

# Toward a Human-Cyber-Physical System for Real-Time Anomaly Detection

Bojana Bajic , Aleksandar Rikalovic , Nikola Suzic , and Vincenzo Piuri , *Member, IEEE*

**Abstract**—In recent years, researchers and practitioners have focused on Industry 4.0, emphasizing the role of cyber-physical systems (CPSs) in manufacturing. However, the operationalization of Industry 4.0 has presented many implementation challenges caused by the inability of available technologies to meet industry needs effectively. Furthermore, Industry 4.0 has been criticized for the absence of focus on the human component in CPSs impacting the concept of sustainability in the long run. Responding to this critique and building on the foundation of the Industry 5.0 concept, this article proposes a holistic methodology empowered by human expert knowledge for human-cyber-physical system (HCPS) implementation. The proposed novel HCPS methodology represents a more sustainable solution for companies that consists of five phases to promote the integration of human expert knowledge and cyber and physical parts empowered by big data analytics for real-time anomaly detection. Specifically, real-time anomaly detection is enabled by industrial edge computing for big data optimization, data processing, and the industrial Internet of Things (IIoTs) real-time product quality control. Finally, we implement the developed HCPS solution in a case study from the process industry, where automated system decision-making is achieved. The results obtained indicate that an HCPS, as a strategy for companies, must augment human capabilities and require human involvement in final decision-making, foster meaningful human impact, and create new employment opportunities.

**Index Terms**—Artificial intelligence (AI), big data analytics (BDA), edge computing, industrial Internet of Things (IIoTs), Industry 5.0, smart manufacturing.

Manuscript received 7 October 2023; revised 26 February 2024; accepted 13 May 2024. Date of publication 12 June 2024; date of current version 20 June 2024. This work was supported in part by the Ministry of Science, Technological Development and Innovation under Grant 451-03-65/2024-03/200156 and in part the Faculty of Technical Sciences, University of Novi Sad through project “Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad” under Grant No. 01-3394/1, and in part by the EC under projects EdgeAI (101097300) and GLACIATION (101070141) and project SERICS (PE00000014) under the MUR NRRP funded by the EU – NextGenerationEU, and by OLINEA bilateral project. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Italian MUR. Neither the European Union nor Italian MUR can be held responsible for them. (*Corresponding author: Vincenzo Piuri.*)

Bojana Bajic and Aleksandar Rikalovic are with the Department of Industrial Engineering and Management, Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia, and also with the Institute for Artificial Intelligence Research and Development of Serbia, 21000 Novi Sad, Serbia (e-mail: bojana.bajic@ivi.ac.rs; a.rikalovic@uns.ac.rs).

Nikola Suzic is with the Department of Industrial Engineering, University of Trento, 38123 Trento, Italy (e-mail: nikola.suzic@unitn.it).

Vincenzo Piuri is with the Department of Computer Science, Università degli Studi di Milano, 26013 Milano, Italy (e-mail: vincenzo.piuri@unimi.it).

Digital Object Identifier 10.1109/JSYST.2024.3402978

## I. INTRODUCTION

THE fourth industrial revolution (i.e., Industry 4.0) marked the beginning of the digital transformation era [1], [2], [3]. It is a widely discussed topic in both academia and industry, with many manufacturing companies expecting it to revolutionize their supply chains, operations, and business models [4], [5]. On the one hand, the implementation challenges of Industry 4.0 [1], [2] have proven to be more complex than expected, hindering companies from fully deploying advanced technologies. While Industry 4.0 has focused on creating smart, connected manufacturing systems using cyber-physical systems (CPSs) based on the industrial Internet of Things (IIoT), artificial intelligence (AI), and big data analytics (BDA), it has been criticized for lacking a human focus as its main shortcoming. This lack of research has been reflected in the low level of expert knowledge integration and data-driven approaches by designing systems that diminish human capabilities and require limited human involvement in decision-making.

With overwhelming evidence that industry change goes beyond mere technological transformation, in 2021, a new concept, Industry 5.0, was introduced, emphasizing human centricity, sustainability, and resilience as the core elements for industrial progress [6], [7]. Industry 5.0 represents a paradigmatic shift that emphasizes the harmonious integration of human and technological capacities. This approach intends to ameliorate the shortcomings of Industry 4.0 by establishing a synergistic relationship between humans and advanced technologies. Thus, Industry 5.0 not only mitigates the historical overemphasis on automation but also augments adaptability and productivity by fostering collaboration and human centricity; thus, the development of human-cyber-physical system (HCPS) based on human-machine integration is proposed.

Achieving human-machine synergy in an HCPS requires a holistic methodology that combines human expert knowledge with advanced technological solutions (i.e., BDA and IIoT) to analyze large volumes of process data [8]. However, this synergy is difficult to achieve under industrial conditions, especially in real-time process data analysis [1], [2] for predictive maintenance, processing, and product quality control. In this article, we refer to real-time process data analysis as a time frame in which the system operates with minimal or specifically defined delay, ensuring timely and accurate handling of information to meet the demands of industrial processes and control systems.

Furthermore, the challenges in human-machine integration revolve around the need to change the human mindset and

resistance to accepting and working with advanced technologies [2]. Additionally, the lack of human trust in a technological system is an issue since workers do not believe in data reliability [9]. Therefore, these challenges raise the question of whom to believe—the data or expert experience.

Thus, the main objective of this article is to support system designers by developing a holistic HCPS design methodology for real-time anomaly detection in a smart manufacturing environment. Notably, the methodology aims to support the acceptance of advanced technology by implementing user-friendly solutions, reducing employee resistance to change and incorporating human expert knowledge into each phase.

This novel holistic system design methodology (named the *HCPS design methodology for real-time anomaly detection*)<sup>1</sup> aims to create a smart manufacturing system empowered by human-cyber-physical integration. The methodology development is inspired by the Industry 5.0 data mining methodology and informed by the research team’s practical experience in implementing smart manufacturing systems. The implementation of the HCPS design methodology can help manufacturing companies improve their operational efficiency, reduce waste, and enhance product quality in real-time. These effects are made possible through the integration of human expert knowledge and cyber and physical parts via BDA, edge computing and the IIoT.

Finally, the developed HCPS design methodology is applied under industrial conditions in a case company in the process industry sector (i.e., vinyl flooring production). In the case study, automated real-time anomaly detection for monitoring and control of manufacturing systems was achieved.

The rest of this article is organized as follows. Section II describes related work on the use of HCPS in Industry 5.0 and BDA for anomaly detection, emphasizing the most commonly used advanced analytical methods based on an imbalanced dataset. Section III presents the proposed system design methodology. Section IV provides the details on the settings in which the developed methodology was tested. Section V describes the results of the developed system design methodology testing in the industrial surroundings. Finally, Section VI discusses the results, summarizes the contributions, and derives the conclusions of this article.

## II. RELATED WORK

The present section provides an overview of an HCPS within the Industry 5.0 context. The role of BDA is explored via different anomaly detection models in smart manufacturing.

### A. Human-Cyber-Physical Systems in Industry 5.0

Before Industry 4.0 emerged, the human factor was central to the design, maintenance, and supervision of manufacturing systems [10]. However, the global Industry 4.0 hype somehow neglected human factor and moved its focus to completely automated manufacturing [10], [11]. The goal of moving from traditional to smart manufacturing was the deployment of advanced

technologies [1], [2]. This interconnected bundle of advanced technologies imagines CPSs with closely integrated physical objects and software via the IIoT to enable information exchange based on process data analysis among different components [1], [2]. However, both industry and academia have frequently pointed out the challenge related to the limitation of the human aspect of the manufacturing process [12].

Hence, Industry 5.0 puts the human aspect back at the center of manufacturing processes [6], [7], [13], proposing the development of the HCPS [14]. The literature provides a definition of HCPS within Industry 4.0 [15]. However, information on exactly what the word “human” refers to in this term is lacking. Some authors argue that the term “human” in an HCPS refers to different roles of operators who should integrate and collaborate with machines [16], [17]—such as collaborative robots [11], [18], [19]—while others argue that the term “human” refers to human knowledge implemented within a CPS [20], [21]. Moreover, the role of data scientists as important links connecting cyber and physical components with human expert knowledge based on BDA has been neglected in the literature. Responding to this lack of comprehensiveness in current definitions, we define an HCPS as *an advanced intelligent system comprising various roles that humans may have in companies (i.e., managers [22], experienced engineers [21], [23], and skilled shop floor workers [24], [25]), and data scientists integrating expert knowledge into a CPS to achieve smart manufacturing systems by using BDA and manufacturing data [21].*

### B. Big Data Analytics for Anomaly Detection

Smart manufacturing has revolutionized various industries by providing enhanced efficiency and innovative solutions. This is made possible by processing the abundance of data generated during production for precise decision-making. However, smart manufacturing implementation encounters challenges when dealing with massive amounts of data [1], [2]. Thus, the use of BDA is expected to continue, as it enables independent decision-making by production machines and the creation of intelligent, flexible, and self-adaptive manufacturing systems [26], [27].

When dealing with large datasets, selecting only the relevant process parameters becomes critical due to the complexity of BDA methods. Choosing an appropriate BDA method that effectively provides insight into the real state of manufacturing systems depends heavily on the set of relevant process data. The process data should consist of a balanced set of data containing equal samples from when the manufacturing system functioned smoothly and when problems occurred [28].

However, in companies where continuous improvement of manufacturing systems is implemented [29], [30], anomalies created during the manufacturing process are largely eliminated. As a result, it becomes difficult to establish a balance between the number of process data samples that did not lead to anomaly occurrences and those that caused those anomaly issues. This often leads to an imbalanced set of data where most of the collected process data samples belong to one specific class without providing information on the specific anomalies that need to

<sup>1</sup>In further text for readability reasons referred to as *HCPS design methodology*.

be detected. Thus, detecting anomalies based on an imbalanced dataset has attracted significant attention, highlighting that the use of classic analytical techniques can be a critical error when studying such problems [31], [32]. Notably, even though deep learning techniques are used in many state-of-the-art applications for anomaly detection in Big Data [33], [34], dealing with highly imbalanced real-world datasets, especially when acquiring data in industrial settings, can pose significant challenges. Classifying such datasets with imbalances may lead to issues such as overfitting in minority categories and the dominance of majority categories [35]. According to the relevant literature [32], [33], [36], [37], [38], the advanced analytical methods that are most commonly used for anomaly detection based on an imbalance dataset include the Mahalanobis–Taguchi system (MTS), the one-class support vector machine (OCSVM), the isolation forest (IF), the local outlier factor (LOF), and the robust covariance (RC).

- 1) The Mahalanobis–Taguchi system represents a pattern recognition technology that integrates the Mahalanobis distance (MD) and robust engineering via the Taguchi method and is mainly used for process and product quality control [39], [40], [41]. The MD is calculated based on the correlation between parameters and different patterns that can be identified and analyzed concerning a reference dataset

$$\text{MD}_j = \frac{1}{k} Z_{ij}^T A^{-1} Z_{ij} \quad (1)$$

and  $Z_{ij}$  is determined by  $Z_{ij} = \frac{x_{ij} - m_i}{s_i}$

where  $k$  refers to the total number of parameters;  $i$  is the serial number of parameter ( $i = 1, 2, \dots, k$ );  $j$  represents the number of samples ( $j = 1, 2, \dots, n$ );  $T$  is the transposed vector;  $A^{-1}$  is the inverted correlation matrix;  $Z_{ij}$  represents the standardized vector of normalized features of  $x_{ij}$ ;  $x_{ij}$  is the value of the  $i$ th parameter of the  $j$ th observed sample;  $m_i$  refers to the mean value of the  $i$ th parameter; and  $s_i$  is the standard deviation of the  $i$ th parameter.

- 2) The OCSVM is an unsupervised machine learning (ML) technique that involves fitting a hyperplane to most of the training data. The OCSVM identifies anomalies by minimizing the hyperplane of a single class of examples in the training data and considers all the other samples outside the hypersphere to be outliers or out of the training data distribution [42]. The optimal hyperplane parameters are found by solving an optimization problem

$$\begin{aligned} \min_{w, \rho, \xi} \quad & \frac{\|\omega\|^2}{2} + \frac{1}{\nu m} \sum_{i=1}^m \xi_i - \rho \\ \text{s.t.} \quad & \omega \cdot \varphi(x_i) \geq \rho - \xi_i, \quad i = 1, 2, \dots, m \\ & \xi_i \geq 0 \end{aligned} \quad (2)$$

$\varphi: R^d R^d \rightarrow R^n R^n$  is a nonlinear mapping function from the input to the feature space;  $w$  and  $\rho$  are the parameters of the hyperplane; the slack variables  $\xi = [\xi_1, \dots, \xi_m]$  allow the presence of anomalous examples in the training set; and  $\nu$  limits the fraction of training examples classified as anomalies [43].

- 3) The IF method is an unsupervised ML technique used for anomaly detection in a dataset. The algorithm works by randomly partitioning data units into binary trees until all the data units are isolated in their leaf nodes. The anomaly score for a data unit is then calculated as the average path length from the root node to the leaf node in which the data units reside [44]. The equation that provides a scoring system facilitating the classification of samples as either normal or abnormal is

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (3)$$

where for a given  $x$ ,  $h(x)$  corresponds to the length of the path from the root node to the outer node that isolates  $x$  in a given forest tree,  $E(h(x))$  is the average length of paths in the forest, and  $c(n)$  is the average length of the IF for  $n$  samples, calculated as

$$c(n) = 2H(n-1) - (2(n-1)/n) \quad (4)$$

where  $H$  is a harmonic number [45].

- 4) The LOF is an unsupervised ML technique used for anomaly detection in datasets. It measures the local deviation of a data unit with respect to its neighbors. The basic concept behind the LOF is that outliers will have fewer neighbors in their local neighborhood than nonoutliers [46], [47]. The mathematical formula for the LOF can be expressed as [46]

$$\text{LOF}(p) = \frac{1}{k} \sum_{q \in N_k N_k(p)} \frac{\text{density}(p)}{\text{density}(q)} \quad (5)$$

where  $p$  is the point of interest for which the LOF score is calculated;  $N_k(p)$  is the set of  $k$ -nearest neighbors of  $p$  (including  $p$  itself); and  $\text{density}(p)$  is the local density of  $p$  defined as the inverse of the average distance between  $p$  and its  $k$ -nearest neighbors, that is, [46]

$$\text{density}(p) = \frac{1}{\frac{1}{k} \sum_{q \in N_k(p) N_k(p)} d(p, q)} \quad (6)$$

$d(p, q)$  is the distance between points  $p$  and  $q$ . The LOF score of  $p$  reflects how much more or less dense its local neighborhood is compared to the neighborhoods of its neighbors. A point with an LOF score greater than 1 is considered an outlier, a score of 1 indicates a typical point, and a score less than 1 indicates a point that is denser than its neighbors [47].

- 5) The RC is an unsupervised ML technique that estimates the covariance matrix of a set of variables that is less sensitive to the presence of outliers or nonnormality in the data than traditional covariance estimation methods are. The method identifies outliers by utilizing maximum likelihood estimators (MLEs) to determine the mean and covariance matrix of the normal data. Observations with unusually high MD (1) are considered outliers. The mean ( $\hat{\mu}$ ) and covariance matrix ( $\hat{\Sigma}$ ) estimated using MLEs are

given as [47]

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n a_i \quad (7)$$

$$\widehat{\Sigma} = \sum_{i=1}^n (a_i - \hat{\mu})(a_i - \hat{\mu})^T / (n - 1) \quad (8)$$

where  $A = \{a_1, \dots, a_n\}$  is an  $n$ -sample dataset of  $d$  dimensions and each observation  $a_i = \{a_{i1}, \dots, a_{id}\}$ .

### III. DEVELOPED HCPS DESIGN METHODOLOGY

This section proposes a system design methodology for real-time product anomaly detection in the HCPS environment. The system design methodology emphasizes real-time quality control improvement through the integration of human expert knowledge and cyber and physical parts empowered by BDA and connected with the IIoT.

The developed methodology is inspired by human-cyber-physical integration based on fuzzy expert systems and a data mining approach. Notably, the methodology development is informed by the practical field experience of the research team in the implementation of smart manufacturing systems in the industry. The developed HCPS system design methodology consists of five phases (see Fig. 1).

- 1) *Phase 1*: Assessment of company readiness for smart manufacturing implementation.
- 2) *Phase 2*: Edge computing integration of physical and cyber components.
- 3) *Phase 3*: Edge computing dataset optimization.
- 4) *Phase 4*: Anomaly detection model establishment.
- 5) *Phase 5*: Real-time anomaly detection model deployment.

Noticeably, different types of human expert knowledge play a critical role throughout all stages of the proposed methodology. Specifically, human experts are called upon to give their input, conduct analysis, and provide feedback on various phases of the HCPS methodology. In this way, human expert knowledge adds value to HCPS integration through the entire architecture for real-time anomaly final decision-making (see Fig. 2). This acknowledgment underscores the indispensable nature of human expertise in guiding and informing decision-making processes at various stages of the proposed methodology's implementation. By incorporating different types of human expert knowledge throughout the entire methodology, a holistic approach is proposed, enabling a more accurate and insightful assessment of anomalies, the formulation of effective strategies, and the generation of valuable insights into anomaly occurrence. The continuous involvement of human expertise guarantees that the methodology remains adaptive, responsive, and aligned with the evolving needs and challenges of the manufacturing system, ultimately leading to more reliable outcomes. In the following paragraphs, we describe the phases of the proposed HCPS design methodology.

*Phase 1. Assessment of Company Readiness for Smart Manufacturing Implementation:* This phase applies a fuzzy expert system defined through the following criteria: acquisition of human expert knowledge, usage of data analytics, application

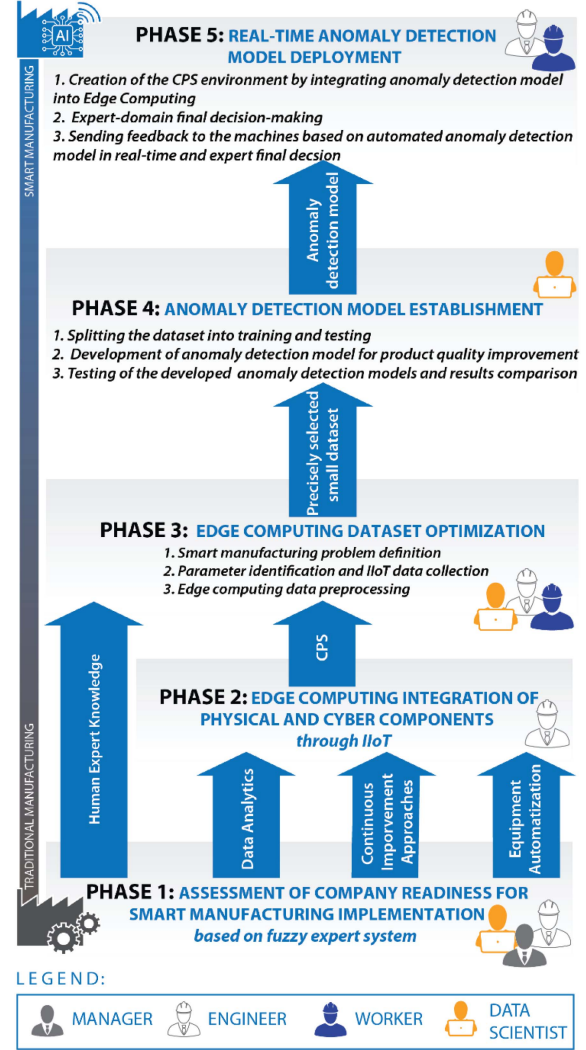


Fig. 1. HCPS design methodology for real-time anomaly detection in a smart manufacturing environment.

of continuous improvement approaches, and level of equipment automatization (see Fig. 1). The four specified criteria are depicted as triangular membership functions because triangles serve as the fundamental membership function for fuzzification. These criteria are derived from the pillars of Industry 5.0, which are resilience, sustainability, and human centricity. Each pillar encompasses both technological and managerial solutions [48]. Technological solutions that contribute to enhancing resilience in companies include edge computing. Sustainability is fostered by technological solutions such as BDA and the IIoT, while managerial solutions involve continuous improvement approaches, such as lean, total quality management, and world class manufacturing (WCM). Furthermore, human centricity involves organizing manufacturing processes with humans (i.e., engineers, workers, and data scientists) at the center, where managers derive the final decision regarding the acceptance of the implementation process. This aspect is facilitated by managerial solutions such as expert experience and industry-academia collaboration. Drawing upon the managerial and technological solutions of Industry 5.0, the criteria for the maturity model are established. These criteria are designed to be general,

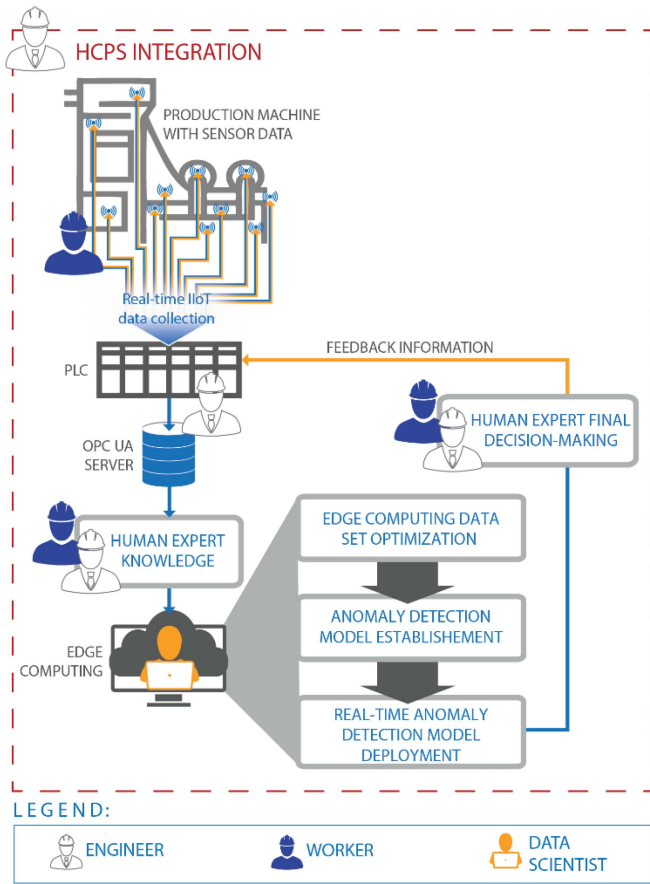


Fig. 2. HCPS integration architecture of the real-time anomaly detection system for quality improvement.

ensuring the model's ease of use. Each criterion represents the organizational aspect of the company as a prerequisite condition to proceed with subsequent phases of the developed methodology. If the company reaches an acceptable level of readiness (above 0.5—based on the designed fuzzy expert system—the company is 50% ready to start the implementation of smart manufacturing), it can move on to the next methodology phase. If not, changes must be made first to meet the defined criteria. Therefore, this phase is the eliminatory phase of the proposed methodology.

**Phase 2. Edge Computing Integration of Physical and Cyber Components:** This phase involves the enhancement of interactions among machinery, devices, sensors, and software [1], [2]. This integration is achieved using edge computing technology and an IIoT network, enabling cyber-physical system creation for real-time information exchange (see Fig. 2). Moreover, since this research relies on BDA and the IIoT, the data privacy and security (e.g., encrypting sensitive and confidential information, controlling access to data to repel the cybersecurity threats, etc. [49], [50]) are additionally reinforced by the use of edge computing. Thus, the use of industrial edge computers within the proposed HCPS guarantees the protection of sensitive and confidential information and reduces the possibility of cyber-security threats since these solutions are deployed within industrial settings (i.e., using secure Ethernet), preventing the need for data transfer via the Internet.

**Phase 3. Edge Computing Dataset Optimization:** This phase focuses on optimizing the dataset through three steps (see Fig. 1): smart manufacturing problem definition for a clear understanding of the processes and activities; parameter identification based on expert knowledge and experience; and preprocessing of the data using edge computing. The main aim of this phase is to streamline Big Data optimization and improve production resilience by rationalizing power and processing resources using edge computing technology. This is achieved by reducing the amount of Big Data to a smaller, precisely selected dataset that still contains crucial information from the original dataset. The selected dataset serves as the foundation for the anomaly detection model establishment phase.

**Phase 4. Anomaly Detection Model Establishment:** This phase comprises three steps (see Fig. 1): dividing the dataset into 80:20 training and testing datasets; developing models for product quality improvement using state-of-the-art anomaly detection methods (namely, MTS, OCSVM, IF, LOF, and RC); and testing the anomaly detection models to compare the results. The final step evaluates the performance of the anomaly detection model through a confusion matrix and measures the following metrics based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values [36], [51], [52]:

- accuracy*, which measures the proportion of correct predictions out of the total number of predictions;
- recall*, which measures the proportion of TPs out of the total number of actual positives; and
- F1 score*, which is the harmonic mean of precision and recall. In addition, the F1 score is a good measure of overall model performance, especially when the classes are imbalanced [53].

These metrics are among the conventional evaluation methods [54] used in classification. The best-performing model is implemented for the real-time validation of product anomaly detection in the case study as the last phase of methodology implementation.

**Phase 5. Real-Time Anomaly Detection Model Deployment:** This phase involves deploying the model in real-time for product quality detection at a specific manufacturing company. The deployment process includes three steps: creating a CPS environment by integrating an anomaly detection model into edge computing; integrating human expert knowledge into the anomaly detection model for final decision-making to create the HCPS environment; and sending automated feedback from the anomaly detection model to the programmable logic controller (PLC) and from the PLC back to the machine sensors in real-time through the IIoT (see Fig. 2). This integration leads to advanced automation and real-time decision-making and the realization of smart manufacturing by creating a closed loop in the HCPS environment.

#### IV. CASE STUDY

The present section provides a detailed explanation of the case study setup for developed methodology testing. Specifically, it describes the case company, manufacturing equipment, and generated dataset.

### A. Case Company

We applied the developed HCPS design methodology to manufacturing data collected from a case company in the vinyl flooring industry. A vinyl flooring company was selected for application of the developed methodology for multiple reasons: the process industry is characterized by high productivity, high production complexity, high levels of automation, and high volumes; the company has successfully implemented a high level of WCM methodology for continuous improvement of their manufacturing system; and the process industry sector (as one of the most automated) is expected to be the first to embrace the implementation of Industry 4.0 as well as to face Industry 4.0 implementation challenges.

Vinyl flooring is produced in the form of a roll on a continuous production line. The specific production line is 850 m long and consists of 12 machines. The number of machines increases the complexity of the manufacturing process and, consequently, the complexity of the proposed methodology. To reduce this complexity and make the whole production process manageable, the production line is divided into five clusters.

A cluster is a part of the production line that includes one or more machines to simplify the observed anomaly problems of the product being of poor quality in the case company. Notably, experience has shown that inconsistencies in the process parameters of one machine can potentially result in a poor-quality product on a subsequent machine, which in turn means that the first machine can influence the quality of the products exiting from the last machine in the production line. Machines must be grouped in the sequence of production, which is a limiting factor for overall product quality attainment. Therefore, locating the occurrence of poor-quality products is determined based on expert knowledge following defined gaps and concerns. Thus, it is concluded that poor-quality products in the manufacturing system most often occur at the beginning of the production line. The beginning of the line includes Cluster 1, which is composed of three machines, namely, coating 1, coating 2, and printing. Given that no errors occur on the printing machine that affect the insufficient quality of the product, the cause of which is the conformity of the process parameters (on the printing machine, low-quality products occur where the cause is related to human work), this machine is exempt from analysis and does not represent one of the locations of noncompliance. Further analysis of the location of nonconformity (i.e., poor-quality products) led to the conclusion that a negligibly small number of anomalies occur on machine coating 1. Therefore, the location of the nonconformity of the parameters in the production system is machine coating 2.

### B. Manufacturing Equipment for Real-Time Product Quality Control

The methodology is applied by designing the system based on the integration of physical [i.e., machines, sensors, and programmable logic controllers—(PLCs)] and virtual (software for data collection, anomaly detection model development, and real-time monitoring and sending feedback back to the

TABLE I  
SYSTEM DESIGN ENVIRONMENT

Physical assets	Machines, sensors, PLCs, industrial computer based on edge computing
Virtual components	Python – developing ML models, namely: OCSVM, IF, LOF, and RC
	Real-time Statistic Diagnosis Tool – development of MTS model
IIoT network	Real-time Flow Designer – real-time monitoring and sending feedback back to the manufacturing system
	OPC UA

manufacturing system) components through open platform communications unified architecture (OPC UA) as an IIoT standard protocol (see Table I). The equipment used is an industrial computer based on edge computing for real-time data collection and analysis—a MELIPC MI5000 developed by Mitsubishi Electric. The goal of using the MELIPC MI5000 is to systematically reduce the amount of Big Data by optimizing it into a precisely selected small dataset. Additionally, this approach is used for the development of anomaly detection models for real-time quality improvement and for sending feedback information to production lines in real-time.

### C. Dataset

The dataset was generated and collected from the real manufacturing environment of the case company—vinyl floor production. Therefore, the dataset is not publicly available due to privacy restrictions.

The original dataset contained 15 selected influential process parameters as the inputs. All process parameters are numerical and related to various temperatures, line production speeds, pressure(s) between rolling drums, and gap between rolling drums. The dataset was collected for a 33-day period due to the storage limitations of edge computing technology and because most of the products belonging to the group of products with a defined parameter configuration were produced during that period (according to the production plan). During that time, 29 403 000 data units grouped into 6534 separate .csv files were collected.

Notably, output data on final poor-quality products were obtained only after the manufacturing process was completed. Therefore, the output was subsequently added to the original dataset. The output of the process parameters is defined in collaboration with the case company experts and has binary values since the manufacturing problem is defined as classification. Thus, the output has a defined value of “0” when the final product has no defects (i.e., a good quality product). Accordingly, the dependent parameter has a defined value of “1” when the final product has defects (i.e., a poor-quality product). The generation of the dependent parameter was performed in .csv data format.

## V. DEVELOPED METHODOLOGY TESTING

This section describes the application of the developed HCPS system design methodology. The five phases of the methodology are described in detail.

TABLE II  
ASSESSMENT OF COMPANY READINESS FOR SMART MANUFACTURING  
IMPLEMENTATION THROUGH INTERVIEWS WITH EXPERTS

Criterion	Linguistic variable	Membership Function	Assigned value by experts
Level of equipment automatization	Low	(0, 0.3, 0.6)	0.5
	Medium	(0.5, 0.7, 0.9)	
	High	(0.8, 1, 1)	
Usage of data analytics	Non-data-are-analyzed	(0, 0.3, 0.6)	0.75
	Some-data-are-analyzed	(0.5, 0.7, 0.9)	
	All-data-are-analyzed	(0.8, 1, 1)	
Application of continuous improvement approaches	Low	(0, 0.3, 0.6)	0.95
	Medium	(0.5, 0.7, 0.9)	
	High	(0.8, 1, 1)	
Acquisition of human expert knowledge	Low	(0, 0.3, 0.6)	0.75
	Medium	(0.5, 0.7, 0.9)	
	High	(0.8, 1, 1)	

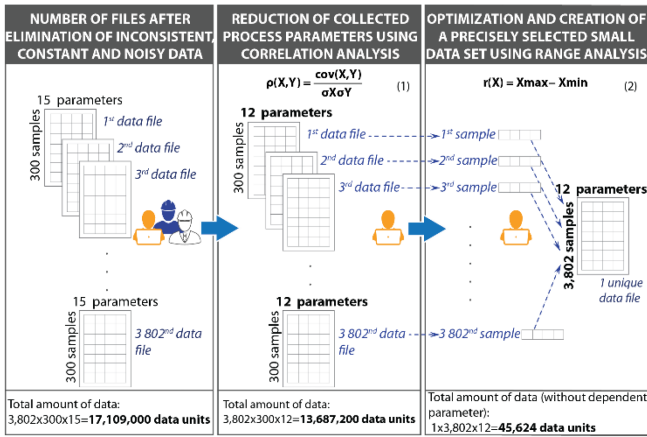


Fig. 3. Phase 3 involved the optimization of the collected data files without losing significant information (adapted from [155]).

### A. Phase 1: Assessment of Company Readiness for Smart Manufacturing Implementation

The assessment of company readiness for smart manufacturing implementation was performed based on a developed fuzzy expert system. The fuzzy expert system serves as a tool for assessing the degree of readiness of the manufacturing system for the development of a model for real-time anomaly detection. By applying the developed expert system, a phased inference system is defined for each decision criterion. The defined criterion was evaluated through interviews with case company experts using linguistic variables to describe the current state of the manufacturing system.

Each criterion is defined by one or more linguistic variables that have quantitatively interpreted values (see Table II). The quantitative value of each linguistic variable is presented using a scale from 0 to 1 based on the knowledge of the manufacturing

system experts. Therefore, by defining the associated functions of the variables, a certain degree of uncertainty appears when defining the boundaries of the area and evaluating the convenience of the obtained values. Observing the uncertainty when defining area boundaries in a fuzzy way allows the interpretation of data by defining intermediate areas and extreme areas. The assessment of company readiness integrates the percentage values of the variables into standardized criteria to support the evaluation of the readiness of the smart manufacturing system to assess its current state. An assessment of company readiness for smart manufacturing implementation was developed for each criterion using the Mamdani fuzzy inference system (FIS). The associated functions were defined based on expert knowledge and data integration using the FIS editor in the MATLAB software package. The developed expert system is highly flexible, making it easy to modify and adapt to different manufacturing systems. This is because all systems share common features in fuzzification and defuzzification, and the main differences lie in the knowledge base. The knowledge base is expressed in terms of phased IF-THEN rules, which form the core of the system. The use of relations such as AND and OR was avoided when defining the knowledge base to simplify it. According to the assigned linguistic variables based on the expert-defined values for each criterion (see Table II), a value of 0.779 was obtained, which indicates that the degree of readiness of the manufacturing system is almost 78%, which is acceptable for the development of an anomaly detection model for real-time product quality control.

### B. Phase 2: Edge Computing Integration of Physical and Cyber Components

Interconnection and intelligent collaboration among basic manufacturing factors, including humans, machines, manufacturing equipment, methods, and the environment, were achieved by implementing edge computing as middleware through an IIoT network. The OPC UA has established itself as an open and platform-independent IIoT standard for data exchange in smart manufacturing use cases. Therefore, implemented edge computing solutions enabled the integration of production machines, industrial computers, and PLC devices as physical components of the system and software solutions for automated analysis of process data as virtual components via the OPC UA server (see Fig. 2).

Furthermore, range analysis, which involved calculating the difference between the maximum and minimum values of each individual .csv file, was employed to assess parameter variability to optimize the dataset. The optimized dataset consisted of one .csv file with 3802 data samples for the 12 most important process parameters, resulting in a final dataset of 45 624 data units (see Fig. 3).

### C. Phase 3: Edge Computing Dataset Optimization

The dataset optimization phase was carried out following the steps outlined in Fig. 1, resulting in the optimization of

29403000 data units of the 15 most important process parameters defined (based on expert knowledge) by 99.73% without losing significant information contained in the big data. The optimization was performed using edge computing technology implemented in the real manufacturing environment of the case company.

Upon visual examination of the data, it was observed that most process parameter values fell within the defined tolerances. Consequently, the collected process parameter values predominantly represented data on the production of high-quality “Class A” products (referred to as “good” data), with only a small number of instances capturing information about poor or inadequate product quality (referred to as “bad” data). As the phase progressed, the monitoring and refinement steps led to a significant reduction in the total number of data units collected. This reduction can be attributed to the following factors.

- 1) Expert knowledge (both engineering and scientific) was utilized to eliminate inconsistent, constant, and noisy data, including the manual removal of timestamp data. This led to an approximately 50% decrease in the total number of data units collected, leaving a remaining count of 17 109 000 data units [55] (see Fig. 3).
- 2) Correlation analysis was also conducted to reduce the number of collected process parameters. Based on the Pearson correlation coefficient, parameters showing high correlation (with  $\rho$  exceeding the values of 0.8 and  $-0.8$ ) were removed, resulting in a decrease from 15 to 12 influential process parameters and a reduction to 13 687 200 data units [55] (see Fig. 3).

Furthermore, range analysis, which involved calculating the difference between the maximum and minimum values of each individual .csv file, was employed to assess parameter variability to optimize the dataset. The optimized dataset consisted of one .csv file with 3802 data samples for the 12 most important process parameters, resulting in a final dataset of 45 624 data units (see Fig. 3). However, the dataset was highly imbalanced because the case company mostly produced high-quality products. This imbalance determines the most suitable technique for developing a predictive model for real-time product anomaly detection.

#### D. Phase 4: Anomaly Detection Model Establishment

The dataset is divided into a training set and a testing set at a ratio of 80:20, which has been proven to be the best division for anomaly detection model development [56]. Here 80% of the optimized dataset is used for training the model, and 20% is used for testing. The predictive models for real-time product anomaly detection were developed based on the process parameters collected from the selected company using methods for anomaly detection on an imbalanced dataset that included MTS, OCSVM, IF, LOF, and RC. To verify the applicability of the proposed system design methodology, we compared the performances of the five anomaly detection models using MTS, OCSVM, IF, LOF, and RC on an imbalanced dataset. To ensure the comparability of the experiments, the other experimental

TABLE III  
EVALUATION PERFORMANCES OF THE ANOMALY DETECTION MODELS

	TP	TN	FN	FP	Accuracy	Recall	F1 score
<b>MTS</b>	<b>747</b>	<b>0</b>	<b>12</b>	<b>1</b>	<b>98.29%</b>	<b>98.41%</b>	<b>99.13%</b>
<b>OCSVM</b>	732	0	28	0	96.32%	96.32%	98.12%
<b>IF</b>	683	0	77	0	89.87%	89.87%	94.66%
<b>LOF</b>	722	0	38	0	95.00%	95.00%	97.44%
<b>RC</b>	684	0	76	0	90.00%	90.00%	94.74%

parameters were kept unchanged. Considering that a model with better performance should be deployed in the manufacturing environment, the developed models are compared based on accuracy, recall, and F1 score. The experimental results are given in Table III.

The results of the accuracy, recall, and F1 score measurements (see Table III) suggest that the MTS model outperformed all the other ML models. Notably, the accuracy, recall, and F1 score are different performance metrics, and they each measure a different aspect of an algorithm’s performance. Additionally, the F1 score is a widely used performance metric in practical applications, especially under industrial conditions when managing an imbalanced dataset. Given the results of the performance evaluation and the significant weight of the F1 score, it is justifiable to implement the MTS model in the manufacturing environment and assess its performance in real-time.

#### E. Phase 5: Real-Time Anomaly Detection Model Deployment

The validation of the developed HCPS methodology for real-time anomaly detection involved deploying a predictive model with the best evaluation performance in a specific manufacturing company. Therefore, the validation of the developed system design methodology for real-time anomaly detection involved the deployment of the predictive model based on the MTS at the case company in the vinyl flooring sector for a defined period of 33 days. The validation focused on a specific group of products and aimed to detect the insufficient quality of the product based on deviations from the defined MD limit. During the validation period, notifications of crossing the defined MD threshold were monitored. Checks were performed on the quality of the resulting errors in the specific group of products in the case company. During the defined validation period, 145 .csv files with 300 data units for each of the 12 influential process parameters were generated. Therefore, the MTS was validated on 522 000 process data units.

The real-time deployment of the anomaly detection model involves automating the analysis of the MTS model at the exact time when the process parameters are collected. The results obtained from this analysis provide new information that is sent back to the machines, referred to as *feedback*. This feedback signal is sent to the PLC device via the OPC UA after real-time detection of any noncompliance with the process parameters using the MTS model. The application of edge computer technology and *Real-Time Flow Designer* software (see Table I)



enabled the sending of feedback back to the machines at the manufacturing site. To facilitate the recording of the new value of the feedback parameter, it was necessary to define it as a binary data type in the edge computing system.

The activation of the feedback parameter is triggered when the MTS model for anomaly detection provides information that the MD has exceeded the defined limit value. At a given moment, the value of the feedback parameter is changed in the OPC UA server; that is, its value changes to 1 so that information is automatically sent back to the PLC device. Modifying the value of the feedback parameter within the OPC UA server serves as an activation element, triggering the resetting of all influential process parameters to their defined mean values in the PLC device. This reset mechanism is necessary because it is possible to detect process parameter noncompliance even when all parameter values are within their specified tolerances. The adjusted parameter values are subsequently sent to the machine to enable automated analysis of process data in the manufacturing system of the case company.

Based on the analysis of the obtained results, it was concluded that during the validation period in the case company, two products were produced that were characterized as poor-quality products that occurred on the observed machine, where the MTS correctly detected the anomalies of those two products that led to the production of products of insufficient quality.

Furthermore, given that the validation of the MTS model was performed based on a real industrial problem, going deeper into the analysis we encountered the problem of the appearance of false alarms. When discussing the results with the company experts, it was concluded that in addition to errors related to poor product quality, system errors are also present in the manufacturing system (referring to the justified occurrence of product quality errors). The developed MTS model predicted not only product quality issues but also system errors, which were initially classified as false alarms, although they refer to quality errors generated during the machine configuration process. However, this does not diminish the fact that system errors lead to the occurrence of product anomalies detected by the MTS model.

## VI. DISCUSSION AND CONCLUSION

The present research is built on two fundamental issues. The first issue is related to the critique of the Industry 4.0 concept, which puts the human factor out of the focus while striving toward technological development. The second issue, important both for industry and academia alike, relates to the lack of examples of fully implemented Industry 4.0, Industry 5.0, and smart manufacturing solutions in practice.

To address these two issues, we developed a new HCPS design methodology for real-time anomaly detection. The developed methodology provides a holistic approach for system designers to implement smart manufacturing solutions in companies that involve various stakeholders (i.e., managers, engineers, workers, and data scientists) by considering human knowledge and experience, cyber systems, and physical objects connected via the IIoT. Since a CPS is not flexible enough when a problem

occurs, the human aspect of an HCPS adds value to the entire decision-making system—a human expert can recognize real mistakes based on expert knowledge and distinguish what is important from what is not.

The proposed methodology for detecting product anomalies in real-time strongly emphasizes enhancing quality control through human-cyber-physical interactions, utilizing the power of BDA, and edge computing in the IIoT environment. Thus, the convergence of all these elements empowers the creation of an HCPS achieved through the implementation of five phases of the proposed methodology described in Section III.

In the following paragraphs, we summarize the main contributions of this article.

*The proposed HCPS design methodology relies on human expert knowledge:* The proposed HCPS design methodology places strong emphasis on harnessing the full potential of different kinds of human expert knowledge, contrary to other reported methodologies that discuss the interaction between human experts and machines [13], [14], [19]. Additionally, the proposed methodology is designed to include expert knowledge in various domains (i.e., managers, engineers, workers, and data scientists) to enable knowledge transfer and thereby strengthen the entire system design process. In addition, this methodology contributes to the continuous participation of human experts, not only in phase 1, where the initial state of the company's readiness for smart manufacturing implementation is assessed, but also through all phases of the system design process. Thus, we expect that the HCPS system design methodology can help system designers leverage expert insights to improve and optimize their manufacturing systems.

*The proposed HCPS system design methodology is holistic:* The proposed HCPS design methodology for real-time anomaly detection in a smart manufacturing environment is holistic. This can also be seen when compared with other focused approaches available in the literature. Although focused approaches have high value for the introduction of real-time anomaly detection in smart manufacturing environments [42], [57], they usually focus on and describe one of the phases based on Industry 4.0 and Industry 5.0 concepts. We expect that the holistic methodology proposed in this article will have high value for industry and researchers connecting the various phases as a coherent whole.

*The proposed methodology was developed based on an imbalanced process dataset collected in industry:* The proposed methodology was developed and tested based on industrial application through process data collection, optimization, model development, and automated data analysis for real-time anomaly detection in a case study. At the time of this article, there is a gap in practical proof in the relevant literature on the developed theoretical methodologies for both Industry 4.0 and Industry 5.0 [2], [5], [58]. Specifically, evidence of the implementation of data analytics, such as Industry 4.0 and Industry 5.0 technology, is scarce [59], [60], [61]. The main reason for the lack of implementation evidence is related to the lack of dataset quality [1]. In other words, an obstacle to implementing smart manufacturing solutions in the process industry is encountered in the form of an imbalanced set of generated data. An imbalanced dataset

indicates homogeneity, where the quality of the process data is significantly reduced. Such cases significantly complicate the application of data analysis methods. To address this gap, in this article, we demonstrated that by using imbalanced process data from a manufacturing system, it is possible to achieve automated real-time anomaly detection analysis in the HCPS environment.

*The traditional anomaly detection model outperformed state-of-the-art ML methods:* During the development of the model, five different state-of-the-art anomaly detection methods were used for imbalanced datasets, namely, MTS, OCSVM, IF, LOF, and RC. Notably, OCSVM, IF, LOF, and RC represent unsupervised ML methods, while MTS is an advanced statistical process control method [62]. Interestingly, the developed MTS model outperformed all the other ML models in terms of accuracy, recall and F1 score. Noticeably, the obtained results indicate that although ML models tend to perform better than advanced statistical methods [63], [64], advanced statistics have been increasingly applied to solve specific engineering problems by analyzing process data to detect and diagnose anomalies in manufacturing systems [40], [42].

*The proposed methodology provided an automated real-time analysis of the MTS model:* Since the current literature lacks real industry testing cases [65], [66], the validation of the HCPS design methodology was performed in an industrial environment. The validation involved the real-time application of the proposed methodology in a process industry company from the vinyl flooring sector. During the validation, the MTS model was deployed as the best-performing model via the integration of hardware and software elements, namely, production machines, industrial computers, PLC devices, and data analytics software, via the OPC UA server as the IIoT standard. Notably, the MTS automated data analysis model correctly detected anomalies in poor-quality products that occurred during the manufacturing process in real-time. Additionally, the obtained results of the poor-quality product based on the MTS model were automatically sent back to the PLC device as an activation element, triggering the resetting of all influential process parameters to their defined mean values.

*The methodology was developed based on researchers' practical experience in smart manufacturing implementation:* The significance of collaboration between industry and academia for developing a real-time anomaly detection smart manufacturing system has been emphasized in the article. Therefore, the HCPS design methodology was developed based on real industry needs for smart manufacturing implementation. Notably, the manufacturing system objectives were considered in the development of the HCPS methodology for real-time anomaly detection. In this specific case, the company noted that it had issues maintaining the good product quality without knowing what caused quality issues. Thus, by combining industry know-how with data-driven predictive models developed by researchers, successful industry–researcher synergy was achieved for the development and implementation of automated real-time data analysis based on the proposed methodology. An additional benefit for the industry lies in the targeted benchmarks, where

we managed to successfully detect 50% of the poor-quality products that occurred during the real-time anomaly detection model deployment phase, thereby contributing to cost savings for the company.

Moreover, we designed the HCPS design methodology to ensure that the use of AI was transparent and accountable. We did this by taking a careful and principled approach to involve humans in decision-making within the HCPS, ensuring the ethical integration of AI and fostering fairness. In our HCPS design methodology, we specifically incorporate the human presence approach in each phase of the model, emphasizing transparency and accountability to uphold ethical standards. Hence, concluding remarks suggest that the implementation of the HCPS design methodology should aim to enhance human capabilities and involve human participation in the final decision-making process for anomaly detection.

The main limitation of the article is that the HCPS design methodology was applied in only one case study. However, with the lack of research that presents real case scenarios of Industry 4.0/Industry 5.0 in the relevant literature, we expect that this application case will be highly valuable to system designers.

To enhance the outcomes of the proposed methodology for real-time anomaly detection implemented in an industrial setting, further research needs to be conducted. Thus, the HCPS system design methodology should be tested in additional companies from different industries. Moreover, continuing the research and analyzing the factors that could impact the performance of the proposed methodology in other manufacturing systems and industrial sectors, as well as in small and medium enterprises are necessary. In this way, the generalizability and scalability of the proposed HCPS system design methodology can be confirmed and further optimized.

## REFERENCES

- [1] A. Rikalovic, N. Suzic, B. Bajic, and V. Piuri, "Industry 4.0 implementation challenges and opportunities: A technological perspective," *IEEE Syst. J.*, vol. 16, no. 2, pp. 2797–2810, Jun. 2022.
- [2] B. Bajic, A. Rikalovic, N. Suzic, and V. Piuri, "Industry 4.0 implementation challenges and opportunities: A managerial perspective," *IEEE Syst. J.*, vol. 15, no. 1, pp. 546–559, Mar. 2021, doi: [10.1109/jsyst.2020.3023041](https://doi.org/10.1109/jsyst.2020.3023041).
- [3] C. J. Turner, J. Oyekan, L. Stergioulas, and D. Griffin, "Utilizing Industry 4.0 on the construction site: Challenges and opportunities," *IEEE Trans. Ind. Inform.*, vol. 17, no. 2, pp. 746–756, Feb. 2021, doi: [10.1109/TII.2020.3002197](https://doi.org/10.1109/TII.2020.3002197).
- [4] J. Reyes, J. Mula, and M. Díaz-Madroñero, "Development of a conceptual model for lean supply chain planning in Industry 4.0: Multidimensional analysis for operations management," *Prod. Plan. Control*, vol. 34, no. 12, pp. 1209–1224, 2021.
- [5] V. L. Da Silva, J. L. Kovaleski, R. N. Pagani, J. D. M. Silva, and A. Corsi, "Implementation of Industry 4.0 concept in companies: Empirical evidences," *Int. J. Comput. Integr. Manuf.*, vol. 33, no. 4, pp. 325–342, 2020.
- [6] E. Commission, "Industry 5.0 - towards a sustainable, human-centric and resilient European industry," 2021.
- [7] European Commission, "Industry 5.0: A transformative Vision for Europe," 2022.
- [8] M. P. Pacaux-Lemoine, D. Trentesaux, G. Z. Rey, and P. Millot, "Designing intelligent manufacturing systems through Human–Machine Cooperation principles: A human-centered approach," *Comput. Ind. Eng.*, vol. 111, pp. 581–595, 2017.

- [9] M. Gil, M. Albert, J. Fons, and V. Pelechano, "Engineering human-in-the-loop interactions in cyber-physical systems," *Inf. Softw. Technol.*, vol. 126, Oct. 2020, Art. no.106349.
- [10] J. Jiao, F. Zhou, N. Z. Gebraeel, and V. Duffy, "Towards augmenting cyber-physical-human collaborative cognition for human-automation interaction in complex manufacturing and operational environments," *Int. J. Prod. Res.*, vol. 58, no. 16, pp. 5089–5111, 2020.
- [11] K. A. Demir, G. Döven, and B. Sezen, "Industry 5.0 and human-robot co-working," *Procedia Comput. Sci.*, vol. 158, pp. 688–695, 2019.
- [12] W. P. Neumann, S. Winkelhaus, E. H. Grosse, and C. H. Glock, "Industry 4.0 and the human factor – A systems framework and analysis methodology for successful development," *Int. J. Prod. Econ.*, vol. 233, 2021, Art. no.107992.
- [13] Y. Lu et al., "Outlook on human-centric manufacturing towards industry 5.0," *J. Manuf. Syst.*, vol. 62, pp. 612–627, 2022.
- [14] B. Wang, P. Zheng, Y. Yin, A. Shih, and L. Wang, "Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective," *J. Manuf. Syst.*, vol. 63, pp. 471–490, 2022.
- [15] D. Romero, P. Bernus, O. Noran, J. Stahre, and Å. F. Berglund, "The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems," in *Proc. IFIP Adv. Inf. Commun. Technol.*, 2016, pp. 677–686.
- [16] E. Coronado, T. Kiyokawa, G. A. Garcia, I. G. Ramirez-alpizar, G. Venture, and N. Yamanobe, "Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an Industry 5.0," *J. Manuf. Syst.*, vol. 63, pp. 392–410, 2022.
- [17] M.-P. Pacaux-Lemoine and F. Flemisch, "Human-industrial cyber-physical system integration: Design and Evaluation methods," in *Digitalization and Control of Industrial Cyber-Physical Systems*, Hoboken, NJ, USA: Wiley, 2022.
- [18] M. Doyle-Kent and P. Kopacek, "Adoption of Collaborative robotics in Industry 5.0. An Irish industry case study," *IFAC PapersOnLine*, vol. 54, no. 13, pp. 413–418, 2021.
- [19] E. Laudante, A. Greco, M. Caterino, and M. Fera, "Human – Robot interaction for improving fuselage assembly tasks: A case study," *Appl. Sci.*, vol. 10, no. 5757, pp. 1–17, 2020.
- [20] P. Pathak, P. R. Pal, M. Shrivastava, and P. Ora, "Fifth revolution: Applied AI & Human intelligence with cyber physical systems," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 23–27, 2019.
- [21] J. Zhou, Y. Zhou, B. Wang, and J. Zang, "Human-Cyber-Physical systems (HCPSs) in the context of new-generation intelligent manufacturing," *Engineering*, vol. 5, no. 4, pp. 624–636, 2019.
- [22] P. Fantini, P. Leitao, J. Barbosa, and M. Taisch, "Symbiotic integration of human activities in cyber-physical systems," in *Proc. IFAC-PapersOnLine*, 2019, pp. 133–138.
- [23] P. K. R. Maddikunta et al., "Industry 5.0: A survey on enabling technologies and potential applications," *J. Ind. Inf. Integr.*, vol. no. 26, 2021, Art. no.100257.
- [24] S. Nahavandi, "Industry 5.0 - A human-centric solution," *Sustainability*, vol. 11, no. 4371, pp. 1–13, 2019.
- [25] K. Jabrane and M. Bousmah, "A new approach for training cobots from small amount of data in industry 5.0," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 634–646, 2021.
- [26] S. De Blasi, M. Bahrami, E. Engels, and A. Gepperth, "Safe contextual bayesian optimization integrated in industrial control for self-learning machines," *J. Intell. Manuf.*, vol. 35, pp. 885–903, 2023.
- [27] C. Li, Y. Chen, and Y. Shang, "A review of industrial Big Data for decision making in intelligent manufacturing," *Eng. Sci. Technol., Int. J.*, vol. 29, 2022, Art. no.101021.
- [28] M. Čosović and S. Obradović, "BGP anomaly detection with balanced datasets," *Tehnicki Vjesnik*, vol. 25, no. 3, pp. 766–775, 2018.
- [29] S. G. Ponnambalam and N. Subramanian, "Implementation of lean manufacturing and lean audit system in an auto parts manufacturing industry – an industrial case study," *Int. J. Comput. Integr. Manuf.*, vol. 31, no. 2, pp. 1–16, 2017.
- [30] W. D. Leong et al., "Adaptive analytical approach to lean and green operations," *J. Cleaner Prod.*, vol. 235, pp. 190–209, 2019.
- [31] Z. Kang, C. Catal, and B. Tekinerdogan, "Machine learning applications in production lines: A systematic literature review," *Comput. Ind. Eng.*, vol. 149, 2020, Art. no.106773.
- [32] J. Klikowski and M. Woźniak, "Employing one-class SVM classifier ensemble for imbalanced data stream classification," in *Proc. Int. Conf. Comput. Sci.*, 2020, pp. 117–127.
- [33] J. Lee, Y. C. Lee, and J. T. Kim, "Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced database," *J. Manuf. Syst.*, vol. 57, pp. 357–366, Oct. 2020.
- [34] W. Zhang, X. Li, X. D. Jia, H. Ma, Z. Luo, and X. Li, "Machinery fault diagnosis with imbalanced data using deep generative adversarial networks," *Meas. (Lond)*, vol. 152, Feb. 2020, Art. no.107377.
- [35] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Learning from class-imbalanced data with a model-agnostic framework for machine intelligent diagnosis," *Rel. Eng. Syst. Saf.*, vol. 216, Dec. 2021, Art. no.107934.
- [36] M. El-Banna, "Modified Mahalanobis-Taguchi System for Imbalance Data Classification," *Comput. Intell. Neurosci.*, vol. 2017, 2017, Art. no.5874896.
- [37] C. S. Cheng, P. W. Chen, and Y. T. Wu, "Phase I analysis of nonlinear profiles using anomaly detection techniques," *Appl. Sci. (Switzerland)*, vol. 13, no. 4, Feb. 2023, Art. no.2147.
- [38] S. W. Yahaya, A. Lotfi, and M. Mahmud, "Detecting anomaly and its sources in activities of daily living," *SN Comput. Sci.*, vol. 2, no. 1, Feb. 2021.
- [39] Z. Sheng and L. Cheng, "A modified Mahalanobis distance metric used to optimize mahalanobis space and improve classification performance of MTS," *Math Probl. Eng.*, vol. 2022, 2022, Art. no.9972666.
- [40] B. Bajic, N. Suzic, N. Simeunovic, S. Moraca, and A. Rikalovic, "Real-time data analytics edge computing application for Industry 4.0: The Mahalanobis-Taguchi approach," *Int. J. Ind. Eng. Manage.*, vol. 11, no. 3, pp. 146–156, 2020.
- [41] K. Arai and T. Ueda, "Data analysis for stabilizing product quality and the Mahalanobis Taguchi (MT) method," 2016.
- [42] K. Zope, K. Singh, S. H. Nistala, A. Basak, P. Rathore, and V. Runkana, "Anomaly detection and diagnosis in manufacturing systems: A comparative study of statistical, machine learning and deep learning techniques," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, 2019, pp. 1–10.
- [43] C. Li, D. Cabrera, F. Sancho, M. Cerrada, R. Sánchez, and E. Estupinan, "From fault detection to one-class severity discrimination of 3D printers with one-class support vector machine," *ISA Trans.*, vol. 110, pp. 357–367, 2020.
- [44] V. Selvaraj and S. Min, "Real time fault identification system for a retrofitted ultra precision CNC machine from equipment's power consumption data: A case study of an implementation," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 10, pp. 925–941, 2023.
- [45] T. Mendes, P. J. S. Cardoso, and J. Monteiro, "Anomaly detection of consumption in hotel units: A case study comparing isolation forest and variational autoencoder algorithms," *Appl. Sci. (Switzerland)*, vol. 13, no. 1, pp. 1–25, 2023.
- [46] K. Chang, Y. Yoo, and J.-G. Baek, "Anomaly detection using signal segmentation and segmentation and one-class classification in diffusion process of semiconductor manufacturing," *Sensors*, vol. 21, 2021, Art. no.3880.
- [47] S. Wada, Y. Ahmad, and L. Mufti, "Detecting Anomaly and its sources in activities of daily living," *SN Comput. Sci.*, vol. 2, no. 1, pp. 1–18, 2021.
- [48] B. Bajic, S. Moraca, and A. Rikalovic, "Fuzzy maturity model for smart manufacturing readiness: Industry 5.0 perspective," in *Proc. IEEE Zooming Innov. Consum. Technol. Int. Conf.*, 2023, pp. 142–147.
- [49] M. Hammad et al., "Security framework for network-based manufacturing systems with personalized customization: An Industry 4.0 approach," *Sensors*, vol. 23, no. 17, Sep. 2023, Art. no.7555.
- [50] V. Terziyan, D. Malyk, M. Golovianko, and V. Branytskyi, "Encryption and generation of images for privacy-preserving machine learning in smart manufacturing," *Procedia Comput. Sci.*, vol. 217, pp. 91–101, 2023.
- [51] T. Watanabe, I. Kono, and H. Onozuka, "Anomaly detection methods in turning based on motor data analysis," *Procedia Manuf.*, vol. 48, pp. 882–893, 2020.
- [52] J. Wang, S. Gao, Z. Tang, D. Tan, B. Cao, and J. Fan, "A context-aware recommendation system for improving manufacturing process modeling," *J. Intell. Manuf.*, vol. 34, no. 3, pp. 1347–1368, 2023.
- [53] G. Hackeling, *Mastering Machine Learning With Scikit-Learn*, 1st ed. Birmingham, AL, USA: Packt Publ., 2014.

- [54] A. Abdullahi, N. A. Samsudin, M. R. Ibrahim, M. S. Aripin, S. K. A. Khalid, and Z. A. Othman, "Towards IR4.0 implementation in e-manufacturing: Artificial intelligence application in steel plate fault detection," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 20, no. 1, pp. 430–436, 2020.
- [55] B. Bajic, N. Suzic, S. Moraca, M. Stefanović, M. Jovicic, and A. Rikalovic, "Edge computing Data Optimization for Smart Quality Management: Industry 5.0 perspective," *Sustainability (Switzerland)*, vol. 15, no. 7, Apr. 2023, Art. no.6032.
- [56] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," *Int. J. Intell. Technol. Appl. Statist.*, vol. 11, no. 2, pp. 105–111, 2018.
- [57] O. Abdelrahman and P. Keikhosrokiani, "Assembly line anomaly detection and root cause analysis using machine learning," *IEEE Access*, vol. 8, pp. 189661–189672, 2020.
- [58] D. Adebajo, T. Laosirihongthong, P. Samaranayake, and P. Teh, "Key enablers of Industry 4.0 development at firm level: Findings from an emerging economy," *IEEE Trans Eng Manag*, vol. 70, no. 2, pp. 400–416, 2021, doi: [10.1109/TEM.2020.3046764](https://doi.org/10.1109/TEM.2020.3046764).
- [59] J. Maritz, "Implementation considerations for Big Data analytics (BDA): A benefit dependency network approach," in *Proc. Int. Federation Inf. Process. IFIP*, 2020, pp. 481–492.
- [60] P. Tabesh, E. Mousavidin, and S. Hasani, "Implementing Big Data strategies: A managerial perspective," *Bus Horiz*, vol. 62, no. 3, pp. 347–358, 2019.
- [61] J. Abasova, P. Tanuska, and S. Rydzi, "Big data—knowledge discovery in production industry data storages - implementation of best practices," *Appl. Sci. (Switzerland)*, vol. 11, no. 16, 2021, Art. no.7648.
- [62] Z. P. Chang, Y. W. Li, and N. Fatima, "A theoretical survey on Mahalanobis-Taguchi system," *Measurement*, vol. 136, pp. 501–510, 2019.
- [63] A. Kumar, R. Shankar, A. Choudhary, and L. S. Thakur, "A Big Data MapReduce framework for fault diagnosis in cloud-based manufacturing," *Int. J. Prod. Res.*, vol. 54, no. 23, pp. 7060–7073, 2016.
- [64] B. Bajic, I. Cosic, M. Lazarevic, N. Sremcevic, and A. Rikalovic, "Machine learning techniques for smart manufacturing: Applications and challenges in Industry 4.0," in *Proc. 9th Int. Sci. Expert Conf. TEAM*, 2018, pp. 29–38.
- [65] J. Zubrzycki and N. Smidova, "Cyber-physical systems technologies as a key factor in the process of Industry 4.0 and smart manufacturing," *Appl. Comput. Sci.*, vol. 17, no. 4, pp. 84–99, 2021.
- [66] M. Park and J. Jeong, "Design and implementation of machine vision-based quality inspection system in mask manufacturing process," *Sustainability*, vol. 14, no. 10, 2022, Art. no.6009.

Open Access funding provided by 'Università degli Studi di Milano' within the CRUI CARE Agreement