monitoring multivariate serially correlated processes. J Qual Technol 53:396–409
39. Zhang C, Tsung F, Zou C (2015) A general framework for monitoring complex processes with both in-control and out-of-control information. Comput Ind Eng 85:157–168

Fault Prediction of Papermaking Process Based on Gaussian Mixture Model and Mahalanobis Distance

Guojian Chen, Zhenglei He, Yi Man, Jigeng Li, Mengna Hong, and Kim Phuc Tran

Abstract Equipment monitoring and process fault prediction are increasingly concerned in the modern industry due to the growing complexity of the production process and the high risk derived from severe consequences on the paper mills in case of production failure. Whereas the paper manufacturing process is continuous that is difficult to be warned early of faults. To address such issues, this Chapter proposes a data-driven approach to predict fault in the papermaking process on the basis of correlation analysis and clustering algorithms. Historical operating data of key variables were acquired in normal operating conditions. The health benchmark dataset was constructed based on the Gaussian mixture model (GMM) and Mahalanobis distance (MD) to evaluate the operating status of the papermaking process. The verification results showed that the proposed model has a fault prediction accuracy of 76.8% and a recall rate of 72.5%, which allows anomalous data to be observed in advance, providing valuable time for subsequent fault diagnosis.

Keywords Fault prediction · Papermaking process · Gaussian mixture model · Mahalanobis distance

1 Introduction

The paper manufacturing process is quite a long and continuous process in which a large number of different machines follow the production sequence. With the globalization and development of the economy, the papermaking industry is faced with increasingly fierce competition, so its productivity is supposed to be improved [1, 2]. A growing number of machines are putting into production in the field, which

G. Chen · Z. He (✉) · Y. Man · J. Li · M. Hong
State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou 510640, China
e-mail: hezhenglei@scut.edu.cn

K. P. Tran
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_5

has increased productivity in recent decades. However, due to the influence of a variety of factors on the production process, such as pulp raw materials, operations of staff, the operational status, and the environment of the equipment, unexpected breakdowns may occur during the papermaking process, resulting in the suspension of production and consequently enhance the production cost, energy consumption, as well as disadvantageous environmental impacts [3]. In paper mills, the tools of traditional system monitoring focused on the detection of the assignable causes of the system's abnormal status as soon after it occurs as possible [4]. Regular inspections are carried out to check for potential faults [5]. However, it is difficult for conventional approaches to detect slight fluctuations before the failure, and early prediction of fault is therefore of great importance [6].

At present, most fault prediction techniques are still in the status of experimental applications. Current methods are still not complete. The fault prediction methods are generally divided into three categories: model-based methods, knowledge-based methods, and data-based methods [7]. Papermaking is a complicated process consisting of a range of components, and there are complex interactions among the components, leading it difficult to establish an accurate mathematical model. The knowledge-based methods rely mainly on expert experience, which is integrated into the computer knowledgebase and the different faults are predicted by computer programs [8]. This approach requires a large amount of extensive experience knowledge and is influenced by expert knowledge. Data-based methods use the record data of the papermaking process, it is not necessary for this approach to access the physical model of the real papermaking process or much expert knowledge about faults.

There has been substantial research and applications in the field of industrial fault prediction based on data-driven approaches [9]. The assessment of the health status of production processes is also an important area of research [13]. The health assessment method based on operational condition recognition, as reported, has achieved the goal of evaluating the online health condition of wind turbines [10]. A new method to identify early defects of wind turbines has been presented based on Dynamical Network Marker (DNM) by adopting the data of supervisory control and data acquisition (SCADA) [11]. A method of health evaluation is proposed by using sensitivity analysis and stability analysis to select health indicators combined with the K-means clustering to estimate the health status of rolling bearings [12]. A deep reinforcement learning technique for predicting potential aircraft failures has been presented by using the Aircraft Central Maintenance System dataset [14]. Furthermore, acoustic and vibration analysis based on the similarity characteristics of filter banks has been applied to the fault detection of machines [15]. The analytical clustering of characteristic variables for industrial processes is regarded as an unsupervised process. In this respect, an unsupervised weld defect classification has been presented by using multivariate generalized GMM with exact computation of mean and shape parameters [16]. The model combining the advantages of long short-term memory (LSTM) networks with statistical process analysis is proposed to achieve the prediction of aero-engine bearing failures with multi-stage performance degradation [17].

In view of the above discussion, this Chapter aims to monitor the key parameters of the papermaking process, and propose a fault prediction model for the process. Data analysis and feature engineering are performed to find out the key variables, and then the health benchmark model is established based on GMM, the threshold is confirmed based on Mahalanobis distance, and these methods are collaboratively used to predict the paper break fault.

The rest of this Chapter is organized as follows: Sect. 2 introduces the process of modeling. Case studies in Sect. 3 validate the proposed method. Finally, the work is concluded in Sect. 4.

2 Process Modeling

In the case of papermaking machines, the occurrence of faults is often a gradual process with the abnormal change of parameters. And there is a correlation between the various parameters as time goes by. As a consequence, subsequent warning situations can be analyzed through time series analysis. The overall process of the modeling and analysis is shown in Fig. 1.

As there are different parts in the papermaking machine, each component has different faults and causes. It is necessary for the collected data to be divided in accordance with the different sections when pre-processing data. The raw data collected by the sensor and control system from the papermaking machine is unable to be used directly for process modeling and the acquired data needs to be pre-processed. Considering the correlation and redundancy amount of multiple variables, multi-dimensional features were selected and dimensional reduction was conducted, and then the reduced multi-dimensional features were divided into different sets according to different failure causes. The features within each subset were processed through sliding windows so that the feature information can be analyzed from two different dimensions in terms of mean and variance. The health benchmark dataset was generated by GMM, and then, a reasonable threshold was set based on the distance metric and kernel density estimation. Thresholds were used to determine whether there was a trend of fault in the actual production process to achieve fault prediction.

2.1 Data Pre-processing

The raw data collected from the system contains anomaly values, absence of data, and other abnormalities, such as labels or time stamps, which may mismatch with the actual data. So it is essential to pre-process the data, eliminate the obvious abnormal values, and combine the production reality of the papermaking process. Meanwhile,

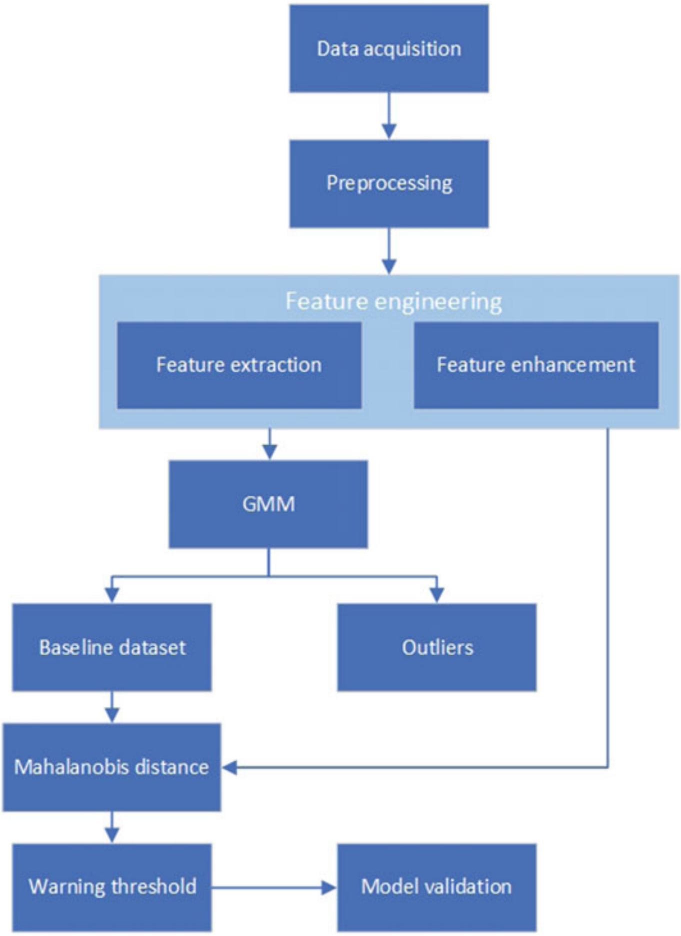


Fig. 1 Flowchart for papermaking process fault prediction

it is removed the records during the production process when the power on the signal is off. Finally, the data after cleaning is standardized according to Eq. (1) as follows:

$$x_i^{'} = \frac{x_i - \mu}{\sigma} \tag{1}$$

where x_i, x_i' denotes the state variable before and after being standardized, μ denotes the mean and σ denotes the standard deviation.

2.2 Feature Engineering

2.2.1 Correlation Analysis

There is a strong correlation between different features in the pre-processed data, therefore, the features of data need to be extracted to obtain a more streamlined subset of features.

1. *Pearson correlation coefficient method.* The Pearson correlation coefficient method is a common method to implement such a function, and it was used to analyze the correlation between two continuous variables. For two sets of variables X and Y, the correlation coefficient r is calculated as follows:

$$\mathrm{r} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{2}$$

   where n is the number of samples, X_i, Y_i is the *i*-th value of *X* and *Y* respectively, and $\bar{X}$, $\bar{Y}$ is the mean of *X* and *Y* respectively. The correlation coefficient *r* ranges from –1 to 1, and the closer the absolute value is to 1, the greater the correlation between the two variables. By correlation analysis, one of the features with high correlation is retained and the other is removed to reduce data redundancy so that the dimensional of the data is supposed to be reduced.
2. *Point two-column correlation method.* The point two-column correlation method is a statistical method used to measure the relevance between data on continuous variables and data on dichotomous variables. The point two-column correlation coefficient takes a value between [–1, 1]. The closer the absolute value is to 1, the stronger the relevance. The calculation formula is shown as follows:

$$r = \frac{\bar{X}_p - \bar{X}_q}{S_X}\sqrt{pq} \tag{3}$$

   for the variable *r*, its values can be expressed in 0 or 1. The rate of taking 1 is p while the rate of taking 0 is q. Where $\bar{X}_p$ means the mean of the data corresponding to the variable p in the continuous variable r; $\bar{X}_q$ means the mean of the data corresponding to the variable q in the continuous variable; S_x means the standard deviation of the continuous variable.
3. *Random forests.* The random forests are able to be used for feature selection or correlation analysis, which use random resampling and random node splitting techniques to construct multiple decision trees. Random Forests are a typical non-linear integrated learning method and are able to analyze the relative importance of each feature, as they are relatively fast to learn, the importance analysis of feature variables is often used as a feature selection indicator for high-dimensional data.

2.2.2 Sliding Windows

Time series data is characterized by non-stationary fluctuations over time, so it is vital for the data to be analyzed by certain methods. Sliding windows are available for time series data extracted to obtain the statistical indicators, which allows for the extraction of the mean, variance, and other characteristics of time series data. The two measures reflect the trend in data concentration and the degree of dispersion respectively. For each dimensional variable, the analysis can be carried out more accurately from different perspectives after the sliding windows process.

2.2.3 Clustering Analysis

The K-means clustering algorithm converges quickly, therefore, that is widely used for data analysis [18]. Before clustering can be performed, the K-value, also known as the number of clusters, is supposed to be determined. There are various metrics for evaluating the effect of clustering. In the case of data without category labels, there is no unique evaluation metric applied. We can evaluate the effect of clustering by the principle of intra-class aggregation and low coupling between classes. In this Chapter, the silhouette coefficient and the Calinski-Harabaz Index (also known as the CH score) are used as indicators for evaluating the clustering effect at different K values.

The silhouette coefficient method combines the cohesion and separation of clusters and is used to assess the effect of clustering with the aim of minimizing the internal distance and maximizing the external distance of each class of the sample. The formula for the silhouette coefficient S(i) is expressed as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

where a(i) is the average distance to the other samples in the cluster and b(i) is the average distance to the other cluster samples. The average silhouette coefficient can be obtained by averaging the contour coefficients of all the samples, and the range of value is [–1, 1]. The closer the value to 1, means the better the clustering performance; conversely, the smaller the value, the worse the clustering performance.

Compared to the silhouette coefficient method, the CH score is faster to calculate, and it aims to cluster as many samples as possible with few categories while obtaining a great clustering effect. The score is calculated by evaluating the variance between and within categories. The smaller the covariance of the data within categories and the larger the covariance between categories, the better the clustering performance. The higher CH score indicates a better clustering effect, and the expression is shown as follows:

$$\mathrm{s(k)} = \frac{\mathrm{tr}\,(B_k)}{\mathrm{tr}\,(W_k)} \frac{m - k}{k - 1} \tag{5}$$

where tr is the trace of the matrix, B_k and W_k is the covariance matrix between and within categories respectively; m is the full sample size and k is the number of categories.

2.3 Health Benchmark Model

2.3.1 Gaussian Mixture Model

The gaussian mixture model is a combination of several single Gaussian models, of which each can demonstrate distribution characteristics of data within subspace. Meanwhile, these sub-models are the hidden variables of the mixture model. Gaussian models are often used to fit non-linear data due to the smoothing property [19]. Gaussian mixture models are able to model the characteristics of high-dimensional spaces. The probability of a Gaussian mixture model is shown as follows:

$$P(x \mid \theta) = \sum_{k=1}^{K} \alpha_k \phi\left(x \mid \theta_k\right) \tag{6}$$

where α_k is the probability that the observed model is the Kth sub-model, which can also be interpreted as the weighting factor ($\sum_{k=1}^{K} \alpha_k = 1$).$\phi\left(x \mid \theta_k\right)$ is the Gaussian distribution density function of the Kth sub-model, whose expression is

$$\phi\left(x \mid \theta_k\right) = \frac{1}{(2\pi)^{\frac{D}{2}} \left|\Sigma_k\right|^{\frac{1}{2}}} \exp\left(-\frac{\left(x-\mu_k\right)^T \Sigma_k^{-1} \left(x-\mu_k\right)}{2}\right) \tag{7}$$

Among them, theμ_k and Σ_k denote the mean and covariance of $\phi\left(x \mid \theta_k\right)$ respectively, and D is the dimensionality of the feature.

Suppose there is a training sample set $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_\mathrm{n}\}$ containing N samples obeying independent identical distribution, then the log-likelihood function of the training samples is

$$\log L(\theta) = \sum_{j=1}^{N} \log P\left(x_j \mid \theta\right) = \sum_{j=1}^{N} \log\left(\sum_{k=1}^{K} \alpha_k \phi\left(x \mid \theta_k\right)\right) \tag{8}$$

The maximum likelihood estimation method can be solved to find the $\log L(\theta)$. the maximum parameter estimation of $\theta = \left(\alpha_k,\ \mu_k,\ \Sigma_k\right)$ The Expectation-Maximum algorithm (EM) is usually used to iteratively update the parameters of the Gaussian mixture model. The parameter estimation process is as follows:

- Initialization: the identity matrix is used as the initial covariance matrix for each Gaussian distribution, $\alpha_k = 1/K$ as the initial weights of the distribution, and the first vector of observations as the initial mean.

- E-step: calculate the probability that each data j comes from sub-model k based on the current parameters γ_{jk}:

$$\gamma_{jk} = \frac{\alpha_k \phi\left(x_j \mid \theta_k\right)}{\sum_{k=1}^{K} \alpha_k \phi\left(x_j \mid \theta_k\right)}, j = 1, 2, \ldots, N; k = 1, 2, \ldots, K \tag{9}$$

- M-step: calculation of model parameters for a new iteration:

$$\mu'_k = \frac{\sum_j^N \left(\gamma_{jk} x_j\right)}{\sum_j^N \gamma_{jk}}, k = 1, 2, \ldots, K \tag{10}$$

$$\Sigma'_k = \frac{\sum_j^N \gamma_{jk}\left(x_j - \mu'_k\right)\left(x_j - \mu'_k\right)^T}{\sum_j^N \gamma_{jk}}, k = 1, 2, \ldots, K \tag{11}$$

$$\alpha'_k = \frac{\sum_{j=1}^N \gamma_{jk}}{N}, k = 1, 2, \ldots, K \tag{12}$$

- Convergence determination: repeat the calculation of E-step and M-step until convergence ($||\theta_{i+1} - \theta_i|| < \varepsilon$, which is a very small positive number, indicating that the parameter changes are very small after the iteration).

After the steps above have been completed, the generated health benchmark dataset is used as a criterion for health assessment. The deviation of the actual production data from the health benchmark is measured by distance to determine whether there is a fault. As Euclidean distance measures only the absolute distance between two points in a multidimensional space, it does not reflect these deviations well. The Mahalanobis distance is a commonly used distance metric that performs better than the Euclidean distance in pattern recognition because it takes into account the magnitude of the feature parameters and the correlation between the features. It is expressed as follows:

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \tag{13}$$

where x is the data point of the actual production process, μ and Σ denote the mean and covariance matrices of the data points in the distribution M respectively.

2.3.2 Determination of Threshold

There are two methods of probability density estimation, the parametric method, and the non-parametric method, parametric density estimation is used to estimate the parameters of distribution when it is confirmed that the data follow a known distribution. However, in fact, the data may not follow a common probability distribution, or not easy to fit with a certain distribution. In such cases, non-parametric estimation methods are generally used. The most common non-parametric method for estimating the probability density function of a continuous random variable is the kernel density estimation (KDE) method.

There are two important parameters in KDE, one is the smoothing parameter which is also known as the bandwidth, and the other is the kernel function. The KDE method is actually combined by using a kernel function to treat the data and bandwidth of each data point as a parameter of the kernel function to obtain N kernel functions and then linearly superimposed to form the kernel density estimation function, which is normalized to the kernel density probability density function.

Suppose $\mathrm{x_1, x_2, \cdots, x_n}$ are independent and identically distributed N sample points, let the probability density function be f, and the formula for kernel density estimation is shown as follow:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(x - x_i\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{14}$$

where K is the kernel function and h is the bandwidth. Once the bandwidth is determined, the estimation results based on different kernel functions are similar. Gaussian kernel functions are generally used as the kernel function for kernel density estimation. The kernel density estimation method is actually generated by using the kernel function to treat the data point and bandwidth of each data as the parameters of the kernel function to get N kernel functions, and then linearly superimposed to form the kernel density estimation function, and then normalized to be the probability density function.

3 Case Studies

The resulting variables were further filtered by using the K-means algorithm to cluster the data, selecting different K values, and using the CH score and silhouette coefficient as evaluation indicators to assess the clustering effect. After determining the optimal number of clusters, the variables were subjected to analysis, and the minority of classes in each variable were labeled as abnormal, and the majority of classes as normal. For instance, the average pressure at the outlet of the sizing pump was subjected to K-means clustering, and the clustering effect analysis is shown in Fig. 2.

The red line represents the silhouette coefficient, while the black line represents the CH score. It can be seen that the CH score generally decreases with an increasing K value of the clusters and then gradually stabilize, while the silhouette coefficient enhances with growing K values and keep falling since K > 4. Therefore, the K value of the clustering can be chosen as 5, and the minority categories are marked as anomalies after the clustering.

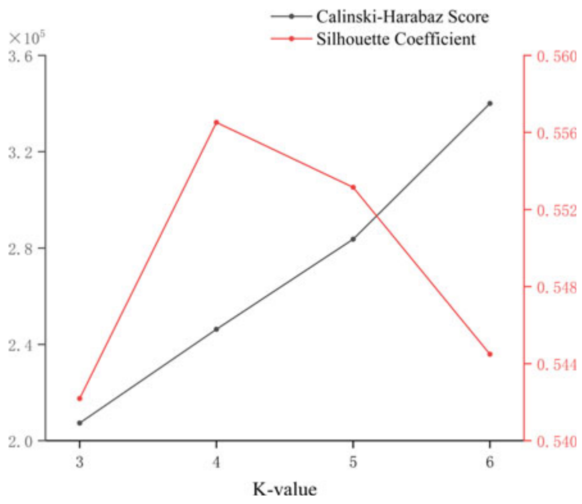


Fig. 2 Clustering effect analysis

Table 1 Point two column correlation coefficient

Status parameters	Correlation coefficient
Average pressure at the outlet of the sizing pump	0.536
The average value of the pulp chest level	0.511
Cleaning squeegee loading pressure variance	0.468
Wrinkle squeegee loading pressure variance	0.460
Paper-breaking squeegee loading pressure variance	0.449

Table 2 Random forest feature analysis

Status parameters	Characterisation
Average pressure at the outlet of the sizing pump	0.229
The average value of the pulp chest level	0.193
Cleaning squeegee loading pressure variance	0.191
Wrinkle squeegee loading pressure variance	0.095
Paper-breaking squeegee loading pressure variance	0.071

The correlation analysis of discrete-continuous data was conducted. Both the point two-column correlation method and the random forests method were used for the subsequent correlation analysis. After sorting the features in importance, 5 of them were retained and the others were discarded. The final results are shown in Tables 1 and 2 respectively.

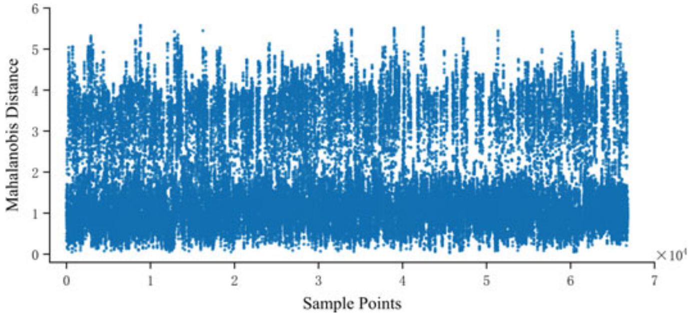


Fig. 3 Distribution of scraper fluctuation Mahalanobis distances

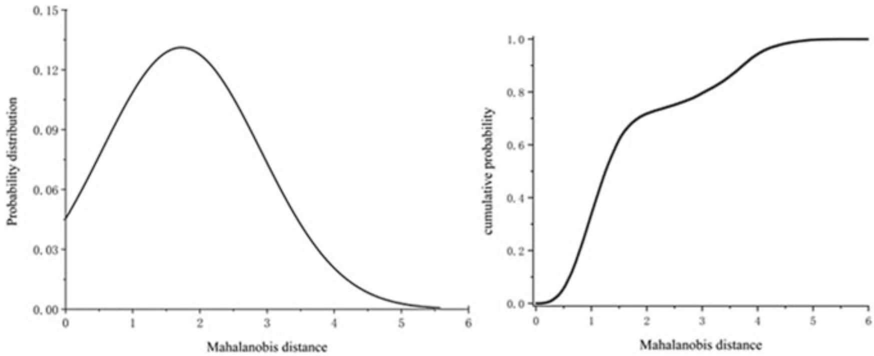


Fig. 4 Probability distribution and cumulative probability distribution curve of all sample points of the Mahalanobis distance

As can be seen from the two tables above, the results from the two different analysis methods are similar through different feature importance analysis methods, with the same type and ranking of the top 5 features in terms of relevance or importance. This also illustrates the reliability of the two analysis methods. The features closely related to the occurrence of paper break fault are pressure fluctuation of scraper and variation of sizing flow at the stocking, and the two subsets of features were established respectively. The distribution of the Mahalanobis distance at the scraper is shown as follows.

As shown in Fig. 3, most of the points are at a lower level, indicating that the Mahalanobis distance from the normal data is small and these points are likely to be normal. The distribution of sample points along the longitudinal coordinates is gradually sparse as the number goes up, which means that the Mahalanobis distance is larger, indicating that the points at the top of the graph are possible abnormal points. In order to describe the state of the process more accurately, a threshold of abnormal values needs to be set. As shown in Fig. 4, The plot on the left shows the distribution curve of the estimated Gaussian kernel density for the Mahalanobis distance, which can be seen as mainly concentrated at smaller coordinates. The actual

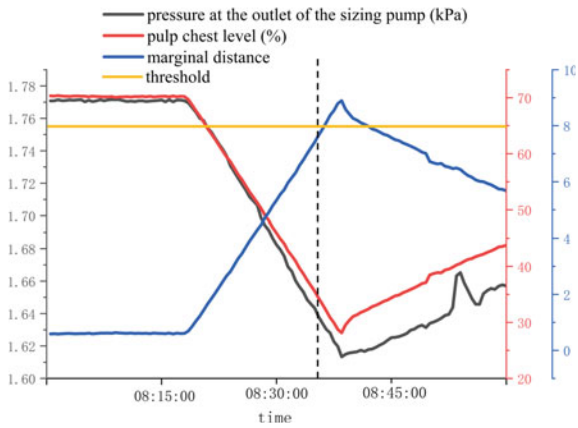


Fig. 5 The relevant variables and MD1 before the failure 1

distribution of the Mahalanobis distance can be seen from the cumulative probability distribution plotted against the growth curve. In general, a confidence interval of 95% is considered to be sensible. According to the plot, the value obtained corresponding to the Mahalanobis distance is set as the threshold, and the value obtained was used as an alert criterion for the early warning of the process.

The health benchmark model was validated with historical data, and the two models were evaluated for two different causes of paper break fault. The papermaking process broke at 09:00 one day and the key parameters were extracted 1 h prior to the breakdown, and it was found that there were no significant changes and fluctuations in the data at the scraper, while the pressure and pulp chest level at the stock changed obviously, as shown in Fig. 5.

It can be seen that the level of the pulp chest and the pressure at the outlet of the sizing pump were normal before 08:20, after which the two characteristic states dropped rapidly, though the level gradually rebound at around 08:40. The whole process was already potentially hazardous, resulting in a failure at 09:00. At this time, Result of the health assessment analysis indicates that its distance metric exceeded the set normal threshold at approximately 08:35, achieving an early warning. The warning time was 25 min earlier than the actual time of failure. A good fault prediction result was achieved.

In another case, the papermaking process broke down at 07:01 one day. Extract the data of the relevant parameters before the fault, and it can be found that, before the failure, the pressure at the outlet of the upper stock pump, and the liquid level of the copying pool did not change significantly, the pressure fluctuation of the scraper was larger than normal. The process is shown in Fig. 6.

The small abnormal fluctuation in the scraper can be seen at around 6:15, and around 6:56. the fluctuation value of scrapers rises rapidly until the Mahalanobis distance from the health benchmark exceeds the alarm threshold, five minutes earlier than when the fault occurred, which effectively provides valuable time to deal with the problem.

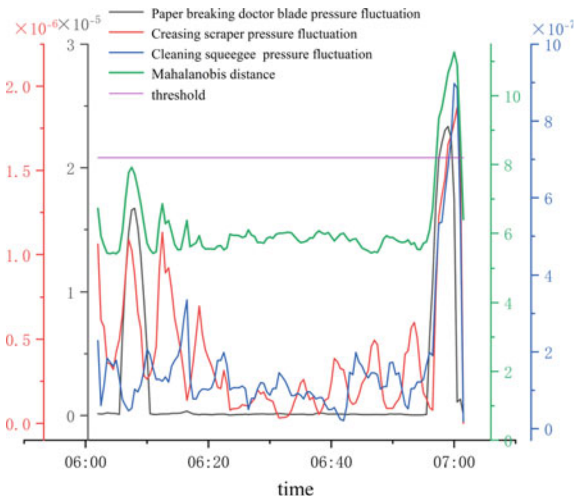


Fig. 6 The relevant variables and MD2 before the failure 2

The health benchmark built from different causes can predict the occurrence of faults more comprehensively, meanwhile identifying the causes of faults more precisely. The two cases above predicted the occurrence of paper break faults from different perspectives, achieving better results and demonstrating the feasibility and validity of the model.

4 Conclusion

This work focuses on the papermaking process and proposes a data-based approach to health assessment and warning of faults. The major contributions in this Chapter can be concluded as:

For fault prediction, a model consisting of 4 steps is proposed. First, the key parameters of the process were extracted by correlation analysis methods and feature engineering. Secondly, the times series characteristics were extracted by sliding window methods, As follows, the feature variables were clustered separately using clustering algorithms. From which the few categories are determined.

The health benchmark was used as a judgment criterion to calculate the deviation of the actual process data from the normal data set. when the deviation exceeds the threshold, alarm signals are sent, so that alarms are raised so that staff in charge of the papermaking machine can be alerted to changes in the relevant modules to avoid faults as earlier as possible. The model was validated with the real-time production data of the paper mill, and it is shown that the model has a fault prediction accuracy of 76.8% and a recall rate of 72.5%, the result shows that the proposed method can accurately monitor the production process and detect faults earlier than existing systems.

References

1. Karikallio H, Mäki-Fränti P, Suhonen N (2011) Competition in the global pulp and paper industries—An evaluation based on three approaches. J Econ 17(1):91–104
2. Li J, Qiu W, Yang L (2021) Comparison and analysis of important product standards and technical indicator systems in the papermaking field-taking facial tissue as an example. IOP Conf Ser Earth Environ Sci 728(1):12012
3. Rahman M, Dahlquist E, Kyprianidis K (2022) Modelling and diagnostics of process faults in continuous pulp digesters. Comput Chem Eng 157:107589
4. Lee W, Seo K (2021) Early failure detection of paper manufacturing machinery using nearest neighbor-based feature extraction. Eng Rep 3(2)
5. Lei Y, Yang B, Jiang X et al (2020) Applications of machine learning to machine fault diagnosis: a review and roadmap. Mech Syst Signal Process 138:106587
6. Lei Y, Li N, Guo L et al (2018) Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech Syst Signal Process 104:799–834
7. Strano S, Terzo M (2019) Review on model-based methods for on-board condition monitoring in railway vehicle dynamics. Adv Mech Eng 11(2):2072151767
8. Zhou Q, Yan P, Liu H et al (2018) Research on a configurable method for fault diagnosis knowledge of machine tools and its application. Int J Adv Manuf Technol 95(1–4):937–960
9. Jeong H, Park B, Park S et al (2019) Fault detection and identification method using observer-based residuals. Reliab Eng Syst Saf 184:27–40
10. Zhang J, Jiang N, Li H et al (2019) Online health assessment of wind turbine based on operational condition recognition. Trans Inst Meas Control 41(10):2970–2981
11. Ruiming F, Minling W, Xinhua G et al (2020) Identifying early defects of wind turbine based on SCADA data and dynamical network marker. Renew Energy 154:625–635
12. Zhao J, Zhang X, Li H et al (2017) Health indicator selection and health assessment of rolling element bearing. IEEE
13. Peng K, Guo Y (2022) Fault detection and quantitative assessment method for process industry based on feature fusion. Measurement 197:111267
14. Dangut MD, Jennions IK, King S et al (2022) Application of deep reinforcement learning for extremely rare failure prediction in aircraft maintenance. Mech Syst Signal Process 171:108873
15. Holguín-Londoño M, Cardona-Morales O, Sierra-Alonso EF et al (2016) Machine fault detection based on filter bank similarity features using acoustic and vibration analysis. Math Probl Eng 2016:1–14
16. Nacereddine N, Goumeidane AB, Ziou D (2019) Unsupervised weld defect classification in radiographic images using multivariate generalized Gaussian mixture model with exact computation of mean and shape parameters. Comput Ind 108:132–149
17. Liu J, Pan C, Lei F et al (2021) Fault prediction of bearings based on LSTM and statistical process analysis. Reliab Eng Syst Saf 214:107646
18. Kuyuk HS, Yildirim E, Dogan E et al (2014) Clustering seismic activities using linear and nonlinear discriminant analysis. J Earth Sci 25(1):140–145
19. Weber CM, Ray D, Valverde AA et al (2022) Gaussian mixture model clustering algorithms for the analysis of high-precision mass measurements. Nucl Instrum Methods Phys Res Sect A: Accel Spectrom Detect Assoc Equip 1027:166299

Multi-objective Optimization of Flexible Flow-Shop Intelligent Scheduling Based on a Hybrid Intelligent Algorithm

Huanhuan Zhang, Zhenglei He, Yi Man, Jigeng Li, Mengna Hong, and Kim Phuc Tran

Abstract With the complexity of the production process, the mass quantification of production jobs, and the diversification of production scenarios, research on scheduling problems are bound to develop in a direction closer to the actual production problems. Considering the combination of workshop scheduling problems and process planning problems, the study of such problems is of great significance for improving the production efficiency of enterprises. Therefore, this chapter studies the intelligent scheduling problem of a flexible flow-shop and establishes a two-stage flexible flow-shop scheduling model. On this basis, the fast non-dominated sorting genetic algorithm II (NSGA-II) and the variable neighborhood search algorithm (VNS) are combined to optimize the established two-stage intelligent scheduling model. Finally, a papermaking production process is taken as an example to comprehensively evaluate the performance of the model and the hybrid intelligent algorithm. The experimental results show that the model and algorithm can effectively solve the presented problem.

Keywords Multi-objective optimization · Flexible flow-shop · Intelligent scheduling · Hybrid intelligent algorithm

H. Zhang · Z. He (✉) · Y. Man · J. Li · M. Hong
State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou 510640, China
e-mail: hezhenglei@scut.edu.cn

K. P. Tran
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_6

1 Introduction

Scheduling problems widely exist in various fields, such as enterprise management, production management, transportation, aerospace, medical and health, network communication, etc. Scheduling performs a vital function in almost all branches of engineering science. It is also one of the key cores in the field of intelligent manufacturing so the research on scheduling problems is very blooming.

The scheduling objects and objectives determine the complex characteristics of the problem, which are prominently manifested in the diversity of scheduling objectives, the uncertainty of the scheduling environment, and the complexity of the process of solving the problem. The specific performance of complex characteristics is as follows:

1. *Multi-objective*. The overall goal of production scheduling is generally composed of a series of scheduling constraints and evaluation indexes. In different types of manufacturing enterprises and different manufacturing environments, the objectives of production schedules are often varied and varied. This largely determines the diversity of production scheduling objectives. For scheduling plan evaluation indicators, the shortest production cycle is usually considered the most. Other than this, objectives may also include the lowest energy consumption, the shortest delay, and the lowest inventory. In actual production, sometimes, not only a certain requirement is considered, it must be comprehensively considered in the process of scheduling plan formulation when various requirements may conflict with each other.
2. *Complexity*. As known, the classical scheduling problem is the extremely complex result of combinatorial optimization [1]. The complex nature of scheduling problems restricts the application and development of related techniques, impeding efforts to find effective methods to fully meet the needs of practical applications in recent years. Whereas, the emergence of intelligent scheduling technology provides new inspirations for tackling scheduling problems [2, 3]. Over the years, research on this problem has attracted a large number of researchers from different fields. They put forward several methods and techniques and made their own contributions to the solution of production scheduling problems to different degrees.

Production scheduling is a very critical sector of the production management system. The research on scheduling problems to improve production efficiency mainly focuses on three workshops, which are a single machine scheduling environment, a flow shop scheduling environment, and a job shop scheduling environment. From the perspective of single-machine scheduling, a lot of research focuses on optimizing the maximum completion time, Inventory minimization, and Maximizing equipment utilization. Some studies consider the actual production and operation of enterprises with cost as the optimization goal. For instance, Feristah et al. [4] studied the impact of customer-related delay costs on optimal machine scheduling under uncertain conditions, and have proposed a harmony search algorithm in the comparison of performance with the exact solution method. The results show that cost savings are not significantly dependent on customer delay costs, but mainly on task volume and lead time flexibility. There are also scholars who have achieved the purpose of reducing energy consumption by changing the speed of machines. Furthermore, Fan et al. [5] studied the single-machine scheduling problem with fixed and variable processing speeds to minimize the total power cost. Mohammad [6] first introduced capable batch processing and picking systems into an integrated production scheduling and distribution job problem, and also proposed a comprehensive integrated production process scheduling model.

From the perspective of flow-shop scheduling, Luo et al. [7] proposed a hybrid flow-shop scheduling problem that considers both production and energy efficiency, with the goal of minimizing the maximum make-time and power costs. And a meta-heuristic algorithm of ant colony optimization is proposed to solve this problem. Miyata et al. [8] proposed a method for assigning PM activities by job sequence for the no-waiting flow-shop scheduling problem. Hasani et al. [9] proposed a scheduling problem with machine-dependent processing stages, using a non-dominated sorting genetic algorithm (NSGA-II) based on a multi-objective problem with the lowest production cost and lowest energy consumption to solve. A flexible flow shop is one of the most extensive manufacturing systems, and the flexible flow-shop scheduling problem (FFSP) exists widely in chemical, metallurgy, textile, logistics, construction, papermaking, and other fields. In order to solve the FFSP problem, researchers have done a lot of research on the solution of FFSP. Tran et al. [10] proposed a water flow algorithm to solve the FFSP with and without waiting for processing workpieces. Ding et al. [11] proposed a new hybrid particle swarm optimization (HPSO) algorithm for solving the flexible shop floor scheduling problem, and the results showed the importance of HPSO in terms of the quantity and quality of non-dominated solutions and computational efficiency. From the perspective of job shop scheduling, Zhang [12] proposed an adaptive NSGA-II algorithm that can change the probability of crossover and mutation operations at different stages of the genetic process to minimize the manufacturing cycle, production cost, and equipment load. Xu et al. [13] studied the energy-saving scheduling problem of order-based manufacturing enterprise workshops under time-of-use electricity prices, and established a mathematical model by introducing switching strategies during equipment idle time to reduce electricity costs in production and assembly.

In general, on the one hand, the research on production scheduling with the optimization goal of energy consumption cost is the most extensive, and few researchers specifically discuss the scheduling that minimizes the number of product switching and the number of production job delays in flexible flow workshops. In actual production, a single objective has been unable to meet the needs of enterprises for optimization to obtain the optimal solution. The traditional multi-objective processing method is to transform the multi-objective problem into a single-objective problem by means of a weighted sum. The advantage of this method is that it is easy to operate, but the disadvantage is that only one solution can be obtained at a time. The multi-objective Pareto method considers multiple objectives simultaneously in the solution process. Generate a set of Pareto optimal solutions, reduce the calculation time and increase the diversity of understanding. In the algorithm, how to satisfy various constraints and effectively find the optimal feasible solution for multiple optimization objectives is a problem that must be considered in the study of flexible flow-shop scheduling. In view of this, this chapter studies the intelligent production scheduling problem of flexible flow workshop establishes the production scheduling mathematical model of flexible flow workshop and constructs a hybrid intelligent optimization algorithm NSGA-II_VNS to solve the intelligent scheduling problem. Finally, taking the chapter industry as a research case, the feasibility and effectiveness of the production scheduling model of the flexible flow shop and the solution performance of the proposed hybrid algorithm in solving the production scheduling problem of the flexible flow shop are verified.

This chapter is organized as follows, and Sect. 2 reviews the methods for solving scheduling problems. The third chapter describes and models the production scheduling problem of flexible flow-shop, and then introduces the NSGA-II algorithm and the hybrid algorithm NSGA-II_VNS. In the fourth chapter, the validity of the NSGA-II_VNS algorithm and the feasibility of this research is proved through experiments. Section 5 is the conclusion of this chapter.

2 Literature Review

The key technologies of scheduling approaches can be roughly divided into three categories, including mathematical programming methods and solvers, heuristic methods, and intelligent optimization methods. Mathematical programming method is an earlier method used to solve shop scheduling [14–16]. Mixed integer programming methods [17, 18], Lagrangian relaxation methods [19, 20] and decomposition methods [21] are several widely used mathematical programming methods. Mixed integer programming methods restrict some decision variables to be integers. In addition, the number of integer variables in the operation of mixed integer programming methods grows exponentially with the size of the problem. The Lagrangian relaxation method uses non-negative Lagrangian multipliers to relax process constraints and resource constraints and finally adds a penalty function to the objective function. It can provide better solutions to complex planning problems in a feasible time. This method has been used to solve job shop scheduling problems. The decomposition method decomposes the original problem into several small easy-to-solve sub-problems and finds the optimal sub-problem. This method has also been used to solve the scheduling problem. The established mathematical programming model can also be solved by a solver. Solvers are a class of packaged optimization algorithm packages that researchers can use to optimize complex problems such as scheduling without having to write algorithm codes themselves [22, 23]. Commonly used solvers include Cplex, Gurobi, MOSEK, etc. Often, different solvers have their own separate mathematical languages for writing established mathematical models. The disadvantage of the solver is that different optimization algorithms need to be called for models with different attributes, and the same mathematical model needs to be written multiple times in different modeling languages, which undoubtedly increases the workload of solving the problem. Moreover, the solving efficiency of the mathematical model is low, and it can only solve small-scale problems. The vast majority of scheduling problems have been shown to be NP-Complete problems. As the size of the problem increases, researchers no longer pursue exact algorithms to find the optimal solution to the problem. Instead, an approximation algorithm seeks a satisfactory solution to the problem in an acceptable time, so heuristics are used to solve the problem. Prioritization rules [24] and bottleneck-based heuristics [25] are typical heuristics. The priority assignment rule has the characteristics of easy implementation and small time complexity and is a common method to solve scheduling problems in practical applications. Priority allocation rules, while very fast, are short-sighted by nature. For example, it only considers the current state of the machine and the quality level of the solution, but cannot comprehensively consider the problem. Commonly used rules include shortest operation time, longest operation time, longest remaining processing time, shortest remaining processing time, and longest operation remaining processing time. Bottleneck-based heuristics generally include bottleneck moving methods and beam search [26]. Although the heuristic method can provide better quality solutions than the priority allocation rule method, the calculation time is long and the algorithm implementation is more complicated.

Intelligent optimization methods are a class of random search algorithms inspired by biological intelligence or physical phenomena. The advantage is that such algorithms generally do not require continuity and convexity of the objective function and constraints, and sometimes even analytical expressions are not required. The intelligent optimization method also has strong adaptability to the uncertainty of the data in the calculation. Due to these unique advantages and mechanisms, intelligent optimization methods have attracted the attention of many scholars at home and abroad, and have been widely used in many fields [27]. Intelligent optimization methods in the field of scheduling mainly include evolutionary algorithms [28], swarm intelligence optimization algorithms [29, 30], local search algorithms [31], and artificial intelligence algorithms [32]. An evolutionary algorithm is a kind of intelligent optimization method that simulates the biological evolution process and is widely used in combinatorial optimization problems such as planning and scheduling. It mainly includes genetic algorithms, genetic programming, evolution strategies, and evolution programming. Among them, the genetic algorithm is the most widely used evolutionary algorithm in the field of scheduling. The swarm intelligence optimization algorithm is mainly a kind of intelligent optimization method constructed by simulating the group behavior of insects, birds, and fish. In the field of scheduling, common swarm intelligence optimization algorithms include particle swarm optimization, ant colony optimization, etc. Local search algorithms are often used in conjunction with evolutionary algorithms to overcome the shortcoming that evolutionary algorithms are easily trapped in local optimum [33]. At present, the core functions of artificial intelligence-related algorithms are mainly used for classification and prediction. Its advantages are large storage space, strong storage capacity, strong self-learning ability, good fault tolerance, and easy classification. The essence of the scheduling problem is the optimization process, and the number of feasible solutions is large, and it is difficult to solve it directly with artificial intelligence.

Therefore, in the scheduling problem, the steps of classification or prediction must be found and replaced by artificial intelligence algorithms to assist in solving them. In addition, artificial intelligence-related algorithms also have problems such as low learning efficiency, slow learning speed, and difficulty in knowledge expression.

On the one hand, with the complication of production processes, the customization of products, and the diversification of production scenarios, the research on scheduling problems will inevitably develop in a direction that is closer to the actual production problems. On the other hand, with the development of smart workshops, the connection between workshop scheduling problems and other production problems is gradually strengthening. For example, the combination of shop scheduling problems and process planning problems will inevitably form a more complex coupling problem. How to use intelligent optimization algorithms to effectively solve scheduling problems has always been a challenging problem and research point.

3 Production Scheduling Model of Flexible Flow Workshop

3.1 Problem Description and Modeling

As an advanced form of manufacturing process organization, a flexible flow workshop (FFS) can simultaneously deal with various requirements such as multi-variety small batch production, large-scale customer customization, and single-piece production. The flexible flow production method can not only smooth the logistics and transportation in the workshop but also provide the possibility for rhythmic and high-efficiency production. The existence of its parallelize improves the brittleness of the production system and provides the possibility for the simultaneous production of diversified and differentiated products. Under the current situation, flexible flow production has become the preferred way for continuous or discrete manufacturing companies to deal with multi-variety, large-scale, and small-batch production and customer customization. It is widely used in automobile manufacturing, semiconductor manufacturing, electronic manufacturing, steel metallurgy and chemical production, and other fields. The FFS problem studied in this chapter is described as: there are N independent jobs that dynamically arrive at the workshop, and the production jobs can be completed in turn through M processing stages according to the same process route. For each job, only one piece of equipment for the operation can be selected for processing in each processing stage. Different types of jobs have different processing times in the same processing stage. Considering the fluctuation of processing time caused by the old and new state of equipment, the processing time of the same type of task may also be different on different equipment in the same stage. The flexible flow-shop scheduling problem is a combination of the flow-shop scheduling problem and the parallel machine scheduling problem. The flow-shop scheduling problem is essentially a series of multiple single-machine scheduling problems. The core of stand-alone scheduling is job sequencing. The core of parallel machine scheduling is device selection. Drawing on the idea of problem decomposition in the literature, this chapter decomposes the original problem into the following three sub-problems through structured analysis:

1. The problem of job selection among its available devices. Due to the difference in machining accuracy, old and new status, and performance between equipment, the same job will have different processing times on different equipment. At the same time, the number of waiting jobs is different for different equipment. Therefore, the target device for the decision artifact needs to be determined.
2. The problem of sorting jobs waiting to be processed on the equipment. Attributes such as priority and delivery time of different jobs are different. How select the next workpiece to be processed from the set of waiting workpieces is particularly important for the overall performance of the production system.
3. Jobs assigned to the same machine are ordered to minimize some optimization objective [34].

Based on the above problem description and structural decomposition, this section designs decision variables from the solution of sub-problems. And give the relevant parameter definitions to build a mathematical model. A real-time scheduling scheme in a dynamic environment is generated by making joint decisions on three sub-problems. The established flexible flow-shop scheduling model is as follows. The selection of the decision variable device is shown in formula 1:

$$O_{(i,j,t)} = \begin{cases} 1 \text{ The job } i \text{ starts processing at time } t \\ 0 \text{ not in} \end{cases} \tag{1}$$

The choice of the decision variable starting time is shown in formula ??: The parameters in this section are defined as shown in Table 1.

The corresponding relationship between production jobs and product types is shown in formula 2.

$$[PT]_{(i,k)} = \begin{cases} 1 & \text{job } i \text{ is of type } k \\ 0 & \text{Does not belong} \end{cases} \tag{2}$$

According to the characteristics of the production problem, the job constraints that the production scheduling problem model established in this section should satisfy are as follows.

$$\sum D_{(i,m,j)} = 1, i = 1, ..., I; m = 1, 2, ..., M \tag{3}$$

$$\sum O_{(i,j,t)} = 1, i = 1, ..., I; j = 1, 2, .., J \tag{4}$$

$$\sum PT(i,k) = 1, i = 1, ..., I \tag{5}$$

$$E(i,j) = S(i,j) + p(i,j), i = 1, 2, ..., I; j = 1, 2, ..., J \tag{6}$$

Table 1 Parameter definition

Parameter	Meaning of parameters
i	Represents the ith production job, $i = 1, ..., I$
j	Represents the jth device, $j = 1, ..., J$
I	Total number of production jobs
J	Total number of devices
t	Represents time, $t = 1, ..., T$
l	Indicates the lth processing position
L	Indicates the total number of machining locations
k	Type of job, $k = 1, ..., K$
m	Indicates the mth processing stage of the job, $m = 1, ..., M$
M	Total number of processing steps
j_m	Indicates that there are j parallel machines in the mth process
$S(i, j)$	Start processing time of job i on the jth device
$O(i, j, t)$	On behalf of the job i can be processed on the jth equipment
$P(j, i, m)$	Represents the processing time of job i on equipment j at stage m
$D(i, j, m)$	The mth process of job i is assigned to the jth device
$PT(i, k)$	The product type of job i is k
$E(j, i, m)$	Represents the completion time of job i on equipment j at stage m
$(ST)_j$	Indicates the number of delays on the jth device

$$E(i, m) \leq S(i, m + 1), i = 1, 2, ..., I; m = 1, 2, .., M - 1 \tag{7}$$

$$\sum E(i, j)D(i, l) \leq S(i, j), j = 1, 2, .., J; l = 1, 2, ..., n \tag{8}$$

$$\sum S(i, j)D(i, j, m) \leq E(i, j)D(i, j, m), j = 1, 2, .., J; m = 1, 2, ..., m_j \tag{9}$$

Constraint 3 means that workpiece i can only be machined once at each stage m. Constraint 4 means that a piece of equipment j can only process one workpiece i at the same time. Constraint 5 means that workpiece i can only belong to one product type k at the same time. Constraint 6 indicates that the completion time of job i on the jth device is equal to the sum of the start time on the jth device of job i and the processing time on the jth device of job i. Constraint 7 means that the end time of the mth process of job i is not greater than the start time of the $(m + 1)$th process of job i; constraint 8 means that the start time of the $(i+1)$th process is not less than the ith process the end time of the job; Constraint 9 means that on the kth device, the start time of job i is less than or equal to the end time of job i. In this chapter, the optimization objective is to minimize the maximum completion time and the number of delays, which can be expressed as:

$$f_1 = \min E(j, i, M) \tag{10}$$

$$f_2 = \min(\sum ST(j, M)) \tag{11}$$

3.2 Method Introduction

In order to improve the efficiency of solving problems and solve large-scale problems, scholars use intelligent optimization algorithms to solve production scheduling problems. The fast non-dominated sorting genetic algorithm II (NSGA-II) was proposed by Srinivas and Deb in 2000 on the basis of NSGA [35]. It has good convergence and distribution. NSGA-II is widely used in many fields and has good performance. The basic principles of the traditional NSGA-II algorithm are:

1. Divide the population into several layers according to the dominance relationship between individuals.
2. Set the set of non-dominated individuals of the evolutionary population as the first layer.
3. Set the set of non-dominated individuals obtained from the population after removing the individuals in the first layer as the second layer.
4. Set the set of non-dominated individuals obtained in the population after removing the individuals in the first and second layers as the third layer.
5. And so on, and then calculate the crowding distance of each individual.
6. The selection operation of the NSGA-II algorithm is: to compare the order of the individual layers and the crowding distance and select the optimal individual to form a new parent population.
7. Finally, compare the new parent population with the child population.

Generation populations are merged, basic genetic operations are performed, and the operations are stopped after meeting the size requirements of the newly evolved population [35]. The time complexity of NSGA-II is MN^2, where M is the number of targets and N is the number of individuals. NSGA-II mainly spends time in three aspects: non-dominated sorting (constructing non-dominated sets), computing aggregation distances, and constructing partially ordered sets. The traditional multi-objective optimization algorithm will be unsatisfactory in the number of solution sets and the value of the objective function when solving the flexible workshop scheduling problem. Therefore, improving the traditional multi-objective optimization algorithm is of great significance for obtaining more solution sets and improving the fast

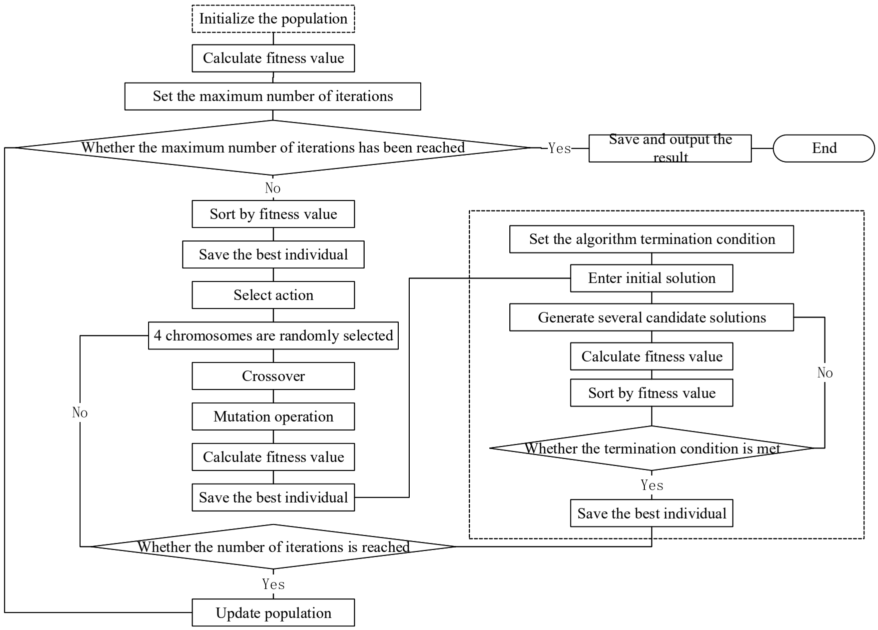


Fig. 1 Flow chart of NSGA-II_VNS hybrid algorithm

convergence of the algorithm. NSGA-II is a landmark algorithm in the field of multi-objective evolutionary optimization. However, considering that NSGA-II does not consider the situation when the crowding degree of individuals is the same and does not consider the impact of changes in the crowding degree of adjacent individuals after an individual is eliminated. Therefore, this chapter proposes a hybrid algorithm combining variable neighborhood search algorithm (VNS) and NSGA-II algorithm (referred to as NSGA-II_VNS algorithm). The flow chart of the NSGA-II_VNS hybrid algorithm proposed in this chapter is shown in Fig. 1. The idea of improvement is: after the mutation operation is completed, the optimal solution obtained by the crossover operation is used as the initial solution of the VNS algorithm; the VNS algorithm is used to conduct a deeper search, and then the optimal solution obtained by the VNS operation is used. The optimal solution is compared; a better solution is screened out and saved. Selection operation, this chapter adopts the roulette selection method [36]. In this method, the selection probability of each individual is proportional to its fitness value. The greater the fitness, the greater the probability of selection. In the crossover operation, the Alternating Position Crossover (APX) method is used in this chapter. In the mutation operation, this chapter uses Swap and adjacent swap (two-element swap), the core idea of which is to exchange two randomly selected elements.

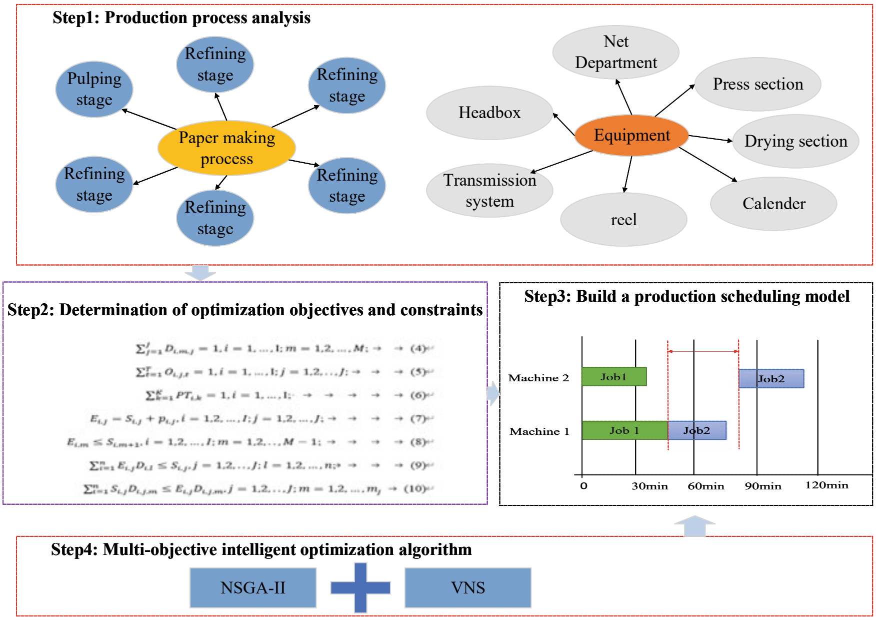


Fig. 2 The technical roadmap of the production scheduling problem

4 Case Study

4.1 Production Process of Papermaking Production Workshop

Figure 2 is the technical roadmap of this section, which can be divided into four steps. First, this section analyzes the production process of the papermaking workshop. Understand the equipment, production process, technology, process constraints, etc. And on this basis, to determine the type of production scheduling problems. Secondly, on the basis of the previous step, the optimization objectives and constraints of the production scheduling model are further determined. The optimization goal is often set as the optimal processing sequence so that the job can be completed as early as possible. For jobs with high order defaults, the optimization goal is often set to meet a large number of delivery dates. In other words, different requirements choose different optimization goals. The third step is to establish the production scheduling mathematical model of the papermaking workshop. The last step is to solve the established production scheduling model using an intelligent optimization algorithm.

The simplified production process of the papermaking workshop is shown in Fig. 3, which are pulping stage, refining stage, pressing stage, stage, rewinding stage, cutting stage, and packaging stage. Commonly used equipment includes pulper,

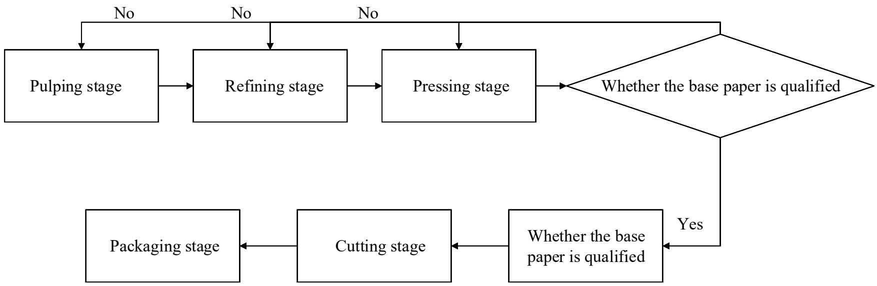


Fig. 3 The production process of the papermaking workshop

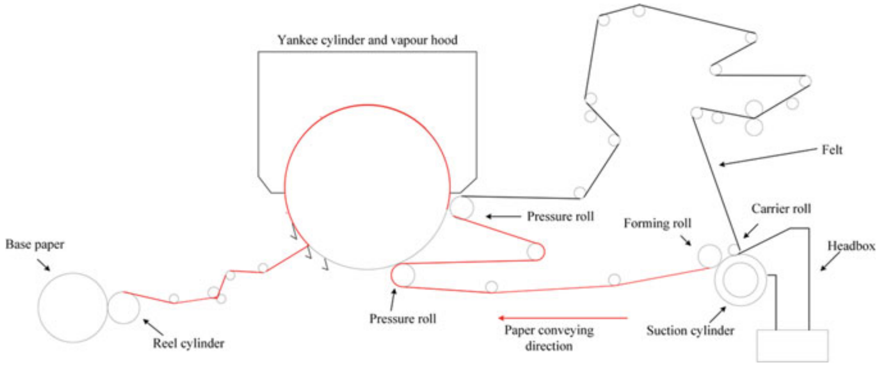


Fig. 4 A typical cylinder paper machine [37]

refiner, papermaking machine, rewinder, baler, and so on. As shown in Fig. 4, the cylinder paper machine mainly consists of five parts, namely the approach flow system, the forming sector, the press section, the dryer section, and the winding section. The pulping stage, refining stage, and pressing stage belong to continuous production, which can be simplified into one stage (referred to as the papermaking stage). The role of the papermaking stage is to turn the pulp into base paper. The rewinding stage, the cutting stage, and the packaging stage can be reduced to one stage (referred to as the packaging stage for short). Its function is to cut the base paper produced in the papermaking stage according to the pre-designed specifications and then package it to meet the requirements of sales.

According to the above analysis, the production scheduling problem of the papermaking workshop can be simplified as a two-stage flexible flow workshop production scheduling model. This section will take the two-stage flexible flow-shop production scheduling problem as the research object, and the established two-stage production scheduling model is shown below.

$$\sum \sum D(i, j), i = 1, 2, ..., n; m = 1, 2, ..., m_1 \tag{12}$$

$$\sum\sum D(i,j), i = 1, 2, ..., n; m = 1, 2, ..., m_2 \tag{13}$$

$$m_1 + m_2 = m \tag{14}$$

$$ST = \sum_{j=1}^{m}\sum_{i=2}^{N_j} st_{i-1,i}^{j} \tag{15}$$

$$f(QT, FT) = QT + FT \tag{16}$$

Constraints 12 and 13 indicate that each job must be processed once in the papermaking and packaging stages, respectively. Equation 14 represents the number of all equipment in the papermaking stage and the packaging machine stage. Equation 15 is the earliest start-up time of the equipment. It should be noted that after the papermaking stage, the base paper produced needs to be inspected. Only after passing the inspection can enter the packaging stage for subsequent processing. Therefore, there is a certain practical interval between the papermaking stage and the packaging stage. Constraint 16 represents the time interval between the papermaking stage and the packaging stage. In actual production, one base paper often corresponds to multiple products of different specifications. When the product type of the previous job and the product type of the next job on the same equipment is inconsistent, before entering the packaging stage and starting processing, the equipment parameters need to be adjusted to meet the production requirements.

4.2 Optimization Objectives

In recent years, the problem of workshop scheduling has also faced the dual pressures of the environment and the economy. It is extremely important to reduce energy consumption by improving the utilization rate of resources for the goal of energy conservation and emission reduction. With the widespread adoption of flexible manufacturing methods in the manufacturing industry, multiple machines can handle the same process; and the machine characteristics are different, increasing the complexity and difficulty of scheduling problems. Therefore, this chapter sets the maximum makespan as one of the optimization objectives. By rationally arranging production through production scheduling, the processing time of the papermaking workshop can be reduced, thereby reducing the processing energy consumption of the papermaking workshop.

When manufacturing products are produced and manufactured, the rationality of their scheduling is a very critical factor affecting the production efficiency of manufacturing. As the global manufacturing industry becomes more competitive, product demands are diversified and personalized. The scheduling problem of the manufacturing systems has been paid more and more attention. Shop scheduling is the core function of the shop manufacturing execution system and has an irreplaceable role in manufacturing. Reasonable scheduling can effectively improve core indicators such as construction period, inventory, delivery time, energy consumption, and cost. In the papermaking workshop, it takes ten minutes to dozens of minutes for production switching to adjust the equipment to adapt to production. In this process, not only a part of standby energy consumption will be wasted. It will also cause the earliest available time of the device to be longer, which may affect the completion time of the job, resulting in an increase in the number of job delays. In the production scheduling process, reducing production switching can reduce energy costs for enterprises, and may also shorten the number of delays in production jobs. Taking the number of delays as the optimization goal can also enable the production workshop to effectively process customer orders, meet customer requirements to the greatest extent, and improve the competitiveness of papermaking enterprises. Therefore, this chapter sets the total number of delays in production jobs as another optimization objective.

4.3 Experimental Data

This chapter takes a papermaking mill in Guangdong Province as the research object and takes the maximum completion time and the number of delays as the optimization goals. The sources of production scheduling data can be divided into two aspects. On the one hand, the used production line, process path, process data, and product type data are extracted from the system database of a papermaking mill in Guangdong Province. The names of the three production scheduling instances and the number of jobs are shown in Table 2. In the cases, job1 and job2, the size of the job is randomly generated according to the range of the actual job size of the papermaking mill and obeys the uniform distribution of [100000, 500000]. The data for Job3 comes from the No. 3, No. 5, and No. 9 production lines of the post-processing guard roll 2 workshop of a papermaking mill. To be more specific, Job3 selects 16 days from July 20, 2019, to August 19, 2019, during which it was powered on for 24 hours, with a total of 96 production jobs. Among them, 96 jobs contain 8 product types.

On the other hand, scheduling job data, production line speed data, and switching data are all derived from simulations based on factory data. Among them, the preprocessing of data is mainly to remove outliers, fill in missing values according to the set value, and match the data of daily production jobs according to the base paper

Table 2 Names and number of jobs of the three production scheduling instances

Instances	Number of jobs
Job1	40
Job2	60
Job3	96

Table 3 Number of production lines

Papermaking stage	Packaging stage
6 production lines	7 production lines

Table 4 The maximum speed of each production line in the papermaking stage

Stage1	Production line speed (m/min)
PL1	1000
PL2	950
PL3	1030
PL4	980
PL5	1020
PL6	1020

Table 5 Maximum speed of each line in the packaging stage

Stage2	Small packing machines (bag/min)
BL1	216
BL2	254
BL3	284
BL4	160
BL5	254
BL6	260
BL7	275

number. As shown in Table 3, the number of production lines in the papermaking stage is 6, and the number of production lines in the packaging stage is 7. The speed range of each production line is shown in Tables 4 and 5, where Stage1 represents the papermaking stage and Stage 2 represents the packaging stage. In the production scheduling case, the size of each case study is randomly generated according to the range of the actual job size of the tissue papermaking enterprise and obeys the uniform distribution of [100000, 500000].

4.4 Parameter Settings

In this chapter, NSGA-II_VNS is used to solve the proposed production scheduling problem with the optimization goal of minimizing the maximum completion time and the number of delays. In the NSGA-II algorithm, important parameters include population size, maximum number of iterations, crossover probability, and mutation probability. The parameters of the NSGA-II algorithm in this chapter are set as follows: the population size is 100, the maximum number of iterations is 100, the crossover probability is 0.9, the mutation probability is 0.1, and the number of objective functions is 2. The parameter settings of the VNS algorithm are: the maximum number of iterations is 100, the size of the neighborhood structure set 1 of VNS is 2; the size of the neighborhood structure set 2 of VNS is 6.

4.5 Results and Analysis

This section mainly solves the production scheduling problem of flexible flow-shop with the minimization of the maximum completion time and the number of switches. To this end, this study designed three groups of production scheduling examples. The first two groups of examples are solved by NSGA-II and NSGA-II_VNS respectively. The third group uses NSGA-II to solve the two-stage flexible flow-shop scheduling problem with minimal maximum completion time and minimum switching times. The performance of NSGA-II in solving the production scheduling problem was verified by comparing it with the results of manual scheduling. The results show that in Job3, the number of product switching in manual production is 8 times, while the number of product switching in the plan obtained by NSGA-II is 6 times, and the result of the intelligent optimization algorithm is 744 minutes shorter than that of manual production scheduling. This fully proves that NSGA-II is feasible and effective in solving the established production scheduling model. In order to further verify the performance of the proposed hybrid algorithm, this chapter uses the data of Job1 to verify NSGA-II and NSGA-II_VNS. Figure 5 shows the maximum makespan iteration graph. In the two graphs in Fig. 5, the horizontal axis represents the number of rank 1s in the Pareto solution set after the algorithm completes one iteration. The vertical axis represents the maximum completion time in hours. In Fig. 5, (1) represents the result obtained by using the NSGA-II algorithm, and (2) represents the use of NSGA-II_VNS. As can be seen from Fig. 5, the maximum make-span obtained by NSGA-II_VNS is shorter than that obtained by using NSGA-II. That is, compared with the production plan obtained by NSGA-II, the production efficiency obtained by using the production plan obtained by NSGA-II_VNS is higher. Pareto dominance thinking is a method for evaluating the pros and cons of multi-objective problem solutions. The Pareto optimal solution is also called the non-dominated solution, and the Pareto optimal solution set is also called the non-dominated solution set. It can also be found from Fig. 5 that in the solution set obtained by the two algorithms in each iteration, the Pareto optimal solution set in NSGA-II_VNS accounts for more. Therefore, the proposed NSGA-II_VNS hybrid algorithm can obtain better solutions. It further proves the feasibility and effectiveness of the proposed NSGA-II_VNS algorithm. Figure 6 shows the iteration graph of product switching times obtained by the two algorithms. The figure on the left in Fig. 6 is NSGA-II, and the figure on the right is NSGA-II_VNS. In Fig. 6, the horizontal axis represents the number of the Pareto solution set at the first level of an algorithm iteration, and the vertical axis represents the number of jobs switching in the production process. As

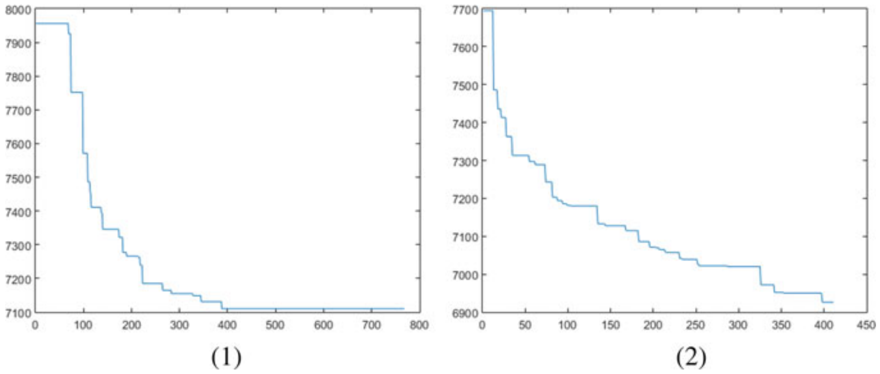


Fig. 5 The maximum make-span iteration graph obtained by (1)NSGA-II and (2) NSGA-II_VN in case Job1 respectively

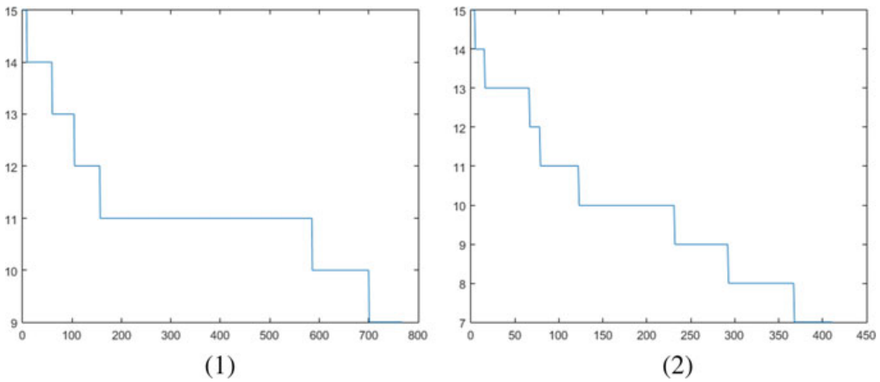


Fig. 6 Iteration diagram of product switching times obtained by (1) NSGA-II and (2) NSGA-II_VNS in case Job1 respectively

can be seen from Fig. 6, the product switching times obtained by NSGA-II_VNS is 7 times, while that obtained by NSGA-II is 9 times. Therefore, the obtained solution of NSGA-II_VNS is better than that of NSGA-II. It is further proved that the proposed hybrid algorithm has good performance for solving the established flexible flow-shop scheduling problem.

Figures 7 and 8 are the results obtained by solving the production scheduling job instances of the cases Job1 and Job2. The horizontal axis represents the maximum completion time, and the vertical axis represents the number of delays. Considering the randomness of the intelligent optimization algorithm, NSGA-II_VNS and NSGA-II are used to solve Job1 and Job2 10 times each. It can be clearly found from Figs. 7 and 8 that the solution sets obtained by NSGA-II_VNS are closer to the coordinate origin. That is to say, using NSGA-II_VNS to solve the established production scheduling problem, the maximum completion time is shorter and the

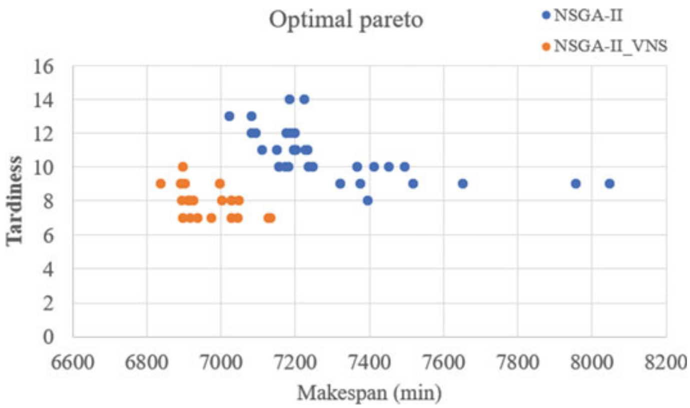


Fig. 7 Two-dimensional scatter plot of case Job1

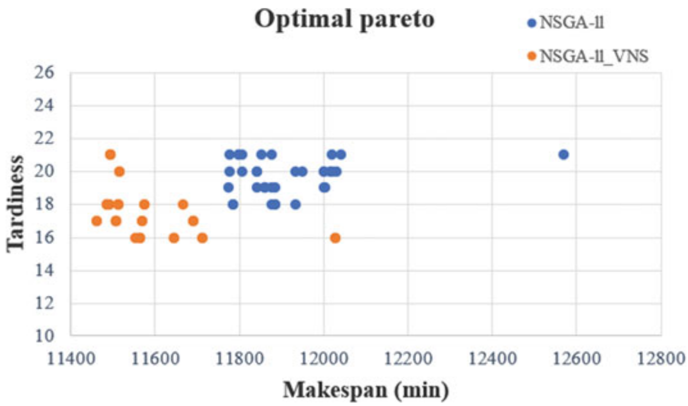


Fig. 8 Two-dimensional scatter plot of case Job2

number of delays is smaller. This further proves that the proposed NSGA-II_VNS algorithm has a stronger optimization ability. Figures 9 and 10 show the comparison of the optimal solutions obtained by the two algorithms, orange represents NSGA-II_VNS, and blue represents NSGA-II. Among them, Fig. 9 is a comparison chart of the number of tardiness, and Fig. 10 is a comparison chart of the makespan. It can also be seen from Figs. 9 and 10 that the makespan and the number of tardiness obtained by the NSGA-II_VNS hybrid algorithm are both smaller.

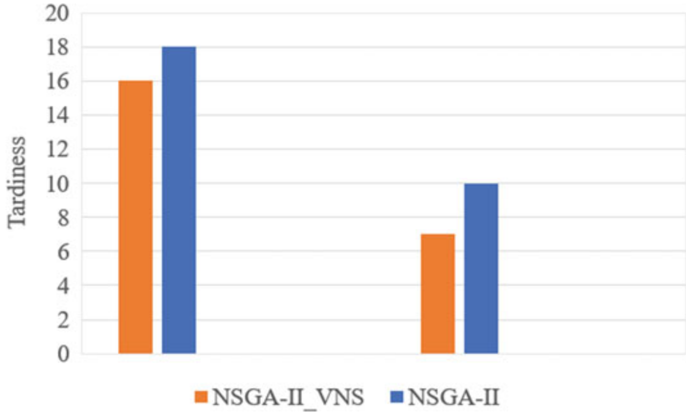


Fig. 9 Comparison chart of the number of tardiness

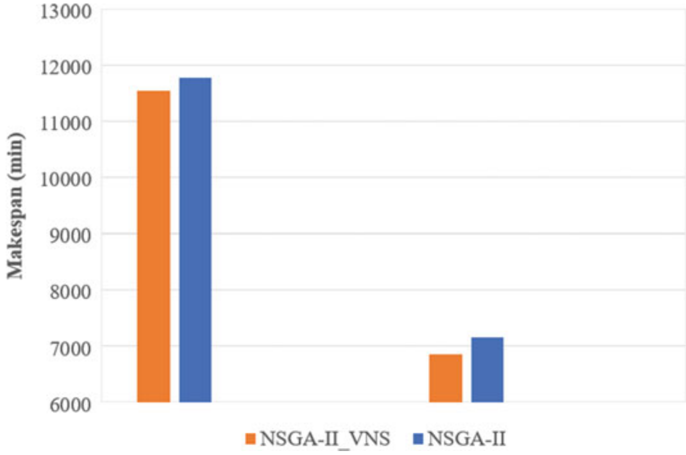


Fig. 10 Comparison chart of the number of makespan

5 Conclusions

This chapter studies the production scheduling problem based on the flexible flow shop and establishes the production scheduling model of the flexible flow shop. To solve the established scheduling model, a hybrid algorithm combining VNS and NSGA-II is proposed. Taking a papermaking mill as a research case, a two-stage flexible flow-shop production scheduling model is established based on the minimization of maximum completion time and the minimization of product switching times. This chapter also compares the NSGA-II algorithm with the manual scheduling scheme of a papermaking mill. Experiments show that, compared with the manual production scheduling scheme, the production scheduling scheme obtained by the NSGA algorithm is 744 minutes shorter than the manual production scheduling scheme, and has fewer product switching times. It is proved that the intelligent optimization algorithm is feasible and effective in solving production scheduling. In addition, NSGA-II and NSGA-II_VNS are used to solve the proposed production scheduling model respectively, and the results show that the NSGA-II_VNS algorithm has better optimization ability when solving the model.

References

1. Wen X, Lian X, Qian Y, Zhang Y, Wang H, Li H (2022) Dynamic scheduling method for integrated process planning and scheduling problem with machine fault. Robot Comput-Integr Manuf 77
2. Wang X, Zhang L, Lin T, Zhao C, Wang K, Chen Z (2022) Solving job scheduling problems in a resource preemption environment with multi-agent reinforcement learning. Robot Comput-Integr Manuf 77
3. Zhang G, Lu X, Liu X, Zhang L, Wei S, Zhang W (2022) An effective two-stage algorithm based on a convolutional neural network for the bi-objective flexible job shop scheduling problem with machine breakdown. Expert Syst Appl 203
4. Ozcelik F, Ertem M, Sarac T (2022) A stochastic approach for the single-machine scheduling problem to minimize total expected cost with client-dependent tardiness costs. Eng Optimiz 54(7)
5. Fan Z, Fang Y, Li W, Cai X, Wei C, Goodman E (2018) MOEA/D with angle-based constrained dominance principle for constrained multi-objective optimization problems. Appl Soft Comput J 74
6. Rostami M (2021) Minimizing maximum tardiness subject to collecting the EOL products in a single machine scheduling problem with capacitated batch delivery and pickup systems. Comput Ind Eng 161
7. Luo J, Fujimura S, El Baz D, Plazolles B (2019) GPU-based parallel genetic algorithm for solving an energy efficient dynamic flexible flow shop scheduling problem. CoRR. arXiv:abs/1903.10741
8. Miyata HH, Nagano MS, Gupta JN (2019) Integrating preventive maintenance activities to the no-wait flow shop scheduling problem with dependent-sequence setup times and makespan minimization. Comput Ind Eng 135
9. Hasani A, Hosseini SM (2020) A bi-objective flexible flow shop scheduling problem with machine-dependent processing stages: trade-off between production costs and energy consumption. Appl Math Comput 386
10. Tran TH, Ng KM (2011) A water-flow algorithm for flexible flow shop scheduling with intermediate buffers. J Sched 14(5)
11. Ding J, Schulz S, Shen L, Buscher U, Lü Z (2021) Energy-aware scheduling in flexible flow shops with hybrid particle swarm optimization. Comput Oper Res 125
12. Zhang L (2022) Multi-objective flexible job-shop scheduling problem researching based on improved adaptive NSGA-II. J Phys: Conf Ser 2235(1)
13. Xu E, Li Y, Liu Y, Du J, Gao X (2022) Energy saving scheduling strategy for job shop under TOU and tiered electricity price. Alex Eng J 61(1)
14. Wang G, Hijazi H (2018) Mathematical programming methods for microgrid design and operations: a survey on deterministic and stochastic approaches. Comput Opt Appl 71(2)
15. Cruz Neto JX, Ferreira OP, Lucambio Pérez LR, Németh SZ (2006) Convex- and monotone-transformable mathematical programming problems and a proximal-like point method. J Glob Optimiz 35(1)

16. Nishimoto K, Mori H, Kawaguchi S, Toyoda M, Tsukimoto K (2002) Development of filler metal for welding of nickel-base superalloy IN738LC by mathematical programming method. Weld World: J Int Inst Weld: J Int Inst Weld 46(9–10)
17. Andreotti S, Klau GW, Reinert K (2012) Antilope-A Lagrangian relaxation approach to the de novo peptide sequencing problem. IEEE/ACM Trans Comput Biol Bioinform (TCBB) 9(2)
18. Karakitsiou A, Migdalas A (2008) A decentralized coordination mechanism for integrated production-transportation-inventory problems in the supply chain using Lagrangian relaxation. Oper Res 8(3)
19. Gu H, Zhang Y, Zinder Y (2022) An efficient optimization procedure for the Workforce Scheduling and Routing Problem: Lagrangian relaxation and iterated local search. Comput Oper Res 144
20. Chatterjee S, Dimitrakopoulos R (2020) Production scheduling under uncertainty of an open-pit mine using Lagrangian relaxation and branch-and-cut algorithm. Int J Min Reclam Environ 34(5)
21. Steinrücke M (2015) Integrated production, distribution, and scheduling in the aluminum industry: a continuous-time MILP model and decomposition method. Int J Prod Res 53(19)
22. Sun Y (2019) Research on optimization scheduling simulation system of electric vehicle charging station based on CPLEX model technology. IOP Conf Ser: Mater Sci Eng 677(3)
23. Poongothai V, Godhandaraman P, Anitha K (2019) Unrelated parallel machine scheduling with multi-maintenance activities using CPLEX. AIP Conf Proc 2112(1)
24. Yuan X, Su A, Yuan Y, Nie H, Wang L (2009) Non-convex dynamic dispatch of generators with prohibited operating zones. Optimal Control Appl Methods 30(1)
25. Wang S, Zhao F, Liu Z, Hao H (2017) Heuristic method for automakers' technological strategy making towards fuel economy regulations based on genetic algorithm: a China's case under corporate average fuel consumption regulation. Appl Energy 204
26. Birgin EG, Ferreira JE, Ronconi DP (2020) A filtered beam search method for the m-machine permutation flowshop scheduling problem minimizing the earliness and tardiness penalties and the waiting time of the jobs. Comput Oper Res 114(C)
27. Liao H, Hu Y, Li Q, Yu Y, Huang S, Chen F, Cai Y, Xiang H (2022) An intelligent optimization method of reload core loading pattern and its application. Ann Nucl Energy 171
28. Łapa K, Cpałka K, Laskowski Ł, Cader A, Zeng Z (2020) Evolutionary algorithm with a configurable search mechanism. J Artif Intell Soft Comput Res 10(3)
29. Kumar A, Rathore PS, Diaz VG, Agrawal R (2020) Swarm intelligence optimization: algorithms and applications. Wiley
30. Li LS (2013) The swarm intelligence optimization algorithm and its application. Adv Mater Res 2393(711–711)
31. Engin BE, Engin O (2020) A new memetic global and local search algorithm for solving hybrid flow shop with multiprocessor task scheduling problem. SN Appl Sci 2(12)
32. Li S, Li B, Wang K, Guo W, Liu Y, Zhao R, Zeng J, Li H, Lin Y (2020) Cloud platform task scheduling strategy for power dispatching automation system based on artificial intelligence algorithm. J Phys Conf Ser 1570(1)
33. Zeng Z, Hong M, Man Y, Li J, Zhang Y, Liu H (2018) Multi-object optimization of flexible flow shop scheduling with the batch process—Consideration of total electricity consumption and material wastage. J Clean Prod 183
34. Hasani A, Hosseini SM (2020) A bi-objective flexible flow shop scheduling problem with machine-dependent processing stages: trade-off between production costs and energy consumption. Appl Math Comput 386
35. Deb K, Pratap A, Agarwal S, Meyarivan TA (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evolut Comput 6(2)
36. Qian W, Chai J, Xu Z, Zhang Z (2018) Differential evolution algorithm with multiple mutation strategies based on roulette wheel selection. Appl Intell 48(10)
37. Huanhuan Z, Li J, Hong M (2021) Machine learning-based energy system model for tissue paper machines. Processes 9(4):655

Personalized Pattern Recommendation System of Men's Shirts

Guillaume Tartare, Cheng Chi, and Pascal Bruniaux

Abstract Commercial garment recommendation systems have been generally used in the apparel industry. However, existing research on digital garment design has focused on the technical development of the virtual design process, with little knowledge of traditional designers. The fit of a garment has a significant role in whether a customer purchases that garment. In order to develop a well-fitting garment, designers and pattern makers should adjust the garment pattern several times until the customer is satisfied. Currently, there are three main drawbacks of traditional pattern-making: (1) it is very time-consuming and inefficient, (2) it relies too much on experienced designers, and (3) the relationship between the human body shape and the garment is not fully explored. In practice, the designer plays a key role in a successful design process. There is a need to integrate the designer's knowledge and experience into current garment CAD systems to provide a feasible human-centered, low-cost design solution quickly for each personalized requirement. Also, data-based services such as recommendation systems, body shape classification, 3D body modeling, and garment fit assessment should be integrated into the apparel CAD system to improve the efficiency of the design process. Based on the above issues, a fit-oriented garment pattern intelligent recommendation system is possible for supporting the design of personalized garment products. The system works in combination with a newly developed design process, i.e. body shape identification—design solution recommendation—3D virtual presentation and evaluation—design parameter adjustment. This process can be repeated until the user is satisfied. The proposed recommendation system has been validated by some successful practical design cases.

Keywords Human-centered · 3D body modelling · Designer's knowledge · Personalized pattern-making · Parametric design · Fit garment · Intelligent recommendation system

---

G. Tartare (✉) · C. Chi · P. Bruniaux
University of Lille, ENSAIT, GEMTEX-Laboratoire de Génie et Matériaux Textiles, Lille, France
e-mail: guillaume.tartare@ensait.fr

C. Chi
e-mail: cheng.chi@ensait.fr

P. Bruniaux
e-mail: pascal.bruniaux@ensait.fr

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_7

1 Introduction

With the rapid development of the economy and the appearance of industry 4.0, more and more consumers require the personalization of products. Apparel companies urgently need a new production model to respond to this requirement. Through this model, all activities in the complete product design and production process will focus more on the consumer's requirements.

Mass customization can satisfy the requirements of a specific population, provide adaptable products or services and realize the process of rapid customization at a low cost. However, the complexity of consumer requirements has led to increased difficulty in garment design. For example, due to rising consumption levels and changes in people's diets, the body sizes of different consumer groups are becoming increasingly different. This leads to the anthropometric data on which the ready-to-wear sizing system is based becoming outdated. Garments produced based on this system do not conform to current body sizes. More designers need to be employed for apparel companies, or the current design methods and processes need to be updated to solve this problem. However, employing more experienced designers would significantly increase costs. Therefore, changing the current design method and process is an effective way to solve the difficulties of modern apparel companies.

To match the design of products and systems to the target users, the body characteristics of the target users should be considered first. Ergonomics research on anthropometry and analysis of the shape of the human body is essential for the design of clothing products. In the apparel industry, body shape is closely associated with the manufacture and fit of garments. The starting point for making comfortable garments is to comprehensively understand and characterize the differences in the shape of the human body, and these change over time. We are no longer the women and men of past centuries.

Currently, most body shape research is based on data obtained from 3D human models. 3D body scanning technology enhances traditional anthropometric body shape descriptions by extracting information on cross-sectional areas and segmental volumes. Most 3D human models used in the apparel industry are rigid geometric surface models in the standing position and do not consider factors such as skeletal structure, body composition (bone, fat, and muscle), and body deformation associated with skin elasticity and body movements. If these factors are not integrated into 3D human models, the products produced on this basis may not be suitable for consumers. In addition, in a natural standing posture, it is difficult to obtain data on hidden areas (e.g., arm roots, snout, crotch) from a 3D scan. This also leads to the generated 3D body model not accurately reflecting the real human body shape. Therefore, it is important to propose a method for correcting 3D human models to solve the above problems. In addition, garment pattern-making is a key part of satisfying the personalization of garments. However, whether through traditional manual garment pattern-making or computer-aided design (CAD) garment pattern-making software, pattern designers need to have a great wealth of professional knowledge and proficient skills to quickly make a better-fitting garment pattern. Meanwhile, these methods also have the following disadvantages:

1. the learning process is particularly long and difficult to promote, and the making process is time-consuming, limiting the improvement of production efficiency. Although the CAD garment pattern-making software can automatically grade, which improves production efficiency, it is only suitable for the overall grading of the standard dimension specifications. It cannot automatically adjust the changes in individual dimensions. Garments produced based on standard dimension specifications cannot satisfy the requirements of consumers for fit;
2. once the garment style changes, the structure drawing of the garment needs to be manually redrawn or adjusted, and inexperienced designers are unable to respond quickly to adjustments in garment patterns. Therefore, it is very for the garment industry to know useful how to develop garment products without experienced designers and pattern makers.

Indeed, the clothing industry can help designers and pattern makers through decision-support systems. Virtual reality technology also plays an important role in the modern mass customization of clothing. Due to the realistic 3D environment and real-time interaction supported by the technology, it has had a profound impact on the garment manufacturing industry. It provides the basic technical conditions for the virtual design of new products and the associated manufacturing. In custom apparel design, designers can use virtual reality technology to assess the fit of apparel products.

In the current clothing recommendation system, most CAD work on clothing focuses on the technical development of virtual clothing assembly and its application to specific human body models. Few involve designer knowledge in the field of design. It is well known that the clothing designer plays a key role in the clothing customization process. The knowledge gained through the designer's experience should be used to develop virtual clothing patterns. When a client or young designer is sure which garment is most suitable for a specific wearer in terms of clothing style and body type, the designer's insights can provide personalized recommendations based on consumer requirements. Therefore, personalized knowledge-based recommendations can provide more effective assistance to a wide range of consumers and young designers in choosing relevant personalized apparel products or design solutions with the knowledge of an experienced designer. For this reason, apparel recommender systems should be well developed to enable the formalization and development of designers' knowledge for the rapid delivery of small quantities of custom apparel design productions and solutions. Clothing recommender systems play an important role in online shopping and CAD-based clothing co-design.

2 Morphological Analysis of The Human Body

The model of the human being carries a specific scale. In the 1st century BC, the Roman architect Marcus Vitruvius proposed a model called homo bene figurants, "the

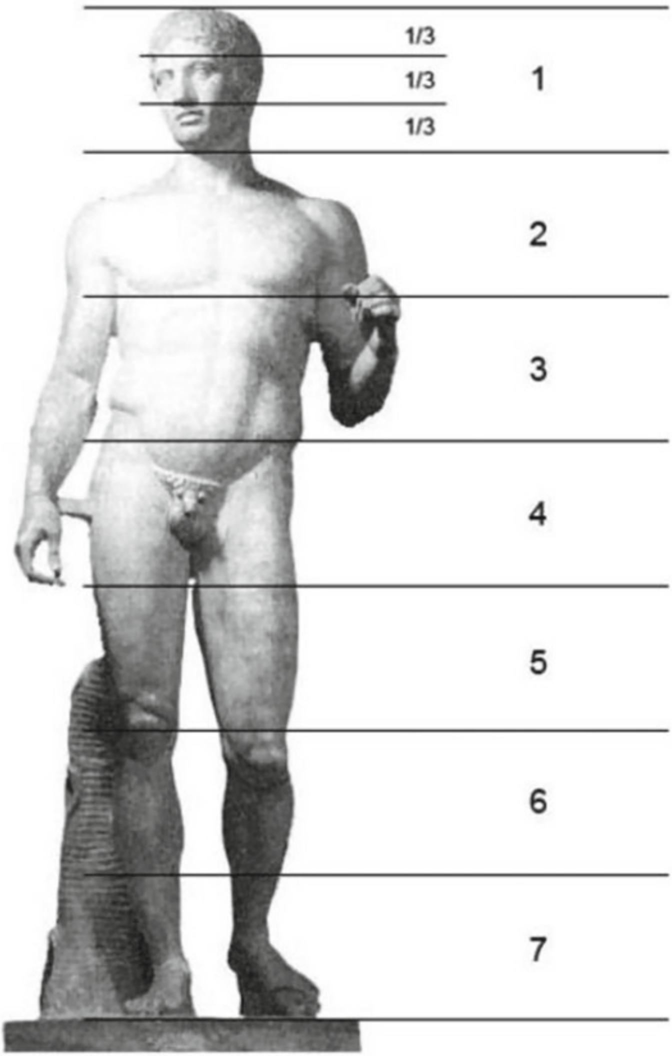


Fig. 1 Seven-head body model

well-trained man" Hersey [14]. The fifth-century Greek sculptor Polykleitos used the mathematical basis to segment the human body into seven equal parts Wetmore and Morgan [27], as in Fig. 1.

In the 15th century, Leonardo da Vinci refined the vague descriptions of Vitruvius in his famous "Vitruvian Man" model, constructing a system of eight head heights Wetmore and Morgan [27], and it is used to this day, as shows Fig. 2. The proportions illustrated by Leonardo's model have been widely accepted as the "golden propor-

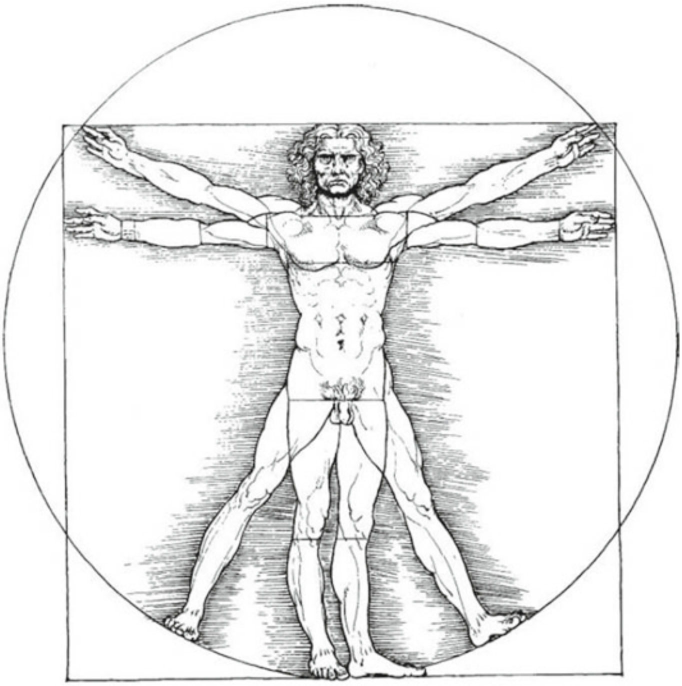


Fig. 2 Vitruviu's man by Leonardo da Vinci

tion" Levin [25]. However, in modern society, most people do not conform to this "golden proportion", and no one is even absolutely symmetrical. Human imperfection and its limitations are part of life and are acceptable. In addition, our society has evolved, (example: less handling has impacted higher stature), which makes these models inaccurate.

To quantify the knowledge about human measurements, anthropometry Sunderland [31] was created. The aim of current anthropometric research is to explore and understand the real human body and its classification to satisfy the needs of the market today. Chaffin and Andersson [5] proposes a most suitable definition of anthropometry. "Anthropometry is a science that deals with the measurement of size, weight, and proportions of the human body. It is empirical in nature and has developed quantitative methods to measure various physical dimensions". It is a knowledge dedicated to describing the quantification of the external geometry of the human body. As an application, the goal of anthropometry is to collect valuable information and make it usable for the designer's purposes. The garment industry is mainly concerned with anthropometric aspects of body dimensions and shapes Zakaria and Gupta [37], Leong and Tsai [16]. Body measurements are the basis of garment design Gupta [13].

The anthropometric data to be incorporated into the anthropometric database must be accurate and interpretable. It is also necessary to create well-fitting garments. Therefore, all steps in the development of anthropometrics need to minimize errors from the beginning.

2.1 Anthropometric Measuring Methods

Anthropometry makes it possible to digitize human body features. Only through anthropometry can specific data on the relevant parts of the human body so that there is a reliable reference for the dimensions of the relevant parts in the garment pattern design, thus ensuring the fit, comfort, and beauty of the garment. Three-dimensional anthropometry appeared in the mid-1980s. It is to measure the dimension of each part of the human body model and the body shape of the human body and then to study the human body's morphological characteristics and technical methods. It is used in human database construction, clothing customization, ergonomics, medicine, and museum display Chen [6]. Technical methods include contact measurement, such as the gypsum wrapping method. As well as non-contact measurement methods, such as mole fringe or laser. At present, this method has leaped forward in development. Using this method and computer, we can get the two-dimensional unfolded drawing of the human body surface from the 3D data of the human body and generate a 2D garment pattern Miyoshi [29]. Compared with the contact method, the noncontact method has the advantages of high automation, short time, and high accuracy Zhu [38]. Noncontact measurement methods can be divided into two categories, active and passive, depending on the measurement method. The active method transforms the image into a point cloud and then measures the point cloud data by the algorithm.

2.2 Human Body Shape Classification

Body shape refers to the physical features and types of the human body. In a multicultural society, body shape can change dramatically depending on cultural and ethnic background. Research has shown that body shape is related to geography, age, marital status, nutritional status, etc. Traditional methods of classifying human body shapes are mainly based on the visual effect of the overall or partial contours of the body, with no clear boundaries between different body shapes. The classification of the human body shape varies from country to country. They give their national criteria based on the physical characteristics of their people. The American Society for Testing and Materials standard uses age, weight, height, and chest girth as a single independent indicator to classify a woman's body shape, rather than girth differences such as chest waist and chest hip differences, reflecting changes in body shape Goldsberry and Reich [9, 10]. Both Japanese and German standards classify body type according to height and then classify the isometric lines on both sides according to the standard hip girth. The difference is that the standards for height classification are different Xia [34]. The Chinese standard classifies the human body shape into four types: Y, A, B, and C, according to the difference between the human chest and waist girth. While these classifications intuitively reflect the body shape, they do not accurately define it. Therefore, the above standards do not fully address the need for clothing design conformity.

In garment design and mass customization, body shape analysis is particularly important to meet the individual needs of the target population. The refinement of the body shape classification has helped to improve the coverage and applicability of products. Lenda Jo Connell et al. have studied the human body using shape analysis theory. They argue that people wearing the same size clothing may have different body shapes. Moreover, people of different body shapes wearing the same clothes may reflect different dressing effects Connell and Presley [24]. Therefore, accurate body shape analysis is an important basis for meeting consumer demand for apparel conformation. Marklin et al. conducted body measurements on U.S. power workers. The results informed the design of power workers' apparel, tools, vehicles, etc., reducing the risk of worker injury and optimizing product performance Marklin and Freier [30]. The study of body shape can guide the production and sale of clothing, building a bridge between the producer and the consumer, with standards to be followed in both the production process as a producer and the purchase process as an end consumer. However, many garment companies only produce less than five body types Marklin and Freier [30]. This does not meet the needs of the average consumer for clothing fit and variety.

In practice, different bodies have different morphological characteristics. The existing classification criteria tend to be rather coarse and do not correspond well to the actual requirements. Researchers have studied the classification of human body shapes from different angles and at different levels. The results of human body shape classification are often not the same due to the different parameters selected for body type characteristics. Therefore, most of the research on the selection of human body type classification is aimed at finding the optimal parameters that can characterize the features of the human body shape Okabe and Yamamoto [22]. These optimal parameters are called critical dimensions. In 1941, O'Brien and Shelton first proposed an anthropometric study of clothing design using critical dimensions Zakaria [36]. Existing classifications of human body shape are mainly implemented using classical statistical methods such as factor analysis, principal component analysis, and regression analysis, and techniques classic data mining techniques such as neural networks Huang [15]. There are two main categories of body shape classification methods: global and partial contours.

In this study, several computer-aided designs (CAD) software are applied to construct a virtual design platform for the proposed design method, permitting 3D scanning, 3D human body modeling, 3D garment construction, 2D pattern design, and 3D virtual tryon (see Fig. 3).

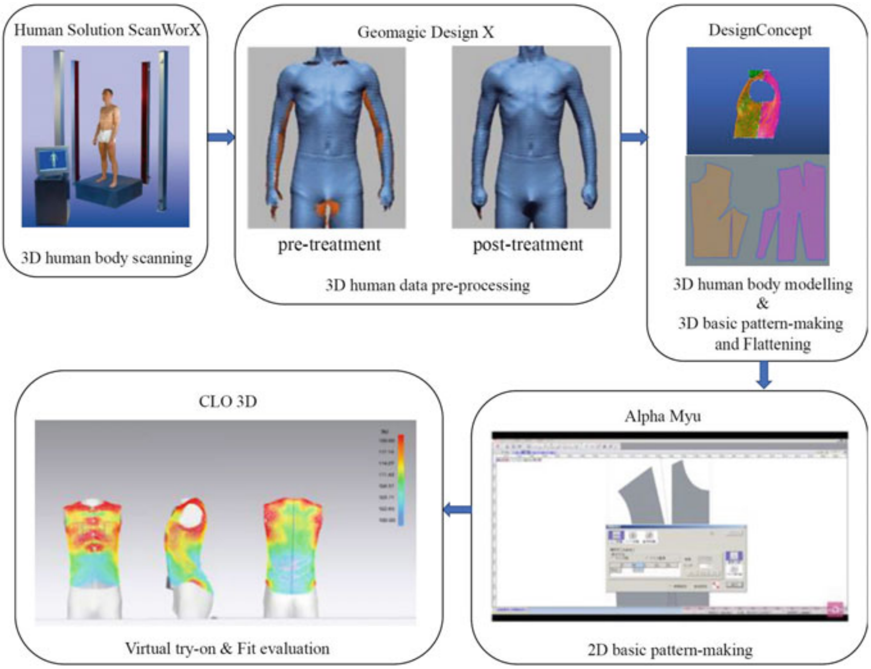


Fig. 3 Different fashion CAD software used and their functions and relations

2.3 *Classification of Knowledge*

Knowledge can be divided into explicit and tacit knowledge according to how it is acquired Masters [28]. Explicit knowledge can be explicitly expressed and obtained from verbal instructions, textbooks, references, software, and databases and is easy to learn Pascual-Leone and Hallett [2]. Tacit knowledge is practical skills or expertise and refers to knowledge that people acquire in terms of skills and understanding, including informal, hard-to-express skills and experiences and insights, intuition, inspiration, etc. Tacit knowledge resides in the brains of experts, and it dominates a variety of human applications Ellis [11]. Tacit knowledge has a significant influence on anthropometry and garment patternmaking. Without the involvement of a designer with extensive professional experience, the accuracy of anthropometric measurements and the fit of garments would be significantly reduced. However, it can take many years to train a designer to become a professional designer. This situation limits the development of fashion houses. All apparel companies would like to be able to extract and formally use the wealth of knowledge that professional designers have and use it freely to ensure the sustainable development of their garment products. At the same time, the extracted knowledge can be used by amateurs who do not have knowledge of fashion design. This can help apparel companies to break away from their dependence on specific designers and make the process of developing garment products more efficient, systematic, and adapted to individual needs in the context of mass customization.

2.4 Human Upper Body Segmentation

To ensure the accuracy and interpretability of the body classification results, the upper body of the subjects was segmented into three parts (the arm root, shoulder, and torso [below the shoulder]) according to the structural characteristics of young male upper bodies and corresponding garment patterns Chi and Tartare [4].
The shoulder shape refers to the shape between the shoulder line, the cross-section where the neck baseline is located, and the cross-section in which the curve of shoulder width is located.
The arm root shape is determined by the arm root line, which is a round curve through the acromion point, shoulder point, axillary anterior point, and axillary posterior point. The shape of this round curve corresponds to the arm root shape.
The torso (below the shoulder) shape refers to the shape between the cross-section where the horizontal line of the waist circumference is located, and the cross-section where the shoulder width curve is located. To achieve the final classification result that includes both global and body part features of the human body, this study proposes to assemble the final classification results of each part to obtain a complete body shape that contains sufficient human features.

2.5 Anthropometric Subject Selection

As this work focuses on young men's shirt pattern-making, it only needs measurements related to men's upper body positions. The general principle can be easily adapted to other body positions and other garment types. The dataset used for this study was based on measurements of 33 young men ages 18–25 years. The impacts of race, socioeconomic status, and lifestyle were considered to some extent. The 33 participants were students (first-year university students to postgraduate students) and are representative of the general population of young men. The cumulative data acquisition time for each sample was approximately four hours.

2.6 Comparing Classification Methods

Many classification methods have been proposed, including clustering based on different classification indicators Olds and Stewart [32], Stewart and Williams [3], and cluster methods Stewart and Williams [21]. For example, the International Standardization Organization (ISO) standard ISO [1] classifies male body shapes (as A, R, P, S, and C) using the chest-waist difference. [33] classifies male body shapes (as Y, A, B, and C) using the chest-waist difference. In Japan, the difference between armpit circumference and waist circumference is used to classify male body types into ten body shapes (as E, EB, BB, B, AB, A, YA, Y, JY, and J) Committee [7]. This means that body shape classification depends on a single variable. However, when classifying body shape, the above methods usually take a global view of the human body, which results in a final clustering result that easily ignores the characteristics of each part of the human body Connell and Presley [24]. Classification result based on the chest-waist difference or the difference between armpit circumference and waist circumference shows that the same shape (after classification) includes different levels of the shoulder, arm root, and torso (below the shoulder) shapes. This classification method thus usually results in fitting problems and the final garment pattern may only fit a few people Alexander and Presley [26]. However, various parts of the body can be effectively classified using SVM. After combining information on the classified body parts, the final classification results include information on the whole and part shape features of the human body. SVM effectively remedies the defects of current body shape classification methods, which do not reflect real body shapes.

2.7 Classification Model for Segmented Upper Body Shapes

Based on the feature measurements obtained, seven variables were selected from the principal component analysis, including Posterior axillary point to the waistline (PAPH), Waist girth (WG), Distance from the shoulder point to the waistline (back) (SWD(B)), Chest arc length (CAL), the Tilt angle of upper body axis (UBAIA), Back arc length (BAL) and Lower chest tilt angle (LCIA). We use the K-means method for body part shape clustering. It has been widely used in various applications. The K-means algorithm requires that the number of classes or categories be determined before running. Feature measurements and classification results for the three body parts will be used to manually label and correct the scanned 3D human body models. When the experimental samples are divided into eight classes, the probability of the F test is <0.05, showing that clustering into eight classes is reasonable. According to the final clustering center, the intermediates of different torso (below the shoulder) shapes are found.

2.8 Feature Measurements Selection

To make the user's measurement easy, a feature measurement from each principal factor is found. Then these items are used for body classification. The selection is based on a high correlation with the corresponding principal factor and relative ease of measurement. Accordingly, feature measurements for the shoulder include SW, SL(F), and FSA. Feature measurements for the arm root include SCD(B), AAL(F), AAL(B), and ARW. Feature measurements for the torso (below the shoulder) include PAPH, WG, SWD(B), CAL, UBAIA, BAL, and LCIA.

1. Shoulder classification
   Three variables were selected from the principal component analysis, including SW, SL(F), and FSA. When the experimental samples are divided into four and five classes, the probability of the F test is <0.05, showing that clustering into three and five classes is reasonable. From convenience for industrial production, we decide to divide the shoulder shapes into three classes. According to the final clustering center, the intermediates of different shoulder shapes are found.
2. Arm root classification
   Four variables were selected from the principal component analysis, including AAL(F), AAL(B), SCD(B), and ARW. When the experimental samples are divided into four and five classes, the probability of the F test is <0.05, showing that clustering into four and five classes is reasonable. From convenience for industrial production, we decide to divide the arm root shapes into four classes. According to the final clustering center, the intermediates of different arm root shapes are found.
3. Torso (below the shoulder) classification
   Seven variables were selected from the principal component analysis, including PAPH, WG, SWD(B), CAL, UBAIA, BAL and LCIA. When the experimental samples are divided into eight classes, the probability of the F test is <0.05, showing that clustering into eight classes is reasonable. According to the final clustering center, the intermediates of different torso (below the shoulder) shapes are found.

3 Parametric Garment Pattern-Making

3.1 Basic Pattern Parametric Pattern-Making Method

In the structural making of the New Bunka Men's upper body basic pattern ITO [17], the key variables include chest girth and back length, as shown in Fig. 4. After the calculation based on these key variables and the structural relationships inherent in the garment pattern, the values of the other secondary parameters are determined. Regarding the darts, the New Bunka Men's upper body basic pattern contains the front chest dart, the back scapular dart, and the waist darts (a, b, c, d, e). The volumes of waist darts are assigned.

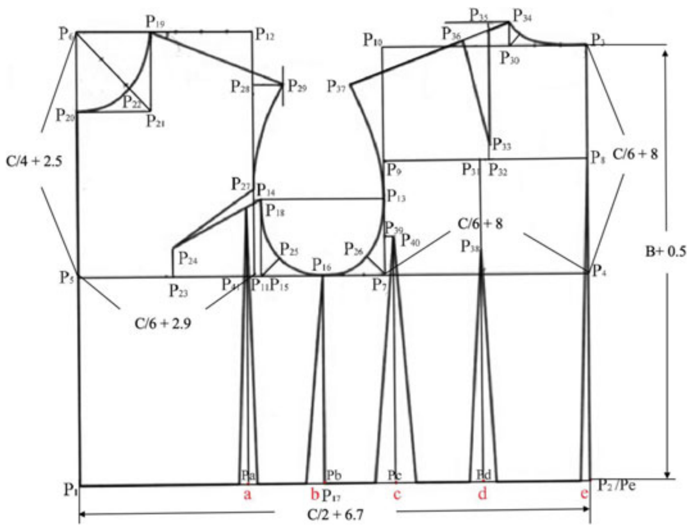


Fig. 4 Structure of New Bunka men's basic pattern. C stands for chest girth and B stands for back length. Red font a, b, c, d, e are waist darts

3.2 Pattern Generation Based on Parametric

The parametric garment pattern-making technology can satisfy the requirements of the apparel industry Liu [19]. Parametric garment pattern-making technology uses the dimensions of the human body feature parts related to the garment structure as constraint values (i.e., parameters, such as chest girth, back length, etc.). By defining the geometric and dimensional relationships between the various geometric objects in a garment pattern, a garment pattern can be quickly adjusted by dimension-driven to generate the new garment pattern Cui [8]. In garment pattern-making, both body shape and dimension are significant for garment fit Liu and Tao [20]. However, current parametric garment pattern-making technology only considers the differences between dimensions Choi and Cui [35]. For this reason, the made garment patterns tend to ignore the individual shape differences, such as the convex chest, hunched back and slipped shoulders, which reduces the custom garment fit Xia and Zhong [12]. Although researchers have proposed some methods of garment pattern adjustment to eliminate body parts unfit, such as shoulder unfit, there is no systematic adjustment method to improve the overall garment fit Kim and Kitazawa [18]. Therefore, it is significant to propose a personalized parametric garment pattern-making method oriented to fit that considers both human body dimension differences and each body part shape differences to help consumers, especially those not in the standard body shape and dimension, achieve the rapid generation of personalized garment patterns. As we have just presented, everyone's body is quite different, in order to limit the study we will focus on the upper male body, but the method is applicable to other parts of the body or populations.

3.3 3D Basic Pattern Design and 2D Basic Pattern Flattening

The basic pattern is relatively easy to fit all shapes and sizes, simplifies the creation of the pattern, and is the basis for all design work. The basic pattern for menswear is mainly for the upper body, which is predominantly on the left side of the body, so only the left side of the body is drawn for pattern making. The basic pattern is drawn based on the chest, as this is the easiest to draw, conforms to the body shape, and provides the best shape. The details are determined by the proportions of the chest measurement, a method known as the bust measurement system. The prototype is semi-fitted with just the right amount of space to allow movement without compromising function. There is additional space to accommodate the length and width of the design and the contours of the design lines. The main advantage is that both chest darts and shoulder darts are clearly defined, which makes it easier to understand how the darts should be deployed and the design developed.
The basic pattern can be divided according to the pattern-making method into the proportional method, the short measure method, and the combination method.

1. The proportional method involves measuring the wearer's chest, back, and arm lengths and using the chest as a base for calculating the pattern dimensions of other parts of the body. The correlation between the chest and other body parts (such as collar width or back width) varies from one individual to another. The main consideration in setting the formula is the average correlation and hence the suitability of the formula.
2. The short measure method, i.e. drawing the dimensions of each part of the human body after taking precise measurements. As the paper pattern is drawn using individual measurements, it is possible to obtain a good-fitting basic pattern if the drawing is correct. However, as the human body is soft and prone to measurement errors, it is important to have the correct measurement data in addition to the correct mapping theory.
3. The combination method, i.e. the combination of the proportional method and the short measure. This method has been used in schools in Japan for a long time.

3.4 Personalized Basic Pattern (Men) Database Construction

After generating a personalized parametric 3D human body model a 3D basic pattern can be designed on the 3D human body model. The top garment pattern is a basic pattern that covers the upper body of the human body and reflects the body shape information. It is a simplification and flat approximation of the complex body surface.

It corresponds structurally to the feature points, curves, and surfaces of the human body. Therefore, the basic pattern structure lines (left part only) are drawn on the model first, based on the correspondence between the basic pattern parts and the body following the body wearing the basic pattern. Then, using the 3D-to-2D flattening technology generates the 2D flattened graph of the basic pattern. Finally, the 2D flattened graph is slightly modified to meet the requirements for industrial production, i.e., Personalized basic pattern (PBP). The PBP obtained by this method will be used for overlapping comparison with the New Bunka basic pattern for the corresponding subject. In addition, the design of the 3D basic pattern must include several darts to generate a 2D flattened graph. The dart locations of the basic pattern are determined according to the structural features of the human body, the common structure of garments in the design, and the location of the common darts and partition lines. After the location of each feature point on the 3D basic pattern has been determined, the basic pattern drawn according to the proportional method is measured to obtain the lengths of the darts. Then, the equal lengths of the darts are drawn on the 3D basic pattern.

3.5 A Regression Model Enabling to Infer from Basic Pattern to Personalized Basic Pattern

From the geometric perspective, a garment pattern can be considered a set of geometric elements, including points, lines, and curves. In this study, the points are considered as the basic elements to construct a basic pattern. By connecting the points, the outline of the basic pattern can be obtained. Therefore, the first step to adjusting the basic pattern is to adjust the coordinate position of the key points that build the basic pattern. The 2D coordinate systems for the front and back pieces have been established separately by using the FCP and BCP as the origin, as shown in Fig. 5.

The coordinates of the corresponding points on the basic pattern and the PBP are measured separately. According to the position of the coordinate points, a linear regression model of the key points on the PBP and the corresponding key points on the basic model are developed. To make the final obtained personalized basic pattern which is oriented to fit, the body parts' feature differences should be considered. The basic pattern is also considered as a combination of three parts in this study based on upper body shape segmentation results. The basic pattern is made of points, straight lines, and curves. Therefore, in this study, the segmentation of the basic pattern should consider the key feature points contained in each part and the coordinate position of the key feature points. Since there is a one-to-one relationship between the body shape and the basic pattern, body part shape differences are considered when the linear regression model is built.

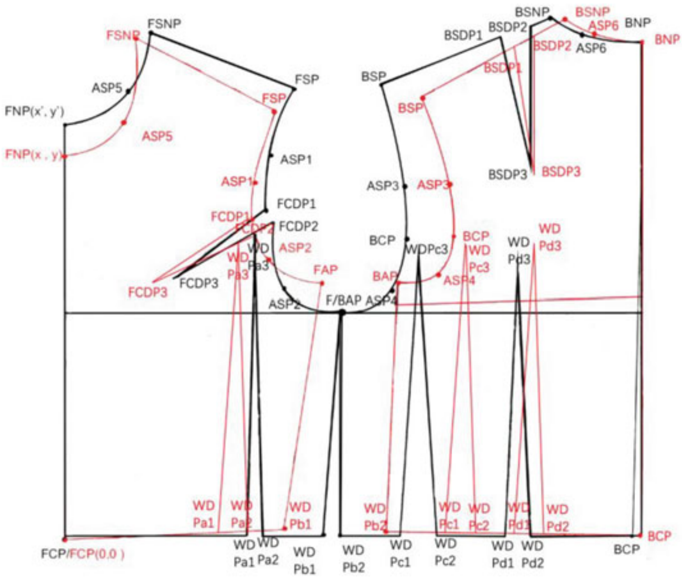


Fig. 5 Correspondence points on the New Bunka basic pattern (Black) and PBP (Red)

3.6 Personalized Shirt Pattern Plotting Method

According to the final coordinates of each key point on the basic pattern and the changing rules, the matpoltlib package of Python is used to draw straight lines on the shirt pattern. Moreover, the Bézier curve model is used to accurately draw the armhole curves, front collar girth, back collar girth, and front, and back hem curves. In this study, a function program for PBP_{shirt} parametric pattern-making is written, named def $\mathrm{PBP_{shirt}}$ (PBP). The initial parameters in the function represent the basic pattern plotting based on chest girth and back length dimensions. When this program is running, both the def PBP and def $\mathrm{PBP_{shirt}}$ are called directly in the Command Window. The values of the initial parameters are entered to obtain the corresponding $\mathrm{PBP_{shirt}}$ directly.

3.7 Garment Fitting Evaluation

The fit of a garment has a significant impact on whether a customer chooses the garment Kim and Damhorst [23]. The evaluation of garment fit can be very useful for clothing buyers if there is no actual designer involved. In this case, this section presents an application based on a garment virtual try-on strain map to evaluate the fit of a garment pattern derived from the recommendation system proposed.

For a specific user, a personalized human body model is used for the garment's virtual try-on. Next, a garment pattern is recommended for this user based on the parametric pattern-making method. Then, the garment is tried on and evaluated virtually in 2D-to-3D using CLO 3D software. The steps are as follows.

1. Import user-specific 3D personalized human body models (without limbs) (Fig. 6a) and personalized garment patterns (front pieces, back pieces, button plackets) (Fig. 6b) respectively, and determine the coordinate orientation and scale of the 3D personalized human body model and personalized garment patterns when importing.
2. Then, the garment patterns are assembled on the personalized human body model (Fig. 6c). Next, the assembled patterns are sewn together to form a 3D virtual garment (Fig. 6d).
3. Finally, the pressure of the digital garment is measured by a strain map (Fig. 6e).

Before evaluating the virtual garment, the user takes a few minutes to understand the method and purpose of the evaluation based on the designer's knowledge. During the evaluation process, the user observes a strain map of the virtual try-on. This observation involves a static view from different angles (front, back, left, right, top and bottom). The fit evaluation of the garment virtual try-on is achieved mainly through the results of the strain map in four areas (chest, back, shoulders, and arm roots). During the fit evaluation process, if the user feels that the design result of the recommended garment pattern is not what he/she expected, he/she can adjust it by adjusting the amount of ease allowance. This procedure is repeated until satisfaction with the result is obtained. Otherwise, the recommended design solution is accepted.

3.8 Design of the System Framework

The system consists of three databases and five models. The three databases include a human body database (Database I), a database of corrected 3D human body parts (Database II), and a personalized basic pattern (PBP) database (Database III). The five models include a relational model between feature measurement items and body shape (Model I), a parametric model for generating a personalized 3D human body (Model II), a relational model between key body dimensions (bust and back length), and basic patterns (Model III), a relational model between basic patterns and PBP corresponding to key points (Model IV), and a relational model between PBP and $\text{PBP}_{\text{shirt}}$ (Model V). The input of the recommendation system is a user interface. The initial input items of the recommendation system are the body dimension constraint parameters (chest girth, back length), and the body feature dimensions (Shoulder width (SW), Front shoulder length (SL(F)), Front shoulder angle (FSA) for shoulder shape; Armhole arc length (front) (AAL(F)), Armhole arc length (back) (AAL(B)), Distance between the back-shoulder point to the chest line (SCD(B)), Arm root width (ARW) for arm root shape; PAPH, WG, SWD(B), CAL, UBAIA, BAL, LCIA for

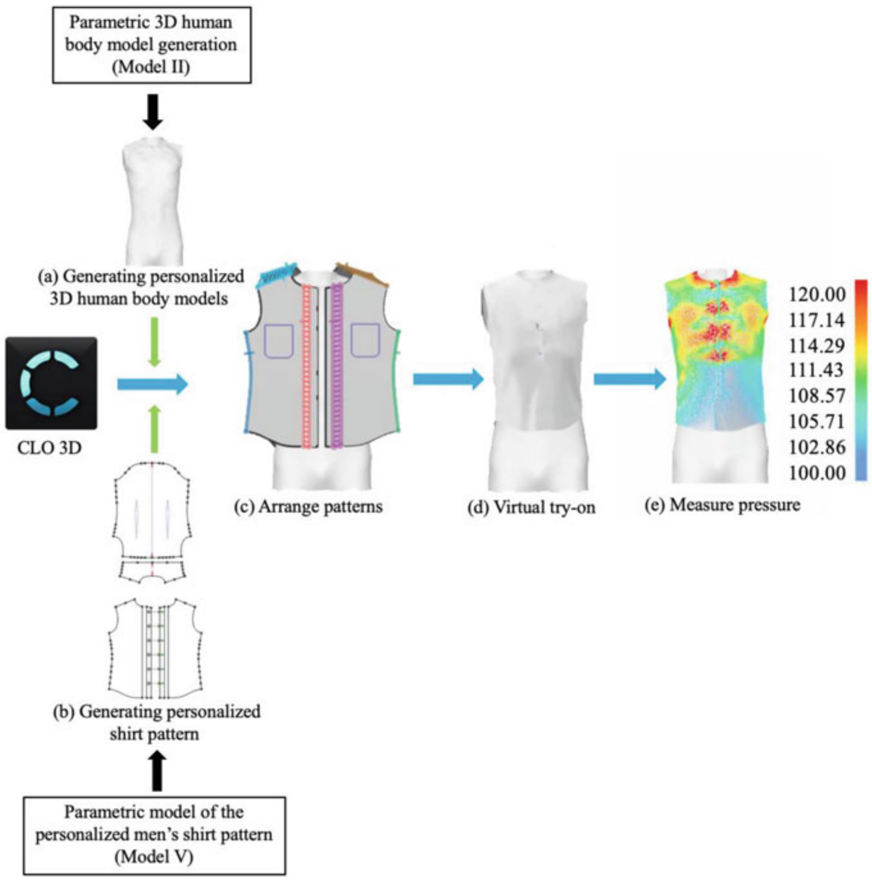


Fig. 6 2D–3D virtual try-on process

torso shape) used to determine each body part shape, as well as three shirt style constraint parameters (slim, regular, and loose), used to determine the style of the final recommended shirt's pattern. Using Model I, the user's body part shapes are identified in Database I based on the input feature measurements. The identified body part shapes will be applied to Model II and Model IV. Based on the results of the determining body shape, the corresponding 3D human body model template is found in Database II. Through Model II, a personalized 3D human body model corresponding to the user is generated. The personalized 3D human body model will be applied in two ways: one for the fit evaluation of the final recommended shirt's pattern (virtual try-on) and one to complement the database III (3D-to-2D flattening technique). Using Model III, it is obtained that the corresponding basic pattern can be generated from the user's chest girth and back length.

The body feature measurements determine the shape of the three body parts. Then, the model is used to generate PBP for the corresponding body part shapes. Depending on the shirt style selected by the user, model V is used to generate a PBP_{shirt} from the PBP. The output of the recommendation system is a fit oriented PBP_{shirt}. Moreover, if the PBP_{shirt} is unsatisfactory after a virtual try-on, four adjustable parameters (front side-seam dart, back side-seam dart, waist dart, and garment bodice length) are designed to adjust the PBP_{shirt} generated by the proposed recommendation system. The user will use the strain map of the virtual try-on to determine if the design suits a specific body. If the user is satisfied with the design, it is then delivered to the production unit for real garment production. Otherwise, the user will obtain a new design solution by adjusting the parameters.

The proposed recommendation system combines the designer's knowledge of manual measurements of the human body, traditional 2D pattern-making methods, and 3D-to-2D flattening techniques to automatically and quickly generate personalized shirt patterns, thus significantly improving the efficiency of pattern-making. The proposed recommendation system allows the development of a new feedback work cycle of recommendation—3D visualization—evaluation—adjustment, which will be repeated until the user is satisfied with the body shape and the recommended shirt's pattern. By successively adding new body shape samples and design cases, the database can become more and more enriched, and recommendation satisfaction will increase. In the next section, we describe the principles of the functional modules used in the recommendation system and the inference process from the initial input parameters to the recommended shirt's pattern, the evaluation of the recommended shirt's pattern, and the adjustment of the recommended shirt's pattern based on the fit evaluation results.

So far, we have mainly introduced a regression model to obtain PBP by adjusting the basic pattern, and parametric modeling methods for the shirt pattern. To prove the feasibility of these methods, in this section, we evaluate the PBP and PBP_{shirt} obtained through these two models separately, in terms of qualitative and quantitative results, and then provide further discussion (Fig. 7).

3.9 PBP's Parametric Pattern-Making Effect

To test the effect of the parametric pattern-making method proposed in the application in this chapter, the chest girth and back length of three people with large differences in body shapes were measured and recorded as an example. Then, these dimensions are input into Python, and the functions written previously are called. The final generated PBPs are shown in Fig. 8. Figure 8 shows a comparison of the strain maps of the virtual try-on effect of the PBP and New Bunka basic pattern. The strain map shows the degree of deformation that occurred after the garment is worn on the virtual model. The area of the color zone represents the garment fit level. If the area shown is red, the garment is stretched by more than 120%. This indicates that the garment strains the body more than the body can handle. The orange to yellow

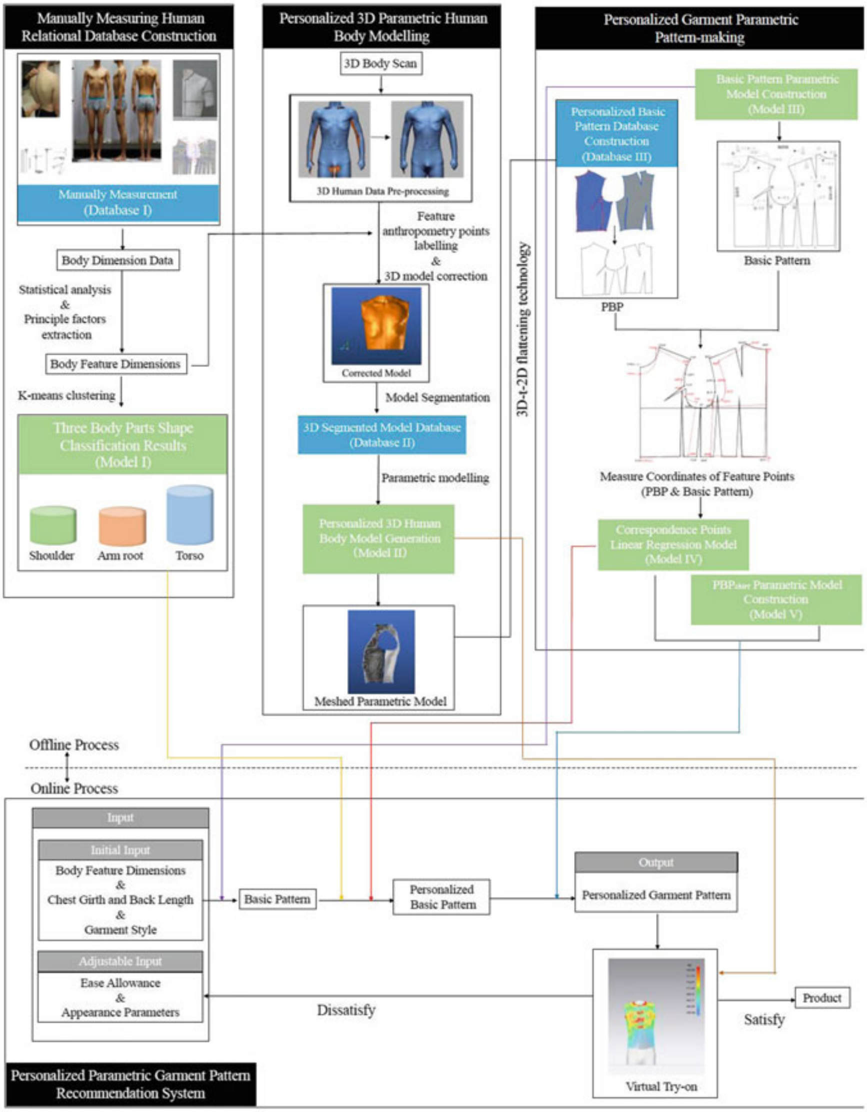


Fig. 7 Overview of the proposed recommendation system

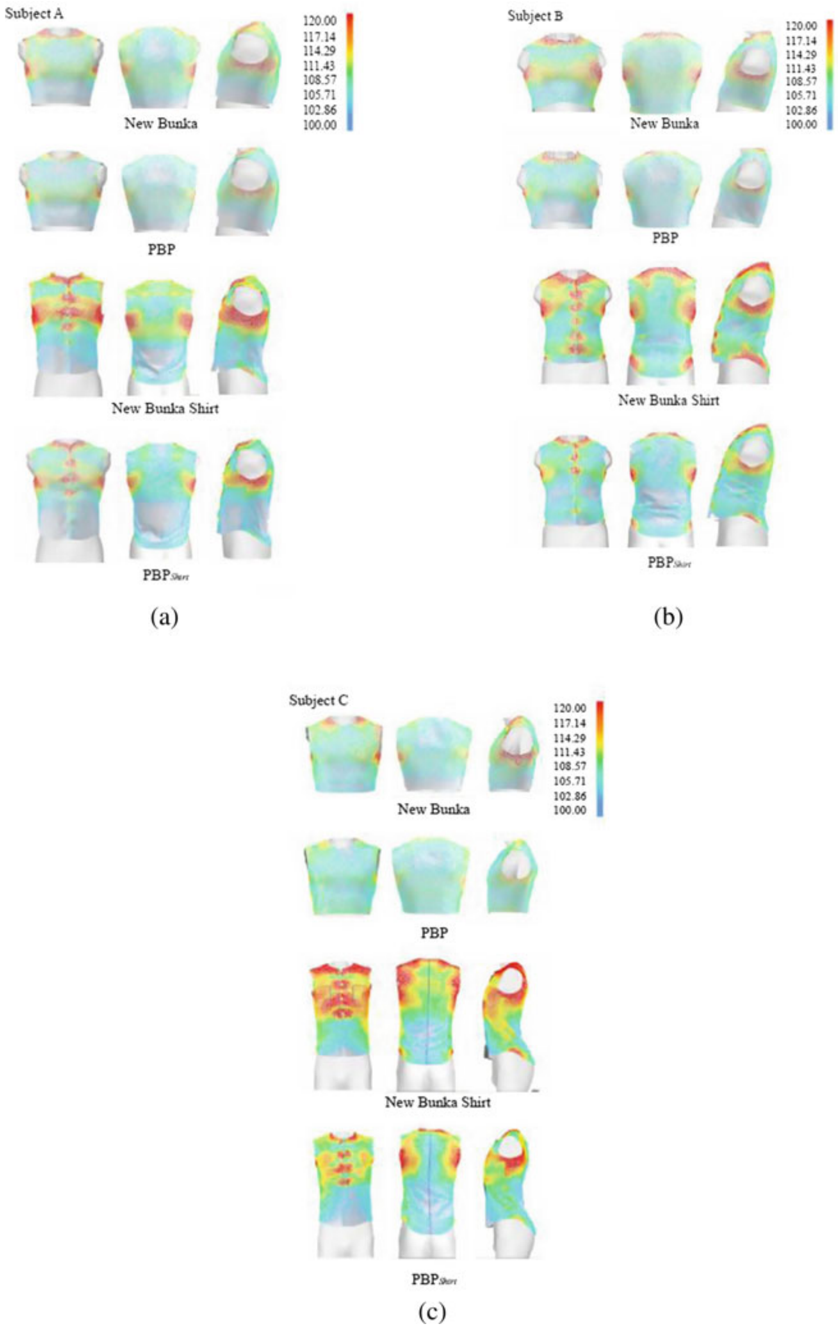


Fig. 8 Virtual try-on strain map for three subjects with different body shapes

area indicates that the garment is stretched between 110 and 120%. This means that although the garment feels tight to the body when worn, it is tolerable. The green area indicates that the garment is in normal condition. This means that the body does not put pressure on the garment, and the person wearing the garment feels comfortable and unrestrained. The light to dark blue area indicates that the garment is stretched between 110 and 100%. This means that there is more space between the garment and the body. Therefore, the fit of the garment is not proper even though there is no tightness. Overall, the fit evaluation standard for judging garments by using the strain map should be that the red area, the orange to yellow area, and the light to the dark blue area on the strain map are only represented in small or no ratio, while the green area is represented in a larger or even all of them.

By comparison, it can be observed in Fig. 8 that the PBP significantly improves the fit of the New Bunka basic pattern. The shoulder fit of subject A is significantly improved in PBP. Subject B's shoulder, arm root, and chest fit are significantly improved in PBP. The shoulder and arm root fit of subject C is significantly improved in PBP. As is shown in Table 1, in the same way, the distribution ratio of the New Bunka basic pattern and PBP of 33 subjects are measured. The red areas on the strain maps of the 33 subjects decrease by an average of 5.6%, the orange to yellow areas decrease by an average of 8.7%, the light to dark blue areas decrease by an average of 15.8%, and the green areas increase by an average of 29.3%. This shows that the PBP has a better fit than the New Bunka basic pattern.

3.10 PBP_{shirt} Parametric Pattern-Making Effect

The PBP_{shirt} proposed in this paper are obtained by using the prototype making method based on the PBP. To test the effectiveness of this method in the application, PBP_{shirt} is made based on the PBP that are obtained in the previous section. Taking the slim fit style as an example, the finally generated PBP_{shirt} are shown in Fig. 8. By comparing the strain maps of the virtual try-on effect of the PBPshirt and the New Bunka shirt version, it can be seen that the PBP_{shirt} fit better than the New Bunka shirt pattern. As is shown in Table 1, compared to the New Bunka shirt pattern, the red areas on the strain maps of the 33 subjects decrease by an average of 10.3%, the orange to yellow areas decrease by an average of 10.3%, the light to dark blue areas decrease by an average of 6.2%, and the green areas increase by an average of 26.8%. This shows that the PBP_{shirt} has a better fit than the New Bunka shirt pattern. Although there is an increase in the light to dark blue areas of subjects like subject C, the increments are not large. Based on the virtual try-on view and designer's experience, this level of increment is not enough to cause the shirt to be excessively loose and thus can be accepted.

Table 1 Comparison of the strain map

Subject	Garment pattern	Strain map distribution ratio			
		Red (%)	Orange/Yellow (%)	Green (%)	Light-dark Blue (%)
A	New Bunka basic pattern	8.9	15.8	19.6	55.6
	PBP	3.8	6.5	46.8	42.5
	New Bunka shirt pattern	17.4	26.4	14.9	41.3
	PBP_{shirt}	9.8	15.7	38.9	35.4
B	New Bunka basic pattern	12.7	22.0	17.5	47.6
	PBP	5.4	9.1	44.7	40.3
	New Bunka shirt pattern	19.7	25.4	13.3	41.6
	PBP_{shirt}	7.2	11.3	46.1	35.2
C	New Bunka basic pattern	4.6	8.1	26.7	60.4
	PBP	0.0	5.7	66.7	27.6
	New Bunka shirt pattern	22.1	31.2	19.7	26.9
	PBP_{shirt}	9.3	20.0	39.2	32.1
All subjects average	New Bunka basic pattern	9.2	14.8	22.4	55.5
	PBP	3.6	6.7	51.7	37.7
	New Bunka shirt pattern	19.1	29.4	15.1	36.3
	PBP_{shirt}	8.8	19.1	41.9	30.1

3.11 The Efficiency of Parametric Pattern-Making

Using parametric garment pattern-making can remove the repeated data calculations and drawings made by the designer after adjusting the parameters. After comparing the results of the above garment pattern-making, it is found that during the pattern-making process, the change of parameter value will cause changes in the relevant structural lines of the pattern. If the pattern is drawn manually, the designer needs to calculate and adjust the new pattern again according to the new parameter values, which increases the designer's workload considerably. Through parametric pattern-making, the position of the feature points and the corresponding structure lines can be automatically adjusted according to the input parameters. Another outstanding effect of using parametric pattern making is that designers can easily handle complex curves in a garment pattern. Designers do not need to calculate the relationship of the structure and position of complex curves because the program can automatically draw complex curves according to the internal structure of the garment and the value of the parameters, which significantly improves the efficiency and accuracy of the pattern design, while in manual pattern making, designers need to recalculate the structure and position of complex curves when the parameter values change. There is no guarantee of the accuracy of the final curve.

4 Conclusion

In the current apparel industry, the existing parametric garment pattern-making models have only considered the differences in human body dimensions. They lack consideration of body shape differences. Therefore, users cannot find suitable clothes often. In this situation, this chapter proposes a personalized garment pattern recommendation system which oriented to fit by integrating the designer's knowledge and 3D measurement. The final generated garment pattern considers the influence of body dimension and body part shape, and the user's requirement of style. The proposed recommendation system can rapidly, accurately, and automatically generate PBP_{shirt} by a linear regression model (from basic pattern to PBP) and parametric model based on the prototype making method (from PBP to PBP_{shirt}). From a comprehensive quantitative analysis, we find that the average of the PBP fit areas has been improved by 29.3% compared with the New Bunka basic pattern. The average of the PBP_{shirt} fit areas has been increased by 26.8% compared with the New Bunka shirt pattern. Meanwhile, the average of PBP and PBP_{shirt} unfit areas (red areas, orange to yellow areas, light to dark blue areas) have been decreased in different degrees. In addition, the proposed method saves plenty of time for designers, and the cost of the shirt's development reduces distinctly. Even users with no pattern-making knowledge can also develop professional shirt patterns by using the proposed system.

References

1. ISO/TR 10652:1991 (1991) Standard sizing systems for clothes
2. Grafman J, Pascual-Leone A, Hallett M (1994) Modulation of cortical motor output maps during development of implicit and explicit knowledge. Science 263(5151):1287–1289
3. Ledingham R, Stewart A, Williams H (2017) Variability in body size and shape of uk offshore workers: a cluster analysis approach. Appl Ergon 58:265–272
4. Bruniaux P, Chi C, Zeng X, Tartare G (2021) A study on segmentation and refinement of key human body parts by integrating manual measurements. Ergonomics 65(1):60–77
5. Chaffin DB, Andersson G (1984) Occupational biomechanics. Wiley, New York
6. Chen CM (2011) Analysis of upper physical characteristics based on angle measurements. Text Res J 81(3):301–310
7. The Japanese Industrial Standards Committee (2001) Sizing systems for men's garments. JATRA/JSA

8. Cui L (2016) The establishment of blouse's parameter constraint database based on pattern automatic generation. In: 22nd international conference on advances in energy, environment and chemical engineering, pp 1–4
9. Shim S, Goldsberry E, Reich N (1996) Women 55 years and older: part i: current body measurements as contrasted to the ps 42–70 data. Cloth Text Res J 14(2):108–120
10. Shim S, Goldsberry E, Reich N (1996b) Women 55 years and older: part ii. Overall satisfaction and dissatisfaction with the fit of ready-to-wear. Cloth Text Res J 14(2):121–132
11. Ellis NC (2005) At the interface: dynamic interactions of explicit and implicit language knowledge. Stud Second Lang Acquis 27(2):305–352
12. Xie H, Xia F, Wu G, Zhong Y (2017) Classification of body shape based on longitudinal section curve. J Text Res 38(6): 86–91
13. Gupta D (2014) Anthropometry and the design and production of apparel: an overview. Anthr Appar Sizing Des 34–66
14. Hersey GL (1997) The evolution of allure: sexual selection from the Medici Venus to the incredible hulk 34(7) (MIT Press, Cambridge, MA)
15. Huang CY (2014) Study on classification of body shape and size grading on young women of Quanzhou district. Adv Mater Res 989:5319–5322
16. Fang JJ, Leong IF, Tsai MJ (2013) A feature-based anthropometry for garment industry. Int J Cloth Sci Technol 25(1):6–23
17. Yumiko ITO (2006) Bunka fashion system lecture (9) fashion design. Bunka Publishing Bureau
18. Takatera M, Narita T, Kanazawa M, Kim K, Innami N, Kitazawa Y (2017) Individualized male dress shirt adjustments using a novel method for measuring shoulder shape. Int J Cloth Sci Technol 29(2):215–225
19. Wang J, Tao X, Xu J, Jiang X, Ren J, Kamalha E, Agrawal TK, Bruniaux P, Liu K, Zeng X (2018) Parametric design of garment flat based on body dimension. Int J Ind Ergon 65:46–59
20. Bruniaux P, Wang J, Kamalha E, Liu K, Zeng X, Tao X (2017) Fit evaluation of virtual garment try-on by learning from digital pressure data. Knowl-Based Syst 133:174–182
21. Tao X, Zeng X, Bruniaux P, Liu K, Wang J, Kamalha E (2016) Fuzzy classification of young women's lower body based on anthropometric measurement. Int J Ind Ergon 55:60–68
22. Yamana N, Okabe K, Yamamoto K (1995) Figure evaluation of the adult females silhouette and relation between the figure and the dress silhouette. J-Jpn Res Assoc Text End Uses 36:45
23. Kim H, Damhorst ML (2010) The relationship of body-related self-discrepancy to body dissatisfaction, apparel involvement, concerns with fit and size of garments, and purchase intentions in online apparel shopping. Cloth Text Res J 28(4):239–254
24. Brannon EL, Alexander M, Connell LJ, Ulrich PV, Presley AB (2006) Body shape assessment scale: instrument development for analyzing female figures. Cloth Text Res J 24(2):80–95
25. Levin EI (1978) Dental esthetics and the golden proportion. J Prosthet Dent 40(3):244–252
26. Connell LJ, Alexander M, Presley AB (2005) Clothing fit preferences of young female adult consumers
27. Morgan MH, Warren HL (1916) Vitruvius: the ten books on architecture, vol 9, no 15
28. Masters RSW (1992) Knowledge, knerves and know-how: the role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. Br J Psychol 83(3):343–358
29. Miyoshi M (2006) Clothing modelling learning. Beijing China Textile Apparel Press
30. Seeley P, Marklin RW, Saginus KA, Freier SH (2010) Comparison of anthropometry of us electric utility field-workers with North American general populations. Hum Factors 52(6):643–662
31. Sunderland E (1995) Anthropometry: the individual and the population 32(7) (Cambridge University Press)
32. Petkov J, Olds T, Daniell N, David Stewart A (2013) Somatotyping using 3d anthropometry: a cluster analysis. J Sports Sci 31(9):936–944
33. (2009) Inspection The State administration of quality Supervision and quarantine of the PRC. In: The People's Republic of China's national standard clothing size men: clothing size mens. China Standard Press

34. Xia F (2017) Research of somatotype based on the sagittal plane silhouette of human body. Donghua University
35. Choi KM, Choi YL, Nam Y, Cui MH (2007) A method for garment pattern generation by flattening 3D body scan data. In: International conference on digital human modeling, pp 803–812
36. Zakaria N (2014) Body shape analysis and identification of key dimensions for apparel sizing systems. In: Anthropometry, apparel sizing and design, pp 95–119
37. Zakaria N, Gupta D (2019) Anthropometry, apparel sizing and design. Woodhead Publishing
38. Zhu J (2015) Human body 3D reconstruction and size measurement based on kinect. Hangzhou Dianzi University

Efficient and Trustworthy Federated Learning-Based Explainable Anomaly Detection: Challenges, Methods, and Future Directions

Do Thu Ha, Ta Phuong Bac, Kim Duc Tran, and Kim Phuc Tran

Abstract Artificial Intelligence (AI) and especially Machine Learning (ML) are the driving energy behind industrial and technological transformation. With the transition from industry 4.0 to 5.0, smart manufacturing proves the efficiency in industry, where systems become increasingly complex, producing massive data, necessitating more demand for transparency, privacy, and performance. Federated learning has demonstrated its effectiveness in various applications, however, there are still exist certain challenges that should be addressed. Thus, in this chapter, a comprehensive perspective on federated learning-based anomaly detection is provided. The problems have posed concerns and should be taken into account when researching and deploying. Then, our perspectives about efficient and trustworthy federated learning-based explainable anomaly detection systems are demonstrated as an end-to-end unified framework. Finally, to provide a complete picture of future research direction, the quantum aspect is introduced in the subject of machine learning.

Keywords Federated learning · Explainable anomaly detection · Blockchain machine learning · Graph transformer network · Quantum computing

1 Introduction

The Internet of things (IoT) aspires to unite the physical and digital worlds into a single system, opening up significant business prospects for the manufacturing industry toward Smart Manufacturing. It has brought about a new paradigm where new processes are being driven by a network of machines and gadgets capable of collaborating and interacting. Smart Manufacturing are systems where the Internet

D. T. Ha · K. P. Tran
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

T. P. Bac · K. D. Tran (✉)
International Research Institute for Artificial Intelligence and Data Science, Dong A University, Danang, Vietnam
e-mail: ductk@donga.edu.vn

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_8

in factories network is connected to the physical world through sensors and can be thought of as the management of a network of devices, home appliances, and vehicles of the IoT. Also difficult due to the dynamic linkage between devices, actors, and resource constraints. Due to its complex environment and a large number of communication technologies, the new IoT paradigm in the Smart Manufacturing context is vulnerable to several security issues that are frequently quite resource-intensive. As a result, there has recently been concern about cybersecurity in this area despite a lack of policy direction and user knowledge values associated with cybersecurity, even though the policy has not been guided by important stakeholder values.

Machine Learning has significantly improved industrial manufacturing. However, ubiquitous distributed manufacturing networks produce enormous amounts of data daily and are becoming more appealing for storing data locally. Moving network computing to the edge, due to the increased computational capacity of these devices combined with concerns regarding communicating private data, is becoming increasingly popular. High-quality ML models require a significant amount of labeled training and test data to achieve these benefits. This information should be kept private and safeguarded from prying eyes. Since these datasets cannot be shared with centralized ML servers in traditional architectures, sharing knowledge amongst cooperating devices that protect privacy is required. Moreover, as the storage and computational capabilities of the devices within distributed networks grow, it is possible to leverage enhanced local resources on each device. This has led to a growing interest in Federated Learning [1].

This chapter discusses the challenges of applying federated learning to anomaly detection and proposes a perspective about an end-to-end efficient and trustworthy federated learning-based explainable anomaly detection system that overcomes the difficulties of existing federated learning. Additionally, this chapter provides future directions for comprehensive studies in the context of smart manufacturing. The remainder of this chapter is organized as follows: Sects. 2 and 3 describe state-of-the-art approaches for anomaly detection and federated learning-based anomaly detection, respectively. In Sect. 4, we analyze the difficulties and challenges of federated learning and apply them to anomaly detection, with the goal of developing an efficient, and trustworthy federated learning-based anomaly detection system. Section 5 presents our end-to-end framework, which provides insights into federated learning-based anomaly detection in the context of smart manufacturing. This framework addresses the aforementioned drawbacks by employing a comprehensive system based on blockchain machine learning and explainable AI, with a robust graph transformer network. Section 6 outlines future directions for a comprehensive study, and finally, Sect. 7 briefly discusses the concluding remarks of this chapter.

2 Anomaly Detection

Anomaly Detection (AD) is a process of identifying instances (anomalies, outliers, exceptions) in data that differ significantly from the expected norm. This is achieved by analyzing the characteristics of normal data. In today's world of distributed systems, effective management and monitoring of system performance are of paramount importance. With a multitude of components to monitor, AD can play a critical role in identifying errors, facilitating root cause analysis, and speeding up technical support. As a result, AD has gained significant traction across a variety of industries, including healthcare, fraud detection, network security, industrial damage detection, and military surveillance [2, 3].

When designing and evaluating anomaly detection systems, it is essential to have a well-defined understanding of both the input and the desired output, as this will have a significant impact on the design and implementation of the anomaly detection system, as well as its effectiveness in detecting anomalies. The type of input data is a crucial consideration, as it can take various forms such as sequence data, spatial data, spatiotemporal data, graph data, etc...

The outputs generated by anomaly detection systems can generally be categorized into two forms: Scores and Labels. The appropriate form of output to use will depend on the specific needs and objectives of the analysis. Scoring techniques assign an anomaly score to each instance in the test data, which reflects the degree to which the instance is considered an anomaly. This enables the analyst to apply a domain-specific threshold to selectively identify the most significant anomalies. Label techniques, on the other hand, provide binary labels (normal or anomalous) to the test instances. While these techniques do not offer direct control over the selection of anomalies to the analyst, the choice of anomalies can be influenced by parameter adjustments within each technique [2].

In terms of the behavior of anomalies, AD could be systematically categorized into three distinct forms based on the behavior of anomalies (i.e. point anomalies, contextual anomalies, and collective anomalies). Point anomalies refer to irregular deviations that might occur at random and do not have a defined interpretation. Contextual anomalies, also known as conditional anomalies, are data points that are deemed abnormal when considered within a specific context. This type of abnormality is identified by considering contextual and behavioral features that may vary based on time and location. Collective anomalies, on the other hand, refer to abnormal groupings of individual data points that may appear normal when considered individually but exhibit unusual properties collectively [4].

Anomaly detection (AD) can be further characterized into two methodologies, including a statistical approach and a machine-learning-based approach. The statistical approach relies on the assumption of data normality and posits that normal data objects are generated by a probabilistic and statistical model with a closed-form probability distribution. The parameters of the model are estimated, and data that deviates from the model is considered an outlier. In manufacturing, Control Charts, an effective tool of Statistical Process Control (SPC), continuously monitor a process as well as detects process abnormalities in order to enhance and optimize the process [5]. For this aim, numerous different forms of Control Charts have been produced [6]. On the other hand, the machine-learning-based approach employs machine-learning techniques to identify abnormalities.

Machine learning-based approaches for anomaly detection can be classified in a variety of ways, depending on the type of algorithm used, the presence of labels, and the form of input data. With respect to the availability of labels, anomaly detection can be further divided into three categories: supervised, semi-supervised, and unsupervised. The process of supervised anomaly detection involves training a supervised binary or multi-class classifier, utilizing the labels of both normal and anomalous data instances. In contrast, unsupervised anomaly detection techniques are based solely on the intrinsic properties of data instances and are used to identify outliers. The difficulty of obtaining labeled data, particularly for anomalous instances, makes the use of unsupervised techniques essential in the automatic labeling of unlabeled data. Semi-supervised anomaly detection might techniques address this challenge by leveraging the available labels of the normal class to differentiate outliers from normal instances in a given dataset. This combination of both labeled and unlabeled data has resulted in the widespread adoption of semi-supervised anomaly detection techniques.

The use of time series data is prevalent in the area of Anomaly Detection, and it can be divided into two main categories: univariate and multivariate time series. Univariate time series is characterized by the variation of a single variable or feature over time. The recent progress in the field of deep learning has presented opportunities to extract more sophisticated features, which have the potential to improve the efficiency of anomaly detection in univariate time series data. In contrast, the detection of anomalies in multivariate time series data presents a complex and compelling challenge, attracting significant research attention [2, 7].

3 Federated Learning-Based Anomaly Detection

Federated Learning (FL), as described in the work of McMahan et al. [1], provides collaborative and privacy-preserving on-device training. FL is a recently developed method for sharing model parameters between edge devices without revealing raw data, such as neural network weights. Models are trained locally on edge devices for this purpose and are subsequently uploaded to an aggregator server that integrates model parameters, for example, by averaging the results. Finally, the model is returned to the clients for evaluation after aggregation. This process is repeated until a predetermined number of communication rounds or a specified degree of quality (such as classification accuracy). Collaborative model training is made possible by federated learning (FL), which avoids explicit data sharing.

Federated Learning (FL) algorithms are appealing to a wide range of applications, especially in healthcare and smart manufacturing, due to their intrinsic privacy-preserving characteristics [8]. By cooperatively developing a joint model using only the parameters of locally learned models, FL aims to give data owners better privacy protection by preventing the release of raw data and enables the collaborative construction of a joint, robust global model [9]. Given these profound advantages, Federated Learning is considered a potential technique in the anomaly detection field, for instance, in Preuveneers et al. [10], Mothukuri et al. [11], Zhao et al. [12], Cui et al. [13]. The integration of blockchain technology with federated learning in the work of [10] aimed to facilitate the auditing of model updates by federated deep learning clients. However, in the context of a specific security use case for federated intrusion detection, the MultiChain blockchain imposed certain restrictions, resulting in limited performance. Subsequently, a Federated Learning (FL) framework based on blockchain technology was presented in [13] to mitigate the threats of poisoning attacks against IoT anomaly detection models, while ensuring accuracy, efficiency, security, and privacy. Authors Zhao et al. [12] utilized federated learning to tackle the data scarcity problem and preserve data privacy by collaboratively training multiple participants in a global mode. In Mothukuri et al. [11], the authors proposed a federated learning (FL)-based anomaly detection approach to proactively recognize intrusion in IoT networks. They used Gated Recurrent Units (GRUs) models and kept the data intact on local IoT devices by sharing only the learned weights with the central server of the FL. With the rise of FL applications, training in heterogeneous and time-series data, as well as potentially massive networks, presents novel challenges that require a fundamental shift in how large-scale machine learning, distributed optimization, and privacy-preserving data analysis are conducted [14].

4 Dificulties and Challenges for Federated Learning

In this section, the core challenges of federated learning-based anomaly detection are analyzed and presented, including Expensive Communication, Systems and Data Heterogeneity, Resource Constraints, Security and Privacy Concerns, and Hyperparameter optimization, which are the main bottleneck problems that limit the expansion of smart manufacturing.

4.1 Expensive Communication

Communication poses significant challenges in federated networks, particularly in the context of 5G and the transition towards 6G. Federated Learning (FL), in which thousands of devices participate in model training, is hindered by communication acting as a critical bottleneck. The requirement for data to remain local due to privacy concerns over raw data transmission exacerbates the communication challenge, especially in federated systems consisting of a large number of devices such as millions of smartphones. In these scenarios, communication within the network may be slower compared to local computation, making it imperative to investigate and implement communication-efficient techniques, particularly for real-time applications.

To make the FL model suitable for networks consisting of massive and heterogeneous devices, it is necessary to develop a communication-efficient method that significantly reduces the number of gradients exchanged between devices and the cloud. To reduce communication overhead even further, two main factors must be considered: decreasing the total number of communication rounds and reducing the number of gradients exchanged in each communication round [15].

4.2 Systems and Data Heterogeneity

In Federated Learning, heterogeneity problems can arise from constraints in both systems (heterogeneity in systems) and data (heterogeneity in statistics). The diverse range of edge devices available, such as the Jetson Nano, smartphones, and Raspberry Pi, may exhibit significant disparities in their storage, computing, and communication capabilities. Such discrepancies can primarily be attributed to differences in their underlying hardware specifications, including variations in CPU and GPU performance, network connectivity options (such as 3G, 4G, and 5G), and power capacity. It is imperative to consider these variations while designing and deploying edge computing applications to ensure optimal performance and efficient resource utilization. It should be pointed out that the dependability of such devices may differ, and it's not rare for an operational device to temporarily disconnect within an iteration due to factors such as connectivity or power restrictions. Therefore, training efficiency may vary significantly across client devices, and considering all clients with the same scale does not provide us with an optimal solution. Given this heterogeneity in systems, it is crucial that researchers and developers take these factors into consideration when implementing and deploying a Federated Learning system.

Data from different users can also exhibit statistical challenges due to variations in data generation and collection processes, leading to non-IID (nonindependent and identically distributed) data. The occurrence of non-IID data in Federated Learning is due to the fact that the training data on each client is heavily influenced by the usage patterns of their local devices. As a result, the data distribution among connected clients may vary significantly from one another. Processing non-IID data presents challenges as it increases the complexity of modeling and evaluation. Federated Learning, which typically uses stochastic gradient descent to train deep networks, is particularly impacted. Ensuring the data is IID is crucial for obtaining an unbiased estimate of the stochastic gradient [16].

4.3 Resource Constraints

The effort to enhance the performance by investing in more complex and deeper models can exist the memory, computational ability, and energy budget of local devices, and be time-consuming. While reduced computational capabilities imply that it takes more time to process data, limited memory capacity makes the device prone to over-flooding. These situations can lead to more expensive communication.

4.4 Security and Privacy Concerns

Privacy is the main concern in FL applications. Federated learning protects data by only sharing model updates (e.g., gradient information) rather than the created raw data on each device. Nevertheless, sharing model updates with other participants in the training process or the central server may expose sensitive information and lack insight into the whole training data. Recent techniques try to increase the privacy of federated learning using techniques like differential privacy or secure multiparty computation, but these methods frequently sacrifice model performance [14]. Besides, FL applications have been limited by security concerns due to various adversarial attacks, such as poisoning attacks. Such attacks attempt to poison the local models and data to manipulate the global models in order to obtain undue benefits and malicious use [17]. Consequently, it is challenging to develop appropriate strategies to address such attacks and balance the trade-off between performance and private federated learning systems, both theoretically and experimentally [9].

4.5 Hyperparameter Optimization

Hyperparameter optimization is a significant component that can influence accuracy, particularly when dealing with varied client data distributions. Standard FL approaches are unstable in these situations and require thorough hyperparameter adjustment to operate at their best. Real-world FL applications cannot use conventional hyperparameter optimization techniques because they require multiple training runs that are frequently unaffordable with low computational budgets [8]. The data distribution influences the choice of the best hyperparameter configuration and suggests that the best hyperparameter configuration for a client might differ from another client based on individual data properties [18].

5 A Perspective about Efficient and Trustworthy Federated Learning-Based Explainable Anomaly Detection System

In the developing context of IoT, new algorithms and approaches in anomaly detection are required to improve the reliability and explainability of leveraging machine learning models. The methodologies used in our study related to the transformer-based [19] and Fourier integral attentions [20] for anomaly detection model, the leveraging of graph attention network for anomaly detection with time series data. Also, Federated learning is the core architecture of the deployment to achieve high accuracy, convergence, efficiency, and trustworthiness in the IoT environment. Transformers are robust neural networks that have achieved tremendous success in many areas of machine learning [21] and have become state-of-the-art models for various applications across different data modalities. In addition to their excellent performance on supervised learning tasks, transformers can effectively transfer the learned knowledge from a pretraining task to new tasks with limited or no supervision [21, 22]. Therefore, Transformer is the potential algorithm to address time-series data problems in the manner of online anomaly detection. The development of technologies entails the complexity of management models, which is the case of social networks, telecommunications and IoT networks, biological networks, or device networks. Numerous exciting tasks entail data that cannot be represented in a grid-like structure and is instead located in an irregular domain. Thus, the organization and modeling of these complex structures are also crucial in improving the network's efficiency in anomaly detection. Applying the Graph presentation method, such data can be represented in graphs and connections. The graph attention network [23] is a novel convolution-style neural network that operates on graph-structured data, leveraging masked self-attentional layers. Besides, federated learning, a promising technique that can reduce the burden of the central server and enhance user privacy, also is integrated to improve performance and applied to different data tasks, e.g., anomaly detection, cyberattack classification, health monitoring, and time-series prediction.

In this section, a new approach for efficient and trustworthy Online Federated Learning-based Explainable Anomaly Detection is proposed by using Continual Federated Learning and Explainable AI with Robust Graph Transformer Networks at edge devices that could eliminate the aforementioned drawbacks. Then, considering communication transmission problems when deploying a system in a real environment depends on hybrid digital-analog neural network transmission for federated learning.

5.1 Rationale and Consideration

Although federated learning has yielded profound benefits, it is imperative to acknowledge the existence of persistent challenges that demand our attention. This section endeavors to present several potential techniques that can effectively address these challenges.

5.1.1 Systems and Data Heterogeneity Problem

The heterogeneous nature of IoT devices frequently generates and collects data in a non-identically distributed (non-IID) manner across the IoT network. The non-IID data paradigm raises the possibility of stragglers, violating the widely accepted independent and identically distributed (I.I.D.) requirements in distributed optimization, and may make modeling, analysis, and evaluations more difficult. The traditional FL depends on SGD aims to train the global model. Although the SGD produced beneficial results when applied to IID data, it is unstable when used with non-IID data. In particular, SGD fails when differential privacy is applied to model parameters to enhance client security and privacy. Therefore, it is necessary to develop new methods that consider the non-IID data issue and strike a balance between the level of privacy and the model utility in FL. The canonical federated learning problem of learning a single global model can also be solved by continuously learning several local models through multi-task learning. There are several existing FL algorithms that were proposed to tackle the data statistical heterogeneity (i.e. FedProx [14], SCAFFOLD [24], and Differential Privacy (DP) noise with FedDyn [25]. However, those methods still have faced some drawbacks, for instance, performance limitations, and high cost. To eliminate those drawbacks, DP-SCAFFOLD [26] could be applied which aims to efficiently train on highly heterogeneous user data and ensure users' privacy by incorporating Differential Privacy (DP) constraints into the SCAFFOLD algorithm.

5.1.2 Robustness Local Hyperparameter Optimization, Global Aggregation Algorithms, Communication Efficiency for Federated Learning

Because FL needs devices to share their ML parameters iteratively, the time it takes to jointly learn a good model relies on the number of training steps and the average time it takes for ML parameters to be transmitted between stages. FL parameter communications frequently occur via networks with constrained resources, such as wireless networks with constrained bandwidth and power. Since wireless channels and other end-user internet connections often operate at lower speeds than intra-datacenter or inter-datacenter lines and might be extraordinarily costly and unreliable, it is now widely accepted that communication can be a significant bottleneck for federated learning. Thus, a noticeable delay that can be orders of magnitude longer than the time required to train an ML model is caused by the repetitive communication of FL parameters from edge devices [27]. As a result, federated learning's communication bandwidth has recently attracted much attention. For example, methods combining Federated Averaging with sparsification and/or quantization of model updates to a few bits have significantly reduced communication costs with little to no impact on training accuracy [28]. It is yet unknown, though, if communication costs may be further decreased and whether any of these approaches, alone or in combination, can provide the best possible balance between communication and accuracy in federated learning. Theoretical statistics has recently shown an interest in characterizing such fundamental trade-offs between accuracy and communication, which was discussed in the works of Konečný et al. [28], Braverman et al. [29], Han et al. [30], Barnes et al. [31], Tang et al. [32], Barnes et al. [33]. The ideal minimax rates for distributed statistical estimation and learning with communication restrictions are described in these studies. However, because they often neglect the impact of the optimization algorithm, it is challenging to draw practical conclusions regarding communication bandwidth reduction from these theoretical works. Utilizing such statistical methodologies to guide actual training methods is still a possibility.

Due to the limited resources of current computing devices, such as computing power, memory, and communication capabilities, there are several compression objectives that are of practical significance. These objectives are motivated by the need to optimize resource utilization and include gradient compression, model broadcast compression, and local computation reduction. Gradient compression aims to reduce the size of the object transmitted from clients to the server for updating the global model, while model broadcast compression reduces the size of the model broadcast from the server to clients, where local training begins. Local computation reduction modifies the overall training algorithm to increase the computational efficiency of the local training procedure. Conducting multiple training rounds with various hyperparameters on an IoT device with limited resources can be challenging. It may lead to excessive use of restricted communication and computing resources for small device populations. However, modern deep neural networks rely heavily on a wide range of hyperparameter choices, including architecture, regularization, and optimization. The selection of hyperparameters significantly affects the performance of machine-learning algorithms. As a result, optimizing local hyperparameters poses new challenges. It requires careful consideration of resource constraints while ensuring that the chosen hyperparameters improve the performance of the machine-learning algorithm. By addressing these challenges, it is possible to achieve efficient and effective machine learning on devices with limited resources [18, 34] in the FL setting. The factor can be the distribution of heterogeneous clients that influences the choice of the best hyperparameter configuration for a client might differ from another client based on individual data properties [8].

In the context of autonomous machine learning, hyperparameter optimization has a long history [35, 36], but its primary focus is on how to increase model accuracy rather than mobile device communication and computation efficiency. Moreover, it is essential to note that Federated learning is contingent upon the integration of updates provided by the participating devices, with the integration process being designed to protect the privacy of the contributed data. However, the traditional approach is susceptible to receiving compromised updates, which may originate from malicious attackers with the intention of compromising the confidentiality of the system or result from the occurrence of hardware malfunctions in inexpensive devices. The susceptibility to corrupted updates poses a vulnerability that warrants the implementation of robust mitigation measures to ensure the integrity and security of the Federated learning process. Therefore, the Robust aggregation algorithm (RFA) [37] is utilized, based on the geometric median and the smoothed Weiszfeld algorithm, to make federated learning more robust to settings and bring benefits about Privacy Preservation, Robustness, and Communication Efficiency.

5.1.3 Interpretability for Federated Learning—Explainable Federated Learning

In IoT security, ensuring system transparency and accountability is crucial for enhancing the security of processes and algorithmic decision-making. In the scope of anomaly detection, while neural network models tend to outperform signature-based methods, deep models are known as "black boxes" that inhibit user interpretation. Therefore, it is important to balance the benefits of deep learning with the need for interpretability and transparency in IoT security. This yields developers time-consuming and vague in interpreting model decisions, causing the unreliability and uncertainty of detection processing [38]. To address this issue, Explainable Artificial Intelligence (XAI) has been developed and become a powerful technique that aids systems reliability. There are plenty of explainable frameworks that appear to adapt to different model structures, for example, Shapley Additive exPlanations (SHAP) [39], deep SHAP [40], Local Interpretable Model-agnostic Explanations (LIME) [41], Deep Learning Important Features (DeepLIFT) [42], Grad CAM [43], attention mechanism, feature importance, rule-based approach, etc…Although XAI can achieve a huge impact on model explanation, there still exists a dilemma between interpretability and accuracy [44]. When performance and complexity coexist, interpretability encounters a previously unavoidable downward slope on this path to performance. However, as more sophisticated explainability methods emerge, that slope may be inverted or canceled. Explanations for machine learning models must be drastic and approximate enough to meet the need of users, while also being representative of the studied model and not oversimplifying its essential feature [45]. XAI techniques are adapted to FL architecture that allows for a balance between performance and interpretability.

To be specific, Shapley values (SV), which are a type of feature interpretation method, are being applied for explanations in Federated Learning. SV has been proven to enhance explainability and interpretability in various conventional machine learning models. However, in the context of FL, a modified version of Shapley values is needed to adapt to differences in system structure and achieve a balance between accuracy and interpretability. The GTG-Shapley algorithm, as presented in Liu et al. [46], holds significant potential for enhancing Federated Learning (FL) processes. The FL server receives gradient updates from each participant during every training round, and the GTG-Shapley method leverages this data to assess the effectiveness of various FL sub-models that emerge from different counterfactual permutations of model aggregation, based on the estimation of Shapley Values (SV). When evaluating a sub-model, GTG-Shapley strategically generates permutation sequences to achieve rapid convergence, assesses the importance of evaluating a sub-model based on its expected marginal gain, and dynamically eliminates less significant sub-models.

5.1.4 Online Federated Learning

To the best of our knowledge, the majority of state-of-the-art works in anomaly detection have been implemented as offline learning algorithms, which means a model is needed to train on the full dataset. If the system wants to update new knowledge on upcoming data (such as a new type of attack), the model needs to be trained on a new version of the system from scratch on the full dataset (not just the new data, but also the old data). This is the main drawback of offline learning in which systems tend to be outdated, especially with the rise of using IoT devices and the increased computing power of computers, the potential size of the data which can be retrieved from a client is continuously updating and increasing. To address a problem, research about Online learning has been conducted, allowing training models on streaming data (data arrives in a sequence). However, online federated neural detection necessitates increased network performance during the detection process as well as tighter computational resource constraints. To address this issue, Zhang [47] proposed Distributed Online Gradient Descent (D-OGD) learning, which refers to a stochastic version of the general DGD framework. This online learning structure offers the benefits of low computation costs per round and low memory requirements for each node, making it well-suited for many real-world scenarios. OGD employs some predefined learning sequences μ_i and decomposes the learning into two phases.

5.1.5 Preventing and Detecting Adversarial Attacks in Federated Learning

Federated Learning (FL) in IoT environments is growing significantly, particularly in applications concerned about privacy, such as healthcare. The federated environment presents obstacles to current privacy-preserving techniques. Beyond offering stringent privacy assurances, it is crucial to provide ways that are inexpensive to compute, effective at communicating, and tolerant of damaged devices without excessively reducing accuracy [48]. The own operation of Federated Learning includes data and behavior auditing, training, and testing, which faces different kinds of security threats such as data poisoning [49], model update poisoning [50], and model evasion attacks [51]. Regarding poisoning attacks [52] where a malicious device modifies its local training data or the model after training it on benign data using carefully designed techniques to cause the performance of the global model. Adversaries may be able to create malicious updates that precisely deliver the desired result, bypassing the approved client loss function or training system. The central aggregator itself is another potential attack point. The trained model can be readily the target of both controlled and uncontrolled attacks if an attacker can corrupt the aggregator. As a prominent trend in current research on data privacy and robustness privacy attack described, differential privacy (DP) [53] was developed to combat the attacks in the FL model. In DP, random noise is injected into the training or testing period to reduce the impact of specific data points. Differential privacy has some compelling advantages as a defensive strategy. First, it offers first-rate defense against a range of threats in the worst-case scenario. Second, various machine-learning activities can be applied to several well-known privacy techniques and safeguards. Finally, it is known that DP is component-closed, meaning that subsequent algorithms' inputs are chosen based on the outputs of earlier ones. DP is a method that has much potential for developing defense mechanisms and detecting attacks in federated learning. Hence, this study proposes avoiding such model and data poisoning attacks. The shared local model gradients should be checked using procedures to ensure they were not trained on anomalous data (noisy, featured poisoning, label poisoning, etc.). In other words, to prevent malicious peers from compromising and influencing the global model, malicious behavior of the locally trained models should be tracked before merging them into the global aggregation. The detection and prevention mechanisms are taken into account to develop a Federated learning-based cybersecurity system. Enhance the security and safety of the federated learning model with advanced techniques such as differential privacy.

5.1.6 Federated Learning for Resource-Constrained IoT Devices

While FL methods achieve significant advantages, almost works assume there are available computational resources and do not consider the limitations of edge device sources. However, the IoT devices, such as robots, drones, and low-cost computing devices, may have limited processing power, bandwidth, and storage capacity. Those challenges are considered addressed, which would arise when implementing resource-constrained devices. There are many works that study and develop federated learning in case of networks are resource unbounded, the detail can be found in the surveys of Li et al. [14], Yang et al. [54], Niknam et al. [55], Veličković et al. [23], Abbas et al. [56]. According to Li et al. [14], Yang et al. [54], Niknam et al. [55], Cui et al. [57], their surveys are mainly focused on FL implementation, architecture design, and components of FL. In addition, surveys on edge computing infrastructure and applications, including but not limited to machine learning, resource management, wireless communication, and security and privacy concerns, have been conducted as exemplified in research of Cui et al. [57, 58]. On the contrary, to the best of our knowledge [59], is the first and only survey on FL for resource-constrained IoT devices, the core challenges of applying FL may be faced such as heterogeneous hardware, fairness, source limited, scalability, communication overhead, scheduling, etc…To address these issues, the technique of Blockchain Federated Learning is utilized, as presented in [60, 61]. In summary, the drawbacks and potential solutions are described in Fig. 1.

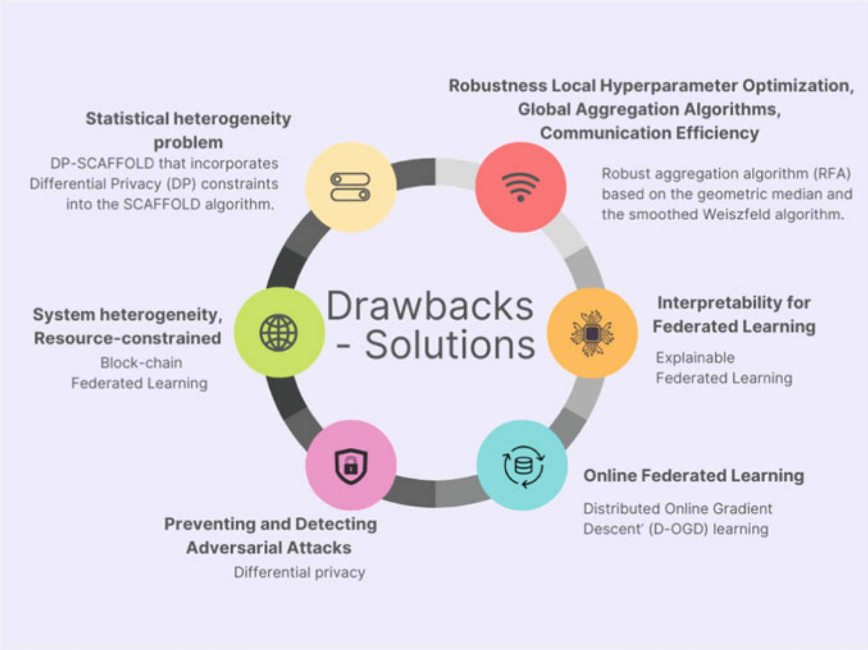


Fig. 1 The existing drawbacks and proposed solutions

5.2 *An End-to-End Efficient and Trustworthy Federated Learning-Based Explainable Anomaly Detection System*

This section relies on the aforementioned solutions and formulates them into an end-to-end architecture for the Federated Learning-based Explainable Anomaly Detection problem. The end-to-end system described in Fig. 2 consists of 5 Layer follow as,

- Layer 1: Asset/IoT Device—data source layer. Users' data is collected from three sources (i.e., user smartphones and IoT sensors). Obtained data in the form of demographic information and real-time technical information of users that constructs comprehend data structure—will be preprocessed before transferring to the next layer.
- Layer 2: Edge Layer—Local training data. The collected data is trained in FL Client, in which data is updated frequently by using Robust Graph Transformer Network—local model. In this layer, turning and optimizing techniques are also applied to support online learning. In addition, Explainable AI is utilized to enhance the transparency and trustworthiness of model decisions.
- Layer 3: IoT Communication Network: Blockchain monitors information sent from different FL clients from the edge layer through secure encryption. This

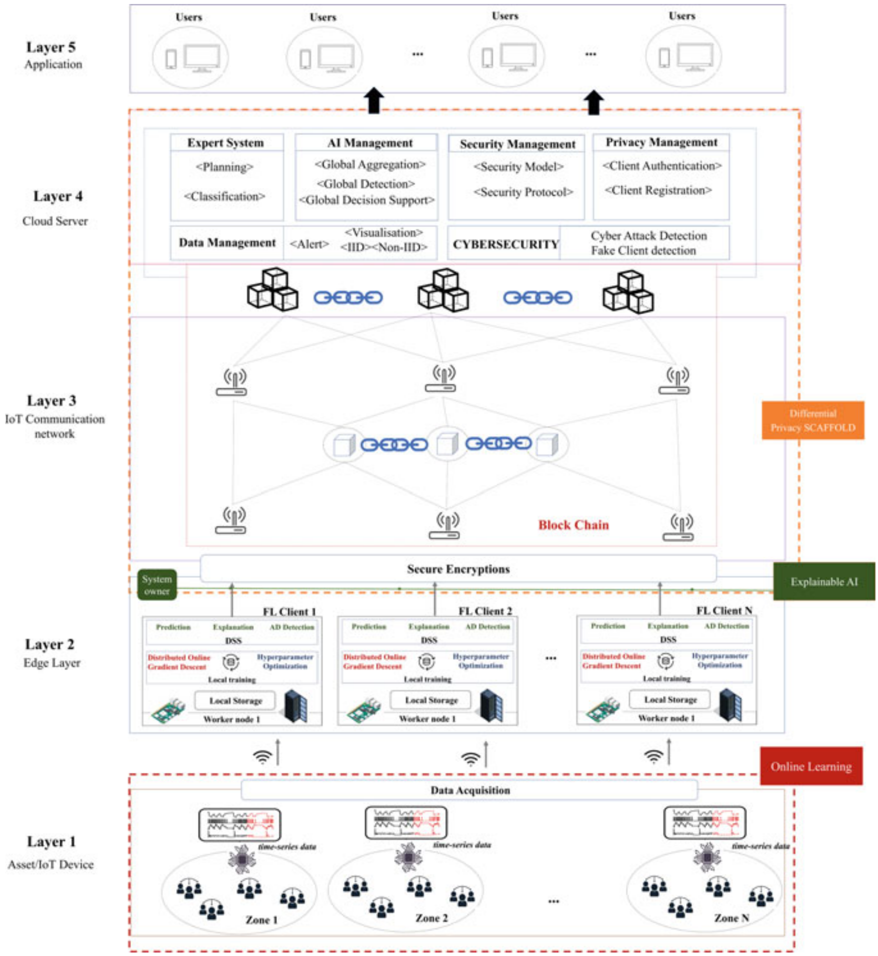


Fig. 2 Federated learning-based explainable anomaly detection

ability allows them not to be copied by any erroneous data. Increased security in the communication network between edge and cloud makes infrastructure more secure than the centralized processing model.

- Layer 4: Cloud Server: Perform centralized management tasks that provide data and compute platforms for applications in the healthcare system. Besides the database, container computation is applied to this layer. It offers flexible computing power and enough network bandwidth to handle the potentially growing data scale and workload of analytic tasks and security models.
- Layer 5: Application: User interfaces. These internal user interfaces will enable communications between users.

6 Future Aspects

In this section, we aim to provide a comprehensive outlook by introducing the future perspectives of analog federated learning and its intersection with quantum aspects.

6.1 Hybrid Digital-Analog Network Transmission for Federated Learning

Most of the existing work on FL research has focused on optimizing model volume without explicitly considering the impact of wireless channels implemented in various applications, for example, the Internet of Things (IoT), autonomous driving, health care, etc... To establish a comprehensive picture of research, it is critical to investigate the effect of the physical characteristics of the wireless medium (e.g., channel distortion, noise), especially in large-scale edge devices. Numerous digital communication-based technologies have recently been implemented to facilitate FL in wireless networks where each edge device is assigned an orthogonal channel to transfer model parameter information. However, due to the interference, distortion, and noise, the performance of FL algorithms across wireless channels is dramatically degraded, especially when the number of edge devices is large and in real condition. Besides, analog federated learning [62] is presented as a promising direction to tackle model performance degradation but often causes significant distortion due to the source signal's immense power. To eliminate those drawbacks and rectify the benefits of both directions, a hybrid digital-analog neural network transmission for federated learning based on [63] could be considered.

6.2 Quantum Computing and Quantum Machine Learning Perspective

In recent years, the evolution of quantum computer hardware and software has fueled an increase in the interconnection of quantum computing and machine learning [64]. Quantum computers are expected to address plenty of limitations that are beyond the capabilities of state-of-the-art, powerful classical computers or even supercomputers. Over the last three decades, advances in quantum computing have sparked significant interest in both academia and industry in the field of Manufacturing, Security, Drug discovery, Optimization, etc... [65]. In reality, advances in quantum hardware development have been critical for empirically assessing the true potential of quantum mechanic phenomena (i.e., Superposition, Entanglement, Parallelism) that accelerate the polynomial in both sources and timely execution. The quantum algorithms can achieve exponential speed-ups over classical approaches [66]. One typical example is one of Google's quantum computers that recently solved a classical computation challenge problem in 200 s that would have taken a classical computer 10,000 years to compute. So far, quantum methods for training neural networks, sampling, and optimization have mostly provided quadratic advantages, and some of these may be implementable on quantum computers.

Furthermore, Quantum machine learning (QML) is an interdisciplinary field that explores the interaction between quantum computing and machine learning. Specifically, it seeks to explore the potential benefits of applying techniques and results from one field to the other in order to solve complex problems. The domain of quantum machine learning investigates the development and implementation of specific quantum software solutions that can provide such benefits. Quantum computers employ phenomena such as quantum coherence and entanglement to perform computations in ways that are beyond the reach of classical computers. Over the last two decades, there has been consistent progress in developing more capable quantum computing systems. In the context of quantum computing, a quantum algorithm is a sequential set of operations executed on a quantum computer to address a specific computational task, such as searching a database [67]. Given that quantum systems are capable of producing counter-intuitive patterns that are not efficiently attainable by classical methods, it is reasonable to hypothesize that quantum computers may exhibit superior performance compared to classical computers in machine learning applications. Nevertheless, there exists a dearth of comprehension regarding the intricacies of quantum computing technology, its present-day capabilities, and its potential impact on communities. Addressing this knowledge gap necessitates obtaining a comprehensive understanding of how to evaluate the performance of quantum computing devices and estimate their potential. However, this undertaking is further compounded by the diverse range of quantum computing models and physical platforms in existence [65].

In recent times, there have been significant breakthroughs witnessed in both directions of influence. Specifically, quantum computing has emerged as a critical tool in facilitating speed-ups for machine learning problems, which holds immense importance in the current era of extensive data analysis. Conversely, machine learning has already pervaded a multitude of advanced technologies and is likely to play an instrumental role in the development of sophisticated quantum technologies. Furthermore, in addition to quantum speed-up in data analysis and classical machine learning optimization in quantum experiments, there have been theoretical demonstrations of quantum enhancements for interactive learning tasks, highlighting the potential of quantum-enhanced learning agents. Finally, research works exploring the utilization of artificial intelligence for the design of quantum experiments and for conducting specific components of genuine research autonomously have reported their initial successes [64].

Another future direction is quantum AD employing quantum machine learning, which has lately become a widespread issue with the development of quantum computers. Quantum computing is used to create quantum algorithms for machine learning. These algorithms are based on quantum matrices, phase estimation, and amplification. To process and communicate quantum information, especially over the cloud or Quantum Internet, anomalous quantum states must be detected because quantum

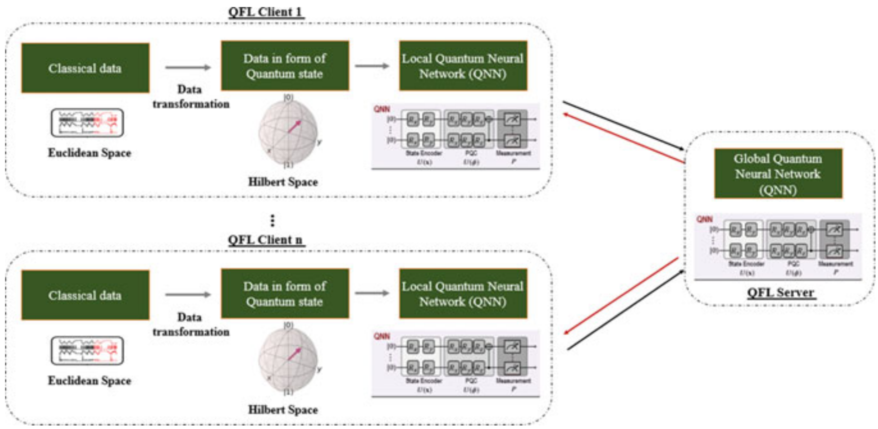


Fig. 3 Quantum federated learning methodology

data (like quantum states) is present in all areas of quantum computation, communication, and simulation. Motivated by the conventional federated learning algorithm, the quantum federated learning framework interconnects multiple quantum nodes with local data to train a global quantum neural network model that could be an innovation for a few decades ahead. The simulation system could be constructed on a cloud platform via a conventional computer (i.e., IBM Qiskit, Google Cirq, Amazon Bracket, Ocean D-Wave, Rigetti Forest, etc.). The methodology is described in Fig. 3. The overall process is analogous to conventional federated learning. First, classical data is preprocessed into quantum state form. Then, the local quantum model uses the processed data as input for the training process. In this step, the quantum neural network model is researched and developed to achieve high performance and accuracy. Next, the model weight parameters are sent and updated on the global FL quantum server.

7 Concluding Remarks

In this chapter, we discussed Federated learning-based Anomaly Detection, specifically for Smart Manufacturing in the new digital era in the future generation of industry, where systems become increasingly complex, producing massive data, necessitating more demand for transparency and performance. Federated learning has demonstrated its effectiveness in various applications; however, several challenges should be addressed, including communication costs, system and data heterogeneity, resource restrictions, security and privacy concerns, and hyperparameter optimization. In addition, several state-of-the-art approaches and methods to address the limitation of FL-based anomaly detection are also presented. After that, perspectives of combining solutions toward efficient and trustworthy anomaly detection systems based on Federated learning are shown through an end-to-end framework. Finally, to provide a complete picture, the quantum aspect is introduced in machine learning as a potential research direction.

References

1. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, PMLR, pp 1273–1282
2. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv (CSUR) 41:1–58
3. Huong TT, Dan NM, Hoang NX, Phung K-H, Tran KP et al (2023) Anomaly detection enables cybersecurity with machine learning techniques. In: Machine learning and probabilistic graphical models for decision support systems. CRC Press, pp 124–183
4. Han J, Kamber M, Pei J (2012) Data mining concepts and techniques, third edn. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University
5. Tran PH, Ahmadi Nadi A, Nguyen TH, Tran KD, Tran KP (2022) Application of machine learning in statistical process control charts: a survey and perspective. In: Control charts and machine learning for anomaly detection in manufacturing. Springer, pp 7–42
6. Tran KP (2022) Introduction to control charts and machine learning for anomaly detection in manufacturing. Springer
7. Truong HT, Ta BP, Le QA, Nguyen DM, Le CT, Nguyen HX, Do HT, Nguyen HT, Tran KP (2022) Light-weight federated learning-based anomaly detection for time-series data in industrial control systems. Comput Ind 140:103692
8. Truong HT, Ta BP, Le QA, Nguyen DM, Le CT, Nguyen HX, Do HT, Nguyen HT, Tran KP (2022) Light-weight federated learning-based anomaly detection for time-series data in industrial control systems. Comput Ind 140:103692
9. Raza A, Li S, Tran K-P, Koehl L (2022) Detection of poisoning attacks with anomaly detection in federated learning for healthcare applications: a machine learning approach. arXiv:2207.08486
10. Preuveneers D, Rimmer V, Tsingenopoulos I, Spooren J, Joosen W, Ilie-Zudor E (2018) Chained anomaly detection models for federated learning: an intrusion detection case study. Appl Sci 8:2663
11. Mothukuri V, Khare P, Parizi RM, Pouriyeh S, Dehghantanha A, Srivastava G (2021) Federated-learning-based anomaly detection for iot security attacks. IEEE Internet Things J 9:2545–2554
12. Zhao Y, Chen J, Wu D, Teng J, Yu S. Multi-task network anomaly detection using federated learning, in: Proceedings of the tenth international symposium on information and communication technology, pp 273–279
13. Cui L, Qu Y, Xie G, Zeng D, Li R, Shen S, Yu S (2021) Security and privacy-enhanced federated learning for anomaly detection in iot infrastructures. IEEE Trans Ind Inform 18:3492–3500
14. Li T, Sahu AK, Talwalkar A, Smith V (2020) Federated learning: challenges, methods, and future directions. IEEE Signal Process Mag 37:50–60
15. Liu Y, Yuan X, Xiong Z, Kang J, Wang X, Niyato D (2020) Federated learning for 6G communications: challenges, methods, and future directions, China. Communications 17:105–118
16. Zhang K, Song X, Zhang C, Yu S (2022) Challenges and future directions of secure federated learning: a survey. Front Comput Sci 16:1–8
17. Huong TT, Dan NM, Hoang NX, Phung K-H, Tran KP et al (2023) Anomaly detection enables cybersecurity with machine learning techniques. In: Machine learning and probabilistic graphical models for decision support systems. CRC Press, pp 124–183
18. Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, Bonawitz K, Charles Z, Cormode G, Cummings R et al (2021) Advances and open problems in federated learning. In: Found Trends® Mach Learn 14:1–210

19. Voita E, Talbot D, Moiseev F, Sennrich R, Titov I (2019) Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. arXiv:1905.09418
20. Nguyen T, Pham M, Nguyen T, Nguyen K, Osher SJ, Ho N (2022) Transformer with fourier integral attentions. arXiv:2206.00206
21. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I et al (2019) Language models are unsupervised multitask learners. OpenAI blog 1:9
22. Zhuang L, Wayne L, Ya S, Jun Z. A robustly optimized BERT pre-training approach with post-training. In: Proceedings of the 20th Chinese National conference on computational linguistics, pp 1218–1227
23. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. arXiv:1710.10903
24. Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT. Scaffold: stochastic controlled averaging for federated learning. In: International conference on machine learning, PMLR, pp 5132–5143
25. Acar DAE, Zhao Y, Navarro RM, Mattina M, Whatmough PN, Saligrama V (2021) Federated learning based on dynamic regularization. arXiv:2111.04263
26. Noble M, Bellet A, Dieuleveut A. Differentially private federated learning on heterogeneous data. In: International conference on artificial intelligence and statistics, PMLR, pp 10110–10145
27. Chen M, Shlezinger N, Poor HV, Eldar YC, Cui S (2021) Communication-efficient federated learning. Proc Natl Acad Sci 118:e2024789118
28. Konečnỳ J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D (2016) Federated learning: strategies for improving communication efficiency. arXiv:1610.05492
29. Braverman M, Garg A, Ma T, Nguyen HL, Woodruff DP. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In: Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, pp 1011–1020
30. Han Y, Özgür A, Weissman T. Geometric lower bounds for distributed parameter estimation under communication constraints. In: Conference on learning theory, PMLR, pp 3163–3188
31. Barnes LP, Han Y, Özgür A (2020) Lower bounds for learning distributions under communication constraints via fisher information. J Mach Learn Res 21:9583–9612
32. Tang H, Lian X, Qiu S, Yuan L, Zhang C, Zhang T, Liu J (2019) Deepsqueeze: parallel stochastic gradient descent with double-pass error-compensated compression. arXiv:1907.07346
33. Barnes LP, Inan HA, Isik B, Özgür A (2020) rTop-k: a statistical estimation approach to distributed SGD. IEEE J Sel Areas Inf Theory 1:897–907
34. Khodak M, Tu R, Li T, Li L, Balcan M-FF, Smith V, Talwalkar A (2021) Federated hyperparameter tuning: challenges, baselines, and connections to weight-sharing. Adv Neural Inf Process Syst 34:19184–19197
35. King RD, Feng C, Sutherland A (1995) Statlog: comparison of classification algorithms on large real-world problems. Appl Artif Intell Int J 9:289–333
36. Kohavi R, John GH (1995) Automatic parameter selection by minimizing estimated error. In: Machine learning proceedings 1995. Elsevier, pp 304–312
37. Pillutla K, Kakade SM, Harchaoui Z (2019) Robust aggregation for federated learning. arXiv:1912.13445
38. Huong TT, Bac TP, Ha KN, Hoang NV, Hoang NX, Hung NT, Tran KP (2022) Federated learning-based explainable anomaly detection for industrial control systems. IEEE Access 10:53854–53872
39. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144
40. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems, vol 30
41. Mishra S, Sturm BL, Dixon S. Local interpretable model-agnostic explanations for music content analysis. In: ISMIR, vol 53, pp 537–543

42. Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. In: International conference on machine learning, PMLR, pp 3145–3153
43. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618–626
44. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: an overview of interpretability of machine learning. In: 2018 IEEE 5th international conference on data science and advanced analytics (DSAA). IEEE, pp 80–89
45. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R et al (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115
46. Liu Z, Chen Y, Yu H, Liu Y, Cui L (2022) Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning. ACM Trans Intell Syst Technol (TIST) 13:1–21
47. Zhang C (2018) Online federated learning over decentralized networks. Ph.D. thesis. Nanyang Technological University
48. Liu P, Xu X, Wang W (2022) Threats, attacks and defenses to federated learning: issues, taxonomy and perspectives. Cybersecurity 5:1–19
49. Biggio B, Nelson B, Laskov P (2012) Poisoning attacks against support vector machines. arXiv:1206.6389
50. Bhagoji AN, Chakraborty S, Mittal P, Calo S. Analyzing federated learning through an adversarial lens. In: International conference on machine learning, PMLR, pp 634–643
51. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv:1412.6572
52. Shafahi A, Huang WR, Najibi M, Suciu O, Studer C, Dumitras T, Goldstein T (2018) Poison frogs! targeted clean-label poisoning attacks on neural networks. In: Advances in neural information processing systems, vol 31
53. Dwork C, McSherry F, Nissim K, Smith A (2006) Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography: third theory of cryptography conference, TCC 2006, New York, NY, USA, Proceedings 3. Springer, pp 265–284
54. Yang Q, Liu Y, Chen T, Tong Y (2019) Federated machine learning: concept and applications. ACM Trans Intell Syst Technol (TIST) 10:1–19
55. Niknam S, Dhillon HS, Reed JH (2020) Federated learning for wireless communications: motivation, opportunities, and challenges. IEEE Commun Mag 58:46–51
56. Abbas N, Zhang Y, Taherkordi A, Skeie T (2017) Mobile edge computing: a survey. IEEE Internet Things J 5:450–465
57. Cui L, Yang S, Chen F, Ming Z, Lu N, Qin J (2018) A survey on application of machine learning for internet of things. Int J Mach Learn Cybern 9:1399–1417
58. Mach P, Becvar Z (2017) Mobile edge computing: a survey on architecture and computation offloading. IEEE Commun Surv Tutor 19:1628–1656
59. Imteaj A, Thakker U, Wang S, Li J, Amini MH (2021) A survey on federated learning for resource-constrained IoT devices. IEEE Internet Things J 9:1–24
60. Li D, Luo Z, Cao B (2022) Blockchain-based federated learning methodologies in smart environments. Cluster Comput 25:2585–2599
61. Wang Z, Hu Q (2021) Blockchain-based federated learning: a comprehensive survey. arXiv:2110.02182
62. Xia S, Zhu J, Yang Y, Zhou Y, Shi Y, Chen W. Fast convergence algorithm for analog federated learning. In: ICC 2021-IEEE international conference on communications. IEEE, pp 1–6
63. Fujihashi T, Koike-Akino T, Watanabe T (2022) Federated airnet: Hybrid digital-analog neural network transmission for federated learning. arXiv:2201.04557
64. Dunjko V, Briegel HJ (2018) Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Rep Progr Phys 81:074001
65. Fedorov A, Gisin N, Beloussov S, Lvovsky A (2022) Quantum computing at the quantum advantage threshold: a down-to-business review. arXiv:2203.17181

66. Hughes C, Isaacson J, Perry A, Sun RF, Turner J (2021) Quantum computing for the quantum curious. Springer Nature
67. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. Nature 549:195–202

Multimodal Machine Learning in Prognostics and Health Management of Manufacturing Systems

Sagar Jose, Khanh T. P Nguyen, and Kamal Medjaher

Abstract Prognostics and health management (PHM) is a crucial enabler to reduce maintenance costs and enhance the availability and reliability of manufacturing systems. In the context of Industry 4.0, these systems become more complex and can be monitored by different types of sensors. The quality and completeness of data are crucial factors for the success of any PHM task in this paradigm. Here, we investigate the possibility of exploiting additional data sources in manufacturing besides monitoring sensors, e.g. production line cameras or maintenance reports. We first present the terminologies of multimodal learning and the potential it holds for industrial PHM. We then further explore the development and notable works in this field applied to other domains, look at the relevant works in PHM, and finally present a case study to demonstrate how multimodal learning can be performed to improve PHM processes.

Keywords Multimodal data · Multimodal learning · Prognostics and health management · Deep learning

1 Introduction

Maintenance in manufacturing comprises the actions taken during the life cycle of a production system to allow it to continue performing its intended function. Maintenance activity can be performed correctively (after failure) or preventively (before failure). Preventive maintenance can be either periodic or condition based [1]. A subset of condition based maintenance, called predictive maintenance, is the main

S. Jose (✉) · K. T. P. Nguyen · K. Medjaher
The National Engineering School of Tarbes - INP Toulouse, Production Engineering Laboratory, 65000 Tarbes, France
e-mail: sagar.jose@enit.fr

K. T. P. Nguyen
e-mail: thi-phuong-khanh.nguyen@enit.fr

K. Medjaher
e-mail: kamal.medjaher@enit.fr

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_9

focus of this chapter. Successful implementation of a predictive maintenance policy allows for reduction of maintenance costs and increasing availability and reliability of manufacturing systems. However, the effectiveness of predictive maintenance is significantly dependent on the development of a reliable process for prognostics of system failures.

Early studies of prognostics are based on developing a mathematical model to describe the system degradation processes, called model-based prognostics. Although model-based approaches can provide accurate long-term Remaining Useful Life (RUL) predictions, deriving models from real physical systems is very challenging due to the increasing complexity of the manufacturing systems in the light of Industry 4.0. In addition, the advanced sensing techniques and data analysis tools emerging from machine learning and deep learning disciplines enable the rise of data-driven prognostics, which is an alternative solution to overcome the limitations of model-based prognostics. The core principle of data driven methods is to use data—this includes but is not limited to past operation data of the machine—to derive the behavior function of the machine, and to identify parameters without first studying the physical principles describing the system behavior.

In literature, data-driven prognostics is a fast developing field [2] with numerous bench-marking datasets being developed [3]. The most common data used for data-driven PHM in industry include vibration, temperature [4–6], electric current [7, 8], sound [9, 10], pressure [4], speed [11, 12], and voltage signals [13]. Table 1 synthesizes the benchmark data sets which are mostly used in the field of PHM. One can see that most of the datasets contain uni-modal data, i.e., one dimensional numeric data. However, in the manufacturing industry, other data such as operation videos, operator reports, and more information are collected. Such data can become a valuable additional source to improve the performance of PHM process [14]. In fact, such multimodal data are widely used in the healthcare industry [15–17]. A detailed overview of the application of multimodal data in healthcare is given by Cai et al. [18]. Those references demonstrate the benefits of exploiting multimodal data to improve the efficiency of deep learning models. In this view, it is necessary to investigate the question of how to seek supplementary sources of manufacturing information to complement the machine health indicators obtained from sensors and therefore improve PHM performance. The potential for using multimodal data for PHM purposes, such as RUL prediction, was demonstrated by the study by Yang et al. [14]. However, the complexity of these data in terms of structure, as well as the requirements of high computation resources, pose considerable challenges that need to be solved before these data can be exploited to support the RUL prediction.

Although both academia and industry pay great attention to mining multimodal data for improving PHM performance, to our best knowledge, none of the existing papers provide a comprehensive review of multimodal learning in PHM and an insightful investigation of its performance. Therefore, this chapter aims to fill this literature gap. In Sect. 2, we introduce the terminologies of the field. In Sect. 3, we present a brief chronological walkthrough of the development of the multimodal learning discipline, look at its impact in some related fields, discuss some of the main challenges in the domain, and identify the main tools and techniques to handle

Table 1 Benchmark datasets for PHM

Dataset name	Data type	Purpose
CWRU bearing dataset [19]	• Drive end accelerometer data • Fan end accelerometer data • Base accelerometer data	• Motor bearing condition assessment • Fault diagnosis
Tennessee eastman process dataset [20]	• Reactor pressure • Reactor level • Reactor temperature • Stripper level • Stripper pressure • and many other measurements	• Fault detection
SEU bearing dataset [21]	• Vibration signals • Fault positions	• Fault detection
NASA bearing dataset [22]	• Vibration signals	• Anomaly detection • RUL prediction
PHM2012 data challenge dataset [23]	• Vibration signals • Temperature	• RUL prediction
Airbus Helicopter accelerometer dataset [24]	• Vibration signals	• Anomaly detection • Fault detection
Numenta anomaly benchmark [25]	• Artificially generated numerical data	• Anomaly detection
NASA turbofan dataset (CMAPPS) [26]	• Total temperature at fan inlet • Total temperature at LPC outlet • Total temperature at HPC outlet • Total temperature at LPT outlet • Pressure at fan inlet • And other numerical data	• Anomaly detection

multimodal data challenges. In Sect. 4, a brief review of the existing works using multimodal data for PHM is discussed. Then, in Sect. 5, we look at a case study. Finally, we conclude the chapter in Sect. 6.

2 Terminologies of Multimodal Machine Learning

2.1 Modality

The word modality has multiple definitions. The first comes from the word ‘mode’, which refers to the point of maximum frequency in a distribution. The term multimodal in this space refers to a population distribution that has multiple local maxima in the probability density function [27].

Another definition refers to the way information is perceived and understood [28]. This definition of modality is more relevant to our study. This is illustrated by the way our human brain receives information from the world and processes it into an understanding of a scenario [29]. In detail, we perceive the world through our five senses (sight, hearing, taste, smell and touch). These are the sensory modalities.

Further, in the context of computing, modality of data refers to the structure in which a computer program receives the data and the way the data are processed to gain knowledge [30]. In computing, the most common modalities are vision, audition, language, proprioception, haptics, and so on.

2.2 Multimodality

Multimodality refers, in the context of information and data, to the existence of multiple modalities in the same set of data [31, 32]. A key concept to understand here is that a dataset is called multimodal when it contains information of multiple modalities to describe features of the same function.

An example comes from the study of communicative behaviors. In-person communication between people consists of three types of communicative behaviors: verbal, vocal and visual. It is important to understand that even as a person is speaking verbally, information can be conveyed at the same time through vocal expression such as intonation, laughter, etc. [33]. Visual information such as gestures, body language and expressions add to the information. Also to be noted is that within the verbal modality are features such as the lexicon (choice of words), the choice of grammatical structures, and so on.

This leads to an important idea: multiple modalities of data can serve one of two purposes. It can either reinforce the information conveyed through one modality, or it can provide complementary information.

2.3 Multimodal Versus Multimedia

There is much overlap between the usage of the terms multimodal and multimedia. Multimedia data is data including media data types such as text, images, video, audio, drawings, and so on. Multimodal data can include also non-media data such as proprioception, point clouds, etc. In summary, multimedia data can be considered a subset of multimodal data.

2.4 Multimodal Versus Heterogeneous Data

Heterogeneous data refer to data that differ in some property. Among the possible differences, one is structural heterogeneity. For our purpose of studying data processing in PHM, structurally heterogeneous data are the same as multimodal data.
However, the two terms multimodal data and heterogeneous data are different in some particular contexts. For example, if two sets of data have the same structural representation format, but differ in their population distribution, the term "heterogeneous" is more relevant than the term "multimodal". Particularly, data coming from two sensors, e.g., temperature and pressure, that have similar numerical structure, are considered as unimodal data.

The definitions and approaches to understanding the term multimodality have been compiled by Parcalabescu et al. [34]. As seen so far, the definitions of multimodal data and multimodality are not conclusive in the literature yet. However, in this study, the definition of multimodal data as structurally heterogeneous data is preferred.

2.5 Multimodal Learning

Multimodal learning is defined as an activity of extracting useful knowledge from multimodal data, while giving due consideration to cross-modal influences. Learning from multiple modalities is important because the information in the real world often involves more than one modality.

In fact, in a dataset containing data of different modalities, one modality could carry information that is not available from the other modality. An example is an image of a city with its caption mentioning the name of the city [35]. Without the textual information, the name of the city could be hard or impossible to deduce from the image alone.

3 Overview of Multimodal Machine Learning Studies

This section aims to give an overview of multimodal machine learning. It begins with a brief look at the historical evolution of multimodal learning in Sect. 3.1, and the impact of multimodal data in different scientific fields, particularly life sciences in Sect. 3.1.1 and robotics in Sect. 3.1.2. Then, Sect. 3.2 briefly discusses the challenges involved with multimodal learning. Finally, Sect. 3.3 explores the existing tools and techniques in the literature to handle the aforementioned challenges.

This section studies multimodal learning from a global perspective with the purpose of giving the reader a general understanding of the field. This would pave the way towards a focused discussion on multimodal learning for PHM later in Sect. 4.

3.1 Brief Literature Review of the Evolution of Multimodal Machine Learning

In the literature, the evolution of multimodal learning is seen to be chronologically separated into four time periods [36, 37]: 1970–1980, 1980–2000, 2000–2010, and after 2010.

In fact, the attempt to combine information from multiple modalities originated from the discipline of behavioral studies, in the 1970s. This consisted of the fields of psychology and linguistic studies. For illustration, Blank [38] studied the connection between linguistic development of children, sensorimotor skills, and visual spatial information. Several of the early multimodal studies were concerned with understanding linguistic development in the early childhood period [39, 40].

A new paradigm rose in the mid-1980s, when studies into the processing of multiple modalities of data via a computational approach were attempted [41, 42]. This trend was followed by developments in affective computing with an increasing focus on more abstract tasks such as emotion recognition [43]. Being close after the end of the second Artificial Intelligence (AI) winter [44], this period was marked by renewed attempts at asking questions such as the role of emotions in the development of a truly intelligent computer. It led to an increase in interest in affective computing [45], and in turn the interest in multimodal learning. It was also during this period that multimedia computing questions were asked, such as the possibility of searching within the content of a video, for example. Chang et al. [46] proposed a method to search the contents of a video with segmented object tracking and spatiotemporal queries instead of keyword based queries.

Another major shift in the research trend occurred near 2000, when the focus was given to studying not just understanding the communication of one subject, but studying the interaction between multiple people. In the Handbook of Virtual Environments [47], the tradeoff between added value from multimodal data and cross-modal effects, and the increase in computational complexity and cost in human-computer interaction systems is discussed. By the end of this period, Zara et al. [48] presented a protocol for collecting and annotating a dataset of multimodal human-human interactions in a game context. This indicates the shift towards the latest era of multimodal learning, the deep learning era.

Near 2010, the latest trend began to emerge where this field, along with many others, were looked through the lens of neural architectures. This was the beginning of the deep learning era. Ngiam et al. [49] demonstrated the use of deep neural networks to learn features from both audio and video, and proved that neural representations effectively facilitate extracting useful features from both modalities. Srivastava et al. [35] used a deep Boltzmann Machine to create fused representations of bi-modal image-text and audio-video data. Xu et al. [50] proposed an extended use of attention mechanism [51] to perform cross-modality attention for image caption generation.

This sudden shift from multimodal research to multimodal machine/deep learning is due in large part to the following factors. Firstly, the creation and free sharing of new large scale multimodal datasets. Easy availability of cheap data storage and ease of sharing data through the internet contributed to this. Faster computers and GPU (Graphical Processing Unit) development enabled researchers and developers to implement deep neural networks and train them on large datasets. These two factors—availability of data and high computing capacity—have been touted as the reasons for the renewal of neural architectures in general [44]. The third reason is that very high dimensional data such as vision and language could now be represented in a uniform neural encoding in the form of vectors. In case of vision, the success of convolutional networks in representing features [52] was an influential milestone for deep learning. Then, vision and language being two modalities around which a large part of multimodal research was oriented, these advancements were key to the rise in research trend on multimodal deep learning.

As the scientific development of multimodal learning evolved, the interest and therefore the production of research volume in this field also rose rapidly. In the rest of this section, we will look quantitatively at the development of multimodal learning. As shown in Fig. 1, with a remarkable success of deep learning methods in the field of computer vision and natural language processing, the interest in the field of multimodal learning has risen rapidly in the recent years.

When investigating the fields where most of the work in multimodal learning is done, one can see that a large part of it is in computer science and AI research. The breakdown of the fields is shown in Fig. 2.

Looking at the distribution shown in Fig. 2, it can be noted that other than computer science and AI domains, a significant share of the work is done in medical science and related fields. It can be inferred that most of the work in computer science and AI would involve the development of algorithms and tools for working

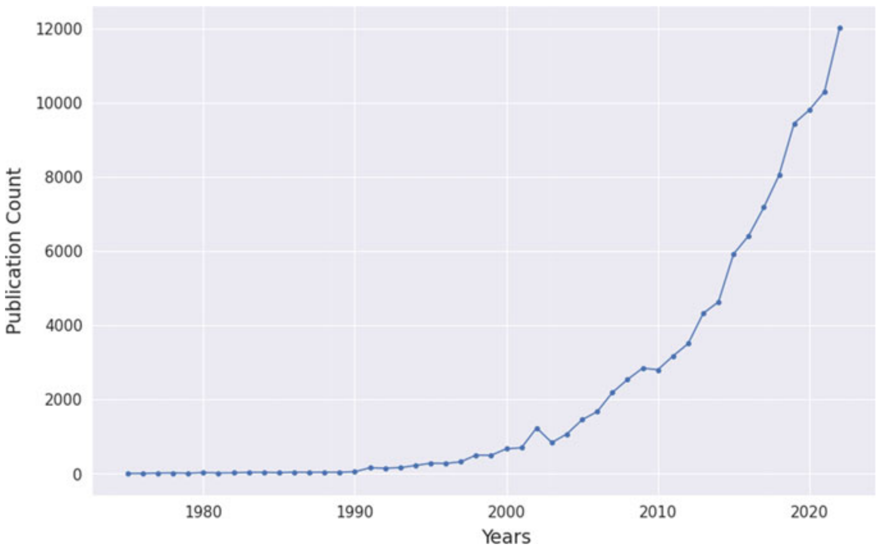


Fig. 1 Trend of publications on multimodal learning over the years

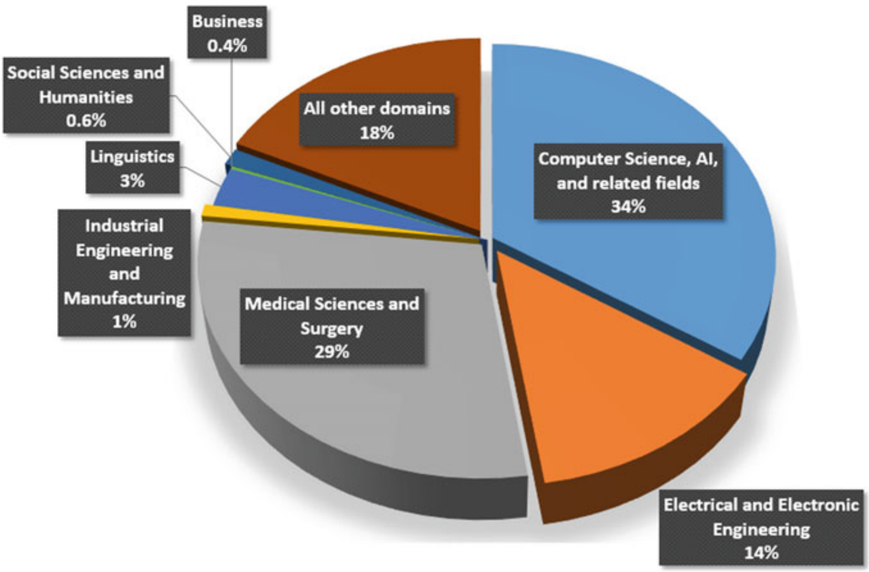


Fig. 2 Main domains where multimodal research is conducted

on multimodal data, whereas research on medical sciences and robotics would be application oriented. In the next subsections, we take a global look at the existing works in those fields that apply multimodal data to solve their specific problems.

3.1.1 Multimodal Learning in Life Sciences, Medicine, and Related Fields

From an intuitive point of view, certain analogies can be drawn between healthcare and industrial maintenance. In healthcare, the health state of a human being is observed and treatments are administered as and when necessary to prolong his or her life in the best condition. This is similar to the health management activity of manufacturing systems in industry. Therefore, observations made from studying the application of multimodal data in healthcare could potentially be exploited to apply multimodal data in PHM.

The use of multimodal data along with machine learning techniques is gaining importance in life sciences, medical science, psychology and other related fields. The associated science is progressing at a rapid pace, with several reviews published every year. Some of the notable papers include [53–56].

Multimodal data in these disciplines not only include medical imaging, data from various scans such as computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and so on, but also omics data, clinical data such as various measurements, demographic information, real time signals such

Table 2 Comparison of early, feature level and late fusion

Comparison of fusion strategies			
Scenario	Early fusion	Feature fusion	Late fusion
Prediction without all modalities	×	×	✓
Feature level interaction	✓	✓	×
Cross modal compatible feature extraction	×	✓	×
Training on sparse data	×	×	✓
Training on only one model	✓	✓	×
Ease of model design	✓	×	×
Input concatenation at different abstraction levels	×	✓	×

Adapted from [54]

as electrocardiogram (ECG), electroencephalogram (EEG), etc. Depending on the type of disease or condition, other types of data can also be available.

Spasov et al. [57], Yala et al. [58] and Yoo et al. [59] used CNNs for medical images and fused the learned features with clinical records to identify a medical condition. Spasov et al. [57] and Yala et al. [58] used simple concatenation to fuse the multimodal data. Yoo et al. reported duplicating the clinical information to solve the dimensionality difference problem between features from image and clinical records.

In [54], the authors synthesized fusion techniques used according to characteristics of the problem to solve. The findings compiled in Table 2 are not limited to medical data and can be potentially used for industrial maintenance as well.

Cao et al. [60] discussed the use of Auto-GAN to synthesize data and address the problem of data sparsity. Li et al. [61] introduced a GAN for retinal disease diagnosis with multimodal images. Hervella et al. [62] used a U-Net for retinal vessel segmentation using multimodal data. Chen et al. [63] used an attention based method for prognosis of breast cancer from omics and clinical data. This work is particularly interesting to our study because it discusses the design of an architecture for prognostics of the future health state of the system under study.

Maghdid et al. [64] introduced transfer learning with X-Ray and CT images from a network trained on pneumonia data to detect Covid-19. Lassau et al. [65] used a deep learning model to extract features from CT images, and then concatenated them with lab tests and other clinical data to input to a logistic regression model for predicting case severity of Covid-19 patients. A notable observation to be made here is that deep learning methods are hard to replace when image modality is involved. A detailed overview of deep learning architectures that have been used with multimodal data in medicine is given in [56].

Another less explored method is modality translation. This involves translating the information from one modality to another (ex: image caption generation), and then treating all the data as the same modality. Wang et al. [66] introduced TieNet in which radiology images are converted to language embedded reports by converting the image modality to text. Though this application is not directly analogous to the problem of PHM, the technique and approach used can potentially be adapted. A prerequisite of most of the methods of this approach is that paired data from two modalities are necessary to train the translator model.

One significant advantage provided by the comparative maturity of multimodal study in a field such as medicine is the availability of more datasets. Notable datasets include:

- MIMIC-CXR dataset [67] containing 227,835 imaging studies for 65,379 patients along with free-text radiology reports.
- PADCHEST dataset [68] containing chest X-rays with multi-label annotated reports.
- ImageCLEF challenges [69] datasets containing image and text for multimodal information retrieval.

From this section, it can be concluded that the use of multimodal data is thriving in life sciences and medical fields, and this is one of the drivers of multimodal deep learning research. The techniques identified in this field could be adapted to other fields, particularly PHM, with promising results.

As discussed in the beginning of this section, healthcare and industrial maintenance can be compared in certain aspects. Therefore, the increasing use of multimodal data in healthcare provides a promising perspective of multimodal learning in industrial health management.

3.1.2 Multimodal Learning in Robotics, Affective Computing and Other Domains

Multimodal data are particularly important for human robot interaction, where the visual, auditory, language, and proprioception modalities, at the least, have to be combined. Even though the scale of this field cannot be compared to the medical science domain, the scientific advancements made here are significant. A comprehensive review has been made by Spezialetti et al. [70]. This study, which focuses on emotion recognition for human robot interaction, is closely tied to affective computing. Data such as thermal facial images and brain activity signals were studied.

In literature, several works demonstrate the use of CNN type networks on image data. Barros et al. [71] used a cross channel CNN to extract features from face expression and body motion data. Val-Calvo et al. [72] also used a CNN variant for emotion recognition from facial images, EEG, Galvanic Skin Response (GSR) and blood pressure. Filintsis et al. (2019) implemented a ResNet and DNN joint network to fuse facial images and body posture information.

Robot manipulation task failures are studied by Inceoglu et al. [73], where the authors present a multimodal dataset comprising of RGB images, depth images, and audio from robots. The dataset is then used to train a multimodal neural network to detect incomplete or failed task scenarios. The network design is particularly inspiring for fault detection tasks in PHM domain. In the network structure proposed by the authors, called FINO-Net, data from comparable modalities such as RGB and depth images are stacked on top of each other and input to the same convolutional path. A separate path for audio data begins with a log mel spectrogram rendering of the audio data which converts the audio into mel frequency spectral coefficient representation. Features from this representation are input into a convolution block. The separate paths are later fused with a dense layer. This philosophy of designing individual paths suited for the treatment of each data modality and fusing the features near the end decision level can be adapted to PHM purposes, as will be shown in Sect. 5.2.

In addition, two noteworthy advancements in multimodal deep learning are given below:

- **CLIP**: Introduced by OpenAI, Contrastive Language-Image Pretraining (CLIP) is a method of learning visual models with natural language supervision [74]. This work is scientifically significant for its discovery of neurons that respond identically to image and text representations of the same concept. Once pre-trained with this method, models were seen to be able to efficiently transfer to other visual classification tasks without need for fine tuning.
- **Flamingo**: Introduced by Deepmind, is a family of visual language models that are made to adapt to numerous visual language tasks with only few annotated examples [75]. It is an example of few shot learning. The capacity of Flamingo models include both open ended text tasks such as visual question answering and closed tasks such as multiple choice question answering.

3.2 Challenges of Multimodal Machine Learning

In this section, we look at the core technical and scientific challenges that arise when we attempt to perform machine learning or deep learning on multimodal data. According to the studies [37, 76, 77], one can cite five principal challenges of multimodal learning: representation, alignment, fusion, co-learning, and translation. Among them, representation and alignment are crucial challenges that need to be solved in any task involving multimodal deep learning. The other three challenges are not common to all multimodal deep learning problems, but depend on the particular problem addressed.

3.2.1 Representation

This is the challenge of joining data from multiple modalities in some uniform representation space. The two main approaches to solve this challenge is to either create a joint representation, or a coordinated representation. In the joint representation approach, the multiple modalities of data are all transformed to a different representation, which is suited to represent the combined information from all of them.

In the coordinated representation approach, each modality is transformed to its own representation, and a coordination spectrum mapping (strong coordination to weak coordination) is defined.

The bimodal deep belief network [49] developed by Ngiam et al. [49] performs the joint representation approach effectively. They demonstrated that by representing both images and text as vectors, it is possible to perform arithmetic operations on image and text. Other notable examples include the deep Boltzmann Machine for image captioning by Srivastava et al. [35], and the audio-visual emotion recognition by Kim et al. [9].

Representation is a tradeoff problem, where the optimal solution tries to minimize the information loss from each of the data modalities. With multimodal data, the challenge is to learn how to summarize and represent the data in a way that exploits the complementarity and redundancy, as needed.

3.2.2 Alignment

Alignment is the problem of identifying direct relations between elements or sub-elements from two or more different modalities. This problem is very important to temporal multimodal data, because synchronicity between modalities is difficult to achieve. While several factors contribute to this, the differences in sample collection rate, sequence length and so on between data from different modalities are crucial.

Explicit alignment, where the task is to directly find correspondences between elements of different modalities, are done by techniques such as deep canonical time warping [78]. An example of use case is event reconstruction from multiple partial video, text and audio descriptions.

Implicit alignment is the internal latent alignment of modalities that is done as an intermediate step on the way to solving another problem. In deep learning, it is necessary to investigate if it is possible to "encourage" a model to align the data when solving a problem. One approach to tackle this challenge is by using context information methods, such as attention mechanism [51].

3.2.3 Fusion

In the case of fusion, the challenge is to bring different modalities together in order to infer some information at a higher level of abstraction. An example is to take the visual, audio, and language modalities from a video of a person speaking, and infer a higher level information such as emotion of the speaker.

In multimodal data, the challenge is to know at what level to fuse one modality with another. Taking the example of image and text, the raw information level for an image is a pixel, and for text, is a word. But fusing at the level of pixels and words does not give useful features. In the case of images, useful features emerge after multiple levels of convolution and pooling.

The difficulty lies in learning at what level a feature is mature or insightful enough to benefit from the information from another modality. If done too soon, the other modality may just become noise. If the fusion is done too late, it may not be suited to the task at hand. For example, in some cases, feature fusion may be more useful than decision fusion.

The fusion techniques can be categorized in two groups: model-agnostic and model-based approaches [37]. For model-agnostic approaches, one can refer to the review by D'Mello et al. [79]. Model-based approaches include deep neural networks [49, 80], kernel based methods [81], and graphical methods [82].

3.2.4 Co-learning

Co-learning is the challenge of transferring learning between modalities, including their representations and predictive models. To illustrate, a model may be trained with data from multiple modalities. The challenge is to train the model such that it is able to perform its intended task even if it is only given one modality as input at test time. This is important because, in many cases, a model may be able to learn useful information from one modality when it is trained, but that modality may not be available at use time, particularly when data in the test time are limited.

Analogical to coordinated representation, co-learning can also be done in two approaches, where in one approach a strong pairing is defined between two datasets, and in another only a weak pairing is considered. The close relationship between representation and co-learning is demonstrated by Pham et al. [83] where the authors demonstrate learning joint representations by cyclic translations between modalities.

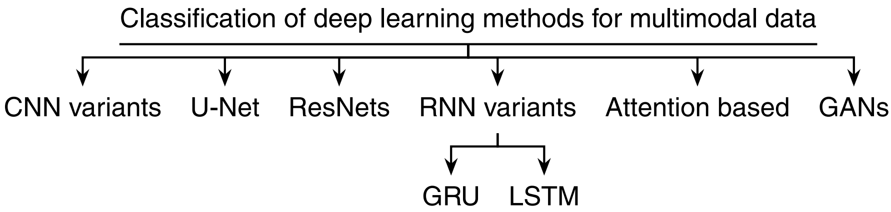
3.2.5 Translation

It is the task of translating data from one modality to another. An example is generating caption given an image. In this case, the objective function may be a text generator [84].

The translation can be example-based, where a mapping is made from source examples to targets. A classic method is the nearest neighbor approach. Another approach in translation is model-based, where a model learns rules to generate a translation. There is significant overlap between the two approaches, as sometimes the model rules can be learned by example-based learning. An example of multimodal translation is presented by Ahuja et al. [85] where the authors propose a model to forecast the pose of a person from language.

3.3 Tools and Techniques in Multimodal Deep Learning

This section summarizes the most commonly used tools and techniques in multimodal deep learning to solve the challenges described above. The classification is based on the principle behind the network mechanism.



- CNN variants: CNNs are among the best architectures to work on image data. The architecture consists of a convolution operation followed by a pooling and, usually, a fully connected layer. CNNs can capture spatially correlated features in an image, and also from any data that can be represented as an image [86].
- U-Net: The U-Net architecture was primarily developed for image segmentation [87]. It consists of a convolutional "shrinking" path where an image is shrunk to a small dimension, followed by a deconvolutional "expanding" path where the small dimension feature representation is expanded back to the original large dimension. U-Net has been used in multimodal applications in medicine where segmentation is required [62].
- ResNets: ResNet stands for residual network [88]. It was originally developed to allow neural networks to be very deep without causing the vanishing or exploding gradient problems. This was done by adding residual or skip connection, which is a parallel connection between one layer and the layer after the next one, skipping the layer in between. ResNets have since been used to create many pretrained networks, which have then been used for transfer learning with multimodal data [89].
- RNN variants: RNNs are a class of neural networks that work best to model sequence data. Therefore, this is particularly useful for PHM, where the majority of data are temporal sequences [90]. The basic principle of RNN is to parse items in an input series one after the other, while updating a hidden state that stores the history of what it has seen before. In this way, RNNs succeed to capture the sequential relationship in the data.
  However, RNNs are not very good at capturing long sequence data. Gated Recurrent Unit (GRU) [91] and Long Short Term Memory (LSTM) [92] are extensions of RNN which overcome some of its disadvantages.
- Attention based: Attention mechanism was originally implemented to solve the natural language translation problem [51]. It is based on the premise of human visual attention. To put it simply, attention mechanism assigns weights on the data depending on how important a piece of data is to the task at hand. Attention

based methods, particularly transformer attention and variants, have succeeded in achieving state of the art performance in both image and text related tasks.
- GANs: Generative Adversarial Network was created as a new method of training a generative network [93]. A GAN consists of two neural networks, called the generator and the discriminator. The generator trains to generate new data that are similar to the training dataset. The discriminator trains to distinguish the 'fake' data generated by the generator. As both generator and discriminator are trained, the output from the generator begins to more accurately resemble the original dataset, thereby generating believable data. Variants of GANs have been used in modality translation tasks and for data synthesis [60].

4 Review of Multimodal Machine Learning in PHM

In this section, we look at the existing studies in PHM that use multimodal data. This section is divided into subsections based on PHM purposes such as fault detection and diagnostics, prognostics, and prescriptive maintenance. To the best of our knowledge, the studies referenced in this section are all the existing works that use multimodal data to solve a PHM problem.

4.1 Multimodal Machine Learning in Fault Detection and Diagnostics

Fault detection and diagnostics (FDD) is typically performed after a fault has occurred in the system. Fault detection involves finding if an anomaly occurred, fault isolation aims to identify where exactly the fault occurred, and diagnostics involves analysing why it happened.

Table 3 presents the existing works that use multimodal data to solve fault detection and diagnostics. One can see that almost all studies investigate a combination between numerical and image data for multi-modal learning, and propose a deep learning model to fuse data for fault detection, isolation and diagnostics.

4.2 Multimodal Machine Learning in Prognostics

Prognostics is the activity of projecting the health state of a machine or system into the future. This projection is used to anticipate failures and take proactive actions as needed. Prediction of remaining useful life (RUL) of a machine is one of the key activities in prognostics. Table 4 presents the existing works that use multimodal data for prognostics. We observe that all deep learning based works use CNN in

Table 3 Research on multimodal learning for fault detection and diagnostics

Problem	Method/tool	Data	Application
Data fusion for fault diagnosis [94]	M-CNN	Vibration, IR images	Rotor system
Data fusion for fault diagnosis [94]	M-ResNet-DCA	Vibration, IR images	Rotor system
Network fault isolation [95]	LSTM	Network metrics, customer complaints	IPTV network
Data fusion for fault diagnosis [96]	RBM-AE	Electric signals, images	Power transformers and circuit breakers
Data fusion for diagnosis [97]	DNN, AE, CNN	Vibration, image of vibration signal	Bearing platform
Data fusion for fault detection [98]	MLP, CNN, GRU	Temperature, Operation details	Plastic molding

Table 4 Research on multimodal learning for prognostics

Problem	Method/tool	Data	Application
Data fusion for RUL prediction [14]	CNN, MLP	Inspection records, signal images, maintenance history	Steam generator
Data fusion for RUL prediction [98]	CNN-LSTM, ResNet-28	Process parameters, tool images	Machining
Data visualization [99]	ccPCA, ccMCA, UMAP + DBSCAN	Network metrics, operation parameters, machine status description	Maintenance log analysis
Wear condition prognosis [100]	CNN, RNN	Process parameters, tool images	Cutting tool

combination with other architectures. It should also be mentioned that the study [99] does not use deep learning methods, but instead studies visualization and clustering of maintenance data to support preventive maintenance.

4.3 Multimodal Machine Learning for Prescriptive Maintenance

Prescriptive maintenance is the term given to all activities related to decision support in maintenance. This goes beyond studying the health state of the system, and extends to specifying what actions can be taken for optimal maintenance. Table 5 presents the existing works that use multimodal data for prescriptive maintenance. Digital twin and associated methods appear to be the dominant methods when observing the table. The papers presented in Table 5 define frameworks for the maintenance process and

Table 5 Research on multimodal learning for prescriptive maintenance

Problem	Method/tool	Data	Application
Decision support framework design [101]	Digital shadow	Maintenance records, machine parameters	Cyber physical production systems
Failure prediction for decision support [102]	Dynamic Bayesian networks	Maintenance records, machine parameters	Cyber physical production systems
Maintenance framework design [103]	Digital twin, rule based model	(Only concept presented in the paper)	Refurbishment of industrial equipment

specify suitable techniques for each part of the respective framework. However, it should be noted in advance that the papers do not present case studies with failure prediction mechanism. On the contrary, the methods are only recommended as a potentially suitable solution for the prediction task in the framework presented in the papers.

5 Detailed Example of Multimodal Machine Learning in PHM of Manufacturing Systems

5.1 Case Study Description

The case study consists of a simulated dataset on 50 steam generators (SG) used in power plants. This dataset was introduced in the paper by Yang et al. [14]. It contains the following data:

- images of WRL signals, which indicate degradation
- text records from inspections
- numerical data such as:
  - time from last maintenance
  - time till next inspection
  - number of performed chemical cleanings
  - number of mechanical cleanings after last chemical cleaning
  - machine degradation level.

Beside photograph examinations and eddy current tests, technicians perform periodic inspections on the steam generators and provide textual descriptions of the health condition of the unit based on their experience.

Two failure mechanisms are simulated in this dataset: tube support plate clogging and tube fouling. Both are caused by the deposit of impurities. The wide range level (WRL) signal shown in the image is a measure of the pressure difference between the top and bottom of the SG downcomer.

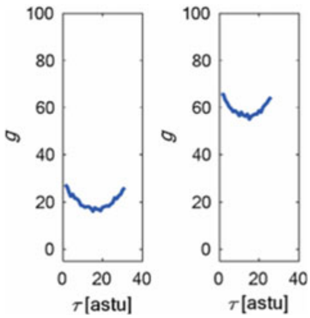


Fig. 3 Images of the WRL signal. Adapted from [14]

A snapshot of the WRL signal in the form of image can be seen in Fig. 3.

There are two types of maintenance: chemical and mechanical cleaning. Mechanical cleaning is an imperfect maintenance procedure, and it is not capable of removing all the impurities from the SG. Each mechanical cleaning is less efficient than the one before. However, this procedure can be performed any number of times. Chemical cleaning, on the other hand, is an almost perfect maintenance. However, it is costly, time consuming, and could damage some parts of the steam generator. It can only be performed a fixed number of times.

5.2 *Proposed Methodology*

Based on the numerical, text and image data, a method is proposed to predict the degradation level at next inspection. It consists of two training steps, as shown in Fig. 4. In the initial step, we train one neural network each to predict the degradation with only a single modal input. In detail, we train a convolutional network on the image, another network on the text, and a third one on the numerical data. In the next step, we connect the output layer of all three networks while freezing all layers except for the final fully connected layer. Then we train this dense layer. When the training is finished, the network will be able to predict the degradation of the SG with all three modalities of information.

5.2.1 Training on Image Data

The first branch of the proposed model aims to predict the degradation level at next inspection time from only the image of WRL signals. Its architecture based on a convolutional neural network is shown in Fig. 5.

At the input of this model, the images are presented in a matrix of pixels, where a pixel stores the value of the color at the position of the pixel in the image. Convolution is used in image processing due to its capability of extracting the spatial features in

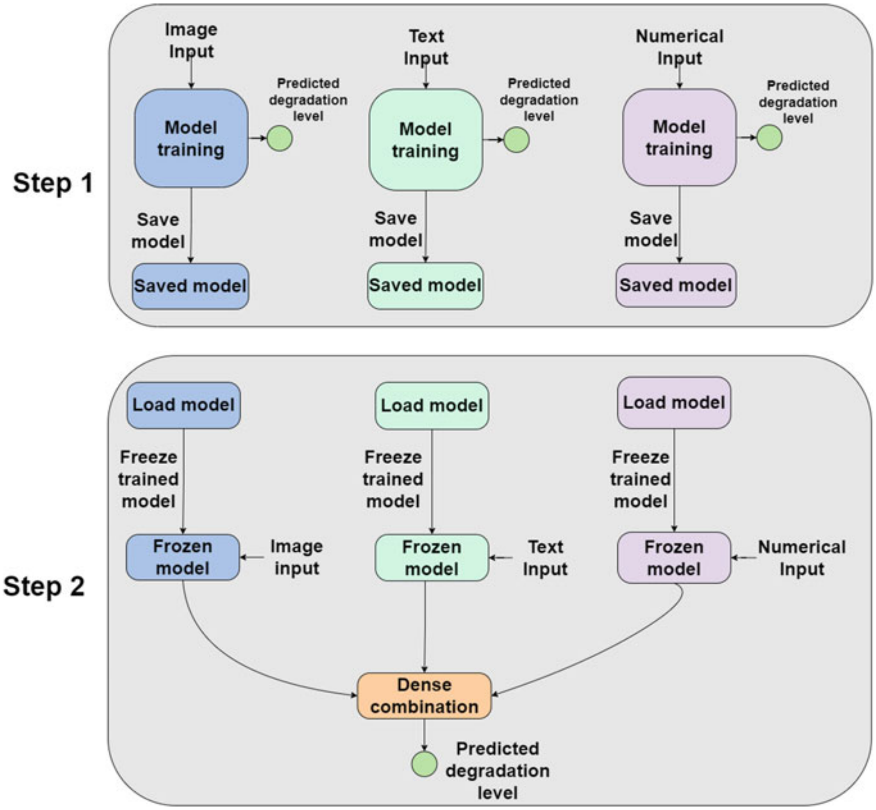


Fig. 4 Proposed multimodal learning method

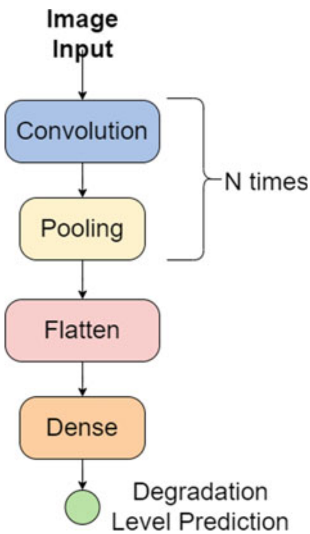


Fig. 5 Image branch network

the image. It is the operation of passing a small matrix over the larger image matrix while performing a mathematical operation at each stride of the smaller matrix. This latter one is called a convolution window. Its values are designed to extract a feature from the image [104]. The output of the calculation performed between the smaller matrix and the region of the image that it is laid over indicates the presence or absence of the feature that the window is meant to extract. An example of convolution can be seen in Fig. 6.

Next, pooling is the operation performed to reduce the dimension of the convolution output. There are several types of pooling methods. In this case study, maxpooling is used, which is simply selecting the highest value in a region (a group of pixels) and assigning that value to represent that region. Finally, flatten and dense layers allow to transform a 3D-tensor at the output of pooling layer to a prediction value (1D tensor) of degradation level at the model output.

5.2.2 Training on Text Data

Figure 7 presents the proposed architecture for text-processing branch. At the input, an embedding layer is used to transform text data into matrix format, which can be treated as an image. Word embedding can be considered as a class of techniques to create a numerical representation of words. The embedding layer allows learning the most appropriate feature representation of words implicitly while solving the supervised NLP (natural language processing) task.

In this model, one-hot-encoding technique is used to embed the words in numerical format. It converts each word into a feature vector of a size k. These features are learned by the network to produce the most suited word embedding for solving the problem. The output of the word embedding layer is a matrix of dimension $N \times k$,

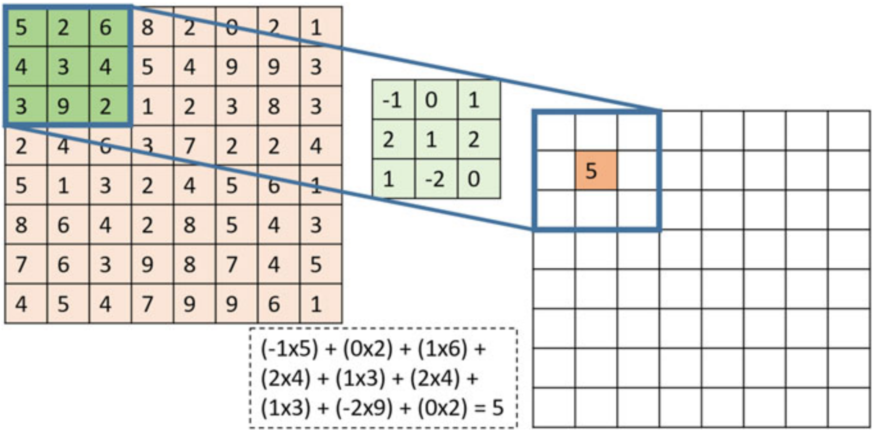


Fig. 6 Convolution example

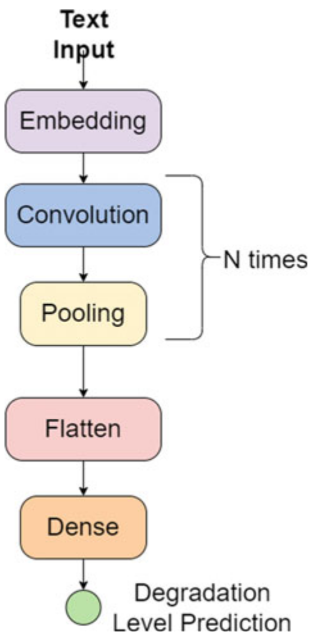


Fig. 7 Text branch network

where N is the vocabulary size. For a large text corpus, this can become a very huge matrix where the majority of values are zeros.

Based on the supervised task that the whole network is learning, words that have similar meanings will have similar (small cosine distance) feature vectors. Thus, the word embedding layer converts the text in the data to numerical representation while not losing word similarity information. While one-hot-encoding of each word can serve as an elementary method to convert text to numerical representation, word embedding attempts to capture the relationship between similar words in the training vocabulary, as well as reduce the computational inefficiency of one hot encoding. For more information about word embedding, readers can refer to [105].

5.2.3 Training on Numerical Data

Numerical data in this dataset are simple enough in terms of their underlying function such that they can be treated with a multilayer perceptron. A fully connected layer, or a dense layer, first multiplies the input with a weight matrix and then adds a bias vector. The architecture of the numerical branch is shown in Fig. 8.

One can note that in the proposed model, no recurrent memory model have been used to train on the data, even though an RNN or LSTM would be standard choices for time-series data. This is because the methodology design at that level is done to accommodate the later combination step. At combination time, computation is faster if none of the three branches are recurrent.

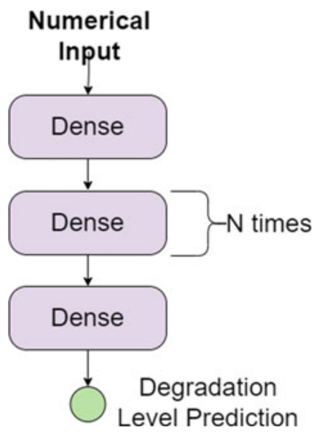


Fig. 8 Numerical branch network

There are several ways of combining multiple neural networks. Here, we present the simplest of them—combination of the output of each of the network branches. To do this, we freeze the trained network branches and add a dense connection to join their outputs as shown in Sects. 5.2.4 and 5.2.5.

5.2.4 Combining Two Modalities

In this section, we attempt to use two modalities together. In this initial experiment, we freeze the trained models on image (Fig. 5) and on text (Fig. 7), and connect their outputs to a dense layer. Then, the dense layer is the only layer whose weights are trained in this phase. The set up is shown in Fig. 9.

Besides the combination between image and text data, the combination between image-numerical data and text-numerical data is also investigated. Their combination approach is similar to the one presented in Fig. 9.

5.2.5 Combining All Three Modalities

In this subsection, a combination of three data modalities is investigated. A dense layer is added at the outputs of three network branches (Fig. 10). It is trained on all three of the data modalities, so that it will learn to assign the appropriate weights to each data branch.

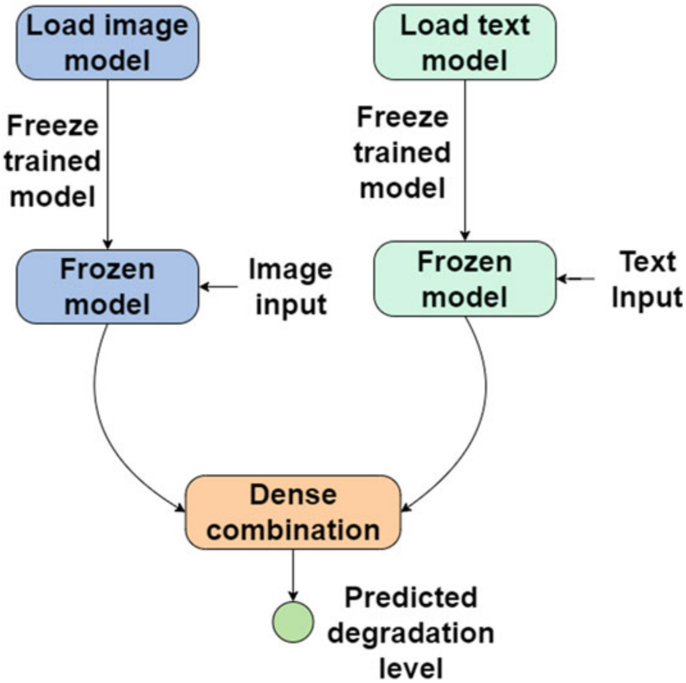


Fig. 9 Combining image and text branches

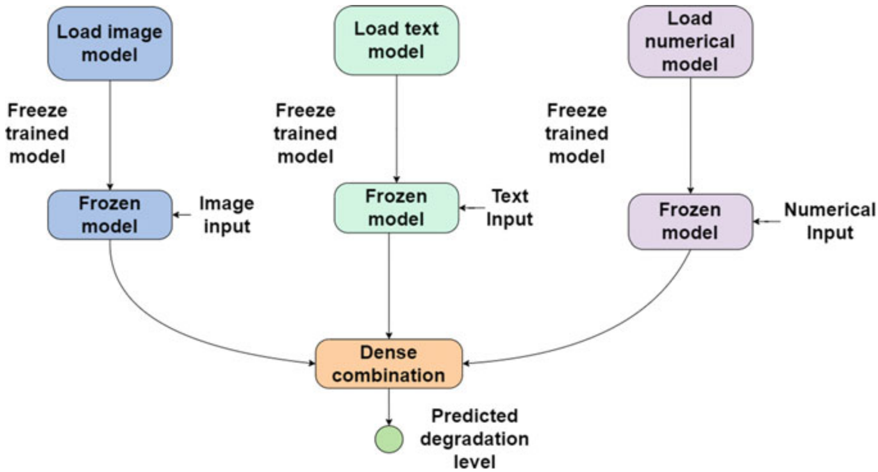


Fig. 10 Combined network for multimodal learning

Table 6 MAE of training on each modality in the dataset

Input data modality	MAE (ours)	MAE ([14])
Image	5.37	16.74
Text	3.83	4.91
Numerical	6.62	8.31
Image + Text	3.56	6.90
Image + Numerical	4.34	7.74
Text + Numerical	2.66	3.66
Image + Text + Numerical	2.52	3.20

5.3 *Result Discussion*

Table 6 presents the performances of the proposed models when using unimodal data and when exploiting multimodal data. The mean absolute error (MAE) of the predicted degradation level is calculated for each network branch presented above and for each combination. Two comparisons are made in this table. Firstly, the performance of the proposed method on using each data modality is compared with the performance of the method on combinations of data modalities. In detail, the performance of three single modality models, three two-modality combinations, and one three-modality combination will be investigated. Secondly, the results of the proposed method is compared with the results of the existing work [14].

One can observe that the proposed method gives better results than the existing ones in all cases. In addition, the performance of multimodal learning has been highlighted when compared to the results of unimodal learning. In detail, the mean absolute error obtained when combining two data modalities together is better than when only one modality is used. Furthermore, we see that the network trained on all three data modalities has the best performance. Those results give weight to the claim that multimodal data can improve model performance and can better support PHM activities.

6 Conclusion and Perspectives

In this chapter, a new and rising trend for data-driven PHM approach in industry was studied. The concept of multimodal data and their properties, their advantages, and the challenges involved when using them were introduced. Then the chapter looked at the evolution of multimodal learning, how multimodal data are exploited in other domains, and how they are used in PHM. The next part of the chapter exploited a case study and reviewed a methodology to use multimodal data for prediction of machine degradation. Those results highlighted that using multiple modalities of data allows improving the performance of the prediction model. However, for effectively exploiting multimodal data, it is necessary to pay attention to the following challenges:

- Finding a suitable representation for all the modalities. This is mostly solved by using a neural representation. It is not a simple "one size fits all" solution, but instead requires some careful engineering to implement. However, several studies have shown that a neural representation is a sufficient method that can represent the information in multiple modalities;
- Identifying the best method of fusion to suit the problem and the data;
- Aligning event indicator signals from the multiple modalities;
- Training a model with multiple modalities, but enabling its predictability in the absence of one or more modalities;
- Enhancing model interpretability. While model interpretability is a hard problem in deep learning, it is very relevant in PHM, where it is crucial to know the origin of the signal(s) that led to the reasoning behind the model's prediction.

As multimodal deep learning has its own domain, and its applications in other fields are maturing, these challenges will become less unsurmountable. Therefore, the use of multimodal data is an avenue of exploration that holds much potential for the future development of industrial maintenance as a discipline.

Future research directions should include establishing a clearer picture on how multimodal data can be used for PHM. Furthermore, while the case study presented in this chapter explores a simple yet basic method to fuse multimodal data, experimentation is required to develop more sophisticated methods that better take advantage of cross modal learning.

Another future direction is the development of methods or tools for explaining the behaviour of multimodal machine learning models. Ultimately, the purpose of developing such tools in the PHM domain is to provide decision support for maintenance activities, which can be costly. Therefore, the ability to explain the operation of the model is crucial for trust. In multimodal models, this problem has multiple axes. On one hand, the interactions between multiple modalities make the system behaviour more complex. The more intertwined the network structure, the more difficult it is to isolate the origin of behavioural deviations. On the other hand, the availability of multiple modalities may offer unexpected avenues of explainability, such as attention visualization for images, model behaviour observation in the absence a modality, and so on.

The chapter can be concluded with a final recall of the advantages of using multimodal data for PHM. The advantage brought by using multimodal data depends on the way it is used. The most common mode of usage is to consider the multiple modalities as additional sources of information. This could be supplementary or complementary information. A failure mode of a machine may not have any response in one modality, but may have easily recognizable indications in another modality. Another situation is when the input data are not available in sufficient quantities to train a model. In this case, a model may be trained on a more easily available modality of data, and then transferred to the data that are actually available for online use of the model.

In the near future, it is reasonable to expect much increase in the interest on multimodal data for industrial maintenance. Therefore, it is vital to allocate sufficient research efforts to solve the pertinent challenges.

References

1. Gouriveau R, Medjaher K, Zerhouni N (2016) From prognostics and health systems management to predictive maintenance 1: monitoring and prognostics, vol 4
2. Xu G, Liu M, Wang J, Ma Y, Wang J, Li F, Shen W (2019) Data-driven fault diagnostics and prognostics for predictive maintenance: a brief overview. In: 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), pp 103–108. https://doi.org/10.1109/COASE.2019.8843068, ISSN: 2161-8089
3. Jia X, Huang B, Feng J, Cai H, Lee J (2018) A review of PHM data competitions from 2008 to 2017 (2018)
4. Zhao P, Kurihara M, Tanaka J, Noda T, Chikuma S, Suzuki T (2017) Advanced correlation-based anomaly detection method for predictive maintenance. In: 2017 IEEE International Conference on Prognostics and Health Management (ICPHM). https://doi.org/10.1109/ICPHM.2017.7998309
5. Falk C, van de Sand R, Corasaniti S, Reiff-Stephan J (2021) A comparison study of data-driven anomaly detection approaches for industrial chillers. In: TH Wildau Engineering and Natural Sciences Proceedings 1 (2021). https://www.tib-op.org/ojs/index.php/th-wildau-ensp/article/view/33. https://doi.org/10.52825/thwildauensp.v1i.33. Accessed 23 May 2022
6. Yan K, Ji Z, Shen W (2017) Online fault detection methods for chillers combining extended Kalman filter and recursive one-class SVM. Neurocomputing 228:205–212. https://doi.org/10.1016/j.neucom.2016.09.076
7. Tian J, Azarian MH, Pecht M (2014) Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. In: PHM society European conference, vol 2, no 1. https://papers.phmsociety.org/index.php/phme/article/view/1554. https://doi.org/10.36001/phme.2014.v2i1.1554
8. Hendrickx K, Meert W, Mollet Y, Gyselinck J, Cornelis B, Gryllias K, Davis J (2020) A general anomaly detection framework for fleet-based condition monitoring of machines. Mech Syst Signal Process 139:106585. https://doi.org/10.1016/j.ymssp.2019.106585. arXiv:1912.12941 [cs, eess, stat]
9. Lu G, Liu J, Yan P (2018) Graph-based structural change detection for rotating machinery monitoring. Mech Syst Signal Process 99:73–82
10. Lu G, Zhou Y, Lu C, Li X (2017) A novel framework of change-point detection for machine monitoring. Mech Syst Signal Process C 533–548. https://doi.org/10.1016/j.ymssp.2016.06.030
11. Pittino F, Puggl M, Moldaschl T, Hirschl C (2020) Automatic anomaly detection on in-production manufacturing machines using statistical learning methods. Sensors 20(8):2344. https://doi.org/10.3390/s20082344 (Multidisciplinary Digital Publishing Institute)
12. Schlechtingen M, Santos I (2011) Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mech Syst Signal Process 25:1849–1875. https://doi.org/10.1016/j.ymssp.2010.12.007
13. Bzymek A (2017) Application of selected method of anomaly detection in signals acquired during welding process monitoring. Int J Mater Prod Technol 54:249–258
14. Yang Z, Baraldi P, Zio E (2021) A multi-branch deep neural network model for failure prognostics based on multimodal data. J Manuf Syst 59:42–50. https://doi.org/10.1016/j.jmsy.2021.01.007

15. Tekin C, Atan O, Van Der Schaar M (2015) Discover the expert: context-adaptive expert selection for medical diagnosis. IEEE Trans Emerg Top Comput 3:220–234. https://doi.org/10.1109/TETC.2014.2386133
16. Yoon J, Davtyan C, van der Schaar M (2016) Discovery and clinical decision support for personalized healthcare. IEEE J Biomed Health Inform 21:1133–1145
17. Rahimi SA, Jamshidi A, Ruiz A, Aï-Kadi D (2016) A new dynamic integrated framework for surgical patients' prioritization considering risks and uncertainties. Decis Support Syst **88**:112–120
18. Cai Q, Wang H, Li Z, Liu X (2019) A survey on multimodal data-driven smart healthcare systems: Approaches and applications. IEEE Access 7:133583–133599. https://doi.org/10.1109/ACCESS.2019.2941419
19. Download a data file | case school of engineering | case western reserve university (2021). https://engineering.case.edu/bearingdatacenter/download-data-file. Accessed 23 May 2022
20. Chen X (2019) Tennessee Eastman simulation dataset. https://doi.org/10.21227/4519-z502
21. Shao S (2022) Mechanical-datasets. https://github.com/cathysiyu/Mechanical-datasets, original-date: 2018-01-16T19:12:43Z
22. Lee J, Qiu H, Yu G, Lin J (2007) Rexnord technical services, bearing data set, IMS, university of Cincinnati. In: NASA AMES prognostics data repository, NASA Ames, Moffett Field, CA
23. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Morello BC, Zerhouni N, Varnier C (2012) PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In: IEEE International conference on prognostics and health management, PHM'12., Denver, Colorado
24. Sas A (2020) Airbus helicopter accelerometer dataset. https://www.research-collection.ethz.ch/handle/20.500.11850/415151. https://doi.org/10.3929/ethz-b-000415151, accepted: 2020-05-19T12:16:26Z publisher: ETH Zurich type: dataset
25. Ahmad S, Lavin A, Purdy S, Agha Z (2017) Unsupervised real-time anomaly detection for streaming data. Neurocomputing 262:134–147. https://doi.org/10.1016/j.neucom.2017.04.070
26. Saxena A, Goebel K (2008) Turbofan engine degradation simulation data set. In: NASA Ames prognostics data repository, pp 1551–3203
27. Silverman BW (1981) Using kernel density estimates to investigate multimodality. J R Stat Soc: Ser B (Methodol) 43:97–99
28. Leahy W, Sweller J (2011) Cognitive load theory, modality of presentation and the transient information effect. Appl Cogn Psychol 25:943–951
29. Norris S (2019) Systematically working with multimodal data: research methods in multimodal discourse analysis. Wiley
30. Lahat D, Adali T, Jutten C (2015) Multimodal data fusion: an overview of methods, challenges, and prospects. Proc IEEE 103:1449–1477
31. Caesar H, Bankiti V, Lang AH, Vora S, Liong VE, Xu Q, Krishnan A, Pan Y, Baldan G, Beijbom O (2020) Nuscenes: a multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11621–11631
32. Chen C, Jafari R, Kehtarnavaz N (2015) Utd-mhad: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In: IEEE international conference on image processing (ICIP). IEEE 2015:168–172
33. Tsiourti C, Weiss A, Wac K, Vincze M (2017) Designing emotionally expressive robots: a comparative study on the perception of communication modalities. In: Proceedings of the 5th international conference on human agent interaction, pp 213–222
34. Parcalabescu L, Trost N, Frank A (2021) What is multimodality? arXiv:2103.06304
35. Srivastava N, Salakhutdinov RR (2012) Multimodal learning with deep boltzmann machines. Adv Neural Inf Proc Syst 25. https://papers.nips.cc/paper/2012/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html. Accessed 31 May 2021
36. Morency L-P, Liang PP, Zadeh A (2022) Tutorial on multimodal machine learning. In: Proceedings of the 2022 conference of the North American chapter of the association for computational linguistics: human language technologies: tutorial abstracts. Association for Computational Linguistics, Seattle, United States, pp 33–38. https://aclanthology.org/2022.naacl-tutorials.5. https://doi.org/10.18653/v1/2022.naacl-tutorials.5

37. Baltrušaitis T, Ahuja C, Morency L-P (2018) Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell 41:423–443
38. Blank M (1974) Cognitive functions of language in the preschool years. Dev Psychol 10:229
39. Roeper T, McNeill D (1973) Review of child language. Ann Rev Anthropol 2:127–137
40. Keller-Cohen D (1978) Context in child language. Ann Rev Anthropol 7:453–482
41. McNeill D (1985) So you think gestures are nonverbal? Psychol Rev 92:350–371. https://doi.org/10.1037/0033-295X.92.3.350 (American Psychological Association, US)
42. Butterworth B, Hadar U (1989) Gesture, speech, and computational stages: a reply to McNeill
43. Picard RW (2000) Affective computing. MIT Press. Google-Books-ID: GaVncRTcb1gC
44. Toosi A, Bottino AG, Saboury B, Siegel E, Rahmim A (2021) A brief history of AI: how to prevent another winter (a critical review). PET Clinics 16:449–469
45. Vesterinen E et al (2001) Affective computing. In: Digital media research seminar, Helsinki, Citeseer
46. Chang S-F, Chen W, Meng HJ, Sundaram H, Zhong D (1998) A fully automated content-based video search engine supporting spatiotemporal queries. IEEE Trans Circuits Syst Video Technol 8:602–615
47. Popescu GV, Burdea GC, Trefftz H (2022) Multimodal interaction modeling. In: Handbook of virtual environments. CRC Press, pp 475–494
48. Zara A, Maffiolo V, Martin JC, Devillers L (2007) Collection and annotation of a corpus of human-human multimodal interactions: emotion and others anthropomorphic characteristics. In: International conference on affective computing and intelligent interaction. Springer, pp 464–475
49. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY (2011) Multimodal deep learning. In: ICML
50. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: neural image caption generation with visual attention. In: International conference on machine learning, PMLR, pp 2048–2057
51. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30
52. LeCun Y, Haffner P, Bottou L, Bengio Y (1999) Object recognition with gradient-based learning. In: Shape, contour and grouping in computer vision. Springer, pp 319–345
53. Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, Tavares JMR, Bradley A, Papa JP, Belagiannis V et al (2018) Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, volume 11045. Springer, Granada, Spain
54. Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP (2020) Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med 3:1–9
55. Heiliger L, Sekuboyina A, Menze B, Egger J, Kleesiek J (2022) Beyond medical imaging—A review of multimodal deep learning in radiology. https://www.techrxiv.org/articles/preprint/Beyond_Medical_Imaging_-_A_Review_of_Multimodal_Deep_Learning_in_Radiology/19103432/1. https://doi.org/10.36227/techrxiv.19103432.v1 (TechRxiv)
56. Behrad F, Abadeh MS (2022) An overview of deep learning methods for multimodal medical data mining. Expert Syst Appl 117006
57. Spasov SE, Passamonti L, Duggento A, Lio P, Toschi N (2018) A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease. In: Annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Annual international conference 2018, pp 1271–1274. https://doi.org/10.1109/EMBC.2018.8512468, PMID: 30440622
58. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R (2019) A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 292:60–66. https://doi.org/10.1148/radiol.2019182716. PMID: 31063083

59. Yoo Y, Tang LY, Li DK, Metz L, Kolind S, Traboulsee AL, Tam RC (2019) Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome. Comput Methods Biomechan Biomed Eng: Imaging Vis 7:250–259
60. Cao B, Zhang H, Wang N, Gao X, Shen D (2020) Auto-GAN: self-supervised collaborative learning for medical image synthesis. In: Proceedings of the AAAI conference on artificial intelligence, vol 34, no 07, pp 10486–10493. https://doi.org/10.1609/aaai.v34i07.6619
61. Li X, Jia M, Islam MT, Yu L, Xing L (2020) Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis. IEEE Trans Med Imaging 39:4023–4033. https://doi.org/10.1109/TMI.2020.3008871
62. Hervella ÁS, Rouco J, Novo J, Ortega M (2019) Self-supervised deep learning for retinal vessel segmentation using automatically generated labels from multimodal data. In: 2019 international joint conference on neural networks (IJCNN). IEEE, pp 1–8
63. Chen H, Gao M, Zhang Y, Liang W, Zou X (2019) Attention-based multi-NMF deep neural network with multimodality data for breast cancer prognosis model. BioMed Res Int 2019:e9523719. https://doi.org/10.1155/2019/9523719 (Hindawi)
64. Maghdid HS, Asaad AT, Ghafoor KZ, Sadiq AS, Khan MK (2020) Diagnosing COVID-19 pneumonia from x-ray and CT images using deep learning and transfer learning algorithms. Tech Rep. http://arxiv.org/abs/2004.00038
65. Lassau N, Ammari S, Chouzenoux E, Gortais H, Herent P, Devilder M, Soliman S, Meyrignac O, Talabard M-P, Lamarque J-P et al (2021) Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. Nat Commun 12:1–11
66. Wang X, Peng Y, Lu L, Lu Z, Summers RM (2018) TieNet: text-image embedding network for common thorax disease classification and reporting in chest x-rays. Tech Rep. http://arxiv.org/abs/1801.04334, https://doi.org/10.48550/arXiv.1801.04334
67. Johnson A, Pollard T, Mark R, Berkowitz S, Horng S (2019) Mimic-cxr database. PhysioNet 10:13026 (C2JT1Q)
68. Bustos A, Pertusa A, Salinas J-M, de la Iglesia-Vayá M (2020) Padchest: a large chest x-ray image dataset with multi-label annotated reports. Med Image Anal 66:101797
69. Abacha AB, Hasan SA, Datla VV, Liu J, Demner-Fushman D, Müller H, VQA-Med: overview of the medical visual question answering task at ImageCLEF 2019. CLEF (Working Notes) 2
70. Spezialetti M, Placidi G, Rossi S (2020) Emotion recognition for human-robot interaction: Recent advances and future perspectives. Front Robot AI 7. https://www.frontiersin.org/article/10.3389/frobt.2020.532279. Accessed 01 June 2022
71. Barros P, Weber C, Wermter S (2015) Emotional expression recognition with a cross-channel convolutional neural network for human-robot interaction, pp 582–587. https://doi.org/10.1109/HUMANOIDS.2015.7363421
72. Val-Calvo M, Álvarez-Sánchez JR, Ferrández-Vicente JM, Fernández E (2020) Affective robot story-telling human-robot interaction: exploratory real-time emotion estimation analysis using facial expressions and physiological signals. IEEE Access 8:134051–134066
73. Inceoglu A, Aksoy EE, Ak AC, Sariel S (2021) Fino-net: a deep multimodal sensor fusion framework for manipulation failure detection. In: 2021 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 6841–6847
74. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. arXiv:2103.00020
75. Alayrac J-B, Donahue J, Luc P, Miech A, Barr I, Hasson Y, Lenc K, Mensch A, Millican K, Reynolds M et al (2022) Flamingo: a visual language model for few-shot learning. arXiv:2204.14198
76. Gao J, Li P, Chen Z, Zhang J (2020) A survey on deep learning for multimodal data fusion. Neural Comput 32:829–864. https://doi.org/10.1162/neco_a_01273
77. Gaw N, Yousefi S, Gahrooei MR (2021) Multimodal data fusion for systems improvement: a review. IISE Trans 1–19

78. Trigeorgis G, Nicolaou M, Zafeiriou S, Schuller B (2016) Deep canonical time warping, pp 5110–5118. https://doi.org/10.1109/CVPR.2016.552
79. D'mello SK, Kory J (2015) A review and meta-analysis of multimodal affect detection systems. ACM Comput Surv (CSUR) 47:1–36
80. Wöllmer M, Kaiser M, Eyben F, Schuller B, Rigoll G (2013) LSTM-modeling of continuous emotions in an audiovisual affect recognition framework. Image Vis Comput 31:153–163
81. Liu F, Zhou L, Shen C, Yin J (2013) Multiple kernel learning in the primal for multimodal alzheimer's disease classification. IEEE J Biomed Health Inform 18:984–990
82. Lafferty J, McCallum A, Pereira FC (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data
83. Pham H, Liang P, Manzini T, Morency L-P, Poczos B (2019) Found in translation: learning robust joint representations by cyclic translations between modalities. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 6892–6899. https://doi.org/10.1609/aaai.v33i01.33016892
84. Marsella S, Xu Y, Lhommet M, Feng A, Scherer S, Shapiro A (2013) Virtual character performance from speech. In: Proceedings—SCA 2013: 12th ACM SIGGRAPH/Eurographics symposium on computer animation. https://doi.org/10.1145/2485895.2485900
85. Ahuja C, Morency L-P (2019) Language2pose: natural language grounded pose forecasting, pp 719–728. https://doi.org/10.1109/3DV.2019.00084
86. Zhang Y, Wallace B (2016) A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. Tech Rep. arxiv:1510.03820
87. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation, arXiv:1505.04597
88. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence
89. Zhang T, Shi M (2020) Multi-modal neuroimaging feature fusion for diagnosis of alzheimer's disease. J Neurosci Methods 341:108795. https://doi.org/10.1016/j.jneumeth.2020.108795
90. Pascanu R, Mikolov T, Bengio Y (2013) On the difficulty of training recurrent neural networks. arXiv:1211.5063
91. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling
92. Staudemeyer RC, Morris ER (2019) Understanding LSTM—A tutorial into long short-term memory recurrent neural networks. arXiv:1909.09586
93. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets, vol 27. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html. Accessed 01 June 2022
94. Yuan Z, Zhang L, Duan L (2018) A novel fusion diagnosis method for rotor system fault based on deep learning and multi-sourced heterogeneous monitoring data. Meas Sci Technol 29:115005. https://doi.org/10.1088/1361-6501/aadfb3 (IOP Publishing)
95. Kao H-Y, Wang Y-Y, Huang C-M, Hsu C-P (2019) Heterogeneous data ensemble learning in end-to-end diagnosis for IPTV. In: 2019 20th Asia-pacific network operations and management symposium (APNOMS), pp 1–6. https://doi.org/10.23919/APNOMS.2019.8892990. ISSN:2576-8565
96. Ma Y, Guo Z, Su J, Chen Y, Du X, Yang Y, Li C, Lin Y, Geng Y (2014) Deep learning for fault diagnosis based on multi-sourced heterogeneous data. Int Conf Power Syst Technol 2014:740–745. https://doi.org/10.1109/POWERCON.2014.6993854
97. Zhou F, Yang S, He Y, Chen D, Wen C (2021) Fault diagnosis based on deep learning by extracting inherent common feature of multi-source heterogeneous data. Proc Inst Mech Eng Part I: J Syst Control Eng 235:1858–1872. https://doi.org/10.1177/0959651820933380 (IMECHE)
98. Marei M, Li W (2021) Cutting tool prognostics enabled by hybrid CNN-LSTM with transfer learning. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-021-07784-y

99. Zhang X, Fujiwara T, Chandrasegaran S, Brundage M, Sexton T, Dima A, Ma K-L (2021) A visual analytics approach for the diagnosis of heterogeneous and multidimensional machine maintenance data. https://doi.org/10.1109/PacificVis52677.2021.00033
100. Wang P, Liu Z, Gao RX, Guo Y (2019) Heterogeneous data-driven hybrid machine learning for tool condition prognosis. CIRP Annal 68:455–458. https://doi.org/10.1016/j.cirp.2019.03.007
101. Ansari F, Glawar R, Nemeth T (2019) Prima: a prescriptive maintenance model for cyber-physical production systems. Int J Comput Integr Manuf 32:482–503. https://doi.org/10.1080/0951192X.2019.1571236
102. Ansari F, Glawar R, Sihn W (2020) Prescriptive maintenance of CPPS by integrating multimodal data with dynamic Bayesian networks. In: Technologien für die intelligente automation. Springer, Berlin, Heidelberg, pp 1–8. https://doi.org/10.1007/978-3-662-59084-3_1
103. Zacharaki A, Vafeiadis T, Kolokas N, Vaxevani A, Xu Y, Peschl M, Ioannidis D, Tzovaras D (2021) Reclaim: toward a new era of refurbishment and remanufacturing of industrial equipment. Front Artif Intell 3:570562
104. Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: International conference on engineering and technology (ICET). IEEE, pp 1–6
105. Goldberg Y (2017) Neural network methods for natural language processing. Synth Lect Hum Lang Technol 10:1–309

Explainable Articial Intelligence for Cybersecurity in Smart Manufacturing

Ta Phuong Bac, Do Thu Ha, Kim Duc Tran, and Kim Phuc Tran

Abstract Industry 4.0 was first presented in 2011 and has revolutionized manufacturing in enormous applications by integrating artificial intelligence (AI), the Internet of Things (IoT), cloud computing, and other leading technologies. As technology continues to grow and expand, the concept of a new Industry 5.0 paradigm could be investigated. Industry 5.0 aims to transform the manufacturing sector into a more sustainable, human-centric, and resilient manufacturing industry. In this chapter, we demonstrate research for Cybersecurity in Smart Manufacturing in Industry 5.0 by leveraging AI and Explainable Artificial Intelligence (XAI) techniques. This chapter especially presents several essential perspectives for a potential approach of XAI to enable Smart manufacturing in the Industrial Revolution 5.0. There also is an illustrative example demonstrating the XAI approach for anomaly detection in the cyber network of an Industrial Control System in the Smart Manufacturing context.

Keywords Explainable artificial intelligence, Cybersecurity, Smart manufacturing, Internet of things, Industrial control systems

1 Introduction

Today, Industry 4.0 is known as a fact of the integration of modern technologies and innovations, in which Artificial Intelligence (AI) is a crucial driving component of the development process. AI applications can be used in many areas, such as Smart Cities, Smart Healthcare, Human-Computer Interaction (HCI), and Predictive Maintenance. The four critical areas of Industrial 4.0 are as follows:

T. P. Bac · K. D. Tran (✉)
International Research Institute for Artificial Intelligence and Data Science, Dong A University, Danang, Vietnam
e-mail: ductk@donga.edu.vn

D. T. Ha · K. P. Tran
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_10

Cyber-physical system and cobots: A CPS is a physical object that can connect to the internet, like a turbine or motor, and is incorporated with computing and control elements. Cobots, or robots with direct physical connection with a human operator within a shared office, are being quickly embraced by manufacturers. Just like standard industrial robots, Cobots comprise a mechanical arm that may be programmed to execute activities in a production facility, such as handling materials, assembling, quality control, and packing, while working alongside humans. However, Cobots may pose some concerns, even though safety risks have decreased recently. Therefore, a risk assessment is necessary to safeguard staff and equipment, given their close coordination with people.

Manufacturing based on Cloud Computing: Resources and capabilities can be turned into services in the manufacturing industry by using internal and external cloud applications [1]. Cloud manufacturing can offer secure, dependable, high-quality, affordable, and on-demand services for the whole manufacturing lifecycle. A "manufacturing cloud" of integrated and interconnected virtualized manufacturing resources and capabilities makes up the parallel and networked system. This also includes the ability to manage intelligently and use services on-demand to offer solutions to all different types of users involved in product manufacturing.

Internet of Things and Big Data: A network of devices that includes the hardware, software, firmware, and actuators necessary for the devices to connect, communicate, and effectively share data through Internet. IoT attempts to connect "things," such as machinery and items, to the internet and, ultimately, to one another. Big data makes it possible to interpret and store data from physical devices like sensors more effectively and efficiently than IoT. Big data and IoT together enable the gathering of information for improved productivity.

System Automation: Traditionally, connection technology separates systems via hardware, and software offers transparency, supports real-time decision-making, and decentralizes judgment inside technological systems to diminish the occurrence of human influence.

In the era of Industry 4.0, industrial systems are capable of intelligent decision-making through real-time interaction and collaboration across "manufacturing objects" [2], allowing flexible manufacture of high-quality, individualized products at scale [3, 4]. To advance the effort and guarantee a coordinated, cross-sectoral strategy. Industry 4.0 aims to solve issues like resource and energy sustainability, urban manufacturing, societal needs, and changing demographics [5].

However, with the increased connectivity of smart machinery, the more devices become connected, the more vulnerable they are to cyberattacks because hackers have many entry points to discover and exploit. Hackers can target the related devices that generate data, the networks that carry it, the factory infrastructure that hosts it, or the information systems of the manufacturing system. Numerous new cyber risks are brought for which the industry is unprepared. Developing a fully integrated strategic approach to cyber risk is critical to manufacturing value chains. AI and Machine Learning have the potential to make cybersecurity more efficient, proactive, and responsive against ever-increasing threats in manufacturing facilities and improve the cybersecurity posture of the manufacturing management system. Cybersecurity should never be an obstacle to progress in manufacturing toward Smart Manufacturing, Automation, and Self-healing.

In other aspects, the groundwork for the human-machine coworking era, which involves growing human-smart system collaboration to address mass personalization while enhancing job efficiency, has been laid by the digital progress of the last ten years. However, Industry 4.0 put more focus on digitalization and AI-driven technologies to improve the efficiency and flexibility of production and less on the basic values of fairness and sustainability. Although, Industry 4.0 may not be seen as a human-centric project in which human-machine collaboration or operator assistant technologies, socio-technical approaches, and work-life balance are not to be ignored. Therefore, the core idea of "Industry 5.0" shifts the attention and significant figure of investigation and invention in assisting the enterprise in delivering long-term services to people within the bounds of the planet. Cybersecurity has always been an important issue that needs to receive much attention. Enhance security for manufacturing systems in industrial 5.0 infrastructure to ensure the system's performance in timely and accurate detection of network cyberattacks on the system and more transparency. All operations and support of AI and ML techniques must be human-oriented, enhancing transparency and clearness. Therefore, XAI is a potentially powerful method for improving AI-driven systems in manufacturing infrastructure toward smart manufacturing in Industry 5.0. In this chapter, we demonstrate research in cybersecurity by leveraging AI and XAI techniques. Especially, we present a potential approach of XAI to enable cybersecurity in manufacturing toward Smart manufacturing in the Industrial Revolution 5.0.

This chapter will benefit researchers and practitioners to enrich their understanding of cybersecurity, particularly in manufacturing. In addition, the use of Machine learning supports decision-making in anomaly detection and cyberattack detection in the Cybersecurity of Industrial Systems. The real-world case study with direction and recommendations for the practical applications of XAI in manufacturing cybersecurity for explainable anomaly detection in industrial control systems is introduced. We also provide state-of-the-art research with pioneering approaches and solutions in cybersecurity manufacturing. XAI trends and opportunities in addressing cybersecurity in Industry 5.0 are discussed in this chapter. From there, points out potential future directions and prospects will be discussed, making it familiar for beginners and young researchers.

The following sessions of this chapter will describe the Cybersecurity and Manufacturing cybersecurity problems in general, along with the application of AI in this manner. Next, we go ahead with the research on XAI with the methods and techniques used to explain the AI model. Moreover, with the discussion of the responsibility of XAI to facilitate cybersecurity in Smart Manufacturing in Industrial 5.0 to carry a view of potential points and the open-up research directions. Then we extend our understanding by presenting a case study with the XAI approach for anomaly cyberattack detection with the experimentation on the real-world industrial dataset (WUSTL-IIoT-2021).

2 Cybersecurity and Manufacturing Cybersecurity

Thanks to the global Internet, people, services, and technologies can communicate with each other anywhere. Along with the development and increasing dependence of life and production on cyberspace, cyber-attacks are increasing on a larger scale. Defined by the International Organization for Standardization [6], cybersecurity is the protection and assurance of any cybersystem's confidentiality, integrity, and accessibility. With the development of the industrial revolution 4.0 in manufacturing, there is an increased demand for cybersecurity and the need for the overall development of cyberattack and anomaly detection solutions in manufacturing systems [7, 8]. Traditionally, manufacturing can be defined as the process of fabricating or assembling components into completed products on a large factory scale, with one of the most critical goals being to produce more products with high quality and minimum optimized costs [9]. The 4.0 revolution has been changing traditional conceptions of related aspects in manufacturing by bringing the leading technologies such as IoT (IIoT), AI and Machine Learning, and Cloud Computing into the core of the factory to be able to automate the operational processes as well as manage monitoring in manufacturing to maximize efficiency and increase sustainability.

Enabling by IIoT, the massive connectivity of manufacturing will allow the gathering of a considerable amount of data from multi data sources (for example, from a large number of tracking sensors mounted on equipment in factories). It is a goldmine of information that helps to generate critical values for various purposes in manufacturing.

On the one hand, IIoT is paving innovation in the manufacturing sector toward smart manufacturing generation. However, on the other hand, this also affects the system's security when a large number of connections are opened in the system and factories. There are many vulnerabilities with cybersecurity which is a potential risk for cyber-attacks on the manufacturing system. According to the 2021 Manufacturing Cybersecurity Threat Index report—June 2021 [10], one out of every five manufacturing enterprises experienced cybercrime in the US and UK. As mentioned in the report, 24% of those reported weekly cyberattacks, and leading manufacturing has become one of the most often attacked sectors since the epidemic began.

Therefore, cybersecurity is always a top concern of manufacturers of manufacturing systems when the scale of production is getting more extensive, and the system is complicated by heterogeneous devices and sensors as well as the mean of communication via wireless network or wired ethernet, making manufacturing systems are becoming more accessible and challenging to manage and operate [11].

Currently, for common cyberattack detection solutions, we can quickly mention well-known systems such as firewall systems, anti-virus, intrusion detection system, or industrial control systems that can detect abnormal performance in the manufacturing environment. However, as cybersecurity issues get more complicated, especially in the industrial environment, it requires that monitoring systems and anomaly mechanisms need to be smarter with adaptive mechanisms that can discover cybersecurity issues and cyberattacks accurately and promptly. In addition, it can be deployed with existing infrastructure in the manufacturing environment without requiring specialized equipment, which is also a critical challenge in manufacturing cybersecurity with compatibility with the manufacturing environment.

AI is an emerging technology used to enhance manufacturing efficiency with several related tasks such as general process improvement, asset management optimization, inventory and supply chain management, predictive maintenance, and manufacturing cybersecurity is the most important and challenging [12]. AI manufacturing cybersecurity systems can obtain more extraordinary performance than human and traditional solutions. For instance, that can allow monitoring in a large factory with many kinds of equipment and sensors, also is complicated network connectivity.

We can see that AI plays a potentially important role in making predictions that aid decision-making in management tasks. However, the interpretation and reliability of these AI model outcomes are still a dilemma causing a lack of interoperability between the model and manufacturers. Most machine learning algorithms cannot allow the freedom to recognize precisely how a decision was made and why it was made after building the model. For example, the administrator often struggles to understand the logic of AI models due to many parameters that should be optimized and tuned in the model development phase. Therefore, it leads to more difficult decision-making and a lack of trust in cases where the AI model gives false positive results that harm the system's integrity. In addition, to trust the decisions made by AI systems, managers must understand how the AI model makes decisions to explain the phenomenon in the system and provide solutions to fix the problem accurately and promptly. Models need to provide clear explanations and justifications for the decisions they make.

Explainable Artificial Intelligence (XAI) [13] is the enabler to tackling these problems that support perceiving, understanding, interpreting, and explaining the results and decisions generated by AI models. Manufacturing cybersecurity is one of many areas where XAI can benefit as the model's outcome become interpretable [14]. In the next session in this chapter, we will present the overview of XAI, its definition, techniques, and methods.

3 Explainable Artificial Intelligence

Artificial intelligence (AI) has achieved impressive performance in a diverse range of domains (i.e. finance, defense, industry, medicine, security, etc...). Nevertheless, many of these systems are unable to interpret their autonomous predictions or actions to users due to the black-box structure of machine learning models, making it difficult for them to trust and analyze decisions and recommendations of models Machlev et al. [15]. To address this challenge, a set of new techniques and principles, commonly referred to as eXplainable Artificial Intelligence (XAI) [13], have been developed in recent years. The primary aim of XAI is to enhance the interpretability of machine learning models, enabling better comprehension of data and insights to be gained.

Explainable Artificial Intelligence (XAI) utilises various techniques and algorithms to understand and interpret the behavior of black box models, as well as to examine decision-making systems. Visualization techniques aid to represent model and decision explanations in a more transparent, understandable, trustworthy, and accurate way. There is widespread agreement that there is no exact definition of XAI because it has evolved and also been modified based on specific applications, scenarios, and researcher expertise [16]. Following [17], XAI is a technique that generates details or explanations to make its operation understandable or explicit. With the assistance of appropriate size, an XAI enables a user to learn transparent, relevant, and reasoned information at the proper time [18]. According to [19], XAI is an area of artificial intelligence that supports a collection of tools, methods, and algorithms that can produce superior interpretable, intuitive, human-understandable justifications for AI judgments. More than a formal technical notion, the term “XAI” is often used to refer to the movement, initiatives, and efforts in response to concerns about AI transparency and confidence [20].

XAI could improve machine learning algorithms and models in terms of three factors, transparency, trustability, and model bias understanding [19]. Transparency is relevant to the need to characterize and observe the mechanisms by which decisions of AI systems are made and learn to adapt to their environment, as well as the governance of the data used to create those decisions. Using visual techniques makes human-understandable justification and provides an accurate explanation for the model behaviors [20]. Besides, XAI could improve transparency and fairness by creating a human-understandable justification for the decisions and could find and deter adversarial examples [21]. Trustability refers to the need to explain and justify model decisions and judgments in order to increase end-user confidence in decision-making. Because of the adverse effects of overfitting and biasing, learning behavioral patterns using XAI approaches for different input data distributions might lead to a more humane comprehension of these input data [19].

Transparency and trustability are critical and important factors in cybersecurity in which XAI has enormous potential in predicting cyber-attacks. XAI provides a thorough explanation, essential reason, and threat identification in order to reduce breach risk and robust security system. XAI can also aid in risk identification and prioritization, incident response coordination, and malware threat detection [22].

In general, XAI could be categorised based on scope, methodology, and usage purpose [19]. **Scope** of explanations answers to the question “Where is the XAI method focusing on?” and can be considered to be local or global explanations. Generally, *Locally explainable methods* generate explanations by focusing on a single input data instance from the data population and utilizing the various data attributions. *Globally explainable models* give sense into the model’s overall judgment, enabling one to comprehend the attributions for various input data.

Methodology is an answer to the question “what is the algorithmic approach?”, focusing on changes to input data (perturbation-based), as opposed to those that focus on model architecture and parameters (backpropagation-based).

Perturbation-based explainable algorithms change a particular input instance's feature set by employing conditional sampling, occlusion, filling operations, masking, or partially substituting features with other features.

Backpropagation-based methods, the explainable method does one or more forward passes via the neural network, producing attributions at the backpropagation step using partial derivatives of the activation functions [19].

Based on the way in which the XAI method is developed, the categorization of XAI can be determined by its intended usage, namely intrinsic and post-hoc classification. Intrinsic Explainable Artificial Intelligence (XAI) methods incorporate a variety of techniques, including decision rules, attention mechanisms, reasoning paths, masks, and/or graphs, among others, directly into the model. These methods are used during both the model's training and execution stages, thereby allowing for real-time explanation of model outcomes. Conversely, post-hoc XAI techniques, such as visualizations, counterfactual analysis, surrogate models, concept importance, LIME, and SHAP, are model-agnostic approaches that are typically employed to explain various components of a model after it has converged [22].

The SHapley Additive exPlanations (SHAP) framework is designed to justify and explain the results of prediction models. It utilizes game theory to demonstrate a relationship between optimal credit distribution and local explanations, employing traditional Shapley values. SHAP uses a bar-length and color-coding approach, which simplifies the understanding of the size and direction of an impact. It can be applied to explain basic machine learning (ML) algorithms, such as linear regression, logistic regression, and tree-based models, as well as more complex models, such as deep learning models for image classification and captioning, and various natural language processing (NLP) tasks, including sentiment analysis, translation, and text summarization. Similar to SHAP, Local Interpretable Model agnostic Explanations (LIME) generate a list of explanations, indicating the role of each feature in the prediction of a specific data sample. LIME, however, is much faster than SHAP. It can explain any black-box classifier that implements a function, taking in raw text and producing a probability for each class. Currently, LIME can provide predictions for tabular data, images, and text classifiers [22].

4 Explainable Artificial Intelligence Enable Cybersecurity for Smart Manufacturing in the Industry 5.0

This section presents the state of the art as well as perspectives on Explainable Artificial Intelligence enable Cybersecurity for Smart Manufacturing in the Industry 5.0.

4.1 *The State-of-the-Art*

Detection of abnormal situations and cyberattacks are significant problems in smart manufacturing. Several methods exist that target XAI in the cybersecurity system of the manufacturing industry, such as intrusion detection systems, cyber-physical systems, and industrial control systems. An intrusion Detection System (IDS) is an essential tool in the network system to provide a secure network environment for the manufacturing process, where the equipment, machines, and factories are connected entirely through the Internet. The goal of the IDS system is to detect unauthorized intrusions into the system and abuse and hijack the system, causing abnormal activities that are harmful to the operation process and the productivity of the manufacturing system. In [23], the author builds an explainability intrusion detection solution to explain the anomalies detected in the network. In particular, detecting anomalies and new attacks in the network lead to the use of more complex AI models. As a result, the manufacturer can hardly understand the decision of AI-based solutions. The author proposes a framework that can provide local and global explanations for any IDS. Furthermore, the solution uses the SHapley Additive exPlanations (SHAP) method to enhance the transparency of the IDS system so that administrators can better understand the cyberattacks. Evaluations were performed on the NSL-KDD set, a dataset containing common types of attacks that can occur in a manufacturing system's network. XAI is approached in local and global directions, whereby the local explanation presents why the model predicts output on the specific input. Meanwhile, the global explanation extracts influential features from IDS to show the relationship between main features and attack types, making the attack classification results more reliable. As a result, localization attacks easily and quickly can improve the effectiveness of the attack mitigation process. However, these explanatory solutions are taken offline because the SHAP method still does not work in real time.

In the same direction, the author in [24–27] also applies XAI to overcome the "black-box properties" of deep neural network models in IDSs systems. However, to increase the explanation's efficiency, the author proposes to use a combination of SHAP and LIME models to generate explanations. SHAP is used for global explanations, and LIME is used for local explanations.

Khan et al. [28] proposed using LIME methods to make an explanation for the autoencoder-based detection model that uses a combination of CNN and RNN to be able to detect cyber threats in connected networks. The author claims that the model can deal with known and new attacks, such as zero-day attacks. The experimentation is performed on the real data set of the manufacturing system, the gas pipeline system. The dataset includes system records of network packets used to communicate and control the gas pipeline. XAI technique is used not only to explain the model decision but also to improve the efficiency in trust management of manufacturing cybersecurity systems. Managers can interpret casual reasoning and root cause analysis. However, the study still has challenges to solve in the future, such as evaluating larger data sets and considering model performance with multivariate time series.

For cybersecurity in general and manufacturing cybersecurity specifically, root cause analysis and diagnosis are not a small challenge when anomaly detection systems stop at detection. In [29], the author proposes to use XAI for fault diagnosis in the overall architecture to detect anomalies and diagnose the root cause of the rotating machinery system. The data used includes various faults of rotating machinery, such as rotors, rolling element bearings, and gears. SHAP-based XAI is again proposed to be used to support the classification and analysis of the source related to different types of anomalies.

Another aspect of manufacturing cybersecurity systems we can mention is Industrial Control System, ICS is an automation control and monitoring system in industrial processes. In which the problem of applying XAI is still a big challenge for researchers. This issue is addressed in pioneering research in [30, 31]. In [30] the authors proposed adopting XAI to explain abnormalities detected from Bi-LSTM models in ICS environments. This paper focus on the context of a smart factory that uses steam-turbine power generation, and pumped-storage hydropower generation generates electricity using steam turbines, and pumped-storage hydropower is the focus of the ICS. A hardware-in-loop-based augmented ICS Security Dataset [32] simulator emulates a smart factory with many processes, such as boiler, turbine, water treatment, and HIL simulation. In this model, the author leverages SHAP to support prompt, reasonable action and operations when an anomaly is detected. Furthermore, by calculating the relative importance of each attribute in a complex model, SHAP generates predictions for test samples that can be understood. The following mainstream research in this field is presented in [31]. In this research, the proposed Fedex architecture focuses on cybersecurity in the ICS ecosystem to build a cybersecurity solution deployed on edge layers to optimize resource consumption while ensuring accuracy and efficiency. The hybrid FedVAE-SVDD model performs the anomaly detection task, and the architecture also provides interpretability based on calculating the shapely values-SHAP method to give explanations. This study shows that the authors have pioneered detecting and interpreting anomalies in distributed environments using federated learning techniques. Furthermore, XAI is applied separately in each small factory area (zone), quickly responding to abnormal points based on the relationship between the features and corresponding physical components at each manufacturing zone. The experimental results were conducted on the SCADA data set [33], which simulates a liquid storage system submitting an automated production line in an ICS system and evaluated on the SWAT dataset [34]. These datasets are typical for cybersecurity manufacturing.

4.2 Toward Smart Manufacturing in the Industry 5.0

Industry 4.0 has had a highly transformative impact thanks to data-driven and connectivity technologies. However, in the development wave of industry 4.0, there are also transformative effects on society. Humans are a crucial factor when the production process is automated, making human people's role in production lines change or even threatened. The changing role of people in manufacturing requires new skills to respond to complex technologies. Increasing automation makes the social part of an industry increasingly weaker. This challenge is the impetus for innovation in the Industry 5.0 concept. One of the core values is to leverage benefits and extend the benefits of Industry 4.0 by taking a human-centric approach. The concept of industry 5.0 has been discussed in various studies [35]. Accordingly, Industry 5.0 is thought to be utilizing the power of industry to achieve social goals exceeding job creation and expansion by enabling production to respect the limits of our world and place the wellness of industrial workers at the heart of the manufacturing operation [35]. Industry 5.0 was born from the belief that Industry 4.0 solely emphasizes digitization and AI-based technology to boost production efficiency and flexibility, eliminating social justice and integrity. Because of this, Industry 5.0 emphasizes the value of research and innovation to support the industry in providing long-term services to humanity within natural limitations [35]. The main tenets of Industry 5.0 are human-centricity, sustainability, and resilience.

- Human-centricity: human needs are given precedence over the production process in a human-centered approach. Manufacturers need to consider what technology can do for the workforce and how it can change to meet those demands rather than the other way around. In addition, technology must not impact concerns like autonomy and privacy.
- Sustainability: the development of circular processes that reuse, and recycle resources is necessary for manufacturing to be sustainable. In addition, reduced environmental repercussions are required. Utilizing technologies like AI and additive manufacturing, sustainable producers may maximize penalization while minimizing waste and resource use.
- Resilience: To better defend themselves against disruptions and crises like COVID-19, manufacturers must increase the robustness of manufacturing output.

As discussed in other sections, XAI is an assistive technique in providing operators and managers with human-readable information showing how to input data into a decision, specifically as an output of AI models. In addition, it is useful in accountability and audit to identify defects in the manufacturing process and where the overall system needs to improve if the AI system makes faulty decisions.

In Industry 5.0, with the goal of human-centric industrial development, more and more workers are supported by cybersecurity control systems in the manufacturing system automatically based on artificial intelligence. Therefore, AI needs to effectively detect cyberattacks and network anomalies in manufacturing environments with huge factories and reliably move towards human-centered smart manufacturing, increasing interoperability in the human-machine interaction and making it possible for humans to understand the decision-making of AI models thanks to XAI. In a similar way to how people interact with people, trust is expressed when we can understandably explain to each other how to make decisions. In addition, XAI can help us recognize and trust the AI-based anomaly detection and cyberattack techniques deployed in the cybersecurity system of the factory, ensuring the network security of factories and manufacturing processes in Industry 5.0.

Consequently, XAI will be an important factor in establishing trust and interaction between people and machines, which is the foundation to unlock the true potential of AI and automation for manufacturing cybersecurity in particular and all other industries in Industry 5.0 where humans and AI are working together.

Resilience is another crucial issue for manufacturing cybersecurity in the evolution to Industry 5.0, which is witnessed as the ability to cope flexibly with change from different network actors. Manufacturing is a heterogeneous environment and is face-down to change, such as changes in equipment and equipment condition, geographic location, and ambient conditions of factories, more than variability unpredictable changes of different attack forms and production systems. Thus, the cybersecurity for smart manufacturing in Industry 5.0 system needs to be equipped with mechanisms to quickly adapt to changes in the environment and new attack sources to ensure efficiency, timeliness, and transparency.

XAI comes across as a tool to help promote the resilience of cybersecurity systems in smart manufacturing and is a sustainable engine for industry 5.0 prosperity. Thanks to its explainability and demonstration, XAI can help in tracing and root cause analysis to deal with vulnerabilities that can occur at various levels in Smart manufacturing, for instance, factory, communication network, and physical industrial system level.

5 Perspectives for Explainable Cybersecurity for Smart Manufacturing in the Industry 5.0

In this section, we will present potential perspectives for applying explainable cybersecurity in smart manufacturing in the Industry 5.0 sector. The primary consideration is as follow: Designing and implementing of XAI-based Anomaly detection system over cloud-edge computing, a new AI strategy in manufacturing cybersecurity-Explainable Augmented Intelligence for cybersecurity and emerging manufacturing topics in 6G and beyond 5G cybersecurity.

5.1 Designing of XAI-Based Anomaly Detection System Over Edge Computing

The decision interpretability of black-box models, based on deep learning-related algorithms with many parameters, offers another viewpoint. Although these algorithms' prediction results can attain excellent performance, it is difficult for engineers to examine and understand them. Explainable Artificial Intelligence (XAI) can be used to overcome this problem, especially in ICSs, where engineers must carefully monitor their IoT systems. XAI makes it easier for people to comprehend why models anticipate a decision. Modern XAI frameworks allow us to analyze and explain outcomes' dependability adequately. Thus in the future, we should use them to understand the choices made by anomaly detection architecture.

Another aspect is, Edge intelligence is a main approach in smart manufacturing that uses a variety of resources, such as storage, caching, computing and networking. For performance improvement, the edge server is located near the end factory or in the local area of the factory in the manufacturing system. Robust cross-disciplinary algorithms, such as deep reinforcement learning, data mining, NPL, and ML, are routinely put into the edge server [36]. XAI adds a multidimensional construct explanation-to edge intelligence, expanding its potential [37]. Access to edge caching is required for creating and storing pre-model explanations and providing security at the edge and IoT layers.

Due to additional storage spaces, new devices or edge/cloud computing resources will be required. Edge computing prospers in ways that lower costs, latency, and bandwidth utilization. However, it has come with some concerns since a new threat surface might make it vulnerable to intrusions. Therefore, besides developing XAI-based anomaly detection solutions, the deployment architecture of the edge/cloud computing infrastructure is an important factor that should be considered and evaluated accordingly. These solutions can be deployed on a tiered cloud-edge architecture [38, 39], ensuring optimal performance and security of the solution for the system.

5.2 Explainable Augmented Intelligence for Cybersecurity

The fast-moving digital revolution has given us a glimpse of when we have improved the efficiency of cybersecurity manufacturing operations using XAI and AI and the risks that come with it. In regulated industries such as manufacturing, technologies need to be explainable and trustworthy, keeping people accountable for critical decisions.

Thus, Industry 5.0 has led to the development of a new AI strategy for manufacturing cybersecurity, designed to solve the problem of ever-greater complexity, which puts humans at the center of cybersecurity systems. These evolutions should be to search for "Augmented Intelligence," which is the other AI in a cybersecurity manner [40, 41]. Augmented Intelligence is a combination between humans and machines, a humanized AI that human potential and explainability of machine learning models in the making of Industry 5.0. In addition, Augmented Intelligence integrates various technologies such as data fusion, machine learning and deep learning, and human off-body sensing techniques. Instead of using raw data, augmented intelligence solutions acquire, combine, fuse, and correlate diverse streams of Intelligence, giving analysts the crucial context they need to make judgments. Assets, Vulnerabilities, Threats, Phenomena (events and anomalies), Consequences (such as commercial impact), and Options/Countermeasures in cyberspace and time are all automatically connected by the context.

Even though AI has recently made significant strides in cybersecurity, it is still impossible to completely and automatically adapt a system to the changing environment and comprehend all uncertainty like risks, threats, and attacks. And autonomously choose and apply appropriate remedial actions to reduce the risks and defend against these attacks and threats. Therefore, Augmented Intelligence Cybersecurity resulted in a strong interdependence and collaboration between AI systems and humans to improve cybersecurity effectively [42, 43].

Furthermore, relying on XAI, an expert cybersecurity analyst should be a part of the AI analytical procedure to help encourage and improve decision-making processes. Therefore, hybrid augmented Intelligence is critically created and studied to strengthen the use of AI and XAI in cybersecurity manufacturing.

5.3 Cybersecurity for the Applications of 5G, 6G in Smart Manufacturing

Information technology has advanced quickly in recent years. As the driving energy behind the industrial and technological transformation, 5G is rapidly demonstrating its ability and being leveraged in the manufacturing sector [44, 45]. According to definition [46], 5G has three main application scenarios: massive machine type communication (mMTC), which refers to enormous connectivity communication; ultra-reliable low latency communication (uRLLC), which refers to high reliability and low latency; and enhanced mobile broadband (eMBB), which refers to high-speed and high-throughput communication. Breakthroughs in machine cooperation, remote maintenance and control, and other fields have been made possible by fully integrating 5G technology with AI, IoT, cloud computing, big data, and edge computing.

5G has been used in the industrial and manufacturing field [45, 47, 48] for remote operation and maintenance, manufactory monitoring, precision assembly, equipment inspection, robot vision, and other things. Therefore, the research and exploitation of the potential of 5G in smart manufacturing have become the shared objective of academics and businesses. Cybersecurity and the implementation of 5G still have significant issues [48, 49]. For instance, research continues to be conducted on integrating 5G technology with manufacturing infrastructure, especially cybersecurity systems of the smart manufacturing sector.

Manufacturing uses cyber-physical systems—intelligent systems with high levels of autonomy, self-awareness, self-judgment, and self-regulation. It can achieve seamless data integration in cyberspace, organized physical entity collaboration in physical space, and the connection and interaction between cyberspace and physical space without human involvement [50]. All new features of intelligent production may be covered by CPS and manufacturing technology integration. However, completely integrating 5G with CPS in manufacturing is a complicated system that needs more research in the field of communication and manufacturing, and cybersecurity.

On the other hand, the explosion of research on 6G is the foundation for the expansion and improvement of 5G [51, 52]. Complex 6G networks in the industrial sector may connect intelligently to many heterogeneous dense sub-networks (private factory networks) and present some trust and security issues at various network connection levels. Manufacturing 6G communication systems must be capable of autonomous operation, intelligent risk reduction, and automated detection of proactive risks. As a result, AI-based trust and security become attractive options for automatically identifying possible risks and taking swift action [53]. To achieve complete automation, AI-enabled 6G might expose ML systems to several data- and model-based security vulnerabilities [54], including data insertion, modification, model evasion, and model change. In addition to multi-threat analysis on large datasets from industrial systems, integrating and deploying AI/ML algorithms for resource-constrained devices still requires more effort. Furthermore, the accountability for AI mistakes, the scalability and reliability of AI models in various storage and computing infrastructures, and the security of AI models in distributed systems are some additional challenges that may result from AI-driven security solutions [55].

6 Case Study—Explainable Anomaly Detection for Industrial Control Systems

In the Smart Manufacturing concept, Industrial Control Systems (ICSs) are an important element that plays a critical role in managing the operation of many manufacturing systems, such as power stations, water supply systems, and gas. Historically, these systems worked on proprietary hardware and/or software physically that are connected by wireless communication to enrich connectivity over the Internet, therefore increasing the risk of cybersecurity vulnerabilities [56].

In addition to the solutions already mentioned for dealing with network intrusion issues, such as the firewall and Intrusion Detection System (IDS), AI-based anomaly detection has also been researched and developed to handle innovative and complex threats to ICSs. The results of anomaly detection (AD) algorithms' predictions are frequently satisfactory [57, 58]. AD is linked to pursuing data without an anticipated pattern or exhibiting a pattern distinct from the usual stream. Hence, AD approaches outperform traditional classifiers because they are effective despite new attacks or unusual behavior.

Despite the remarkable effectiveness of AI-based anomaly detection methods, the model dependability may be compromised if the source of the model's projected findings is not given, typically due to the black-box problem with such technologies. It is crucial to explain to such AD models since a projected anomaly may originate from other technical issues rather than a real cyberattack. An example of a typical abnormality is from faulty sensors (temperature, vibration, etc.), where anomalous sensor data might signal oncoming problems. As a result, it takes time to examine how the AD model makes decisions when anomalies occur. As the interpretability of the model's discovered data has not been a major focus of most prior studies, this issue also poses a substantial obstacle to cyber-attack detection in ICSs. The

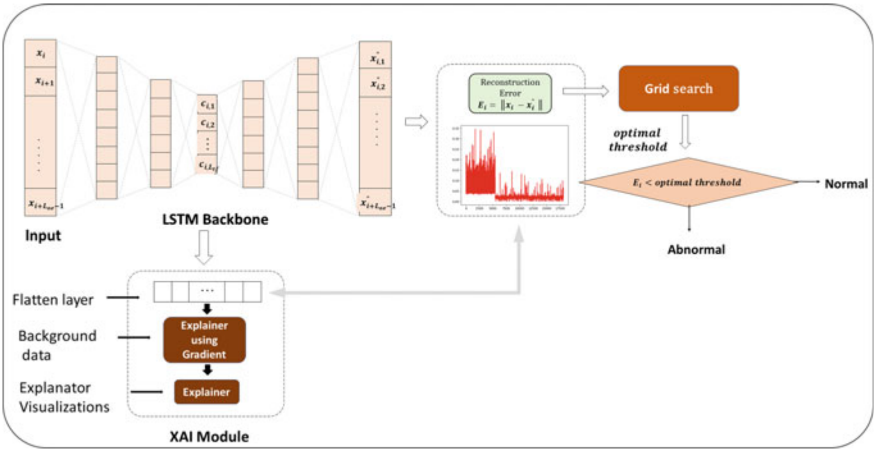


Fig. 1 The XAI integration method for anomaly detection paradigm

use of Explainable AI (XAI), a technique that aids individuals in comprehending or outlining the operation of an AI system, is one approach to the issue. The application of XAI in cracking anomaly detection results has recently been discussed in the literature; for examples, see [58, 59], for example.

In the following section, we present a case study in the field of explainable anomaly detection using deep learning methods to solve a problem with cybersecurity problem. In particular, the LSTM-Autoencoder model is leveraged for anomaly detection in combination with a grid-search method and SHAP-based XAI module to explain the detected anomalies to improve the perspective and reliable results of the LSTM-Autoencoder network. Then, the method's performance is evaluated based on a real-world industrial network dataset named WUSTL-IIOT-2021 [60].

6.1 Anomaly Detection Using the Hybrid Autoencoder LSTM Model

In this section, Autoencoder LSTM combines an autoencoder network using LSTM as a core for the encoder and decoder. First, the autoencoder intends to learn a compressed representation of input by encoding and decoding. Then, by using the LSTM cell for both the encoder and decoder to handle input sequences, the Autoencoder LSTM model outperforms the standard autoencoder. Utilizing LSTM for anomaly detection are discussed in several works, such as [61–64].

More specifically, in an Autoencoder network, the input is particularly compressed in the encoder section to generate a code, and it is then decompressed in the decoder part to reconstruct the information. Then, the generated output will be compared with the input used to calculate the backpropagation and reconstruction loss. If the reconstruction loss is greater than a certain threshold, a sample can be considered abnormal. In this process, grid search is utilized to find an optimal threshold that effectively determines which difference is large enough.

Suppose the Autoencoder-based LSTM model has been trained with the normal sequences data. The the incoming stream data during inference is $X = \{x_1, x_2, \ldots, x_n\}$, where $x_j \in \mathbb{R}^m$, $j = 1, 2, \ldots, n$; and m is the number of attributes. As demonstrated in Figure 1, X is then splited into $n - w + 1$ overlapping sequences $X_i = \{x_i, x_{i+1}, \ldots, x_{i+w-1}\}$, $i = 1, 2, ..., n - w + 1$, by using a sliding window of size w. The difference between X_i and its reconstructed representation $\hat{X}_i$ in the training phase is passed through the Grid search to find the optimal threshold.

The anomaly detection task could be done by using Autoencoder-LSTM. However, this hybrid model is a "black box" that is hard to interpret and lacks transparency. For that reason, Explainable Artificial Intelligence (XAI) is applied to figure out the process of deciding an anomaly.

6.2 Integration of Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) is the term for the algorithms that make AI models understandable to humans, fostering their confidence in the results of the model's application and their efficient management of the model's advantages. In other words, it meets our expectations for an AI model by performing as predicted, offering clear explanations, and being apparent in its work. An AI model, its anticipated effects, and any biases may be defined via XAI.

The SHapley Addictive exPlanations (SHAP) method, a game-theory-based methodology for globally or locally understanding the output of any ML or DL model proposed in [65], is one of the state-of-the-art XAI strategies in that field. The average estimated marginal contribution of a feature over all potential mixture sets or coalitions, named Shapley values, are used in this approach to assess the contribution level of each feature to the projected outcomes. The fundamental concept is to make a complex model f more understandable by using an approximate model g. Similar to LIME in [66], explanation models frequently use input variables that have been simplified x', mapping to the original samples x through a function h_x that satisfies $x = h_x(x')$, and local explanation methods that guarantee $f(h_x(z')) \approx g(z')$ whenever $z' \approx x'$. The model $g(z')$ can be constructed as follows:

$$g(z') = \phi_0 + \sum_{k=1}^{M} \phi_k z'_k, \tag{1}$$

In which $z' \in \{1, 0\}^M$ is a coalition including the simplified values z_k of input features, the value $z'_k = 1$ indicates the presence of the feature in the coalition and vice versa $z'_k = 0$. M is the number of simplified features. And $\phi_k \in \mathbb{R}$ denotes the Shapley values of each kth feature. The value ϕ_k can be computed by:

$$\phi_k = \sum_{\mathcal{S} \subseteq N \setminus \{k\}} \frac{|\mathcal{S}|!(|N| - |\mathcal{S}| - 1)!}{|N|!} \big(v(\mathcal{S} \cup \{k\}) - v(\mathcal{S})\big) \tag{2}$$

where, the set of features is denoted by N contains subsets $\mathcal{S}$ excluding the k^{th} feature. In the subset $\mathcal{S}$, $v(\mathcal{S})$ is the prediction value for features. The contribution of features at the global view is provided based on the Shaley values ϕ_k, which are calculated in each of its samples on a sliding window. Accordingly, to model type f the SHAP framework can use a range of computing techniques to explain the model, such as (Kernel SHAP [65]), the model-agnostic approximation methods (e.g., Linear SHAP [67]), and model-type-specific approximation methods (e.g., Tree SHAP [68]), (Deep SHAP [65]), and (Gradient SHAP).

The Deep SHAP and the Gradient SHAP are two of the SHAP framework's approaches that are computationally set up for deep learning model types, particularly those that use three-dimensional data as input. The Gradient SHAP corresponds to SHAP values by calculating the expectancies of gradients by randomly selecting from the distribution of baselines or references. In contrast, the Deep SHAP uses a method for estimating SHAP values. Hence, the Gradient SHAP will be used in this investigation since the Deep SHAP takes a longer time to process.

The AI model output for the Gradient SHAP approach is anticipated to take the shape of a vector or a single value. The LSTM result, however, does not reflect reality. As a result, we create a new model in the XAI module using the weights from the previous pre-trained model, but also somewhat changing the structure by attaching a flattened class on top of this model, as illustrated in Fig. 1. This model plus a background dataset from the training set serve as the inputs for explaining Gradient SHAP. In addition, the built-in explainer may be provided the anomalous window when an anomaly is predicted to produce definitive explanations. The sequel will show how the Autoencoder LSTM hybrid model identifies anomalies based on industrial network datasets. The Gradient SHAP is then used to interpret the output. We may learn the proportion of the features' influence on the prediction values and get certain XAI advantages by visualizing the calculated Shapley values.

6.3 Illustrative Performance Evaluation

This section aims to illustrate the performance and explanation of the hybrid anomaly detection approach based on a data set named WUSTL-IIOT-2021 [60], gathering the network traffic of the industrial system. Real cyberattacks conduct this dataset against the system, including normal and attack traffic, as specified in Table 1

The industrial system of the dataset illustrates a part of the water treatment and distribution system that monitors the water level and turbidity quantity of the water storage tank. The testbed includes components in an industrial system such as HMI, historical log, and PLC. These emulated components make the system similar to the real industrial control system in the context of smart manufacturing. The detailed

Table 1 The statistic of WUST-IIOT dataset

Dataset	WUST-IIoT
Number of observations	1 194 264
Number of features	41
Number of attack samples	87 016
Number of normal samples	1 107 448

Table 2 The description of several features in the dataset

Features name	DataType	Descriptions
Mean (Mean flow)	Float	The average duration of the active flows
Sport (Source port)	Integer	Source port number
Dport (Destionationport)	Integer	Destination port number
SrcPkts (Source packets)	Integer	Source/destination packet count
Dpkts (Destination packets)	Integer	Destination/source packet count
Tpkts (Total packets)	Integer	Total transaction packet count
Sbytes (Source bytes)	Integer	Source/destination bytes count
Dbytes (Destination bytes)	Integer	Destination/source bytes count
Srate (Source rate)	Float	Source packets per second
Drate (Destination rate)	Float	Destination packets per second
Proto (Transaction protocol)	Char	transaction protocol
sTtl (Source TTL)	Float	Source $\rightarrow$ Destination TTL value
DAppBytes (Destination app byte)	Integer	Destination $\rightarrow$ Source application bytes
TotAppByte (Total app byte)	Integer	Total application bytes

illustrations of the testbed can be found in [69]. Each record in the data has 41 features, which are typical of network flows and also vary under attack. Table 2 shows the 14 features and their description as an example.

In this experiment, 80% of the normal samples (886,000 records) are fed into the model for the training phase, and the remaining 20% (221,500) is extracted with all abnormal samples (87,016) for the testing phase, with the percentage of anomalies of roughly 9.8212%. The performance of the Autoencoder LSTM approach for anomaly detection is evaluated in terms of Precision, Recall, and F1-Score. We also used the Grid-search method in combination with the Autoencoder LSTM model to find out a threshold for classifying normal and abnormal cases, as mentioned in Sect. 6.1. In the context of manufacturing cybersecurity, seeing all actual anomalies tends to be more crucial since any anomaly may yield it. As a result, optimizing the Recall and

Table 3 The performance of the LSTM-based anomaly detection method evaluating on WUSTL-IIOT-2021 dataset

Method	Precision	Recall	F1-score
Autoencoder LSTM	0.7019	1.0	0.8235

F1- score metric is more significant than the value of the Precision metric since a trade-off with a small number of false alarms is acceptable compared to the expenses associated with the injuries inflicted by anomalies.

Table 3 shows the performance of the hybrid model. It is depicted that the hybrid autoencoder LSTM method achieves a Recall of 1.0% and in terms of F1-score of 0.8235% .

Although the results above show that the Autoencoder LSTM model has a respectable performance for anomaly detection, a decent learning model should guarantee high performance and openness. In order to find out why the results are predicted, the following steps must be taken. In this case study, Gradient SHAP is used to visualize the reasons for the observed anomalies and to illustrate how the model functions and the impact of the most important features on the features affected by the detected abnormalities. Manufacturers might be expected to benefit from the insights provided by Gradient SHAP-based visualizations in discovering and verifying abnormalities quickly and visually. As a result, the cure solution may be successful and allow the manufacturer to complete the maintenance duties more quickly.

The SHAP summary plot provides an overall description of a model that combines feature importance with feature effect. On a summary plot, the shapely value for a feature and specific sample is shown as a point. Shapely values are on the X-axis, and features are on the Y-axis. Colors represent low/high values. The summary plot's features are organized in order of relevance, with the top feature being the most essential and the bottom one being the least.

Figures 2 and 3 show how each feature's impact on the model output from a broad viewpoint is explored in one sequence (i.e., one sliding window). This effect includes anomalous patterns. In this representation, the characteristics are arranged in descending significance ranking, and the horizontal line represents whether the influence of that feature (i.e., SHAP value) is associated with a higher or lower indicated value for that sampling. As a result, a positive SHAP value says that a higher feature value will result in a higher predicted value, while a negative SHAP value communicates that a higher feature value will result in a lower projected value. As can be observed in Fig. 2, "lower" "DstBytes" and "Dstjitter" feature values significantly affect the expected anomaly, whereas this is correlated with higher "sTtl" feature values. In other words, the engineers might learn more about their smart manufacturing system and the logical limitation between the features and anomaly's corresponding physical aspects.

Engineers must perform a full system assessment to determine if the anomaly results from a real cyberattack since threats can develop from various sources that are

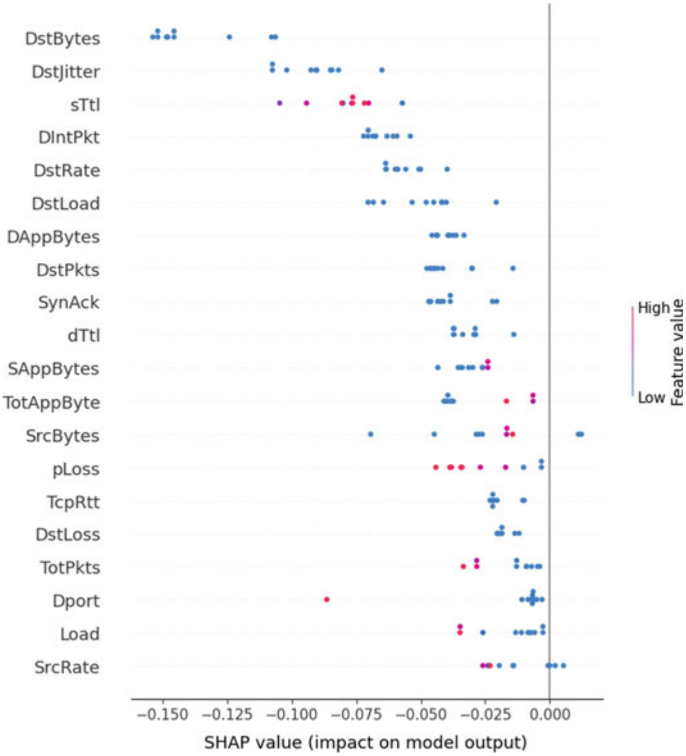


Fig. 2 SHAP summary plot—the low value of 'DstBytes' increases the probability of attack whereas the high value of 'sTtl' decreases the probability of attack

not cyber-intrusion occurrences. This procedure needs time, effort, and even money. Despite this, it is possible to overcome it based on the visualization in Fig. 3. The three characteristics in this figure that have the most influence on the projected anomaly are "DstBytes," "sTtl," and "Dstjiter." It can be concluded from the descriptions of each form of assault that the behavior of the "DoS attack" and the DstByte—Destination byte count should be given priority, in addition to looking up the location of the relevant host. Thus, our proposal successfully assists engineers or cybersecurity specialists in examining abnormalities predicted by the black-box AD model indicates. Furthermore, it not only contributes to time and maintenance cost savings but also satisfies the model's transparency, which encourages engineers or operators to trust the suggested AD model. These encouraging interpretations indicate that XAI will eventually be incorporated in more black-box-related AD investigations for ICSs.

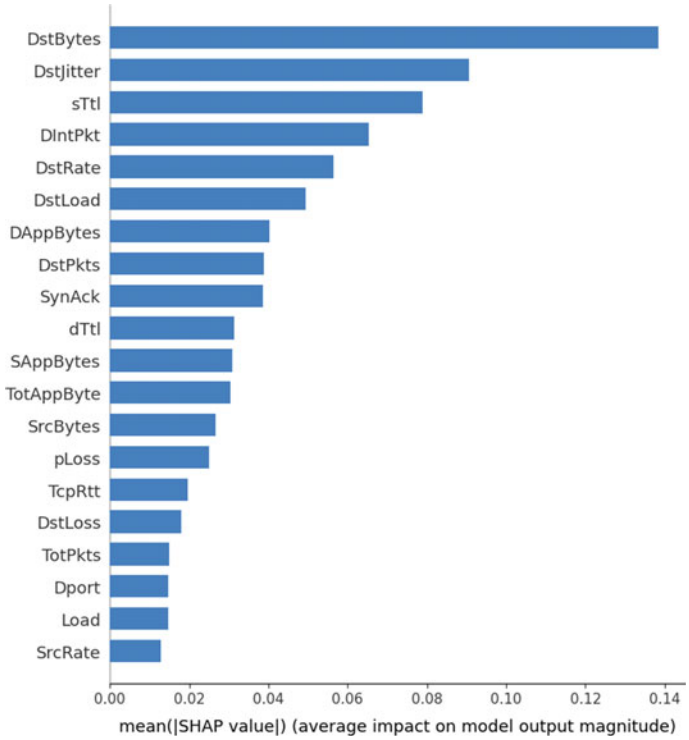


Fig. 3 The model's most crucial affecting factors

7 Concluding Remarks

In this chapter, we discussed cybersecurity for manufacturing manners, especially for Smart Manufacturing in the new digital era in the next generation of industrial (Industrial 5.0). Although cybersecurity has always been an important issue in the manufacturing sector, many potential threats exist. Therefore, machine learning techniques are widely applied to improve the efficiency of anomaly detection and attack detection in the network, which are shown to be efficient approaches for anomaly and cyberattack detection. Furthermore, XAI is selected, integrated, and optimized correctly for each manufacturing system characteristic, as illustrated by the state-of-the-art research on XAI's application in cybersecurity manufacturing in this chapter. This chapter delivers the potential of the new approach of XAI as an enabler to comprehend cybersecurity problems in the manufacturing network of Industrial 5.0 and move towards human-centric smart manufacturing.

In addition, we covered the cybersecurity highlights of XAI-based smart manufacturing in industrial 5.0. Accordingly, the combined deployment of new XAI solutions requires the corresponding design and change of infrastructure with core technologies such as cloud-edge computing. Secondly, it is necessary to develop solutions for a new AI strategy in cybersecurity—Explainable Augmented Intelligence for cybersecurity. Finally, 6G and beyond 5G have always been the leading areas showing tremendous potential with applications in the manufacturing sector. Therefore, researching and fully integrating 5G and 6G applications into manufacturing infrastructure is one of the challenges that need to be cracked to take full advantage of the new generation telecommunications network.

In conclusion, we provide a case study of the application of XAI in the field of explainable anomaly detection for an Autoencoder LSTM-based model to handle cybersecurity issues in manufacturing. All experiments are performed based on a real industrial dataset.

References

1. Griffor ER, Greer C, Wollman DA, Burns MJ (2017) Framework for cyber-physical systems: volume 1, overview
2. Lu Y, Xu X, Wang L (2020) Smart manufacturing process and system automation-a critical review of the standards and envisioned scenarios. J Manuf Syst 56:312–325
3. Wang S, Wan J, Zhang D, Li D, Zhang C (2016) Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Comput Netw 101:158–168
4. Zhong RY, Xu X, Klotz E, Newman ST (2017) Intelligent manufacturing in the context of industry 4.0: a review. Engineering 3:616–630
5. Kagermann H, Wahlster W, Helbig J et al (2013) Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 working group, Forschungsunion, Berlin, Germany
6. ISO/IEC-27032:2012 I (2012) Information technology security techniques guidelines for cybersecurity. https://www.iso.org/standard/44375.html
7. Mahesh P, Tiwari A, Jin C, Kumar PR, Reddy ALN, Bukkapatanam STS, Gupta N, Karri R (2021) A survey of cybersecurity of digital manufacturing. Proc IEEE 109:495–516
8. Mullet V, Sondi P, Ramat E (2021) A review of cybersecurity guidelines for manufacturing factories in industry 4.0. IEEE Access 9:23235–23263
9. Britannica E (2020) Manufacturing. https://www.britannica.com/technology/manufacturing
10. MORPHISEC (2021) Understanding the highly targeted manufacturing threat landscape. https://engage.morphisec.com/2021-manufacturing-cybersecurity-threat-index
11. Tuptuk N, Hailes S (2018) Security of smart manufacturing systems. J Manuf Syst 47:93–106
12. Rai R, Tiwari MK, Ivanov D, Dolgui A (2021) Machine learning in manufacturing and industry 4.0 applications. Int J Prod Res 59:4773–4778
13. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z (2019) XAI—Explainable artificial intelligence. Sci Robot 4:eaay7120
14. Srivastava G, Jhaveri RH, Bhattacharya S, Pandya S, Maddikunta PK, Yenduri G, Hall JG, Alazab M, Gadekallu TR (2022) XAI for cybersecurity: state of the art, challenges, open issues and future directions
15. Machlev R, Heistrene L, Perl M, Levy K, Belikov J, Mannor S, Levron Y (2022) Explainable artificial intelligence (XAI) techniques for energy and power systems: review, challenges and opportunities. Energy AI 100169
16. Alicioglu G, Sun B (2022) A survey of visual analytics for explainable artificial intelligence methods. Comput Graph 102:502–520
17. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R et al (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

18. Schoenborn JM, Althoff K-D. Recent trends in XAI: a broad overview on current approaches, methodologies and interactions. In: ICCBR Workshops, pp 51–60
19. Das A, Rad P (2020) Opportunities and challenges in explainable artificial intelligence (XAI): a survey. arXiv:2006.11371
20. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160
21. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv:1412.6572
22. Srivastava G, Jhaveri RH, Bhattacharya S, Pandya S, Maddikunta PKR, Yenduri G, Hall JG, Alazab M, Gadekallu TR et al (2022) XAI for cybersecurity: state of the art, challenges, open issues and future directions. arXiv:2206.03585
23. Wang M, Zheng K, Yang Y, Wang X (2020) An explainable machine learning framework for intrusion detection systems. IEEE Access 8:73127–73141
24. Mane S, Rao D (2021) Explaining network intrusion detection system using explainable AI framework. arXiv:abs/2103.07110
25. Wang Y, Xu L, Liu W, Li R, Gu J (2022) Network intrusion detection based on explainable artificial intelligence
26. Amarasinghe K, Kenney K, Manic M. Toward explainable deep neural network based anomaly detection. In: 2018 11th international conference on human system interaction (HSI), pp 311–317
27. Hoang NX, Hoang NV, Du NH, Huong TT, Tran KP et al (2022) Explainable anomaly detection for industrial control system cybersecurity. IFAC-PapersOnLine 55:1183–1188
28. Khan IA, Moustafa N, Pi D, Sallam KM, Zomaya AY, Li B (2022) A new explainable deep learning framework for cyber threat discovery in industrial IoT networks. IEEE Internet Things J 9:11604–11613
29. Brito LC, Susto GA, Brito JN, Duarte MA (2022) An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. Mech Syst Signal Process 163:108105
30. Hwang C, Lee T (2021) E-SFD: explainable sensor fault detection in the ICS anomaly detection system. IEEE Access 9:140470–140486
31. Huong TT, Bac TP, Ha KN, Hoang NV, Hoang NX, Hung NT, Tran KP (2022) Federated learning-based explainable anomaly detection for industrial control systems. IEEE Access 10:53854–53872
32. Shin H-K, Lee W, Yun J-H, Kim H. HAI 1.0: HIL-based augmented ICS security dataset. In: Proceedings of the 13th USENIX conference on cyber security experimentation and test, p 1
33. Laso PM, Brosset D, Puentes J (2017) Dataset of anomalies and malicious acts in a cyber-physical subsystem. Data Brief 14:186–191
34. Goh J, Adepu S, Junejo KN, Mathur A (2016) A dataset to support research in the design of secure water treatment systems. In: Critical information infrastructures security: 11th international conference, CRITIS 2016, Revised selected Papers 11. Springer, Paris, France, pp. 88–99
35. Breque M, De Nul L, Petridis A (2021) Industry 5.0: towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Luxembourg, LU
36. Li Y, Yu Y, Susilo W, Hong Z, Guizani M (2021) Security and privacy for edge intelligence in 5G and beyond networks: challenges and solutions. IEEE Wirel Commun 28:63–69
37. Xu D, Li T, Li Y, Su X, Tarkoma S, Jiang T, Crowcroft J, Hui P (2020) Edge intelligence: architectures, challenges, and applications. arXiv:2003.12172
38. Huong TT, Bac TP, Long DM, Thang BD, Binh NT, Luong TD, Phuc TK (2021) Lockedge: low-complexity cyberattack detection in IoT edge computing. IEEE Access 9:29696–29710
39. Huong TT, Bac TP, Long DM, Luong TD, Dan NM, Thang BD, Tran KP et al (2021) Detecting cyberattacks using anomaly detection in industrial control systems: a federated learning approach. Comput Ind 132:103509

40. Naik B, Mehta A, Yagnik H, Shah M (2021) The impacts of artificial intelligence techniques in augmentation of cybersecurity: a comprehensive review. Complex Intell Syst 1–18
41. Khan T, Alam M, Akhunzada A, Hur A, Asif M, Khan MK (2019) Towards augmented proactive cyberthreat intelligence. J Parallel Distrib Comput 124:47–59
42. Zheng NN, Liu ZY, Ren PJ, Ma YQ, Chen ST, Yu SY, Xue JR, Chen BD, Wang FY (2017) Hybrid-augmented intelligence: collaboration and cognition. Front Inf Technol Electron Eng 18:153–179
43. Cichocki A, Kuleshov AP (2021) Future trends for human-AI collaboration: a comprehensive taxonomy of AI/AGI using multiple intelligences and learning styles. Comput Intell Neurosci 2021
44. O'Connell E, Moore D, Newe T. Challenges associated with implementing 5G in manufacturing. In: Telecom, volume 1, MDPI, p 5
45. Mogyorósi F, Revisnyei P, Pašić A, Papp Z, Törös I, Varga P, Pašić A (2022) Positioning in 5G and 6G networks—A survey. Sensors 22:4757
46. Chen S-Z, Kang S-L (2018) A tutorial on 5G and the progress in china. Front Inf Technol Electron Eng 19:309–321
47. Cheng J, Yang Y, Zou X, Zuo Y (2022) 5G in manufacturing: a literature review and future research. Int J Adv Manuf Technol 1–23
48. Chehri A, Zimmermann A. 5G assisted smart manufacturing and industrial automation. In: International conference on applied human factors and ergonomics. Springer, pp 385–393
49. Mohan JP, Sugunaraj N, Ranganathan P. Cyber security threats for 5G networks. In: 2022 IEEE international conference on electro information technology (eIT), pp 446–454
50. Rojas RA, Rauch E (2019) From a literature review to a conceptual framework of enablers for smart manufacturing control. Int J Adv Manuf Technol 104:517–533
51. Noor-A-Rahim M, Firyaguna F, John J, Khyam MO, Pesch D, Armstrong E, Claussen H, Poor HV (2022) Toward industry 5.0: intelligent reflecting surface in smart manufacturing. IEEE Commun Mag 60:72–78
52. Wang B, Zheng P, Yin Y, Shih A, Wang L (2022) Toward human-centric smart manufacturing: a human-cyber-physical systems (HCPS) perspective. J Manuf Syst 63:471–490
53. Mucchi L, Jayousi S, Caputo S, Panayirci E, Shahabuddin S, Bechtold J, Morales I, Stoica R-A, Abreu G, Haas H (2021) Physical-layer security in 6G networks. IEEE Open J Commun Soc 2:1901–1914
54. Porambage P, Gür G, Osorio DPM, Livanage M, Ylianttila M. 6G security challenges and potential solutions. In: 2021 joint european conference on networks and communications & 6G summit (EuCNC/6G Summit). IEEE, pp 622–627
55. Tang F, Kawamoto Y, Kato N, Liu J (2019) Future intelligent and secure vehicular network toward 6G: machine-learning approaches. Proc IEEE 108:292–307
56. Sakhnini J, Karimipour H, Dehghantanha A, Parizi RM, Srivastava G (2021) Security aspects of internet of things aided smart grids: a bibliometric survey. Internet Things 14:100111
57. Zaidi F, Castagliola P, Tran K, Khoo M (2019) Performance of the hotelling t 2 control chart for compositional data in the presence of measurement errors. J Appl Stat 46:2583–2602
58. Tran KP (2016) The efficiency of the 4-out-of-5 runs rules scheme for monitoring the ratio of population means of a bivariate normal distribution. Int J Reliab Qual Saf Eng 23:1650020
59. Morichetta A, Casas P, Mellia M. Explain-it: towards explainable AI for unsupervised network traffic analysis. In: Proceedings of the 3rd ACM context workshop on Big data. Machine learning and artificial intelligence for data communication networks, pp 22–28
60. Zolanvari M (2021) Wustl-IIoT-2021
61. Provotar OI, Linder YM, Veres MM. Unsupervised anomaly detection in time series using lstm-based autoencoders. In: 2019 IEEE international conference on advanced trends in information theory (ATIT). IEEE, pp 513–517
62. Pereira J, Silveira M. Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 1275–1282

63. Principi E, Rossetti D, Squartini S, Piazza F (2019) Unsupervised electric motor fault detection by using deep autoencoders. IEEE/CAA J Autom Sinica 6:441–451
64. Nguyen H, Tran KP, Thomassey S, Hamad M (2021) Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with the applications in supply chain management. Int J Inf Manag 57:102282
65. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems, vol 30
66. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144
67. Štrumbelj E, Kononenko I (2014) Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 41:647–665
68. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2:56–67
69. Zolanvari M, Teixeira MA, Jain R. Effect of imbalanced datasets on security of industrial iot using machine learning. In: 2018 IEEE international conference on intelligence and security informatics (ISI), pp 112–117

Wearable Technology for Smart Manufacturing in Industry 5.0

Tho Nguyen, Kim Duc Tran, Ali Raza, Quoc-Thông Nguyen, Huong Mai Bui, and Kim Phuc Tran

Abstract The innovation of wearable Internet of Things devices has fuelled the transition from Industry 4.0 to Industry 5.0. Increasing resource efficiency, safety, and economic efficiency are some of the main goals of Industry 5.0. Herein, wearable Internet of Things devices is parallel to humans to optimize human tasks and meet a new Industry's requirements. Integrating artificial intelligence algorithms and IoT into wearable technologies and the progress of sensors has created significant innovations in many fields, such as manufacturing, health, sports, etc.. However, wearable technologies have faced challenges and difficulties such as security, privacy, accuracy, latency, and connectivity. More specifically, the increasingly massive and complex data volume has dramatically influenced the improvement of the limits. However, these challenges have created a new solution: the federated Learning algorithm. In recent years, federated learning has been implemented with deep learning and AI to enhance powerful computing with big data, stable accuracy, and ensure the security of edge devices. In this chapter, the first objective is to survey the applications of wearable Internet of Things devices in industrial sectors, particularly in manufacturing. Second, the challenges of wearable Internet of Things devices are discussed. Finally, this chapter provides case studies applying machine learning, deep learning, and federated learning in fall and fatigue classification. These cases are the two most concerning work efficiency and safety topics in Smart Manufacturing 5.0.

Keywords Wearable technology · Smart Manufacturing · Industry 5.0

T. Nguyen · K. D. Tran (✉)
International Research Institute for Artificial Intelligence and Data Science, Dong A University, Da Nang, Vietnam
e-mail: ductk@donga.edu.vn

A. Raza · K. P. Tran
University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

Q.-T. Nguyen
HIGHFI Lab, Sofft Industries, Lyon, France

H. M. Bui
University of Technology, VNU-HCM, Vietnam

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_11

1 Introduction

In recent decades wearable devices have continuously improved with new versions in response to market trends and requirements in Industry 4.0 and novel Industry 5.0; now, they are undergoing a significant increase. The rapid proliferation has been attributed to billions of new Internet of things devices, the growth of the electronic sensor industry [1] and the development of advanced Machine Learning (ML) algorithms, Deep Learning (DL), Federated Learning (FL), and Artificial Intelligence (AI), as well as. Apart from these, a massive demand for applying wearable devices with different goals in terms of medicine (devices assist doctors in monitoring patients' electrocardiograms), sports (athletes, coaches, and doctors may now measure functional motions and workload by using wearable sensors to boost performance and eliminate the risk of injury thanks to technological improvements) is an indispensable factor [2, 3]. Smart wearables are like conventional clothes, equipped with connected sensors to capture internal and/or ambient information without interfering with the user's activity. In industrial scenarios, wearable sensors are used not only in collecting data from the human body but also in the environmental workplace. They are used in monitoring and analyzing ergonomic risk factors in workspace [4]. For transportation, Huang et al. [5] combined an accelerometer, gyroscope, and magnetometer to detect unsafe driving activity (visual and manual distractions, unsafe lane change and turn, and drowsy state). If there is any dangerous behavior of drivers, the system notices early warnings before dangers occur. This device is advantageous for drivers who are frequently fatigued or drowsy from driving a long distance. For data analytics based on smart wearables, Althobaiti et al. [6] used ML with a triaxial accelerometer to detect types of falls and activities of daily life, Zhou et al. [7] applied DL, and wearable devices to catch students' learning states, Raza et al. [8] applied Explainable Artificial Intelligence (XAI) to classify arrhythmias by electrocardiography (ECG) data. In recent years, there has been an explosion of research on wearable devices related to human activity recognition (HAR), which can also be regarded as a precursor to the fifth Industrial Revolution.

On the other hand, though there are more studies about WIoT devices and new methods of analyzing data, there have still been many difficulties and challenges, such as privacy, big data, robust computation, live feedback, limited battery life, and users' agreement. Moreover, many new studies are still conducted in laboratories so that researchers can handle confounding factors, such as reducing noisy data while analyzing. A WIoT device is built with solid computing (AI, ML), and many tasks will consume more energy from the battery. Thus, computational optimization and long-life battery remain the bottleneck for wearable IoT devices now [9]. The latency challenge is also of concern for real-time feedback, and it is improved and optimized in many aspects, such as computation and communication. This way can keep users safe and save their lives in urgent accidents [10].

In short, smart wearable sensors are used in many fields with the general goal of an intelligent assistant: to increase human productivity, safety, and health management. The device's features are drastically innovated towards each culmination of each aspect in each field. Although challenges have always existed, many recent studies have made progress in solving them. Thus, when the difficulty is solved, all companies in the Industry 4.0 area will probably begin to put wearable sensors into the company's operating system. As such, an Industrial Revolution 5.0 would be witnessed clearly when the wearable space is its culmination. The chapter is organized as follows: in Sect. 2, we briefly describe the development of wearable technology. State-of-the-art wearable Internet of Things devices is presented in Sect. 3. Section4 contains a literature review related to wearable IoT technology for smart manufacturing in Industry 5.0. The difficulties and challenges when applying wearable technology for manufacturing in Industry 5.0 are described in Sect. 6. Section 5 presents a short review of Human Cyber-Physical Systems, and the smart wearable role of the system in Smart Manufacturing. Case studies being relevant to proposed frameworks using ML based on DL and DL for wearable technology are given in Sect. 7. Finally, Sect. 8 provides some concluding remarks.

2 Wearable Technology

Ensuring workplace safety remains a pressing challenge in the industrial sector, with far-reaching impacts on employee and employer productivity [11]. Nonfatal occupational injuries in manufacturing alone accounted for approximately 373,300 cases in 2020, according to the Bureau of Labor Statistics [12]. However, there are potential technological solutions that can help mitigate this issue. For example, smart wearables have emerged as a promising tool to improve workers' physical activity and biological well-being, enhancing workplace safety [13].

Research on fall prevention using wearable sensors has shown reliable performance in detecting fall risks among workers, with high precision, ease of access, and feasibility of administration [14, 15]. For instance, a study of environmental sensing-based systems by Taramasco et al. [16] achieved an accuracy rate of 93%, with 95% accuracy detecting non-fall activities and 96.67% for fall activities [17]

Other studies have explored the application of wearable technology in monitoring proper body mechanics during work to prevent and reduce low back pain in nurses. Ferrone et al. [18] demonstrated the potential of wearable technology combined with machine learning to achieve the highest accuracy of 91.66% in identifying proper body mechanics. The research highlighted the usefulness of wearable technology in driving behavior shifts through biofeedback.

In addition, the development of intelligent safety helmets equipped with Inertial Measurement Units (IMU) and electroencephalography (EEG) sensors, which can analyze data using artificial intelligence algorithms, has shown promising results in reducing accident risks in industrial facilities [19, 20] reported on the use of intelligent helmets that detect harmful gases beyond acceptable limits in the mining industry. Other examples of wearable safety technology include sensor-equipped safety vests, smart eyewear, and smartwatches used for mining operations [21].

Wearable technologies are becoming increasingly prevalent in healthcare and are being utilized for various purposes related to (cardiovascular diseases, hypertension) and neurocognitive disorders (Parkinson's disease, Alzheimer's disease) [22, 23]. Specifically, cardiovascular diseases (CVDs) have emerged as a significant global health issue, with CVDs being the leading cause of death. In 2019, mortality related to CVDs was estimated at 17.9 million cases, with heart attacks and strokes accounting for 85% of these deaths [24]. Fortunately, real-time wearable technology can assist medical staff in managing and monitoring patients and reducing emergent cases, such as stroke and acute heart failure. For example, pulse oximeters are a specific type of wearable device that can be used to monitor blood oxygen levels, which is critical for patients with COVID-19, as their blood oxygen levels frequently drop to less than 92%. These devices allow patients to monitor their health status and avoid emergencies. Wearable technology has also shown significant promise in improving rehabilitation outcomes and mobility in orthopedic surgery, specifically total knee arthroplasty rehabilitation. Medical doctors can use biofeedback from various parameters, such as knee angulation moment, vertical ground reaction force, and weight-bearing balance, to better understand patients' gait and rehabilitation progress [25–28]. Wearable technology represents a promising approach to addressing critical health challenges and improving patient outcomes across various medical conditions.

In sports, injuries are common in competition and practices, which poses low permanent performance despite rehabilitation. Therefore, there is a high demand for smart wearables that assist athletes and coaches in management, prevention, and rehabilitation. An oft-cited example is wrist-worn wearables in sports are used for four purposes: monitoring of activities [29–31], classification of activities [32, 33], performance improvement [34] and preventing injury [35].

The weakness of wearable sensors is the need for more connection, information, and associated features that are only beneficial for individuals, leading to difficulty monitoring and managing the entire manufacturing. Thus, the Internet of Things is integrated into wearable sensors to connect all users' wearable sensors in a network, from which managers can have holistic information of the entire system. Section 3 will review the combination of IoT with wearable sensors to bridge the gaps that have existed for wearable devices in an industrial area.

3 Wearable Industrial Internet of Things

In the last decade, wearable devices without IoT had a bigger size and less memory, processing, and networking capability than wearable IoT devices. The reason for these issues is that they must operate with multitasks, from collecting data and storage to exporting results. By multitasks and its initial default, wearable devices faced obstacles in updating, which steadily depleted accurate models and results. Thereby, the Internet of Things is applied to overcome problems. Particularly, wearable IoT devices can connect all wearable sensors, process and analyze data on the cloud, and then send it back to wearable sensors, which makes the significant innovation in optimizing efficient energy, memory, and size [36]. As an intermediary hub between wearable devices and IoT servers, the IoT gateway enables real-time bidirectional communication between users and machines (manufacturing), patients, and medical staff (hospital) [37]. In a connected world, computers, intelligent gadgets, wearable devices, and other items can communicate and exchange information at any time and distance [38]. In recent years, WIoT devices have been applied in various fields to assist users in making intelligent decisions [39].

3.1 Application of Wearable Industrial Internet of Things

Wearable applications without integrating IoT have been presented in Sect. 2. In this subsection, wearable sensors combined with IoT is given, and state-of-the-art of WIoT devices for monitoring, classification, and predicting in various field.

In logistics, end-to-end manufacturing supply chain management is possible by implementing IoT sensors. Manufacturers may track the movement of vehicles transporting supplies and commodities, view extensive information on things in warehouses, and regulate the conditions (temperature, humidity) in which products are stored or moved. In addition, WIoT devices support managers and workers in monitoring remote processes.

In transportation, the World Health Organization has predicted that road traffic collisions will move to be the fifth leading cause of death in 2030 [40]; most accidents involve manual distraction (e.g., hands off the wheel), inattentiveness (e.g., eyes off the road), tiredness and intoxication [41]. Studies have approached applications of WIoT technologies in monitoring driver's distractions and alerting a warning harm behaviors in real-time if occurred to resolve this problem [5]. Detects erratic handle movement using an intelligent watch [42, 43]. Research by Huang et al. [5] used a driver monitoring system, "MagTrack" with a single magnetometer sensor; monitoring is based on tracking magnetic tags with activities: bimanual steering, visual and manual distractions, lane changes and turns, and finally, warnings are alerted through a safe driving app. Authors tested MagTrack devices with over 500 driving trials and over 500 min of 10-object road driving; results are specific (87%) and accurate (90%) in detecting dangerous driving behavior. Detecting sleep deprivation, automatically sending a location of the accident to the registered phone number by a message [44].

In manufacturing, safety, and productivity are the essential criteria in the industry. Though workers equipped with personal protective equipment [45] and guided to work the right way to avoid accidents in the workplace, almost occupational accidents were seen in the industrial sector [46].

Though IoT-enabled wearables are applied in various areas, they are found chiefly in sectors concerned with physical and psychological issues, such as healthcare and sports. In the healthcare sector, WIoT technologies proved superior features in terms of real-time biofeedback, leading to early diagnosis and efficient treatment To et al. [47], early detection of fever patients by IoT smart thermometers Meraj et al. [48], predicting and warning of the most suspicious areas, where there are most COVID-19 cases in Meraj et al. [49]. In the sports sector, wearable and IoT devices support tracking athletes' internal and external workloads and mental situations Passos et al. [50]. A study by Li et al. [51] has shown that WIoT technologies obtained the highest accuracy (98.22%) in monitoring the sports person's health.

3.2 Artificial Intelligence in Wearable Technology

The data analysis task's first steps for WIoT are classification and detection. ML became an inevitable algorithm to predict specific features. The success of ML applications for wearable sensors and IoT systems in various fields is identified in the examples below. Research by Borthakur et al. [52] used unsupervised cluster analysis in analyzing pathological speech data of Parkinson's disease (PD) patients. Rodríguez-Martín et al. [53, 54] applied Support Vector Machine (SVM) in classifying freezing of gait (FOG) of PD patients with a triaxial accelerometer worn at the waist, the results showed that 84.1% sensitivity, 83.4% specificity, and 85.0% accuracy for detecting FOG. The authors concluded that these results could be higher when analyzing large data. In safe transportation, Mehdizadeh et al. [55] used ML techniques to predict unsafe driving risk among commercial truck drivers through dependent variables in terms of driver's characteristics, weather conditions, and day/time categories. Their results showed that they could predict safety-critical events 30 minutes earlier; among learning algorithms (Generalized Linear Model, Lasso, Ridge, Classification And Regression Tree, Random Forest, Naive Bayes, Neural Network, SVM, Extreme Gradient Boost (XGBoost)), the model of XGBoost algorithm gives the best result with an area under the curve (76.5%). In lean management, Hofmann et al. [56] used long-term short-term memory networks (LSTM) to classify HAR to production processes to detect wasteful motion, increasing production in a factory. Manjarres et al. [57] applied Random Forest algorithms in wearable devices to classify the activity being performed and track the physical workload. The results achieved an accuracy of 97.5% for data from nine subjects and an accuracy of 92% during real-time testing for 20 subjects. Using only one classification model in HAR detection may have limitations for a specific situation. Thus Nguyen et al. [58] proposed an improved ensemble machine learning algorithm that combines multiple machine learning models to enhance the performance of HAR. The ensemble algorithm that recognizes human activities from wearable sensor data is achieved better accuracy and can be applied in various fields, particularly in healthcare and manufacturing.

Though ML algorithms are applied widely to detect signals, the AI technique is the higher degree of detection and classification task for wearable IoT devices. The wearable IoT system becomes bigger, leading to immense data, while using conventional ML algorithms may not be an excellent choice to obtain the highest accuracy results. To address this issue, AI is a more robust data processing algorithm to acquire the best performance. Combining AI and IoT wearables is crucial in improving accuracy, resulting in better decision-making. This trend has been found in some examples such as healthcare [59], detecting and managing adverse health conditions in aging populations (chronic conditions, falls, disabilities), cardiac disease detection [60], ECG detection [8], ecological environment [61]. The success of AI applications is not only a breakthrough in powerful computing but also creates new applications in the future.

IoT devices generate massive amounts of data to enhance accurate results of analytic techniques for decision-making while several organizations together collect data to achieve big data. Organizations would not fully join the program when they upload sensitive data on the same server. Facing the obstacle, in 2016, Google researchers first proposed a feasible and compelling solution to protect the private data of that organization, namely "Federated learning" [62]. This access opened a new path for organizations to collaborate on training models with the final goal of developing high-performance models while still keeping initial data in local servers, which preserves organizations' privacy. The DL architecture involves multiple federated nodes and one aggregator agent. The mechanism of DL is to compute average model weights (from clients) and then update them on a server.

In contrast, revised models will be sent back to clients for training on newly generated data. Some examples highlight the applications of DL in some areas. In manufacturing, Hegiste et al. [63] successfully experimented with DL to detect the quality of faulty images in a production line. Clients exchanged local model weights with servers in a DL system to achieve a global model. The results were that the accuracy of local test data remained the highest (99.9%) and the shorter average time for live classification. For an intelligent transportation system, Manias and Shami [64] applied DL to solve two main problems of detecting proper objects by image processing task. That is to ensure model performance in cases of traffic imbalance (i.e. in low traffic areas and high traffic areas) and the different conditions (i.e. light, wind) such as areas with many trees and shade. DL can distribute nodes' data and a completed model to nodes that need more data to train local models. In healthcare, Raza et al. [65] experimented with comparing the detection of poisoning attacks without DL and with DL; the global model with DL enabled it to perform stably in detecting poisoning attacks while increasing the error rate. For the healthcare index (i.e. ECG), Raza et al. [8] proposed using XAI and CNN in DL to classify ECG. As a result, the accuracy of the proposed framework acquired 98.9% (clean data) and 94.5% (noisy data). Other applications in healthcare informatics [66], sport [67]. Combining DL in the WIoT system is a great promise for developing models that classify and forecast with the highest accuracy, live feedback, and preserve the privacy data of organizations, which is more likely to properly prepare and equip for reaching industry 5.0 in the long term.

3.3 Internet of Things Wearable Sensors

A billion sensors are progressively married to WIoT technologies in various areas. The sensor is a pipeline for recording and transmitting data to storage or analytic bodies (cloud). It means that they are the focal point in real-world data, reliable data, and a part of correct outputs depending on the reliable sensors. In contrast, poor sensor quality poses error data (i.e. miss data, bias), consequently, severe repercussions from processing to making wrong decisions [68]. Much research has shown the critical role of sensor quality in the success of current wearable IoT devices. Examples of sensors and their functions on the market include accelerometers, gyroscopes, and magnetometers to recognize movement and activities of the human body, in industry wearable IoT devices with these sensors can monitor workers' fatigue [69, 70]. Similarly, in transportation, accelerometers, gyroscopes, and magnetometers sensors in IMU devices are used to monitor and classify unsafe behaviors of truckers by head motion and steering one hand [5, 55]. Chieh et al. [71] developed a drowsiness detection system that used an EOG sensor to monitor drivers' eye movements. An ECG sensor records electrical signals of heartbeat [72], an EEG sensor records the electrical signals of activity brain, electromyography (EMG) sensor measures electrical activity in response to a nerve's stimulation of the muscle [73], photoplethysmography (PPG) monitors heart rate [74]. A SpO2 sensor measures blood oxygen levels or oxygen saturation in blood [75]. Amyotrophic lateral sclerosis (ALS) affects nerve cells in the brain and spinal cord, causing a loss of muscle control. Patients with this disease commonly communicate by using eye movement. An Electrooculogram sensor is the best solution for supporting communications by eyes for patients with ALS. A classic example of this success is Stephen Hawking, communicated by EOG assistive technology [76]. In the following parts, we present physical and physiological sensors integrated into WIoT sensors and their applications in manufacturing, healthcare, and sport.

3.3.1 Sensors and IoT Wearable Sensors' Applications in Safety in Manufacturing

Sensors are essential devices of WIoT technology. Some kinds of sensors attached to wearable technologies for safety in manufacturing, such as proximity sensors (where workers are near), movement sensors (detect a collision or fall, fatigue), condition sensors (heat/temperature, heart rate, etc.), exoskeleton suits (assist workers in lifting-putting heavy things, leading to less stress on their bodies). The most common sensors utilized in the workplace will be described in Table 1.

Table 1 Wearable IoT sensors for safety in manufacturing

Sensor	Sensed parameter	Case	References
IMU	Movements	Detect workers' fatigue	Li et al. [19]
EEG	Alpha and beta brainwave	Fatigue detection	Li et al. [19]
Accelerometers	Movement	Fall detection	Musngi [77]
Barometric pressure	Altitude	Fall detector	Lu et al. [78], Pierleoni et al. [79]
Moiton	Movement	Follows OSHA lifting guidelines, analyze the high risk of musculoskeletal injuries	Valero et al. [80]

3.3.2 Sensors and IoT Wearable Sensors' Applications in Healthcare

Sensors are meaningful for WIoT devices used in healthcare; combining sensors and AI can get insight into how the human body reacts to different health states based on physical and psychological conditions. Indeed, this trend was witnessed during the COVID-19 pandemic, when people experienced a tough time due to quarantine and difficulty approaching medical services face-to-face. This issue motivated innovation sensors and advanced IoT wearable devices to meet the demand for remote monitoring, assessment, and consulting of people's health at home. In Table 2, we highlight special sensors and cases that used WIoT sensors for monitoring and rehabilitation goals in healthcare.

Implementing IoT wearables in healthcare is the extreme potential for remote clinical exams and consultants if these devices are applied in specific cases such as chronic diseases (diabetes, hypertension,...), physical rehabilitation [96–98] (a stroke or musculoskeletal disorders [99]), elderly care [100], etc.. Patients with the diseases mentioned above usually need follow-up by medical staff in the assessment and treatment processes. This activity would take a long time, a IoT wearable technology with real-time biofeedback supports medical staff in effective management and monitoring of their patient's health situation, which is crucial in the long term.

3.3.3 Sensors and IoT Wearable Sensors' Applications in Sport

Sport sensors are mainly utilized to monitor functional movements and workload and analyze performance, which aims to eliminate injury and accelerate performance. Physical sensors: IMUs, pedometers, accelerometers/gyroscopes, and Global Positioning System (GPS) devices [101]. Physiologic sensors: heart rate, temperature, strain sensors. The sensors' quality and durability are the keys to collecting per-

Table 2 Wearable IoT sensors in healthcare

	Sensor	Sensed parameter	Case	References
Rehabilitation	IMUs	Movement	Stroke rehabilitation Post-traumatic rehabilitation assessment Rehabilitation assessment of patients with gait disorders	Fenu and Steri [81], Zhao et al. [82]
	Pedometer	Step counts	Chronic pulmonary disease Cardiac risk factors at entry to cardiac rehabilitation	Steele et al. [83], Savage and Ades [84], Ahola [85]
	Accelerometer	Movement	Monitoring of mobility after major surgery Measuring motor abilities following stroke	Hester et al. [86], Rand et al. [87], Cook et al. [88]
	EMG	Muscle activity	Monitoring nervous and muscle response, early detection of Parkinson's and Alzheimer's disease	Golab et al. [73], Papazian et al. [89]
Monitoring	ECG	Heart rate	Monitoring babies in newborn or in patients	[90, 91]
	EEG	Alpha and beta brainwave	monitoring and detecting epileptic seizures	Giansanti et al. [92]
	PPG	Blood pressure	Hypertension management	Konstantinidis et al. [93]
	Glucose	Blood glucose	Continuous glucose monitoring diabetes management	Cappon et al. [94], Zhang et al. [95]

Table 3 Wearable IoT sensors in sport

Sensor	Sensed parameter	Case	References
Pedometer	Walking distance, step counts	Walking distance and step counts	Ahola [85]
Accelerometer	Movement	Analyse athlete's performance Classification of team sport activities	James et al. [102], Wundersitz et al. [103]
GPS	Position, distance	Calculate speed, and position of the receiver Detect fatigue in matches, identify periods of most intense play	Portas et al. [104], Izzo et al. [105], Oxendale et al. [106]
PPG	blood flow-pressure	Monitoring sports-related concussion Assessing heart rate regulation reactions to extreme sport activity	Korobeynikov et al. [107], Bishop et al. [108]
Thermography	Temperature	Monitoring fatigue during exercise	Formenti et al. [109], Hadžić et al. [110]

formance data, and they suffer from solid physical activity. Further description of sensors and their application in the sports field is found in Table 3.

4 Wearable IoT Technology for Smart Manufacturing

In manufacturing, wearable Internet of Things systems is a special type of supported tech that takes advantage of data to boost efficiency and productivity. Resulting in a competitive edge for companies in the market. WIoT devices also have great potential for quality control, traceability, and overall supply chain efficiency. The sensors embedded in workers in the manufacturing machine track their performance, facilitating modification and improvement based on requirements. IoT tasks connect all workers and related machines into one system. To improve process efficiency, companies are creating a system that can be easily managed so that all production lines can be monitored in real-time. In a competitive market for manufacturing, focusing on efficiency for continuous improvement and growth is vital. Some companies have found advantages of wearable devices that help them reach a smart factory by utilizing efficient time and decreasing downtime. Taking the Dortec plant in Ontario as a practiced example, they improved systemic downtime as part of the overall continuous improvement project through wearables, virtual reality (VR), and augmented reality (AR). Finding those individuals by phone, radio, or physical search inside an enormous and loud manufacturing facility can be hard and time-consuming. Their wearables collected intelligence that made waste more visible than in the past. The key waste areas are minimizing transportation, reducing downtime, making rework visible, and better-utilizing skills, capabilities, and talents [111]. Employees also benefit from smart glasses, reducing time for carrying scanners, written papers, and hand-free access to computer-generated information, which increases productivity [112]. For example, in industrial food, employees manage food safety using IoT-based wearable devices that monitor overdue food storage and present notifications to relevant personnel. Data findings are delivered to cloud storage, where management may examine them instantly and take the appropriate corrective action to solve issues. This benefit reduces the time it takes to handle important food safety issues and does away with the need to dig through mountains of paperwork to find issues. Moreover, employees can spend less time collecting data via IoT wearables (automating the process of recording data) than paper systems (writing data on paper). The most important feature is that wearable IoT sensors work well in anomaly detection of food if their parameters fall out of the safe set range. Some types of IoT wearables (tasks) applied in industrial food safety are Smart Eyewear (visualization, communication), Wearable Cameras (photographic or audio video evidence), Smart Swabs (real-time bacterial activity) [113]. A special thing in the manufacturing area is a protocol for third-party confirmation before any worker accesses devices. Therefore, wearable technology can remotely restrict, allow, and location-track worker activities. In 2018, Yang et al. [114] suggested a prototype incorporating textile electrodes, motion sensors, and real-time data processing through a mobile application for autonomous risk assessment of physical activity. Bootsman et al. [115] used sensors placed on clothes to alert the nurses when they are in the wrong posture at work. In the industrial sector and firefighting field, wearable sensors are also applied in the undergarment for monitoring physiological extremes as referred to Salim et al. [116]. The appearance of voice-assisted tech (Google's Assistant, Apple's Siri, and Samsung's Bixby) shows the huge potential when integrating voice assistants into a wearable IoT system. The sector more advantageous than smart devices above would be manufacturing, which is found with line productions. Imaginatively, a worker can control a machine using a wearable IoT device with a voice assistant instead of working with the machine by hand. In addition, loss of time of movement in a company is minimized when staff can remotely order machines, which helps staff work with greater productivity. Thus, voice assistants would contribute to the rapid development of wearable IoT devices to reach Industry 5.0. The higher level of voice assistants and the virtual entity is the human digital twin (HDT), that is, the digital representation of the real-world human or real-world twin [117]. In the Industry 4.0 context, a digital twin that comprises wearable devices, IoT networks, and AI is developed to boost workforce productivity [118]. Moreover, it can simulate the future behavior of operators, reduce physical workload and increase the well-being of operators. Towards Industry 5.0, Human-Cyber-Physical systems (HCPS) are the most concerned with optimized manufacturing levels. Wearable technologies connect humans and machines and supply rich information about the operators, surrounding environment, and machines [119, 120]. Besides concerns related to efficient productivity, many companies use IoT-enabled wearables to boost safety and prevent hazards for workers. Health monitoring and management for manufacturing workers in adverse conditions is not only for their safety but also for cost efficiency [121].

5 Human Cyber-Physical Systems and Smart Garment's Adoptions in Smart Manufacturing

The revolution of intelligent manufacturing has experienced many developing stages and is now considered the Human Cyber-Physical Systems (HCPS) stage (Fig. 1). This system's goal obtained optimized levels in manufacturing perspectives, replacing humans' operator role by working directly with the physical system, such as sensing, analysis, control, and decision-making [119]. A strength point of HCPS is that it can perform self-learning and cognition by using new-generation AI technology and be fed by human knowledge into the system. Thus, it can improve and generate new knowledge by itself to optimize necessary resources (efficient power or computing) [122]. The smart garment is a prerequisite for building an HCPS ecosystem, as it is a connector that helps humans connect easier to physical systems [123]. In smart manufacturing, workers can also use WIoT devices to improve existing processes and functions within industrial and commercial settings. For HCPS, the voice technology of WIoT devices can transform some operations with machines into hands-free devices, which will remove the friction with handheld devices and enable a more seamless workflow, translating to more incredible productivity among workers. With multisensory devices combining VR and AR, HCPS workers can avoid adverse factory conditions directly or remotely through mixed reality [124].

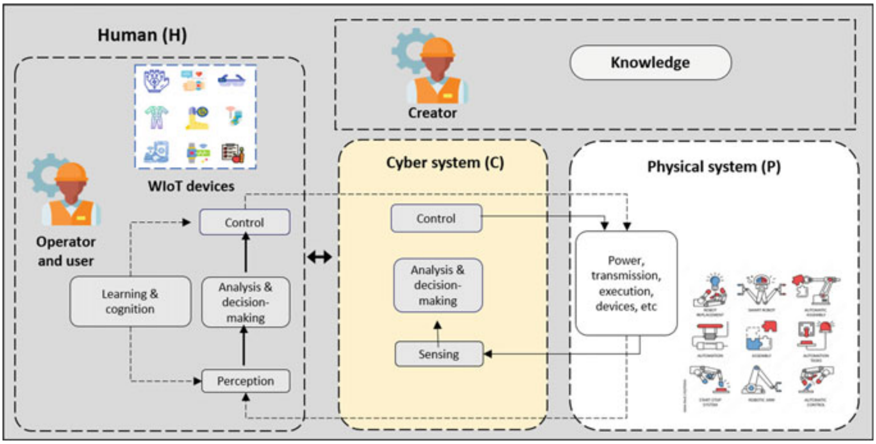


Fig. 1 Human cyber-physical systems and smart Garment's adoptions in smart manufacturing

In HCPS, IoT-enabled wearables ensure safety and prevent hazards for workers at the highest level in intelligent factories, thus augmenting their performance in the challenging environment. As a result, reducing risk factors is the best primary prevention for monitoring and managing workers' health, lowering their financial burden. Wearable sensors are already being applied as part of the research of [125] to recognize emotion, allowing humans and machines to collaborate more efficiently and collaboratively.

The research effort in integrating senior computing techniques into wearable IoT devices obtains the goal of productivity, safety, and efficiency in the manufacturing industry. However, almost all of the current IoT intelligent clothing does not yet deal with the difficulties and challenges mentioned in Sect. 6. Thereby, blockchain or any other Distributed Ledger Technology should be considered to receive, validate, store, and share the collected data to avoid untrusted sources and FL to protect privacy and security [126] and promote wearable applications for HCPS in manufacturing. Industry 5.0 is close when we incorporate all techniques mentioned above into IoT wearable technology.

6 Difficulties, Challenges, and Perspectives for Application of Wearable Technology Techniques for Smart Manufacturing in Industry 5.0

Several papers have reviewed the difficulties and challenges of wearable devices when combining IoT systems in various fields. A survey of Chan et al. [127] showed some challenges of wearable devices, such as efficiency, social inclusion, cost, and ethic. Sabry et al. [128] identified challenges of ML applications on wearable devices from design to deployments, such as storage, power consumption, reliability, communication, security, and privacy. Moreover, many products still need to be applied practically in mass production due to their complexity and economic profit. Insights into the current state-of-the-art and existing challenges related to IoT wearable sensors facilitate the development of a roadmap for a new generation of innovations and breakthroughs.

6.1 Data Security and Privacy

Security and privacy have been one of the leading concerns of engineers and experts in the IoT field. These devices often store patients' sensitive information about personal information, health status, medical information, etc.. Therefore, it is challenging to guarantee the security of users' information and their acceptance of wearable IoT systems. Moreover, privacy information is also the goal of cybercrime [129, 130]. In 2015, 113.2 million healthcare records were stolen in a breach in the healthcare industry [131]. This problem causes a negative cognition for users and organizations [132]. Data encryption solutions for privacy information preservation affect wearable IoT devices' computing power, causing delays in the data processing. Besides, efficient privacy protections must be done while ensuring a low processing delay [133]. The current solution is the FL technique developed to address the privacy leakage problem and data processing delay. However, it is still the experiment duration.

6.2 Data Processing

The demand for enhancing overall productivity, safety, and quality of healthcare management is increasing when wearable IoT systems are widely utilized in the industrial field. Several analytic algorithms are developed to boost accurate results and real-time feedback for users. Successful cases integrating advanced analytical techniques of wearable IoT devices are reviewed in Sect. 3.2. However, most cases were studied in laboratory environments and are still in-control patterns (less heterogeneous data, manageable data, etc). In practice, passive data collection may lead to rising massive data sets, which can generate redundant, noisy, and heterogeneous data. These can cause pitfalls of data storage, extensive data processing, and non-real-time feedback to users [134]. On the one hand, gathering a massive amount of data is a challenge for many fields; therefore, it has yet to be taken advantage of the true potential of big data. Now, rare industries have more chances than healthcare, where almost wearable sensors and IoT are applied to monitor and improve human health and prevent acute diseases (stroke) and accidents (falls). The proof has been witnessed through the COVID-19 pandemic [135]. As mentioned above, the truth is that conventional software and analytic algorithm can not handle big data [136]. Therefore, it should dig deeper into advanced analytical techniques with fast computing power to have a perfectly portable IoT system. Though Big Data is one of the biggest challenges, it is the most incredible opportunity for the coming years for wearable IoT systems when the Industry 5.0 tendency is concerned in recent years.

6.3 Battery Life Expectancy

The life battery feature of wearable IoT technology is indispensable in developing and applying advanced data analysis techniques or user acceptance. Depending on the uses, the loss of a battery in an IoT system might have minimally disruptive to catastrophic consequences. Industrial operations may grind to an abrupt halt, a farmer's agricultural harvest may be disrupted, or self-driving vehicles may lose control. Wearable devices with short battery life can have limited computing capability or stringent latency [137]. Direct communication from a node to sink in Wireless Body Area Network may often lead to rapid energy depletion of nodes [138]. Human context recognition from wearable sensors requires constant monitoring parameters of the body and environment, thereby continuously data collection and communication tasks quickly depleting life battery [139]. However, there have been some solutions to the problem. Djelouat et al. [140] have found that the number of cores speeds up the execution time and further optimizes energy consumption. Sun et al. [137] proposed 5G for wearables' communication, Roy et al. [138] presented enhancing communications in medical IoT using the Markov decision process. Another solution is fog computing, which significantly impacts on energy consumption of IoT sensors. Amiri et al. [141] fog computing can mitigate by offloading computational tasks from layers, therefore not only meeting the wearable IoT devices' limited computational capacity but also enabling the use of local closed-loop energy optimization algorithms to increase the lifespan of the battery. Finally, in the future, the strategy to restrict energy expenditure would be solar energy for sensors, generating power from vibrations. The higher performance efficiencies and lower operational power would incentivize companies to use wearable IoT technology.

6.4 Smart Garment for Human Cyber-Physical Systems in Smart Manufacturing

When the human is a centrality of the HCPS to connect physical systems via the WIoT system, smart garments' components become vital due to errors or accidents stemming from wearable sensors, which impacts not only the local system but also the whole system, leading to massive consequences.

Besides previous challenges of the WIoT system: privacy and security, data processing, and battery life expectancy, in HCPS, challenges of WIoT are forced to the highest level. Particularly data that simulate the human behavior model must achieve the highest accuracy and stability, which will make the physical system operate correctly work. In contrast, errors from data collection or data transformation will pose the physical system to misunderstand human behaviors, resulting in lower learning from the physical system. The physiological, psychological, and behavioral perspectives of humans make it challenging to extend system identification or other modeling techniques to capture human behavior [142]. In Griffor et al. [143], flawed understandings of system behavior affect overall performance criteria.

On the one hand, any attack through input sources or connections through the WIoT system by low cybersecurity will cause catastrophic repercussions for the whole system, including product quality, product performance, and safety, leading to being lost of huge finance or even threatening human life. According to Alguliyev et al. [144], attacks can be hardware tampering of sensors, packet replay attacks, and information theft in communication channels of a closed loop with sensors. From the above scenarios, cybersecurity is a requirement for HCPS.

6.5 Factors Influence User Acceptance of Wearable IoT Devices

The development of IoT wearables clearly shows the number of wearables on the market and users' intentions. However, some factors still hinder user acceptance when wearing wearable IoT technology. This Subsection has comprehensively illustrated these factors. Security and privacy factors are the most important for users. Smart wearables often collect user data through sensors, cameras, etc. These data could contain sensitive data, and customers fear the information protection capability. Some studies have shown that security and privacy are significantly associated with consumers' wearable IoT adoption [132]. A survey by Papoutsi et al. [145] has shown that 79% attendants were concerned about their information via Electronic Health Records, and 47% participants believed Electronic Health Records would be less secure. The social norms factor for IoT wearables is the reaction of the public (beliefs in a population) towards it. The power of social norms in this context could become a crucial determinant of society to wearable technologies [146]. Some research has shown that social norms influenced users' IoT wearables acceptance [132, 147]. The factor "truth" was found to have a statistically significant positive relationship with behavior intention [132, 147, 148]. The technological feature is one of the critical factors that impact users' intention [147, 149].

7 Case Studies

7.1 Case study 1: HAR with Federated Learning

In this case, we apply Convolutional Neural Network (CNN) and a fully connected neural network in a federated setting for HAR. Combining the techniques above is a novelty approach in training algorithms, this technique will be the tendency to align with big data in the future and enhance the accuracy of analysis for HAR studies. The question of the case study is whether implementing FL in classifying human activities shows higher accurate results than without combining FL.

7.1.1 Dataset

The raw WSDM dataset in the research of Kwapisz et al. [150] is used in this case study. The accelerometer sensor of an Android phone collected the data. Twenty-nine volunteers carried a phone in their front pants leg pocket while performing activities. The original raw dataset has 1109942 examples, data sources at: https://www.cis.fordham.edu/wisdm/dataset.php, then steps of visualizing and processing data. Finally, the number of examples is obtained 1,098,204, in which frequency of *walking* is 424,398 (38.6%), *jogging*: 342,176 (31.2%), *upstairs*: 122,869 (11.2%),

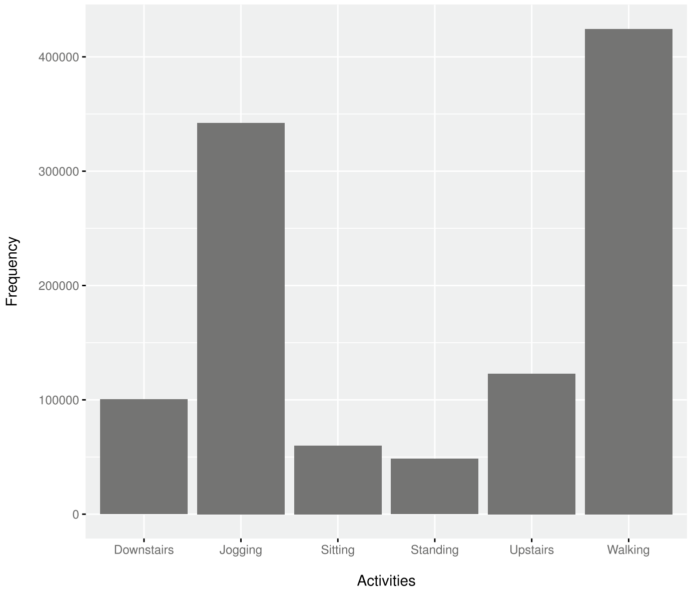


Fig. 2 The distribution of the activities

downstairs: 100,427 (9.1%), *sitting*: 59,939 (5.5%), *standing*: 48,395 (4.4%). The distribution of the activities in the dataset after processing is visualized in Fig. 2. Data was collected every 50ms, thus, having 20 samples per second. Raw time-series accelerometer data is transformed into examples, each reading containing x, y, and z values corresponding to the three axes.

7.1.2 Experiment and Results

The summary results for activity recognition are presented in Table 4 and Fig. 3. The evaluation metrics: precision, recall, F1-score, the accuracy of each client (10 clients), and global servers are given in Table 4. The results show that the proposed model (FL-CNN) achieved global accuracy of up to 96%, and each client has an accuracy above 94% in the 100th round for HAR classification. In the comparison with those of Tran et al. [151] (CNN-SVM), the overall accuracy is 96% which is higher than the former study with 92.5%, and the detection results are stable.

The classification accuracy of recognizing each activity and predictor is shown in the confusion chart in Fig. 3. The chart illustrates the True class versus the Predicted class in each group for all activity combinations in a 6-by-6 grid (in number). The diagonal elements indicate the cases where there are correct estimations for each activity. Our accurate model achieves 98% and more extraordinary for four of six activities. Notably, the third row in the third column shows that walking is recognized with 99% (429/433). According to the third row, only 1% of walking activity

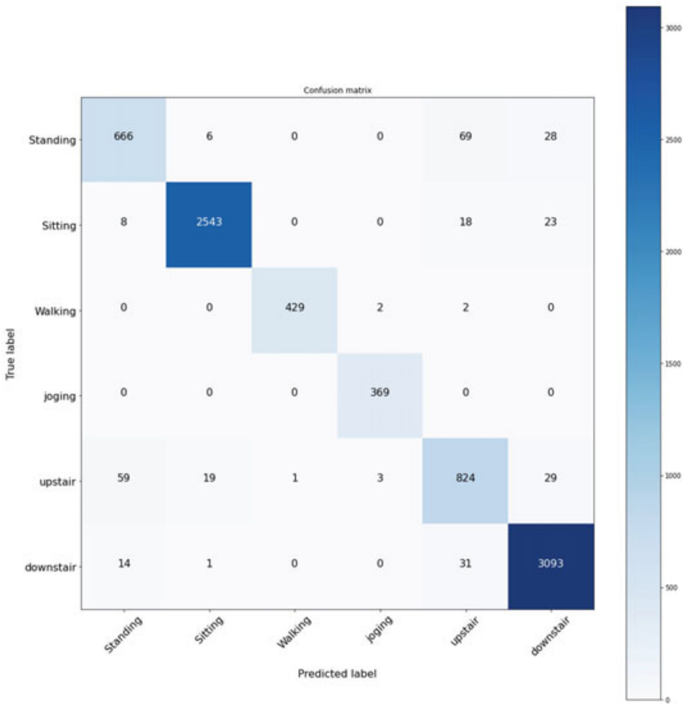


Fig. 3 Confusion matrix for the classification accuracy of 6 activities using FL-CNN

windows are classified falsely as *standing, jogging, upstairs*. The highest accuracy of recognition activity is *jogging* with 99%. For larger samples (*sitting, downstairs*), the accuracy is high at 98%. The misclassification of *upstairs* occurs with relatively high frequency when predicting *standing*. The reason is considered because the similar pose of *standing* and *upstairs* activity leads to sensor responses or may random data noises for features. The misclassified cases are *standing* and *upstairs* activities that are different from classification results of Tran et al. [151] (*upstairs, downstairs*).

7.2 Case study 2: Fall Detection and Classification with Federated Learning

Falling is one of the most severe accidents for the elderly and workers who regularly work at heights far from the ground. The accident causes adverse results, health problems, and family economic burden. Therefore, fall detection has been receiving significant attention due to its value in preventing and reducing the risk of injuries. This case study implements the FL technique in training models from 10-edge devices. We use a CNN-based model to detect *non-fall* and *fall* situations of

Table 4 The classification performance of the proposed framework in % of correctly predicted cases

Class	Precision	Recall	F1-score	Accuracy	Class	Precision	Recall	F1-score	Accuracy
Client 1					Client 2				
Stand	82	89	85		Stand	82	89	85	
Sit	97	98	98		Sit	97	98	98	
Walk	100	99	100		Walk	100	99	100	
Jog	98	100	99	95	Jog	98	100	99	95
Up	93	75	83		Up	93	75	83	
Down	96	99	97		Down	96	95	95	
Client 3					Client 4				
Stand	88	86	87		Stand	84	85	85	
Sit	98	98	98		Sit	96	99	97	
Walk	97	99	98		Walk	99	99	99	
Jog	99	97	98	95	Jog	98	99	99	95
Up	86	84	85		Up	87	82	85	
Down	97	98	97		Down	98	97	97	
Client 5					Client 6				
Stand	87	86	87		Stand	84	86	85	
Sit	98	98	98		Sit	99	97	98	
Walk	100	99	99		Walk	100	99	100	
Jog	99	100	99	95	Jog	99	100	99	95
Up	91	80	85		Up	78	90	83	
Down	96	99	97		Down	99	95	97	
Client 7					Client 8				
Stand	88	88	88		Stand	88	81	85	
Sit	98	98	98		Sit	98	98	98	
Walk	97	99	98		Walk	100	96	98	
Jog	99	96	98	96	Jog	95	100	97	95
Up	83	89	86		Up	87	82	85	
Down	99	96	98		Down	96	99	97	
Client 9					Client 10				
Stand	92	70	80		Stand	82	88	85	
Sit	98	98	98		Sit	98	98	98	
Walk	100	99	99		Walk	100	99	100	
Jog	99	99	99	94	Jog	99	99	99	95
Up	83	84	84		Up	87	81	84	
Down	94	99	96		Down	94	98	98	
	Global Server								
Stand	89	87	88						
Sit	99	98	98						
Walk	99	99	99						
Jog	99	99	99	96					
Up	84	89	87						
Down	94	99	96						

Table 5 The classification performance of the proposed framework for fall detection

Client	Class	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Client 1	Non-fall	100	81	89	
	Fall	76	99	86	88
Client 2	Non-fall	92	96	94	
	Fall	93	87	90	92
Client 3	Non-fall	92	95	93	
	Fall	91	86	88	92
Client 4	Non-fall	92	95	93	
	Fall	92	86	89	92
Client 5	Non-fall	92	95	93	
	Fall	91	86	88	92
Client 6	Non-fall	97	90	93	
	Fall	85	96	90	92
Client 7	Non-fall	95	75	84	
	Fall	70	94	80	82
Client 8	Non-fall	99	83	90	
	Fall	78	89	87	89
Client 9	Non-fall	88	98	93	
	Fall	97	79	87	91
Client 10	Non-fall	89	92	90	
	Fall	86	82	84	88
Global server	Non-fall	94	95	94	
	Fall	92	89	91	93

the human body in federated settings. Similar to the first case study, the motivation for applying the FL-CNN model is to study a feasible and workable framework that can detect fall states with higher accuracy, ensure safety for workers in manufacturing, and save the life of the elderly if they fall by stroke. Furthermore, this proposed framework can be extended in the case of big data in the next few years.

7.2.1 Dataset

In this case, we use the MobiFall dataset,[1] the data was collected from mobile phone sensors data. It has a total of 1278 records that were collected from 24 recruited participants in which, 17 men (age: 22–47 years; height: 169–189 cm; weight: 62–103 kg) and 07 women (age: 22–36 years; height: 160–172 cm; weight: 50–90 kg). Sensor data was from the accelerometer and the gyroscope sensor of a Samsung Galaxy S3 smartphone with the LSM330DLC inertial module.

[1] https://github.com/Luke3D/FallDetection.

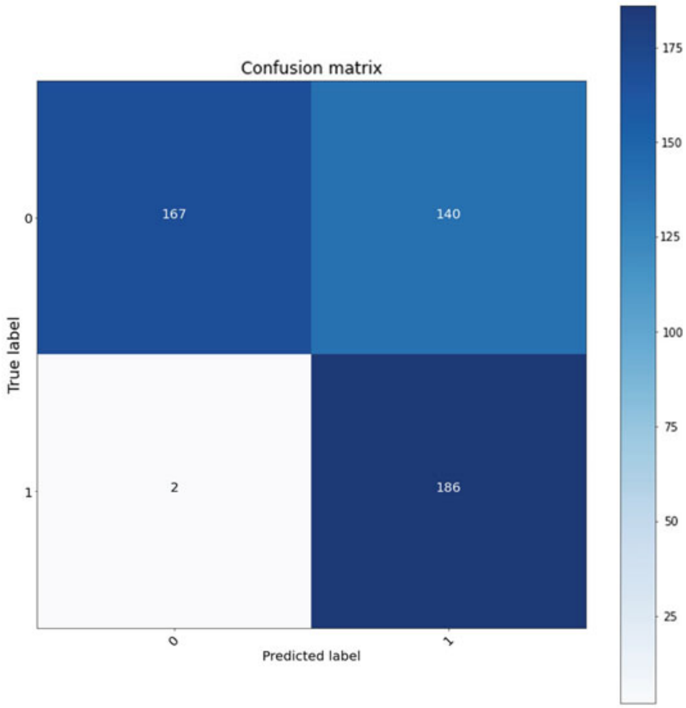


Fig. 4 Confusion matrix for the classification accuracy using FL-CNN

7.2.2 Experiment and Results

Similarly, in the first case study, we use ten clients in FL and evaluate accurate metrics via precision, recall, and F1-score, the accuracy of each client and global servers are given in Table 5. The accuracy of the global server is 93% higher than all clients.

The confusion chart from Fig. 4 demonstrates the True label versus the Predicted label for *fall* and *Non-fall* combinations in a 2-by-2 grid (in number). The correct prediction is presented in the diagonal elements. The global accuracy achieves 93% in binary classification. Particularly, the model predicts non-fall state Precision (94%), Recall (95%), and F1-score (94%). However, the results of the fallen state are lower Precision (92%), Recall (89%), and F1-score (91%). These results consolidate the reliability of the proposed FL-CNN framework and its highlight features in detecting HAR, the early-fall state. The result errors can be reduced by starting with a higher complexity single sensor classifier.

8 Concluding Remarks

The four Sections of this Chapter clearly illustrate the developments of wearable technology and wearable IoT devices in various fields, especially smart manufacturing in industry 5.0, along with difficulties and challenges. The case study Section presented a new approach to applying the FL-CNN framework for wearable IoT systems. The results of case studies also proved the superiority of FL solutions in privacy-preserving, boosting the overall classification and detection performance. Especially when using wearable IoT devices with FL, in reality, companies and organizations will drastically decrease communication costs and increase privacy preservation for raw data. Thus, our proposed framework will emphasize (1) a "win-win" strategy for promoting collaboration among organizations in training the global model to achieve the best one, (2) first steps to develop analytic data techniques that will work with big data in reality, (3) applications of wearable IoT devices for safety in transportation, more accurate diagnosis and managing patients' health in healthcare, (4) integrate wearable IoT system into HCPS or HDT that enhance productivity and safety in intelligent manufacturing. All points mentioned above will help shape future research direction of wearable IoT technology in reaching Industry 5.0.

References

1. Ometov A, Shubina V, Klus L, Skibińska J, Saafi S, Pascacio P, Flueratoru L, Gaibor DQ, Chukhno N, Chukhno O et al (2021) A survey on wearable technology: history, state-of-the-art and current challenges. Comput Netw 193:108074
2. Pantelopoulos A, Bourbakis NG (2009) A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans Syst Man Cybern Part C (Appl Rev) 40:1–12
3. Seshadri DR, Drummond C, Craker J, Rowbottom JR, Voos JE (2017) Wearable devices for sports: new integrated technologies allow coaches, physicians, and trainers to better understand the physical demands of athletes in real time. IEEE Pulse 8:38–43
4. Lamooki SR, Hajifar S, Kang J, Sun H, Megahed FM, Cavuoto LA (2022) A data analytic end-to-end framework for the automated quantification of ergonomic risk factors across multiple tasks using a single wearable sensor. Appl Ergon 102:103732
5. Huang H, Chen H, Lin S. Magtrack: enabling safe driving monitoring with wearable magnetics. In: Proceedings of the 17th annual international conference on mobile systems, applications, and services, pp 326–339
6. Althobaiti T, Katsigiannis S, Ramzan N (2020) Triaxial accelerometer-based falls and activities of daily life detection using machine learning. Sensors 20:3777
7. Zhou Z, Tam V, Lui KS, Lam EY, Yuen A, Hu X, Law N. Applying deep learning and wearable devices for educational data analytics. In: 2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI). IEEE, pp 871–878
8. Raza A, Tran KP, Koehl L, Li S (2022) Designing ECG monitoring healthcare system with federated transfer learning and explainable AI. Knowl-Based Syst 236:107763
9. Qaim WB, Ometov A, Molinaro A, Lener I, Campolo C, Lohan ES, Nurmi J (2020) Towards energy efficiency in the internet of wearable things: a systematic review. IEEE Access 8:175412–175435
10. Khan MA (2021) Challenges facing the application of IoT in medicine and healthcare. Int J Comput Inf Manuf (IJCIM) 1

11. Svertoka E, Saafi S, Rusu-Casandra A, Burget R, Marghescu I, Hosek J, Ometov A (2021) Wearables for industrial work safety: a survey. Sensors 21:3844
12. U. B. of Labor Statistics, Injuries, illnesses, and fatalities (2021). Accessed 27 July 2022
13. Jacobs JV, Hettinger LJ, Huang Y-H, Jeffries S, Lesch MF, Simmons LA, Verma SK, Willetts JL (2019) Employee acceptance of wearable technology in the workplace. Appl Ergon 78:148–156
14. Shany T, Redmond SJ, Narayanan MR, Lovell NH (2011) Sensors-based wearable systems for monitoring of human movement and falls. IEEE Sens J 12:658–670
15. Warrington DJ, Shortis EJ, Whittaker PJ (2021) Are wearable devices effective for preventing and detecting falls: an umbrella review (a review of systematic reviews). BMC Publ Health 21:1–12
16. Taramasco C, Rodenas T, Martinez F, Fuentes P, Munoz R, Olivares R, De Albuquerque VHC, Demongeot J (2018) A novel monitoring system for fall detection in older people. IEEE Access 6:43563–43574
17. Bhandari S, Babar N, Gupta P, Shah N, Pujari S. A novel approach for fall detection in home environment. In: 2017 IEEE 6th global conference on consumer electronics (GCCE). IEEE, pp 1–5
18. Ferrone A, Napier C, Menon C (2021) Wearable technology to increase self-awareness of low back pain: a survey of technology needs among health care workers. Sensors 21:8412
19. Li P, Meziane R, Otis MJ-D, Ezzaidi H, Cardou P. A smart safety helmet using IMU and EEG sensors for worker fatigue detection. In: 2014 IEEE international symposium on robotic and sensors environments (ROSE) proceedings. IEEE, pp 55–60
20. Hazarika P. Implementation of smart safety helmet for coal mine workers. In: 2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES). IEEE, pp 1–3
21. Mardonova M, Choi Y (2018) Review of wearable device technology and its applications to the mining industry. Energies 11:547
22. Pardoel S, Kofman J, Nantel J, Lemaire ED (2019) Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: a review. Sensors 19:5141
23. Iqbal S, Mahgoub I, Du E, Leavitt MA, Asghar W (2021) Advances in healthcare wearable devices. NPJ Flex Electron 5:1–14
24. WH Organization (2022) Cardiovascular diseases. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed 27 July 2022
25. Zeni Jr J, Abujaber S, Flowers P, Pozzi F, Snyder-Mackler L (2013) Biofeedback to promote movement symmetry after total knee arthroplasty: a feasibility study. J Orthop Sports Phys Therapy 43:715–726
26. Christensen JC, LaStayo PC, Marcus RL, Stoddard GJ, Foreman KB, Mizner RL, Peters CL, Pelt CE (2018) Visual knee-kinetic biofeedback technique normalizes gait abnormalities during high-demand mobility after total knee arthroplasty. Knee 25:73–82
27. Hardt S, Schulz MR, Pfitzner T, Wassilew G, Horstmann H, Liodakis E, Weber-Spickschen TS (2018) Improved early outcome after TKA through an app-based active muscle training programme—A randomized-controlled trial. Knee Surg Sports Traumatol Arthrosc 26:3429–3437
28. Oh H-T, Hwangbo G (2018) The effects of proprioception exercise with and without visual feedback on the pain and balance in patients after total knee arthroplasty. J Phys Therapy Sci 30:124–126
29. Margarito J, Helaoui R, Bianchi AM, Sartor F, Bonomi AG (2015) User-independent recognition of sports activities from a single wrist-worn accelerometer: a template-matching-based approach. IEEE Trans Biomed Eng 63:788–796
30. Maijers M, Verschuren O, Stolwijk-Swüste J, van Koppenhagen C, de Groot S, Post M (2018) Is Fitbit charge 2 a feasible instrument to monitor daily physical activity and handbike training in persons with spinal cord injury? a pilot study. Spinal Cord Ser Cases 4:1–10
31. Enomoto K, Shimizu R, Kudo H (2018) Real-time skin lactic acid monitoring system for assessment of training intensity. Electron Commun Jpn 101:41–46

32. Salman M, Qaisar S, Qamar AM (2017) Classification and legality analysis of bowling action in the game of cricket. Data Min Knowl Discov 31:1706–1734
33. Stiles VH, Pearce M, Moore IS, Langford J, Rowlands AV (2018) Wrist-worn accelerometry for runners: objective quantification of training load. Med Sci Sports Exerc 50:2277
34. Wang Y, Zhao Y, Chan RH, Li WJ (2018) Volleyball skill assessment using a single wearable micro inertial measurement unit at wrist. IEEE Access 6:13758–13765
35. Rawashdeh SA, Rafeldt DA, Uhl TL (2016) Wearable IMU for shoulder injury prevention in overhead sports. Sensors 16:1847
36. Mukhopadhyay SC, Suryadevara NK, Nag A (2021) Wearable sensors and systems in the IoT
37. Haghi M, Neubert S, Geissler A, Fleischer H, Stoll N, Stoll R, Thurow K (2020) A flexible and pervasive IoT-based healthcare platform for physiological and environmental parameters monitoring. IEEE Internet Things J 7:5628–5647
38. Raad H (2020) Fundamentals of IoT and wearable technology design. Wiley
39. Dian FJ, Vahidnia R, Rahmati A (2020) Wearables and the internet of things (IoT), applications, opportunities, and challenges: a survey. IEEE Access 8:69200–69211
40. World Health Organization (2009) Global status report on road safety: time for action. World Health Organization
41. Neumann D, Peltonen E. Skadi: heterogeneous human-sensing system for automotive IoT. In: 2022 IEEE international conference on smart computing (SMARTCOMP). IEEE, pp 165–167
42. Berg FJ, Bennett TJ, Davis AC, Riefe RK, Dybalski RH, Phillips TM (2007) Steering system with haptic driver warning. US Patent 7,280,046
43. Bi C, Huang J, Xing G, Jiang L, Liu X, Chen M (2019) Safewatch: a wearable hand motion tracking system for improving driving safety. ACM Trans Cyber-Phys Syst 4:1–21
44. Irsan M, Hassan R, Hasan MK, Lam MC, Hussain WMHW, Ibrahim AH, Ahmed ASAMS (2022) A novel prototype for safe driving using embedded smart box system. Sensors 22:1907
45. World Health Organization et al (2020) Personal protective equipment. World Health Organization
46. Izadi N, Aminian O, Esmaeili B (2019) Occupational accidents in Iran: risk factors and long term trend (2007–2016). J Res Health Sci 19:e00448
47. To KK-W, Tsang OT-Y, Yip CC-Y, Chan K-H, Wu T-C, Chan JM-C, Leung W-S, Chik TS-H, Choi CY-C, Kandamby DH et al (2019) Consistent detection of novel coronavirus in saliva. Clin Infect Dis 71:841–843
48. Meraj M, Alvi SAM, Quasim MT, Haidar SW. A critical review of detection and prediction of infectious disease using IoT sensors. In: 2021 second international conference on electronics and sustainable communication systems (ICESC). IEEE, pp 679–684
49. Chamberlain SD, Singh I, Ariza C, Daitch A, Philips P, Dalziel BD (2020) Real-time detection of COVID-19 epicenters within the united states using a network of smart thermometers. MedRxiv
50. Passos J, Lopes SI, Clemente FM, Moreira PM, Rico-González M, Bezerra P, Rodrigues LP (2021) Wearables and internet of things (IoT) technologies for fitness assessment: a systematic review. Sensors 21:5418
51. Li F, Martínez OS, Aiswarya R (2021) Internet of things-based smart wearable system to monitor sports person health. Technol Health Care 29:1249–1262
52. Borthakur D, Dubey H, Constant N, Mahler L, Mankodiya K. Smart fog: fog computing framework for unsupervised clustering analytics in wearable internet of things. In: 2017 IEEE global conference on signal and information processing (GlobalSIP). IEEE, pp 472–476
53. Rodríguez-Martín D, Samà A, Pérez-López C, Català A, Moreno Arostegui JM, Cabestany J, Bayés À, Alcaine S, Mestre B, Prats A et al (2017) Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. PloS one 12:e0171764
54. Reches T, Dagan M, Herman T, Gazit E, Gouskova NA, Giladi N, Manor B, Hausdorff JM (2020) Using wearable sensors and machine learning to automatically detect freezing of gait during a fog-provoking test. Sensors 20:4474

55. Mehdizadeh A, Yazdi MAA, Cai M, Hu Q, Vinel A, Rigdon SE, Davis K, Megahed FM (2021) Predicting unsafe driving risk among commercial truck drivers using machine learning: lessons learned from the surveillance of 20 million driving miles. Accid Anal Prev 159:106285
56. Hofmann C, Patschkowski C, Haefner B, Lanza G (2020) Machine learning based activity recognition to identify wasteful activities in production. Procedia Manuf 45:171–176
57. Manjarres J, Narvaez P, Gasser K, Percybrooks W, Pardo M (2019) Physical workload tracking using human activity recognition with wearable devices. Sensors 20:39
58. Nguyen HD, Tran KP, Zeng X, Koehl L, Tartare G (2020) An improved ensemble machine-learning algorithm for wearable sensor data based human activity recognition. In: Reliability and statistical computing. Springer, pp 207–228
59. Fouad H, Hassanein AS, Soliman AM, Al-Feel H (2020) Analyzing patient health information based on iot sensor with ai for improving patient assistance in the future direction. Measurement 159:107757
60. Lin Y-J, Chuang C-W, Yen C-Y, Huang S-H, Huang P-W, Chen J-Y, Lee S-Y. Artificial intelligence of things wearable system for cardiac disease detection. In: 2019 IEEE international conference on artificial intelligence circuits and systems (AICAS). IEEE, pp 67–70
61. Silva MC, da Silva JC, Delabrida S, Bianchi AG, Ribeiro SP, Silva JS, Oliveira RA (2021) Wearable edge AI applications for ecological environments. Sensors 21:5082
62. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, PMLR, pp 1273–1282
63. Hegiste V, Legler T, Ruskowski M (2022) Application of federated learning in manufacturing. arXiv:2208.04664
64. Manias DM, Shami A (2021) Making a case for federated learning in the internet of vehicles and intelligent transportation systems. IEEE Netw 35:88–94
65. Raza A, Li S, Tran K-P, Koehl L (2022) Detection of poisoning attacks with anomaly detection in federated learning for healthcare applications: A machine learning approach. arXiv:2207.08486
66. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F (2021) Federated learning for healthcare informatics. J Healthc Inf Res 5:1–19
67. Sun W (2022) Predictive analysis and simulation of college sports performance fused with adaptive federated deep learning algorithm. J Sens 2022
68. Teh HY, Kempa-Liehr AW, Wang KI-K (2020) Sensor data quality: a systematic review. J Big Data 7:1–49
69. Maman ZS, Yazdi MAA, Cavuoto LA, Megahed FM (2017) A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. Appl Ergon 65:515–529
70. Baghdadi A, Cavuoto LA, Jones-Farmer A, Rigdon SE, Esfahani ET, Megahed FM (2021) Monitoring worker fatigue using wearable devices: a case study to detect changes in gait parameters. J Qual Technol 53:47–71
71. Chieh TC, Mustafa MM, Hussain A, Hendi SF, Majlis BY. Development of vehicle driver drowsiness detection system using electrooculogram (EOG). In: 2005 1st international conference on computers, communications, & signal processing with special track on biomedical engineering. IEEE, pp 165–168
72. Rashid N, Al Faruque MA. Energy-efficient real-time myocardial infarction detection on wearable devices. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC). IEEE, pp 4648–4651
73. Golab M, Breedon P, Vloeberghs M (2016) A wearable headset for monitoring electromyography responses within spinal surgery. Euro Spine J 25:3214–3219
74. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H (2018) A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 4:195
75. Akbulut FP, Ikitimur B, Akan A (2020) Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome. Artif Intell Med 104:101824

76. Gitanjali N, Bhargavi P, Sangeetha K, Bhagyashree R et al (2019) A survey on assistive technology for ALS patients. Perspect Commun Embed-Syst Signal-Process-PiCES 3:44–46
77. Musngi MMM (2018) Fall detection algorithms using accelerometers, gyroscopes and a barometric pressure sensor. Ph.D. thesis. Appl Sci: School Mechatron Syst Eng
78. Lu W, Wang C, Stevens MC, Redmond SJ, Lovell NH. Low-power operation of a barometric pressure sensor for use in an automatic fall detector. In: 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 2010–2013
79. Pierleoni P, Belli A, Maurizi L, Palma L, Pernini L, Paniccia M, Valenti S (2016) A wearable fall detector for elderly people based on AHRS and barometric sensor. IEEE Sens J 16:6733–6744
80. Valero E, Sivanathan A, Bosché F, Abdel-Wahab M (2017) Analysis of construction trade worker body motions using a wearable and wireless motion sensor network. Autom Constr 83:48–55
81. Fenu G, Steri G. IMU based post-traumatic rehabilitation assessment. In: 2010 3rd international symposium on applied sciences in biomedical and communication technologies (ISABEL 2010). IEEE, pp 1–5
82. Zhao H, Wang Z, Qiu S, Shen Y, Wang J. IMU-based gait analysis for rehabilitation assessment of patients with gait disorders. In: 2017 4th international conference on systems and informatics (ICSAI). IEEE, pp 622–626
83. Steele BG, Belza B, Cain K, Warms C, Coppersmith J, Howard J et al (2003) Bodies in motion: monitoring daily activity and exercise with motion sensors in people with chronic pulmonary disease. J Reh Res Dev 40:45–58
84. Savage PD, Ades PA (2008) Pedometer step counts predict cardiac risk factors at entry to cardiac rehabilitation. J Cardiopulm Rehab Prev 28:370–377
85. Ahola TM (2010) Pedometer for running activity using accelerometer sensors on the wrist. Med Equip Insights 3:MEI–S3748
86. Hester T, Hughes R, Sherrill DM, Knorr B, Akay M, Stein J, Bonato P. Using wearable sensors to measure motor abilities following stroke. In: International workshop on wearable and implantable body sensor networks (BSN’06). IEEE, pp 4–pp
87. Rand D, Eng JJ, Tang P-F, Jeng J-S, Hung C (2009) How active are people with stroke? Use of accelerometers to assess physical activity. Stroke 40:163–168
88. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C (2013) Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Annal Thoracic Surg 96:1057–1061
89. Papazian C, Baicoianu NA, Peters KM, Feldner HA, Steele KM (2021) Electromyography recordings detect muscle activity before observable contractions in acute stroke care. Arch Rehab Res Clin Transl 3:100136
90. Anton O, Fernandez R, Rendon-Morales E, Aviles-Espinosa R, Jordan H, Rabe H (2019) Heart rate monitoring in newborn babies: a systematic review. Neonatology 116:199–210
91. Rahman A, Rahman T, Ghani NH, Hossain S, Uddin J. IoT based patient monitoring system using ECG sensor. In: 2019 international conference on robotics. In: Electrical and signal processing techniques (ICREST). IEEE, pp 378–382
92. Giansanti D, Ricci G, Maccioni G (2008) Toward the design of a wearable system for the remote monitoring of epileptic crisis. Telemed e-Health 14:1130–1135
93. Konstantinidis D, Iliakis P, Tatakis F, Thomopoulos K, Dimitriadis K, Tousoulis D, Tsioufis K (2022) Wearable blood pressure measurement devices and new approaches in hypertension management: the digital era. J Hum Hypertens 1–7
94. Cappon G, Vettoretti M, Sparacino G, Facchinetti A (2019) Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 43:383–397
95. Zhang S, Zeng J, Wang C, Feng L, Song Z, Zhao W, Wang Q, Liu C (2021) The application of wearable glucose sensors in point-of-care testing. Front Bioeng Biotechnol 9

96. Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel B, Figlin RA, Meinert CL, Piantadosi S, Shinde AM (2018) Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. NPJ Digit Med 1:1–8
97. Alexandre R, Postolache O. Wearable and IoT technologies application for physical rehabilitation. In: 2018 international symposium in sensing and instrumentation in IoT era (ISSI). IEEE, pp 1–6
98. Wang K, Xie S, Rodrigues J (2022) Medical data security of wearable tele-rehabilitation under internet of things. Internet Things Cyber-Phys Syst 2:1–11
99. Vallati C, Virdis A, Gesi M, Carbonaro N, Tognetti A (2018) ePhysio: a wearables-enabled platform for the remote management of musculoskeletal diseases. Sensors 19:2
100. Tun SYY, Madanian S, Mirza F (2021) Internet of things (IoT) applications for elderly care: a reflective review. Aging Clin Exp Res 33:855–867
101. Li RT, Kling SR, Salata MJ, Cupp SA, Sheehan J, Voos JE (2016) Wearable performance devices in sports medicine. Sports Health 8:74–78
102. James DA, Davey N, Rice T (2004) An accelerometer based sensor platform for insitu elite athlete performance analysis. Sensors (IEEE) 1373–1376
103. Wundersitz DW, Josman C, Gupta R, Netto KJ, Gastin PB, Robertson S (2015) Classification of team sport activities using a single wearable tracking device. J Biomech 48:3975–3981
104. Portas MD, Harley JA, Barnes CA, Rush CJ (2010) The validity and reliability of 1-hz and 5-hz global positioning systems for linear, multidirectional, and soccer-specific activities. Int J Sports Physiol Perform 5:448–458
105. Izzo R, D'isanto T, Raiola G, Cejudo A, Ponsano N, Varde'i CH (2020) The role of fatigue in football matches, performance model analysis and evaluation during quarters using live global positioning system technology at 50 HZ. Sport Sci 13:30–35
106. Oxendale CL, Highton J, Smith G, Twist C (2022) Can player tracking devices monitor changes in internal response during multidirectional running?. Res Q Exercise Sport 1–7
107. Korobeynikov G, Korobeynikova L, Iermakov S, Nosko M (2016) Reaction of heart rate regulation to extreme sport activity in elite athletes. J Phys Educ Sport
108. Bishop SA, Dech RT, Guzik P, Neary JP (2018) Heart rate variability and implication for sport concussion. Clin Physiol Funct Imaging 38:733–742
109. Formenti D, Merla A, Priego Quesada JI (2017) The use of infrared thermography in the study of sport and exercise physiology. In: Application of infrared thermography in sports science. Springer, pp 111–136
110. Hadžić V, Širok B, Malneršič A, Čoh M (2019) Can infrared thermography be used to monitor fatigue during exercise? A case study. J Sport Health Sci 8:89–92
111. SamSung (2022) Wearables reduce downtime and boost productivity in manufacturing. https://www.samsung.com/us/business/support/resource-center/case-studies/wearables-reduce-downtime-and-boost-productivity-in-manufacturing/. Accessed 28 July 2022
112. Vujica Herzog N, Buchmeister B, Beharic A, Gajsek B (2018) Visual and optometric issues with smart glasses in industry 4.0 working environment. Adv Prod Eng Manag 13:417–428
113. Bader F, Jagtap S (2020) Internet of things-linked wearable devices for managing food safety in the healthcare sector. In: Wearable and implantable medical devices. Elsevier, pp 229–253
114. Yang L, Lu K, Diaz-Olivares JA, Seoane F, Lindecrantz K, Forsman M, Abtahi F, Eklund JA (2018) Towards smart work clothing for automatic risk assessment of physical workload, IEEE. Access 6:40059–40072
115. Bootsman R, Markopoulos P, Qi Q, Wang Q, Timmermans AA (2019) Wearable technology for posture monitoring at the workplace. Int J Hum-Comput Stud 132:99–111
116. Salim F, Prohasky D, Belbasis A, Houshyar S, Fuss FK. Design and evaluation of smart wearable undergarment for monitoring physiological extremes in firefighting. In: Proceedings of the 2014 ACM international symposium on wearable computers: adjunct program, pp 249–254
117. Miller ME, Spatz E (2022) A unified view of a human digital twin. Hum-Intell Syst Integr 1–11

118. Montini E, Bettoni A, Ciavotta M, Carpanzano E, Pedrazzoli P (2021) A meta-model for modular composition of tailored human digital twins in production. Procedia CIRP 104:689–695
119. Zhou J, Zhou Y, Wang B, Zang J (2019) Human-cyber-physical systems (HCPSS) in the context of new-generation intelligent manufacturing. Engineering 5:624–636
120. Sun S, Zheng X, Gong B, Garcia Paredes J, Ordieres-Meré J (2021) Healthy operator 4.0: a human cyber—Physical system architecture for smart workplaces. Sensors 20:2011
121. Xu X, Zhong M, Wan J, Yi M, Gao T (2016) Health monitoring and management for manufacturing workers in adverse working conditions. J Med Syst 40:1–7
122. Zhou J, Li P, Zhou Y, Wang B, Zang J, Meng L (2018) Toward new-generation intelligent manufacturing. Engineering 4:11–20
123. Romero D, Bernus P, Noran O, Stahre J, Fast-Berglund Å. The operator 4.0: human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In: IFIP international conference on advances in production management systems. Springer, pp 677–686
124. Santoso H, Baroroh DK, Darmawan A (2021) Future application of multisensory mixed reality in the human cyber—Physical system. South Afr J Ind Eng 32:44–56
125. Calero JAM, Marino R, Lanza-Gutierrez JM, Riesgo T, Garcia-Valderas M, Lopez-Ongil C. Embedded emotion recognition within cyber-physical systems using physiological signals. In: 2018 conference on design of circuits and integrated systems (DCIS). IEEE, pp 1–6
126. Nguyen Q-T, Tran TN, Heuchenne C, Tran KP (2021) Decision support systems for anomaly detection with the applications in smart manufacturing: a survey and perspective. In: Machine learning and probabilistic graphical models for decision support systems. CRC Press, pp 34–61
127. Chan M, Estève D, Fourniols J-Y, Escriba C, Campo E (2012) Smart wearable systems: current status and future challenges. Artif Intell Med 56:137–156
128. Sabry F, Eltaras T, Labda W, Alzoubi K, Malluhi Q (2022) Machine learning for healthcare wearable devices: the big picture. J Healthc Eng 2022
129. Luna R, Rhine E, Myhra M, Sullivan R, Kruse CS (2016) Cyber threats to health information systems: a systematic review. Technol Health Care 24:1–9
130. Khan WZ, Aalsalem MY, Khan MK (2018) Communal acts of IoT consumers: a potential threat to security and privacy. IEEE Trans Consum Electron 65:64–72
131. Fuentes MR (2017) Cybercrime and other threats faced by the healthcare industry. Trend Micro 5566
132. Suhluli SA, Ali Khan SMF (2022) Determinants of user acceptance of wearable IoT devices. Cogent Eng 9:2087456
133. Zhou W, Jia Y, Peng A, Zhang Y, Liu P (2018) The effect of IoT new features on security and privacy: new threats, existing solutions, and challenges yet to be solved. IEEE Internet Things J 6:1606–1616
134. Hussain S, Kang BH, Lee S. A wearable device-based personalized big data analysis model. In: International conference on ubiquitous computing and ambient intelligence. Springer, pp 236–242
135. Dwivedi P, Singha MK (2021) IoT based wearable healthcare system: post COVID-19. In: The impact of the COVID-19 pandemic on green societies. Springer, pp 305–321
136. Dash S, Shakyawar SK, Sharma M, Kaushik S (2019) Big data in healthcare: management, analysis and future prospects. J Big Data 6:1–25
137. Sun H, Zhang Z, Hu RQ, Qian Y (2018) Wearable communications in 5G: challenges and enabling technologies. IEEE Veh Technol Mag 13:100–109
138. Roy M, Chowdhury C, Aslam N (2018) Designing transmission strategies for enhancing communications in medical IoT using Markov decision process. Sensors 18:4450
139. Rault T, Bouabdallah A, Challal Y, Marin F (2017) A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications. Pervasive Mob Comput 37:23–44

140. Djelouat H, Al Disi M, Boukhenoufa I, Amira A, Bensaali F, Kotronis C, Politi E, Nikolaidou M, Dimitrakopoulos G (2020) Real-time ECG monitoring using compressive sensing on a heterogeneous multicore edge-device. Microprocess Microsyst 72:102839
141. Amiri D, Anzanpour A, Azimi I, Rahmani AM, Liljeberg P, Dutt N, Levorato M (2020) Optimizing energy efficiency of wearable sensors using fog-assisted control. In: Theory and practice, fog computing, pp 245–268
142. Munir S, Stankovic JA, Liang C-JM, Lin S. Cyber physical system challenges for {Human-in-the-Loop} control. In: 8th international workshop on feedback computing (Feedback Computing 13)
143. Griffor ER, Greer C, Wollman DA, Burns MJ et al (2017) Framework for cyber-physical systems: volume 2, working group reports
144. Alguliyev R, Imamverdiyev Y, Sukhostat L (2018) Cyber-physical systems and their security issues. Comput Ind 100:212–223
145. Papoutsi C, Reed JE, Marston C, Lewis R, Majeed A, Bell D (2015) Patient and public views about the security and privacy of electronic health records (EHRS) in the UK: results from a mixed methods study. BMC Med Inform Decis Making 15:1–15
146. Thierer AD (2015) The internet of things and wearable technology: addressing privacy and security concerns without derailing innovation, Adam Thierer, vol 21
147. Lee SM, Lee D (2020) Healthcare wearable devices: an analysis of key factors for continuous use intention. Serv Bus 14:503–531
148. Yildirim H, Ali-Eldin AM (2019) A model for predicting user intention to use wearable IoT devices at the workplace. J King Saud Univ-Comput Inf Sci 31:497–505
149. Chang C-C et al (2020) Exploring the usage intentions of wearable medical devices: a demonstration study. Interact J Med Res 9:e19776
150. Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. ACM SigKDD Explor Newsl 12:74–82
151. Tran PH, Nguyen QT, Tran KP, Heuchenne C. Wearable sensor data based human activity recognition using deep learning: a new approach. In: Developments of artificial intelligence technologies in computation and robotics: proceedings of the 14th international FLINS conference (FLINS 2020). World Scientific, pp 581–588

Benefits of Using Digital Twin for Online Fault Diagnosis of a Manufacturing System

Ramla Saddem and Dylan Baptiste

Abstract In this work, we illustrate the interest in the use of a digital twin for the online fault diagnosis in a manufacturing system with sensors and actuators delivering binary signals that can be modeled as Discrete Event Systems. This chapter presents an intelligent diagnostic solution to replace traditional solutions, which are often non-industrialized, with a new data-based method learned from the simulation of the plant behaviors and using recurrent neural networks (RNN) with short-term and long-term memory (Long short-term memory, LSTM).

Keywords Digital twin · Online fault diagnosis · Discrete event systems · Automated production systems

1 Introduction

Digital Twin (DT) is one of the tools of the industry of the future. This chapter studies the interest in using a DT for online fault diagnosis in a manufacturing system is illustrated. The concept of DT consists of digitizing a factory and reproducing its behavior. Most industrial solutions allow matching a desired machine's behavior to make virtual commissioning. In this study, the (DT) tool is considered to obtain a large amount of data and inject failures into the digitized system in order to simulate abnormal behavior and validate the training algorithm. We present here an extension of an original online diagnosis approach, based on Machine Learning (ML), and presented in Saddem et al. [12]. All registered data come from a DT, one of the tools of industry 4.0. The concept of DT Kritzinger et al. [9], Tao et al. [14] consists of digitizing a factory and reproducing its behavior.

The rest of this chapter is organized as follows: Sect. 2 presents a brief overview of the state of the art. Section 3 introduces the proposal. In Sect. 4, a description of a case use and related results are presented. And for Sect. 5, a conclusion with some prospects is provided.

R. Saddem (✉) · D. Baptiste
CReSTIC, University of Reims Champagne-Ardenne, Reims, France
e-mail: ramla.saddem@univ-reims.fr

K. P. Tran (ed.), *Artificial Intelligence for Smart Manufacturing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-30510-8_12

2 State of the Art

In the manufacturing industry, saving money and time is the aim of manufacturers' spirit, who are continuously looking for ways, methods, and tools to achieve. In this context, this chapter studies the diagnosis of Automated Production Systems (APS). For "industry 4.0", production systems need to be more flexible and resilient while becoming more complex. In fact, performance requirements (production, quality, safety) lead industries to anticipate various failure occurrences. The DT has the potential to provide information, not only on machine performance but also on the production line, which gives the manufacturer the ability to predict problems earlier. Artificial Intelligence algorithms coupled with DT have the potential for greater accuracy as the machine can hold large amounts of data, needed for performance and prediction. To ensure the safe operation of goods and equipment, the diagnosis task consists of detecting, isolating, and identifying, as accurately and as soon as possible, the slightest failure or deviation from the nominal machine behavior. The systems we are interested in in this chapter are the Discrete Event Systems (DES). The classical DES diagnostic approaches in the literature are mainly based on:

1. Offline studies of the diagnosability of a system (ability to diagnose a fault with certainty in a finite time)
2. Online system observer models to be integrated into the control process.

Although such approaches are well known by the community, a huge amount of expertise is required to obtain high-performance models of the system. Furthermore, these approaches are quickly exposed to the problem of the explosion of the state space to be observed, which also makes the calculation of diagnosers to be implemented online often impossible for complex systems. In this chapter, we are examining the use of DT for online APS data-based fault diagnosis. The diagnosis algorithm uses the simulation of the plant behaviors from DT and uses recurrent neural networks (RNN) with short-term and long-term memory (Long short-term memory, LSTM) to learn. The literature proposes different approaches to dealing with the problem of online APS diagnosis. These approaches can be classified into three classes either according to the dynamics of the APS: the class of continuous systems, the class of DES, and the class of hybrid dynamic systems (HDS), or according to the reasoning used mode: model-based approaches, knowledge-based approaches, or data-based approaches.

Model-based approaches Sampath et al. [13], Zaytoon and Lafortune [17], de Souza et al. [1] are generally used when there is sufficient knowledge of the internal functioning of the system. They are efficient and able to validate the consistency and completeness of the faults to be diagnosed. However, to work properly, these approaches require accurate and deep analytical models of the domain, and the major difficulty is the high cost of implementing the models saddem and Philippot [11]. Indeed, the temporal complexity of implementing most models is exponential.

Knowledge-based approaches have a high diagnosis capacity thanks to the symptoms of faults they model. However, its major limit lies in the formalization of the expert's knowledge and its updates.

Data-based approaches Venkatasubramanian et al. [16], Moosavian et al. [10], Dou and Zhou [2], Vazan et al. [15], Han et al. [5] don't require either knowledge of the internal workings of the system, or an explicit model through a mathematical model. Available historical data is enough to allow them to give predictions. These approaches learn from each experience to improve their performance. They rely on ML techniques to achieve their objectives. However, a data preparation step is required to extract the most relevant data that will be formatted according to the ML technique to be used. In this chapter, we are interested in the diagnosis of APS using the data-based approach.

3 Proposed Approach

3.1 Automated Production Systems Modeled as DES

An APS system (Fig. 1) consists of three parts: the operative part (OP), the control part (CP), and the Human Machine Interface (HMI). The OP represents all material resources that physically operate on the system. The CP is the set of information processing and acquisition means that ensure the piloting and the control of the process. There are two types of information exchanges between the CP and the OP.

1. The CP sends orders to the actuators and pre-actuators of the OP to obtain the desired effects.
2. The OP sends sensor values to the CP. The HMI allows communication between the CP and the human operator. The human operator gives instructions via the HMI and receives various signals from the CP such as sound indicators, light indicators, messages displayed on the screens, etc.

Most APSs are controlled by programmable logic controllers (PLCs) that perform three successive operations. (1) Reading the inputs, which consist of the recording of the states of sensors. (2) Executing the program. (3) Updating the outputs (actuators). These operations are cyclical. The diagnosis, thus, consists in cyclically reading the sensors' values and the CP's orders and analyzing them to detect and isolate faults. In this work, we are interested in the online diagnosis of APS with sensors and actuators delivering binary signals that can be modeled as DE. Four faults are possible for each component: stuck to 0; stuck to 1; an unexpected move from 0 to 1 and an unexpected move from 1 to 0. The monitored APS can be normal, failed, or uncertain. An uncertain state means that the system may be normal or faulty: there are not enough indicators to decide its state. The objective is therefore to return the online status of the plant. If a fault occurs in a component of the plant, the diagnosis returns this fault. Therefore, one needs to have a list of the components of the plant to fix the number and the name of each fault that may occur. A specification of the APS operation allows us to establish a control program for the plant. We assume that

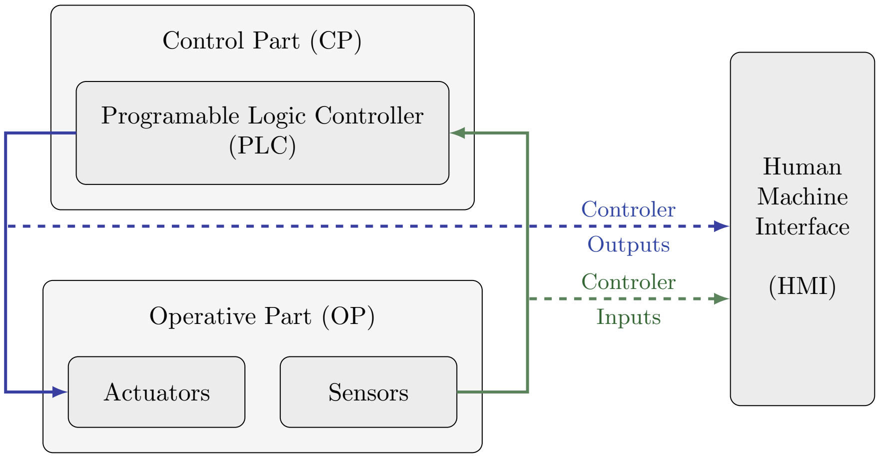


Fig. 1 APS structure

this program does not contain any faults, i.e., if the CP sends an order to the OP, and then the OP correctly receives this order.

3.2 Proposal

The data-based online diagnosis approach involves two phases:

1. Offline Phase, which includes
   - Data acquisition task, which consists of identifying and collecting the needed data to support the problem. This data can come from several sources and can be structured (such as database records, trees, graphs, etc.) or unstructured (such as images, texts, voices, etc.). In this work, we use a digital twin to generate normal and faulty behavior data.
   - Data preparation task, which consists of formatting the data according to the ML algorithm to be used. It includes transformation, normalization, cleaning, and selection of training data.
   - Learning task, which consists of the training and validation of the algorithm. This requires dividing the available data into three parts: training data, validation data, and test data. The algorithm is trained with the training data. Then, validation data is used to fine-tune the hyperparameters defined before the beginning of the training phase. The test data is used for testing and evaluating the algorithm.
2. Online Phase: the online learning phase that diagnoses online possible faults.

In this chapter, the use of a DT is applied for the task of data acquisition and the validation of the learning algorithm. It would be detailed in the following of this chapter.

3.3 Data Acquisition

The aim of this task is to save all input and output data history capturing the evolutions of sensors, actuators, and control program variables on a Database. We have developed a new JavaScript program with the Node-RED software, which allows recovering the data from the PLC and saving them into a database. Each record is composed of the name of the variable, the date of occurrence of the change, starting time, and its value. This task needs information on the characteristics of the PLC and the plant. A connection to the PLC is established to record at each PLC cycle the input/output values and to save them into the Database. While a DT is used to simulate the normal plant's behavior, the PLC program ensures the control of this system. As result, all records are saved on a Data Base. Table 1 shows an extract from the records.

3.4 Data Preprocessing

In this step, we transform data obtained previously, depending on the architecture used in the training step and depending on the number of past steps to be given to our neural network. The date of each state is changed to be relative to the last state, and the data is scaled between 0 and 1 (see Tables 2 and 3). Let N_O be the number of actuators (outputs of PLC) and N_I be the number of sensors (inputs of PLC). Let N_{past} and N_{future} be two strictly positive integers. We give the learning algorithm N_{past} states as input, and we predict the N_{future} future states.

We aim to predict the future state of the system to tell even it is normal or faulty and what fault that has arisen in the case of fault. However, in case of stuck to zero or stuck to one of an output (an actuator), the Acquisition task will save the value send by the CP and not the value of this output in the plant. So, the prediction of actuators

Table 1 Extract from the records

A0	A1	A2	A3	A4	A5	A6	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	Time
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1671112186784
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1671112188584
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1671112189084
0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1671112191059
1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1671112191110

Table 2 Relative time

A0	A1	A2	A3	A4	A5	A6	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	Time
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1800
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	500
0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1775
1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	51

Table 3 Data after rescaling

A0	A1	A2	A3	A4	A5	A6	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	Time
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.000000
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.455696
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.126582
0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0.500000
1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0.012911

is ignored. In different words, even if there is a fault in the actuator, the acquisition system does not transmit the real value of the actuator but the value sent by the CP.

Moreover, some inputs are directly influenced by external factors in the system. For example, a human operator puts a part on a conveyor (that activates a sensor). These sensors' values are not predictable. Let N_E be the number of these not-predictable sensors ($N_E < N_I$) and N_P be the number of predictable sensors. So, $N_I = N_P + N_E$. Let I_P be the vector of sensors to predict. $I_P = [I_{P_1}, I_{P_2}, \ldots, I_{P_{N_P}}]$. So, in this work, we use LSTM to predict I_P values from $T+1$ to $T + N_{future}$. Let $Y_P(t)$ be the binary vector that contains the value of I_P components at the tth time-step. Thus $Y_P(t) = [y_{t,1}, y_{t,2}, \ldots, y_{t,N_P}]$, where $y_{t,\alpha}$ stands for the binary value of I_{P_α} at the tth time-step. Similarly, the prediction vector $\hat{Y}_P(t) = [\hat{y}_{t,1}, \hat{y}_{t,2}, \ldots, \hat{y}_{t,N_P}]$, where $\hat{y}_{t,\alpha} = p\left(y_{t,\alpha} = 1\right)$ is the predicted probability for a given time-step t, that the αth component of I_P is equal to 1. Then $Y_P = [Y_P(t+1), \ldots, Y_P\left(t + N_{future}\right)]$, and $\hat{Y}_P = [\hat{Y}_P(t+1), \ldots, \hat{Y}_P\left(t + N_{future}\right)]$. $\hat{Y}_P$ is the output of the RNN model and Y_P is the ground truth target.

The input of the RNN model is the vector $X = [X(t - N_{past}), \ldots, X(t)]$, where $X(t)$ is the concatenation of the vector of all binary values of $\overrightarrow{I/O}$ at the tth time-step. $X(t) = [x_{t,1}, x_{t,2}, \ldots, x_{t,N_I+N_O}]$, where $x_{t,\alpha}$ stands for the binary value of the αth component of $\overrightarrow{I/O}$ at the tth time-step.

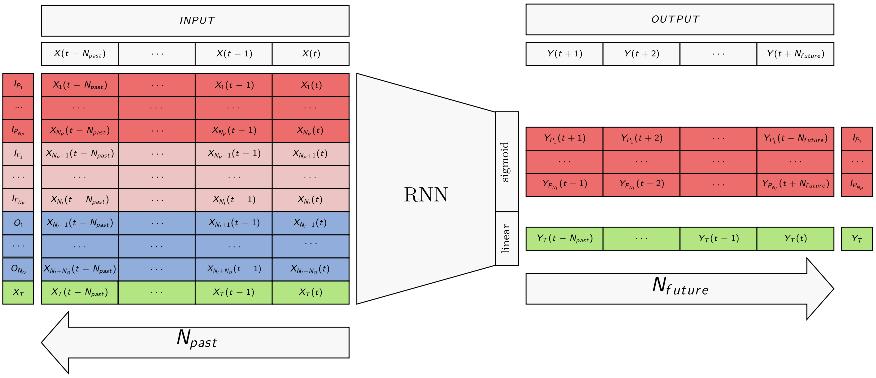


Fig. 2 Input-output architecture of the RNN model

3.5 Model Training

To train the proposed model, we used long short-term memory (LSTM) recurrent neural networks (RNNs) Hochreiter and Schmidhuber [6]. RNN is an effective neural model for a wide range of applications involving temporal or sequential data Karpathy et al. [8] such as video analysis, speech recognition Graves et al. [4], language modeling, handwriting recognition, or its generation, machine translation, image captioning, etc.

Figure 2 represents the architecture of the input and output of the RNN model. The data collected is unbalanced, some components keep true or false values most of the time. The model may be biased toward the most common value for his prediction. To counter this undesirable effect, a weighted binary cross-entropy (WBCE) loss function was used. This function takes into account this unbalanced representation and can compensate for it. Let's define W_P and W_N respectively the weights for the positive (true or 1) and negative (false or 0) binary values for a given future feature. For each element of I_P his W_P and W_N associated weights are proportional to the inverse frequency of the presence of 1 and 0 in the dataset. Let consider I_{P_α} the αth component of I_P. Let n_0 and n_1 be the numbers of occurrences of 0 and 1 for I_{P_α} in the dataset. Then $W_P\left(I_{P_\alpha}\right)$ and $W_N\left(I_{P_\alpha}\right)$ was defined as follows: $W_P\left(I_{P_\alpha}\right) = \frac{n_0+n_1}{2\cdot n_1}$ and $W_N\left(I_{P_\alpha}\right) = \frac{n_0+n_1}{2\cdot n_0}$.

The WBCE also allows another weighting factor to consider the importance of certain time-step over others. This is done by using a vector of time-weight $W_T = [w_1, \ldots, w_{N_{future}}]$ that can be set freely depending on the application.

$$
\begin{aligned}
WBCE(Y_P, \hat{Y}_P) = -\frac{1}{N_{future} \cdot N_P} \sum_{j=1}^{N_{future}} W_T(j) \sum_{i=1}^{N_P} W_P(i) y_{i,j} \log\left(\hat{y}_{i,j}\right) \\
+ W_N(i) y'_{i,j} \log\left(\hat{y}'_{i,j}\right)
\end{aligned}
\tag{1}
$$

where $y'_{i,j} = 1 - y_{i,j}$ and $\hat{y}'_{i,j} = 1 - \hat{y}_{i,j}$.

For the time prediction, which is a classical regression problem, we use the Mean Square Error (MSE) function.

$$MSE(Y_T, \hat{Y}_T) = \frac{1}{N_{future}} \sum_{t=1}^{N_{future}} \left(y_t - \hat{y}_t\right)^2 \tag{2}$$

where $Y_T = [y_t, \ldots, y_{N_{future}}]$ and $\hat{Y}_T = [\hat{y}_t, \ldots, \hat{y}_{N_{future}}]$ with y_t the value of the ground truth relative time between $T + t - 1$ and $T + t$ and $\hat{y}_t$ the predicted value of the relative time between these two time-steps.

The loss used for the training is the sum of both loss functions to predict the future features (WBCE) and the regressive time (MSE).

$$Loss(Y, \hat{Y}) = (1 - \lambda)WBCE(Y_P, \hat{Y}_P) + \lambda MSE(Y_T, \hat{Y}_T) \tag{3}$$

where $\lambda \in [0, 1]$ is a parameter that allows modulating the contribution of the time-regression loss regarding the binary classification loss, $Y = [Y_P, Y_T]$ and $\hat{Y} = [\hat{Y}_P, \hat{Y}_T]$ are the concatenation of the two vectors Y_P with Y_T, and $\hat{Y}_P$ with $\hat{Y}_T$ respectively.

4 Application on the Sorting by Height System

4.1 Use Case Descreption

We apply our approach to the Sorting by Height system [3], a virtual plant simulated using the 3D simulation software Factory I/O (see Fig. 3). The objective of the sorting system is to bring boxes of entry conveyors to the exit conveyor by sorting them according to their heights. The system has 12 sensors (c0–c11) to determine box size (small or large) and the box entry or exit in different conveyors (feeding, intermediate, and evacuation) or turntable. It has also 7 actuators (A0–A6) to activate the various conveyors and the turntable ($N_I = 12$ and $N_O = 7$). A present gantry on the entry conveyor is composed of two sensors c1 and c2 at different heights. If the c2 sensor undergoes a rising Edge the box is considered high enough and will be sent to the left by the rotary conveyor. If only c1 undergoes a rising Edge, then the box will be sent to the right. The rotary conveyor is composed of a motor allowing its rotation at 90° controlled by the actuator A4. The sensors c4 and c5 report their orientation (0° for c4 and 90° for c5). Sensor c6 indicates if a box is loaded on the rotary conveyor. The actuators A2 and A3 allow respectively to move the box forward and backward on the rotating conveyor when it is in its initial position. When the rotating conveyor is oriented at 90°, A2 moves the case to the left and A3 moves the case to the right.

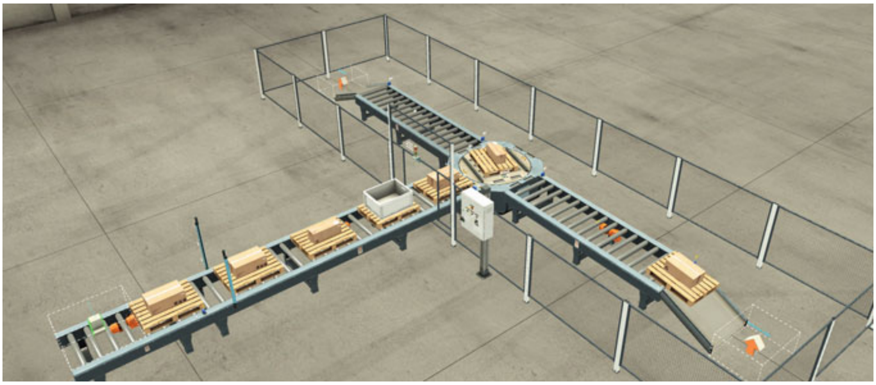


Fig. 3 Sorting by height station

All other actuators (A0, A1, A5, A6) are used to operate the static conveyors. Each conveyor has a sensor at the beginning and at the end of its path.

4.2 Experimental Results

In this paragraph, we present the results of the application of our approach to the Sorting by Height system. For the understanding of the problem step (Sect. 3.2), we have developed the control program for the Sorting System. We used Unity Pro software. For the data acquisition step (Sect. 3.3), we started the system in normal mode. We collected the changes in variable values during several cycles. Three sensors are directly influenced by external factors (their changes depend on the human operator who puts part and retrieves sorted parts). Hence, these three sensors composed the set $I_E = \{c0,\ c2\ c11\}$. Table 4 presents different plant components and classes them by type: actuator, predictable sensors, and non-predictable sensors.

For the training, we used 80% of the dataset with a cross-validation K-fold ($K = 5$). The other 20% of the dataset is kept for the test. We trained on 1500 epochs for $N_{past} = 5$, 10, 15, and 20. N_{future} was fixed to 3. The RNN model used was composed of 3 layers of 50 LSTM cells. The output layer was composed of $N_P + 1$ classical neurons. The first N_P neurons were activated with the sigmoid function to predict $\hat{Y}_P$, and the last neuron was activated with the linear function to predict the next event $\hat{Y}_T$ time. Each model was trained with the Adam optimizer with a learning rate of 0.0015 and a batch size of 128. The loss function used is defined in 3 with $\lambda = 0$ to ignore time prediction (we are not focused on the time prediction for this work). $W_T = [1, 1, 1]$: the same importance was given to each time prediction. W_P and W_N were computed as describe in Sect. 3.5. Table 5 shows W_P and W_N results.

Figure 4 shows loss's value evolution, during the training process, evaluated on the training and validation data sets. We can observe that after a plate between epochs

Table 4 Box sorting components

O	I_P	I_E
A0	c1	c0
A1	c3	c2
A2	c4	c11
A3	c5	
A4	c6	
A5	c7	
A6	c8	
	c9	
	c10	

Table 5 W_P and W_N (round to 4 decimals) for all I_P of Box sorting station

I_P	W_N	W_P
c1	0.9568607068607069	1.0472127417519910
c3	1.7433712121212122	0.7010662604722011
c4	1.0115384615384615	0.9887218045112782
c5	0.708076923076923	1.7014787430683920
c6	0.7735294117647059	1.4139784946236560
c7	0.514533258803801	17.701923076923077
c8	0.5336231884057971	7.9353448275862070
c9	0.5296317606444189	8.9368932038834950
c10	0.5326967592592593	8.1460176991150440

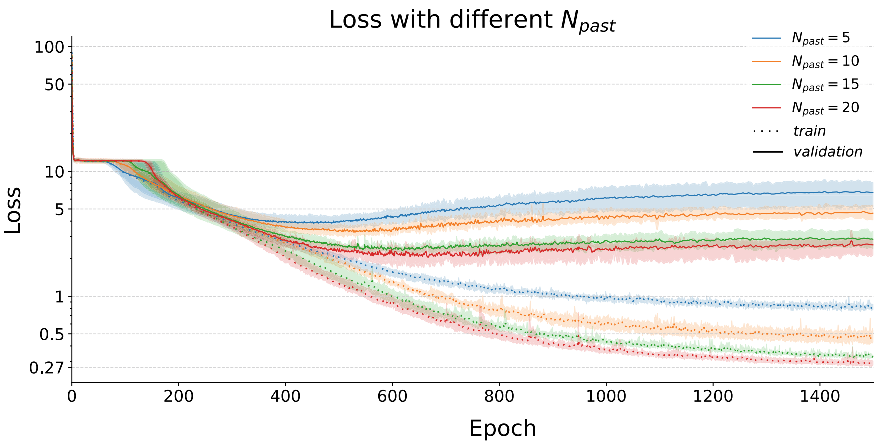


Fig. 4 Evolution of the $WBCE$ on validation and training data for different values of N_{past} (logarithmic scale)

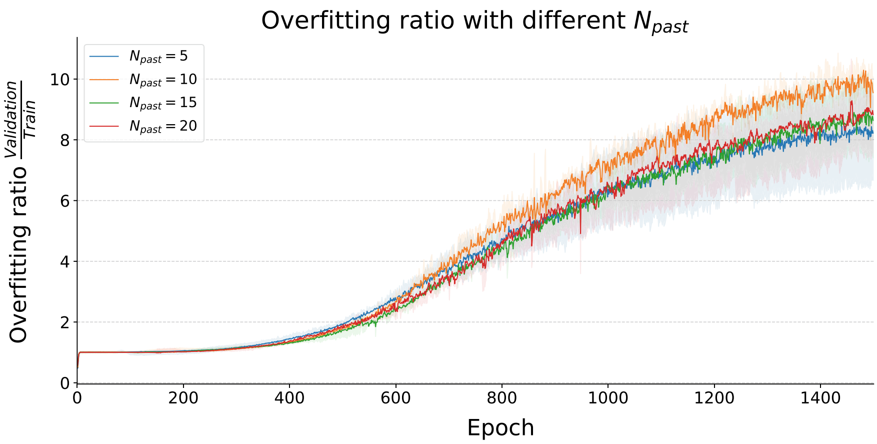


Fig. 5 Overfitting effect for the 4 models

10 and 150, all models converge to a minimum. All the models suffer more or less strongly from the overfitting effect. In general, the more N_{past} is high, the lower the validation error is. We note that models with $N_{past} = 15$ and $N_{past} = 20$ are strongly similar with respect to the difference between models with $N_{past} = 5$ and $N_{past} = 10$. As well as between $N_{past} = 10$ and $N_{past} = 15$. The higher the N_{past} is, the better the model performs on the validation and training data.

Figure 5 underlines the overfitting effect that the models undergo by dividing the error value into the validation data by the error on the training data for each epoch. They all follow the same pattern of overfitting. The model with $N_{past} = 10$ is the most affected.

Figure 6 shows the predicted vector'S accuracy measurement on step $T + 1$. Results are correlated with those observed in Fig. 4. We can note that the training data's accuracy converges perfectly, but the overfitting effect blocks the generalization on the validation data which does not exceed 85%. The accuracy of the validation data does not exceed 85% for the model with $N_{past} = 20$.

Figure 7 shows accuracy results on $T + 1$, $T + 2$, and $T + 3$ on test data with their weights at the 1500th epoch. We note that results are not correlated with those observed previously on validation data. In fact, it could be indicated a lack of training data. As expected, for each model, the more the predicted data are at a distant stage, the lower the accuracy is. The plant state vector is more difficult to predict at $T + 3$ than at $T + 2$ which is harder to predict than at $T + 1$.

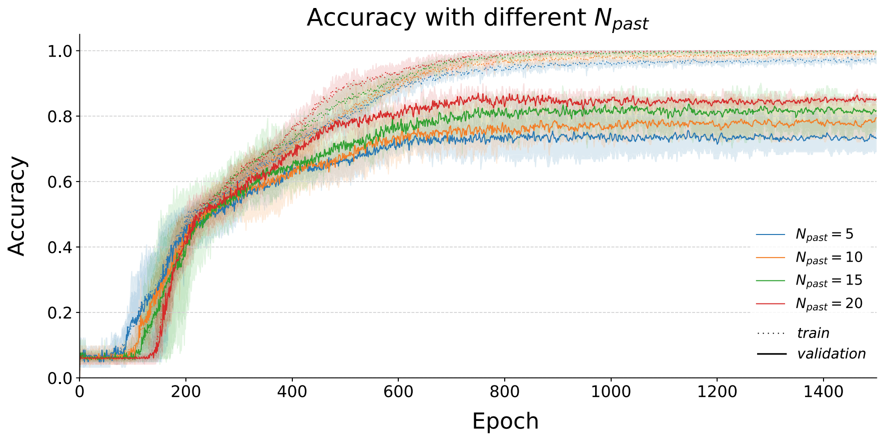


Fig. 6 Evolution of the accuracy of $T + 1$ vector prediction on validation and training data for different N_{past}

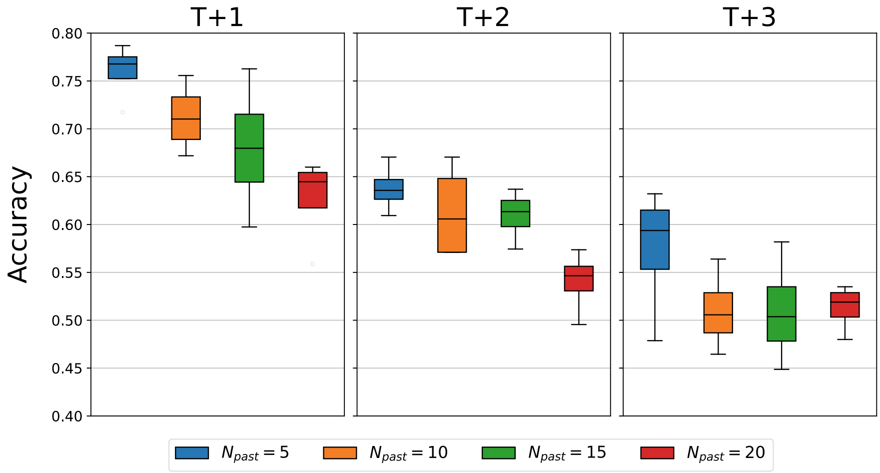


Fig. 7 Accuracy on test dataset on all model with their weights at the 1500th epoch

4.3 Diagnoser Results

The RNN model is trained only on normal behavior data of the APS. We used DT to generate faulty behaviors. The DT allows us to inject faults into the APS (stuck to 1 or stuck to 0 of a component) with the failures panel. To evaluate the diagnoser's ability to diagnose faults, we gave it a faulty behavior and we tested if it detected and identified the fault(s) using Algorithm 1.

Algorithm 1 compares the prediction with the system observation. In the case of normal behavior and good model prediction, no difference exists between the vectors

Algorithm 1: Diagnoser using trained RNN model

Input : RNN Model M,
A stream of SAP change Y_P,
$Pred \in [1, N_{future}]$ the number of steps to predict and wait for the next prediction
Output: I_F the list of failed components

$X \leftarrow [X(0), X(1), \ldots, X(N_{past} - 1)]$
$I_F \leftarrow \emptyset$
$i \leftarrow 0$
while *True* **do**
  $\hat{Y}_P \leftarrow M(X)$
  `// Wait` $Pred$ `SAP change(s)`
  for $j \leftarrow 1$ **to** $Pred$ **do**
    $i \leftarrow i + 1$
    $X \leftarrow [X(i), \ldots, X(i + N_{past} - 1)]$
  end
  `// Compare the prediction with the observations`
  if $\hat{Y}_P \neq [Y_P(i), \ldots, Y_P(i + Pred - 1)]$ **then**
    $I_F \leftarrow \{I_{P_j} \mid Y_P(t)_j \neq \hat{Y}_P(t)_j$ for $t \in [i, i + Pred]$ and $j \in [0,\ N_P]\}$
    return I_F
  end
end

Y_P and $\hat{Y}_P$. In the case of abnormal behavior and good model prediction, a difference exists. The model predicts the state of a subset of system sensors. So it can only detect a deviation from the normal behavior of one or more sensor(s) of this set. Algorithm 1 returns the set of sensors affected by this deviation. However, this deviation does not necessarily mean that the sensor in question is faulty, but that a component connected to it could be faulty. In José Guilherme Castro [7] authors propose a data-based method to find a set of components that are part of a sub-system. As an extension of this work, we will use it to return the set of connected components to sensors affected by the deviation. In case of bad prediction, no deduction is possible.

5 Conclusion

In this chapter, we have illustrated the benefit of using DT for online fault diagnosis of a manufacturing system, with sensors and actuators delivering binary signals that can be modeled as DES. We have proposed an improvement of our approach presented in Saddem et al. [12] in order to extend its limitations. Data acquisition of normal and abnormal behaviors is carried out through a DT, using new software developed in JavaScript with the Node-RED software, which allows for recovering the data from a PLC and saving them into a database. The data preparation consists of the transformation of rows saved on a database into vectors for the proposed RNN model. The results of the application of the proposed method on the Sorting system of Factory IO software show the significant contribution and the interest of this method to detect faults and predict plant future behavior. A comparison between model-based and knowledge-based diagnosis methods will be carried out in the near future. Another perspective of this chapter is to extend the set of predictable events from ignored sensors. Among these ignored sensors, some are overpassed for only one event (rising edge or falling edge) and therefore the second event remains predictable. For example, when a human or a robot puts a part in front of a sensor, the rising edge is not predictable while the falling edge remains predictable. An improvement of this work will take place to allow the prediction of the predictable event. Last but not least, the use of time prediction Y_T inside algorithm 1 to detect system blockage, might be another improvement for this work.

References

1. de Souza RP, Moreira MV, Lesage JJ (2020) A hierarchical approach for discrete-event model identification incorporating expert knowledge. IFAC-PapersOnLine 53(4):275–281
2. Dou D, Zhou S (2016) Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery. Appl Soft Comput 46:459–468
3. FactoryIO (2022) Sorting by height advanced. https://docs.factoryio.com/manual/scenes/sorting-by-height-advanced/
4. Graves A, Mohamed Ar, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 6645–6649
5. Han T, Jiang D, Zhao Q, Wang L, Yin K (2018) Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. Trans Inst Meas Control 40(8):2681–2693
6. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
7. José Guilherme Castro MVM, Viana G (2022) Distributed identification of discrete-event systems with the aim of fault detection. In: 16th IFAC workshop on discrete event systems
8. Karpathy A, Johnson J, Fei-Fei L (2015) Visualizing and understanding recurrent networks. arXiv:1506.02078
9. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W (2018) Digital twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine 51(11):1016–1022
10. Moosavian A, Ahmadi H, Tabatabaeefar A, Khazaee M (2013) Comparison of two classifiers; k-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing. Shock Vib 20(2):263–272
11. Saddem R, Philippot A (2014) Causal temporal signature from diagnoser model for online diagnosis of discrete event systems. In: 2014 international conference on control. decision and information technologies (CoDIT). IEEE, pp 551–556
12. Saddem R, Marrakh A, Baptiste D, Dabounou J (2022) Data based approach for online diagnosis of discrete event system. In: 14th IFAC workshop on intelligent manufacturing systems, vol 55, no 2, pp 289–294
13. Sampath M, Sengupta R, Lafortune S, Sinnamohideen K, Teneketzis D (1995) Diagnosability of discrete-event systems. IEEE Trans Autom Control 40(9):1555–1575
14. Tao F, Qi Q, Wang L, Nee A (2019) Digital twins and cyber—Physical systems toward smart manufacturing and industry 4.0: correlation and comparison. Engineering 5(4):653–661
15. Vazan P, Janikova D, Tanuska P, Kebisek M, Cervenanska Z (2017) Using data mining methods for manufacturing process control. IFAC-PapersOnLine 50(1):6178–6183

16. Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K (2003) A review of process fault detection and diagnosis: part iii: process history based methods. Comput Chem Eng 27(3):327–346
17. Zaytoon J, Lafortune S (2013) Overview of fault diagnosis methods for discrete event systems. Ann Rev Control 37(2):308–320