



Original research article

Real-time Data Analytics Edge Computing Application for Industry 4.0: The Mahalanobis-Taguchi Approach

B. Bajic^a, N. Suzic^b, N. Simeunovic^a, S. Moraca^a, A. Rikalovic^{a*}^a Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia;^b University of Padova, Department of Management and Engineering, Vicenza, Italy

ABSTRACT

Industry 4.0 and its innovative technologies (e.g., Internet of Things, Cyber-Physical Systems, Cloud Computing, Big Data and Artificial Intelligence) represent great promise. Still, companies experience hardship when transforming from reactive to predictive manufacturing systems. The latter, driven by data science development, use predictive models to detect and solve production and maintenance issues before they happen. To eliminate the need for large and varied datasets for development of predictive models, in the present research we propose development of real-time predictive models based on small dataset without faulty data. This is achieved by using Mahalanobis-Taguchi system for fault detection in lack of fault data samples, and by using Edge Computing environment which provides higher responsiveness, better security and decreased costs. Subsequently, two predictive models are developed, tested and compared for the case company from process industry (i.e. the vinyl-floor industry sector). Finally, recommendations for the industry are provided.

ARTICLE INFO

Article history:

Received May 30, 2020

Revised July 8, 2020

Accepted July 9, 2020

Published online July 31, 2020

Keywords:

Real-time data analytics;
Edge computing;
Mahalanobis-Taguchi system;
Industry 4.0;
Predictive manufacturing

*Corresponding author:

Aleksandar Rikalovic
a.rikalovic@uns.ac.rs

1. Introduction

Advances in Internet of Things (IoT), Cyber-Physical Systems (CPS), Cloud Computing, Big Data, and Artificial Intelligence have significantly impacted manufacturing and are driving the fourth industrial revolution, Industry 4.0 [1]-[5]. With the development of these Industry 4.0 technologies, the volume of raw data obtained during manufacturing processes is constantly increasing (i.e., Big Data) [6], [7], so many researchers agree that data is becoming a key for optimizing manufacturing processes and improving competitiveness [8], [9]. Data acquired

across the product life cycle can be converted into manufacturing intelligence and yield positive impacts on all aspects of manufacturing [10]. Nevertheless, Big data also increases complexity of processing and analysis of large amount of information in manufacturing systems. Consequently, Big data stimulates the development of data analytics [11], [12], which unfortunately often has to cope with insufficient processing power of existing software applications and personal computers [8]. However, when big data is exploited in a right way, the discovery of patterns in raw data enables real-time predictive data analytics [13]. Thus, by using predictive models, real-time decisions can

be made transforming effectively a reactive manufacturing system into a predictive manufacturing system [8].

Predictive manufacturing systems enable proactive behavior; that is, by using predictive models, they permit to anticipate an error within the manufacturing system before it occurs and to immediately take appropriate actions to avoid errors [14], [15]. Developing a predictive model depends largely on the nature of the data collected from manufacturing systems over a period of time, and it is necessary to select a dataset that will not have incomplete, homogeneous, or noise data that disrupts the quality of a defined dataset. Notably, precisely selected heterogeneous dataset composed of fault and non-fault data, with emphasis on fault data particularly, is a challenge for industry and academia [16], [17]. Up to now, different predictive data analytics methods have been proposed in the literature [18]–[20]. The review of the relevant literature showed that the available articles focus on developing predictive models based on large and varied datasets (i.e. consisting of both fault and non-fault data) [19]–[21]. On the one hand, Cloud Computing has an important role in Data Analytics process with the accent on Big Data (i.e. Big Data Analytics) as it offers access-based computing infrastructure oriented to subscription, data, and application services [22], [23]. On the other hand, the Edge Computing, unlike Cloud Computing, represents the decentralized computing service for storage, processing and applications. It takes place on the network edge and acting as a middle layer between end user and cloud data centers. In that way it increases system responsiveness by reducing the distance that data must travel on the network while producing minimal delays [24].

To eliminate the need for large and varied datasets for development of predictive models in the present research we develop real-time predictive models based on small dataset without faulty data. This is achieved by using Mahalanobis–Taguchi system (MTS) [25] and Edge Computing. Firstly, for development of the predictive models we apply the MTS. The MTS involves the development of predictive models based on small carefully selected non-fault data samples collected from the manufacturing process together with the company experts. Notably, these data samples, differently from existing methods [20], do not require the presence of fault data so predictive models can be developed observing the process that works without errors. Secondly, to enable handling of large amount of data with high responsiveness, in secure environment and at a low cost, we

apply Edge Computing solution MELIPC MI5000 developed by Mitsubishi Electric Company.

Further, the review of the relevant literature showed that the majority of the research available in the relevant literature deals with models developed and tested in experimental conditions. The present research looks to fill this gap by testing two developed predictive models in a case company from process industry (i.e. vinyl floor industry sector) with product quality issues. Notably, the case company satisfies highest standards of the World Class Manufacturing and is highly automatized. With these characteristics it presents a great testing ground for the developed predictive models.

The present paper is organized as follows. Section 2 provides an overview of the related work for the real-time data analytics applied in manufacturing systems, the MTS for multivariate data, and Edge Computing for multivariate limited-scale data analysis. Section 3 presents details of the research method, the predictive models development and the testing results of the two developed predictive models. Section 4 discusses the results and compares results of the two developed predictive models. Finally, Section 5 derives some conclusions, summarizes the paper's contributions, and suggests opportunities for the future research.

2. Related work

2.1 Real-time data analytics for manufacturing systems

The increasing availability of manufacturing data is changing the way decisions are made in industry regarding predictive maintenance and quality improvement using data analytics methods [26]. The implementation of advanced Industry 4.0 technologies, namely CPS and IoT combined with data analytics, can enable predictive manufacturing and networked production environments [8].

Data analytics, as part of the data science field, represents a practice that reveals hidden information among data collected from various devices by using advanced analytics techniques, including expert systems, machine learning, and advanced statistical analysis [27], [28]. These analytical techniques further enable real-time decision making in manufacturing systems that use real-time data analytics [12].

Real-time data analytics, as a part of data analytics, refers to analytical techniques where data is processed and analyzed as it is generated, in real time

[29] or near to real time [30]. Currently, there are few real-time data analytics applications for manufacturing systems [30]. The existing architectures have been mainly developed for offline data analysis and are not suitable for real-time data processing and analysis [31].

Among the available applications of real-time data analytics, Zang et al. [32] proposed a real-time production performance analysis and exception diagnosis integration model for manufacturing systems. Their model, consisting essentially of a hierarchical timed colored Petri net model and a decision tree model, achieves real-time production performance and exception information for dynamic and stochastic manufacturing processes. Similarly, Oh et al. [33] focused their research on real-time monitoring of quality and on controlling the system with an integrated cost-effective support vector machine (CESVM) algorithm. In their study, an integrated CESVM was developed and installed for the door trim manufacturing process using a kiosk so real-time quality inspection data could be collected, analyzed, and predicted. Finally, Qian et al. [34] proposed the real-time data-driven framework, Intelligent Collaborative Mechanism (ICM), in order to achieve collaborative and effective interactions among assembly stations and operators. The real-time data-driven ICM for a fixed-position assembly system is based on three models, namely the Petri net model of the assembly workflow, the constraint matrix of tasks, and the partitioned structure of the task pools. Using these models, the ICM was able to monitor the assembly progress and select proper tasks for the further matching process of dynamic scheduling in manufacturing.

2.2 Mahalanobis–Taguchi system for multivariate data

A high quality dataset is crucial for developing an effective predictive model. Specifically, for a dataset to be considered high quality, it should consist of samples from periods when the manufacturing system worked without problems, interruptions, and difficulties (non-fault data samples), as well as data from when problems occurred in the system (fault data samples). However, it is very difficult to find a balance between the number of non-fault data samples and the number of fault data samples [20]. Moreover, for the non-fault data that comprises the majority of the data collected, it is not possible to use the most common data analytics methods, which require a balanced use of faulty and non-faulty data (e.g., linear regression, support vector machines, artificial neural

networks etc.) [17]. Additionally, the non-fault data mainly consists of a large number of similar data values that do not provide different types of information about the condition of the manufacturing system. In this case, it is necessary to optimize the collected non-fault dataset in the right way.

The MTS is a relatively new method in the field of diagnosis and prediction using multivariate data [25]. The MTS uses the Mahalanobis distance introduced by Mahalanobis [35] in order to distinguish the pattern of an observed group from other groups [36] based on the correlation between data that is maintained with the Mahalanobis distance and various patterns that can be identified and analyzed in relation to the observed dataset [37].

Taguchi [37] extended the use of the Mahalanobis distance by developing the primary methodology of the MTS [38], [39]. Therefore, the MTS is Taguchi's hybrid systematic diagnosis and forecasting methodology based on the Mahalanobis distance principle for multivariate data [25]. The MTS calculates the distance between observed data and sample data, and quantitatively determines their differences and optimizes the measurement scale of unobserved data [35], [37].

In many scenarios, the MTS is used for monitoring the quality of products and the quality of the manufacturing systems [35]. For example, Su and Hsiao [40] investigated the effect of imbalanced training set using the MTS and other classification techniques, namely stepwise discriminate analysis, decision tree analysis, back-propagation neural network, and support vector machines. The results showed that the MTS has the best classification ability and is the most robust classification technique. Additionally, Huang, Hsu, and Liu [41] integrated the MTS and the artificial neural network algorithm to create a novel algorithm that solves pattern-recognition problems and can be applied to construct a model for manufacturing improvements regarding inspections in dynamic environments. Jobi-Taiwo and Cudney [42] used the MTS for extracting information in a multidimensional system and integrating information from different variables into a single composite metric. The method was evaluated in a case study on a multiple fault class of steel plate manufacturing, which indicates the practicality of the method for improving system quality in industrial applications. Using multi-sensor signals and the MTS, Rizal et al. [43] investigated predictive maintenance for detecting cutting tool wear during manufacturing processes, and the results showed that the medium wear and critical wear stages of the cutting tools can be detected in real time.

Wang et al. [44] focused on developing a method called the adaptive Multiclass Mahalanobis-Taguchi system in conjunction with variational mode decomposition and singular value decomposition, which are employed to detect faults based on variable conditions. Finally, Reséndiz-Flores, Navarro-Acosta, and Hernández-Martínez [45] proposed a novel methodology representing a combination of the MTS and a hybrid binary metaheuristic based on particle swarm optimization and gravitational search algorithm for improving product quality. They did so by performing an optimal feature selection to detect the relevant variables in a real foam injection process in the automotive industry.

2.3 Edge Computing for multivariate limited-scale data analysis in Industry 4.0

As already stated, the volume of acquired data in Industry 4.0 environment is constantly increasing [8]. Due to insufficient processing power [8], big data often cannot be processed and analyzed using the existing software applications and personal computers. Therefore, new technologies (i.e., Cloud, Fog, and Edge Computing) use advanced data analytics techniques to detect hidden information. To overcome the problems of processing and analyzing large amounts of data, various approaches and system architectures have been proposed in the research literature [24], [46]–[48]. However, a large amount of data does not guarantee the good quality datasets required for data analysis. On the contrary, a carefully selected limited-scale dataset that reflects the real state of the manufacturing system can provide a lot of valuable information if the right technologies for the data analysis are selected. The most appropriate technology for limited-scale data analysis is Edge Computing which is mainly used for data storage, processing, and applications that take place on the network's edge [24]. Edge Computing optimizes different technologies, namely Cloud Computing, by performing data analytics as close to the data sources as possible [23]. Thus, we argue that data analytics should begin during manufacturing, using an Edge Computing solution and striving for real time predictive data analytics in order to minimize network traffic [49], increase data security [50], [51], speed up data processing [52] and reduce costs regarding data analysis transfer in comparison to Cloud Computing [53].

Cases of Edge Computing use for data analytics are not numerous in the literature, especially for limited-scale data. For example, Qian et al. [52] proposed an Edge Computing framework for fault

diagnosis and dynamic control of rotating machines and deployed it on a designed edge computing node based on small amounts of data that have machine fault values. Their framework is used for real-time diagnosis of multiple types of electrical and mechanical faults by fusing multiple sensor data. Similarly, Forkan et al. [54] presented an architecture for implementing edge computing in Industrial IoT applications to perform real-time data analysis and data mining of the local database based on limited-scale data collected from a customized candy production line. The experiment showed that the self-organized scheduling mechanism of Edge Computing provides obvious advantages in terms of bandwidth optimization compared to traditional approaches. Finally, Vater, Harscheidt, and Knoll [55] proposed IT architecture based on a combination of Edge and Cloud Computing for prescriptive automation that enables network-based interoperable process control. Their architecture offers the possibility of comprehensive data processing for continuous increases in manufacturing process productivity where real-time data analysis is not possible.

3. Development and testing of the predictive models

3.1 Research method

In the present research we opted for application of MTS for development of the predictive models. This is because the MTS has proved itself as an effective approach for fault detection, diagnosis, and data classification in the context of improving product quality and the quality of manufacturing processes when applied to multivariate datasets in manufacturing environments [56]. These characteristics have led to great interest from both industry and researchers for MTS.

The present research applies a research method based on the data mining approach [57]–[59] - Figure 1. The research method, used for development of the predictive models, contains six phases which result in application the MTS for real-time data analytics in Industry 4.0. We develop two predictive models (Figure 1), namely 1) a random parameter model (RPM); and 2) a parameter configuration model (PCM) (Figure 1). The RPM represents a predictive model developed from randomly collected samples with random process parameters, while the PCM represents a predictive model developed for a group of products with a defined parameter configuration.

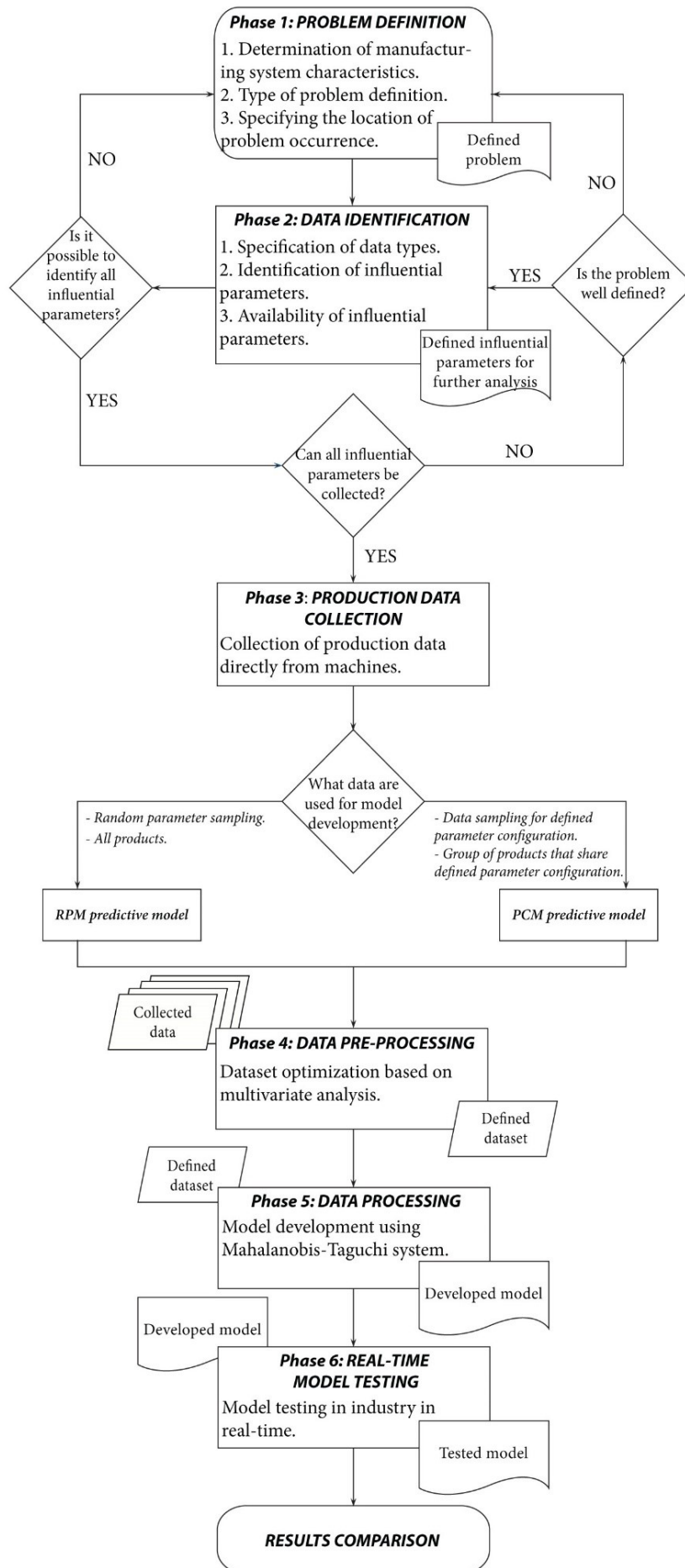


Figure 1. Research method (based on data mining approach [57]–[59])

3.2 Development and testing of the predictive models: the performed steps

We apply the proposed six-phase method in a process industry conditions in a company from vinyl floor sector. First, we developed the RPM and the PCM models. Subsequently we tested them. Finally, we assessed and compared the performances of these two models. Hereafter we describe the six phases of the method and exemplify them, when needed, with what we performed in the case company.

Phase 1: Problem definition – This phase has three parts (Figure 1): determining the manufacturing system characteristics; defining the type of problem in the manufacturing system; and specifying the location of the problem.

Phase 2: Data identification – Based on the defined problem (Phase 1), in this phase, the data that affect the onset of the problem in the manufacturing system are identified and collected (Figure 1). Data identification is performed in three steps: specification of data types, identification of influential parameters, and checking the availability of influential parameters. Influential parameters represent all process parameters that have a significant impact on the occurrence of product quality issues. As a result the influential parameters are defined.

Phase 3: Production data collection – In this phase, two different datasets are collected for the two developed predictive models (Figure 1). The first dataset, used for the RPM predictive model, represents the data collected continuously for 24 h for random parameter samples. The second dataset, used for the PCM predictive model, represents data collected intermittently for 12 days in which the data was collected for a group of nine products that share the same parameter configuration for the observed part of the production line (i.e., Coating 1, Coating 2, and Printing).

Phase 4: Data pre-processing – The dataset optimization based on multivariate analysis (MVA) was done in this phase (Figure 1). The MVA approach is defined based on the range option, where the defined number of collected rows of data, called samples (exactly 120 samples), is grouped into single data for each set of collected data. According to this methodology, the final number of samples used for further analysis for the first dataset is 710, and for the second dataset is 815 samples. The number of influential parameters differs in the datasets because of constant values for certain parameters, so the number of influential parameters for the first dataset is 63, while the number of influential parameters for the

second dataset is 62.

Phase 5: Data processing – In this phase, the pre-processed data is processed using the MTS for each set of collected data separately (Figure 1). As a result, two different predictive models are developed for early detection of quality issues based on principles of mutual distance in the reference data, that is, the Mahalanobis distance. According to Peng et al. [60], the Mahalanobis distance in data values that are characterized as non-fault is less than 2.5, but in certain cases, values over 4 are defined as non-fault [60].

Phase 6: Real-time model testing – In this phase, two developed predictive models are tested in real time in manufacturing environment conditions (Figure 1). The two models are tested in a process industry company from the vinyl floor sector for a defined period of 15 days. Real-time model testing was performed exclusively for the products with defined parameter configuration. As a result of this phase, the performances of the two developed predictive models are summarized and compared.

3.3 Testing of the developed predictive models: context information and results

The developed predictive models were tested in a vinyl floors process industry company. The observed production line consists of 12 machines, and the main problems are product quality issues. The quality issues include wrinkles on the final product, low paint finish quality, stains on the final product, and so on. The company is not currently able to detect root causes of these quality issues.

In order to address the need to prevent these quality issues, an MELIPC MI5000 edge computing solution developed by Mitsubishi Electric Company was implemented in the company. According to the manufacturer, the MELIPC MI5000 can run two operating systems at the same time, namely VxWorks and Windows. VxWorks is used for device control and data collection, while Windows displays analysis results for collected data, allowing superior processing. The pre-installed software enables easier collection of production data. Installing additional software allows easier collection of third-party company production data, allowing off-line model development, real-time data processing, and monitoring of production processes on the shop floor using functions provided at the edge computing level. Finally, the most important characteristic of this Edge Computing solution is its ability to support machine self-configuration by providing corresponding feedback.

The present research is defined as a pilot study

for early detection of product quality issues in a vinyl floors production company. The research has been limited to the first three machines in the production line (i.e. Coating 1, Coating 2, and Printing) due to the limited resources of Edge Computing for real-time model testing and the great complexity of the production line in the manufacturing system. Accordingly, production data was generated and collected on these three machines. Thus, the occurrence of quality issues is limited to the three machines observed.

The two predictive models (RPM and PCM) were developed and tested using the MTS approach according to the six-phase method presented in Figure 1. A comparison of the performances of the two developed predictive models is presented in Table 1, which provides the main results for each phase of the proposed and applied method.

4. Discussion

At the beginning of the research, the review of the relevant literature showed that the available articles focus on developing predictive models based on large and varied datasets (i.e. consisting of both fault and non-fault data) [19]–[21]. However, the needed size of the dataset and the lack of faulty data can both represent a problem for development of predictive models and implementation of predictive manufacturing systems in industry. Thus, in order to respond to this gap, in the present research we focused on development of predictive models based on small dataset for fault detection with lack of fault data samples. Specifically, we proposed development of the predictive models based on MTS (Figure 1) that enables working on non-faulty data. Further, we used Edge Computing (designed to process small datasets) to increase MTS responsiveness, provide security and decrease costs [24]. Subsequently, we developed and operationalized two predictive models using small sets of carefully selected data.

The main difference between the two developed predictive models is the methodology for data sampling that is used for creating the optimal dataset to develop a predictive model. The data sampling methodology is adapted for process industry conditions for real-time fault detection. The dataset for developing the RPM predictive model consists of random parameter configurations sampled over 24 h, while the dataset for developing the PCM model consists of data for a group of products with defined parameter configuration collected intermittently for 12 days.

In the data processing phase (Phase 5, Figure 1), both models demonstrated high accuracy. The PCM

model, with an accuracy of 98.04%, slightly outperformed the RPM model's 97.89% (Table 1). Furthermore, the final testing of the predictive models (i.e., Phase 6. Real-time model testing, Figure 1) was performed in an industrial environment in real time using the Edge Computing solution MELIPC MI5000. The system was designed to give alarms when there are changes in the process parameters that signal the occurrence of a product quality issue. During the real-time testing of both models, only the products pertaining to the defined parameter configuration were used because the PCM model was developed exclusively for defined parameter configuration, while the goal with the RPM model was to understand whether it could be universally applicable to parameter configurations in the production system.

The PCM model outperformed the RPM model regarding the total accuracy of the model by almost 10% during real-time model testing (Table 1). However, the RPM model correctly predicted 5 faults out of 7 (real number of faults provided by company experts) regarding product quality issues, giving better recall performances compared to the PCM model, which correctly predicted 4 faults. Comparing the number of false alarms, even though the number of false alarms was very high in both cases, the RPM model produced almost twice as many alarms as the PCM model, which again gives the advantage to the PCM model.

Deepening the issue of false alarms, while discussing the results with the company's experts, it was concluded that besides faults regarding product quality issues, there were also so-called "system faults" present in the system. These system faults appear, for example, at the start of production of a new batch while the machine calibration is done. Both models predicted not only product quality issues but also system faults, which were originally classified as false alarms. These system faults were left to be remedied in the future and are beyond the scope of the present research.

5. Conclusions

The case company in which the two developed predictive models were tested is from the vinyl floor process industry. For the case company's process, we developed two predictive models based on different sampling for dataset creation, namely 1) random parameter sampling and 2) parameter configuration sampling. Subsequently, the two models were applied in industry conditions, and the performances of these two models were assessed and compared. Both

Table 1. Comparison of the characteristics and performances achieved by the two developed predictive models

Phase No.	Phase name	RPM model	PCM model
Problem definition			
1	1. Determination of manufacturing system characteristics	# Process industry # Vinyl flooring production # Over 40 variations of flooring # Constant reduction in batch sizes and increase of product variety	
	2. Type of problem definition	# Product quality classification problem	
	3. Specifying the location of problem occurrence	# The beginning of the production line consisting of three machines: Coating 1, Coating 2, and Printing	
Data identification			
2	1. Specification of data type	# Real time production data generated from three machines (Coating 1, Coating 2, and Printing) in the manufacturing system	
	2. Identification of influential parameters	# Influential parameters are identified based on experts' knowledge # The number of identified parameters is 65 in total	
	3. Availability of influential parameters	# All influential parameters are generated by sensors implemented on the production line confirming their availability	
3	Production data collection	# Real time production data acquisition and logging every 2 minutes (after 2 minutes, one logging file is generated consisting of 120 samples) • Dataset corresponds to the data collected continuously for 24 h and relates to all products that were produced at a given time in the manufacturing system. During that time, the number of relevant logging files for 65 process parameters is 710 including only non-fault values	• Dataset corresponds to the parameter configuration. Dataset is collected intermittently for 12 days and consists of 815 logging files consisting of the non-fault values
4	Data pre-processing	• Pre-processed dataset consists of: - 710 samples that correspond to the non-fault values - 63 influential parameters	• Pre-processed dataset consists of: - 815 samples that correspond to the non-fault values - 62 influential parameters
5	Data processing	# MTS is used for model development # The Mahalanobis distance is set to a value of 4 • From the total number of 710 non-fault samples: - 695 are true positive - 15 are false negative (error type II) • Total accuracy of the model: 97.89%	• From the total number of 815 non-fault samples: - 799 are true positive - 16 are false negative (error type II) • Total accuracy of the model: 98.04%
6	Real-time model testing	# The number of logging files generated is 2,253 (each logging file contains 120 samples), where 7 faults were detected in product quality for the specified time period (information provided by experts). Thus, 2,246 of the generated logging files are classified as non-fault product quality data. • Total number of 2,253 samples: - 1,741 are true positive - 505 are false negative (error type II) - 5 are true negative - 2 are false positive (error type I) - total accuracy of the model: 77.50% - precision of the model: 0.98% - recall of the model: 71.40%	• Total number of 2,253 samples: - 1,958 are true positive - 288 are false negative (error type II) - 4 are true negative - 3 are false positive (error type I) - total accuracy of the model: 87.10% - precision of the model: 1.40% - recall of the model: 57.10%
Legend:		# Characteristic/performance are equal for both predictive models	• Characteristic/performance are different for each predictive model

models demonstrate high accuracy and applicability to a large number of different products benefiting the company that implements them in two ways: 1) the manufacturing errors are anticipated before they occur, and 2) new hidden relationships among data influencing the occurrence of the errors in production are discovered.

Based on the results obtained and discussed, the parameter configuration model (PCM) outperformed the random parameter model (RPM) in the real-time testing phase in an industrial environment using Edge Computing. This is an indication that the PCM model has strong potential for application in the process industry, with high accuracy and usability of the model based on the small dataset. Notably, PCM model was subsequently applied as a pilot project in the case company since it was recognized by the company top management as a viable way of moving towards digital transformation based on Edge Computing, advanced statistics and artificial intelligence techniques. Moreover, the application of the developed PCM model demonstrated possibilities to immediately eliminate great amount of quality issues on the produced products. The developed PCM model in an Edge Computing environment can provide signal information (alarms) if a critical state is nearly reached for faults regarding product quality issues and system faults. This signal can also be sent to PLCs in order to autonomously make changes inside the manufacturing system and achieve machine self-configuration by changing the parameter configuration.

Although the PCM outperformed the RPM in the real-time testing phase, the limitation of this predictive model is that it is not universal. Specifically, the PCM provides strong performances if it is used for preventing quality issues in products that are produced with a defined parameter configuration. Thus, products that do not fall into the exact parameter configuration cannot use this predictive model. For products that have other parameter configurations, a new PCM should be developed.

It can be argued that the RPM could come close to the performance of the PCM in some specific cases. However, we expect that the PCM will significantly outperform the RPM in a great majority of applications in the process industry. Thus, a recommendation for the process manufacturer would be to develop the necessary number of PCMs to cover the entire product assortment. Accordingly, newly developed products should be fitted in one of the existing PCMs or a new PCM should be developed for new products in order to obtain good results in real-time

fault detection.

The next possible step in the research could be to identify all parameter configurations and to develop PCM models for all present parameter configurations in the vinyl floor process industry sector case company. Testing these models on part of the production line and then extending this testing to the rest of the production line machines is a plausible and meaningful continuation of the present research in the future.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] Z. Li, Y. Wang, and K. S. Wang, "Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario," *Adv. Manuf.*, vol. 5, no. 4, pp. 377–387, 2017.
- [2] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *J. Ind. Inf. Integr.*, vol. 6, pp. 1–10, 2017.
- [3] M. Ghobakhloo, "The future of manufacturing industry: a strategic roadmap," *J. Manuf. Technol. Manag.*, vol. 29, no. 6, pp. 910–936, 2018.
- [4] S. Digiesi, F. Facchini, G. Mossa, and G. Mummolo, "Minimizing and balancing ergonomic risk of workers of an assembly line by job rotation: A MINLP Model," *Int. J. Ind. Eng. Manag.*, vol. 9, no. 3, pp. 129–138, 2018.
- [5] M. Crnjac, I. Veža, and N. Banduka, "From concept to the introduction of industry 4.0," *Int. J. Ind. Eng. Manag.*, vol. 8, no. 1, pp. 21–30, 2017.
- [6] R. Holubek and P. Kostal, "The intelligent manufacturing systems," *Adv. Sci. Lett.*, vol. 19, no. 3, pp. 972–975, 2013.
- [7] B. Li, B. Hou, W. Yu, X. Lu, and C. Yang, "Applications of artificial intelligence in intelligent manufacturing: a review," *Front. Inf. Technol. Electron. Eng.*, vol. 18, no. 1, pp. 86–96, 2017.
- [8] B. Nikolic, J. Ignjatic, N. Suzic, B. Stevanov, and A. Rikalovic, "Predictive Manufacturing Systems in Industry 4.0: Trends, Benefits and Challenges," in *28TH DAAAM International Symposium on Intelligent Manufacturing and Automation*, 2017, pp. 769–802.
- [9] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, 2017.
- [10] M. P. Pacaux-Lemoine, D. Trentesaux, G. Zambrano 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.
- [11] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent Manufacturing in the Context of Industry 4.0: A Review," *Engineering*, vol. 3, no. 5, pp. 616–630, 2017.
- [12] S. S. Kamble, A. Gunasekaran, and S. A. Gawankar, "Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives," *Process Saf. Environ. Prot.*, vol. 117, pp. 408–425, 2018.

- [13] B. Wang, "The Future of Manufacturing: A New Perspective," *Engineering*, vol. 4, no. 5, pp. 722-728, 2018.
- [14] N. K. Ming, N. Philip, and S. Sahlan, "Proactive and predictive maintenance strategies and application for instrumentation & control in oil & gas industry," *Int. J. Integr. Eng.*, vol. 11, no. 4, pp. 119-130, 2019.
- [15] M. Kumari and M. S. Kulkarni, "Single-measure and multi-measure approach of predictive manufacturing control: A comparative study," *Comput. Ind. Eng.*, vol. 127, pp. 182-195, 2019.
- [16] S. J. Qin, "Process Data Analytics in the Era of Big Data," *AIChE*, vol. 60, no. 9, pp. 3092-3100, 2014.
- [17] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, 2016.
- [18] J. C. Kabugo, S. L. Jämsä-Joumelä, R. Schiemann, and C. Binder, "Industry 4.0 based process data analytics platform: A waste-to-energy plant case study," *Int. J. Electr. Power Energy Syst.*, vol. 115, pp. 1-18, 2020.
- [19] J. M. D. Delgado, L. Oyedele, M. Bilal, A. Ajayi, L. Akanbi, and O. Akinade, "Big Data Analytics System for Costing Power Transmission Projects," *J. Constr. Eng. Manag.*, vol. 146, no. 1, pp. 1-10, 2020.
- [20] H. N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran, "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies," *Enterp. Inf. Syst.*, vol. In press, pp. 1-25, 2019.
- [21] J. Ruiz-sarmiento, J. Monroy, F. Moreno, and C. Galindo, "A predictive model for the maintenance of industrial machinery in the context of industry 4.0," *Eng. Appl. Artif. Intell.*, vol. 87, pp. 1-15, 2020.
- [22] C. Wu, R. Buyya, and K. Ramamohanarao, "Big Data Analytics = Machine Learning + Cloud Computing," *Big Data Princ. Paradig.*, pp. 3-38, 2016.
- [23] G. P. Bhandari and R. Gupta, "An Overview of Edge/Cloud Computing Architecture with its Issues and Challenges," in *Advancing Consumer-Centric Fog Computing Architectures*, 2018, pp. 1-37.
- [24] B. Bajic, I. Cosic, B. Katalinic, S. Moraca, M. Lazarevic, and A. Rikalovic, "Edge Computing Vs. Cloud Computing: Challenges And Opportunities In Industry 4.0," in *30th Daaam International Symposium on Intelligent Manufacturing and Automation*, 2019, pp. 864-871.
- [25] E. A. Cudney, D. Drain, K. Paryani, and N. Sharma, "A Comparison of the Mahalanobis-Taguchi System to A Standard Statistical Method for Defect Detection," *J. Ind. Syst. Eng.*, vol. 2, no. 4, pp. 250-258, 2009.
- [26] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Trans. Ind. Informatics*, vol. 11, no. 3, pp. 812-820, 2015.
- [27] Y. Chen, H. Chen, A. Gorkhali, Y. Lu, Y. Ma, and L. Li, "Big data analytics and big data science: a survey," *J. Manag. Anal.*, vol. 3, no. 1, pp. 1-42, 2016.
- [28] S. Tiwari, H. M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries," *Comput. Ind. Eng.*, vol. 115, pp. 319-330, 2018.
- [29] Z. Milosevic, W. Chen, A. Berry, and F. A. Rabhi, "Real-Time Analytics," in *Big Data: Principles and Paradigms*, 2016, pp. 39-61.
- [30] Y. Cui, S. Kara, and K. C. Chan, "Manufacturing big data ecosystem: A systematic literature review," *Robot. Comput. Integr. Manuf.*, vol. 62, no. In press, pp. 1-20, 2020.
- [31] S. Wang, J. Wan, D. Li, and C. Liu, "Knowledge reasoning with semantic data for real-time data processing in smart factory," *Sensors (Switzerland)*, vol. 18, no. 2, pp. 1-10, 2018.
- [32] Y. Zhang, W. Wang, N. Wu, and C. Qian, "IoT-Enabled Real-Time Production Performance Analysis and Exception Diagnosis Model," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1318-1332, 2016.
- [33] Y. G. Oh, M. Busogi, K. Ransikarbum, D. Shin, D. Kwon, and N. Kim, "Real-time quality monitoring and control system using an integrated cost effective support vector machine," *J. Mech. Sci. Technol.*, vol. 33, no. 12, pp. 6009-6020, 2019.
- [34] C. Qian, Y. Zhang, C. Jiang, S. Pan, and Y. Rong, "A real-time data-driven collaborative mechanism in fixed-position assembly systems for smart manufacturing," *Robot. Comput. Integr. Manuf.*, vol. 61, no. In press, pp. 1-13, 2020.
- [35] K. Arai and T. Ueda, "Data Analysis for Stabilizing Product Quality and the Mahalanobis Taguchi (MT) Method," 2016.
- [36] G. Taguchi and R. Jugulum, *The Mahalanobis-Taguchi strategy: A Pattern technology System*. John Wiley & Sons, 2002.
- [37] G. Taguchi, S. Chowdhury, and Y. Wu, *Taguchi's Quality Engineering Handbook*. 2007.
- [38] E. Ghasemi, A. Aghaie, and E. A. Cudney, "Mahalanobis Taguchi system: A review," *Int. J. Qual. Reliab. Manag.*, vol. 32, no. 3, pp. 291-307, 2015.
- [39] W. H. Woodall, R. Koudelik, K. L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, "A review and analysis of the Mahalanobis-Taguchi system," *Technometrics*, vol. 45, no. 1, pp. 1-15, 2003.
- [40] C. T. Su and Y. H. Hsiao, "An evaluation of the robustness of MTS for imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 10, pp. 1321-1332, 2007.
- [41] C. L. Huang, T. S. Hsu, and C. M. Liu, "The Mahalanobis-Taguchi system - Neural network algorithm for data-mining in dynamic environments," *Expert Syst. Appl.*, vol. 36, no. 3, part 1, pp. 5475-5480, 2009.
- [42] A. A. Jobi-Taiwo and E. A. Cudney, "Mahalanobis-Taguchi system for multiclass classification of steel plates fault," *Int. J. Qual. Eng. Technol.*, vol. 5, no. 1, pp. 25-39, 2015.
- [43] M. Rizal, J. A. Ghani, M. Z. Nuawi, and C. H. C. Haron, "Cutting tool wear classification and detection using multi-sensor signals and Mahalanobis-Taguchi System," *Wear*, vol. 376-377, pp. 1759-1765, 2017.
- [44] N. Wang, Z. Wang, L. Jia, Y. Qin, X. Chen, and Y. Zuo, "Adaptive multiclass Mahalanobis Taguchi system for bearing fault diagnosis under variable conditions," *Sensors (Switzerland)*, vol. 19, no. 1, pp. 1-16, 2019.
- [45] E. O. Reséndiz-Flores, J. A. Navarro-Acosta, and A. Hernández-Martínez, "Optimal feature selection in industrial foam injection processes using hybrid binary Particle Swarm Optimization and Gravitational Search Algorithm in the Mahalanobis-Taguchi System," *Soft Comput.*, vol. 24, pp. 341-349, 2020.
- [46] Y. Zuo, F. Tao, and A. Y. C. Nee, "An Internet of things and cloud-based approach for energy consumption evaluation and analysis for a product," *Int. J. Comput. Integr. Manuf.*, vol. 31, no. 4-5, pp. 337-348, 2018.
- [47] D. Wu, S. Liu, L. Zhang, J. Terpenney, R. X. Gao, T. Kurfess, and J. A. Guzzo, "A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing," *J. Manuf. Syst.*, vol. 43, pp. 25-34, 2017.
- [48] J. Um, V. Gezer, A. Wagner, and M. Ruskowski, "Edge Computing in Smart Production," in *Advances in Service and Industrial Robotics*, Springer International Publishing, 2020, pp. 144-152.
- [49] K. Bilal, O. Khalid, A. Erbad, and S. U. Khan, "Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers," *Comput. Networks*, vol. 130, pp. 94-120, 2018.

- [50] S. Garg, A. Singh, K. Kaur, G. S. Auja, S. Batra, N. Kumar, and M. S. Obaidat, "Edge Computing-Based Security Framework for Big Data Analytics in VANETs," *IEEE Netw.*, vol. 33, no. 2, pp. 72-81, 2019.
- [51] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Futur. Gener. Comput. Syst.*, vol. 97, pp. 219-235, 2019.
- [52] G. Qian, S. Lu, D. Pan, H. Tang, Y. Liu, and Q. Wang, "Edge Computing: A Promising Framework for Real-Time Fault Diagnosis and Dynamic Control of Rotating Machines Using Multi-Sensor Data," *IEEE Sens. J.*, vol. 19, no. 11, pp. 4211-4220, 2019.
- [53] T. Francis, "A comparison of cloud execution mechanisms fog, edge, and clone cloud computing," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, pp. 4646-4653, 2018.
- [54] A. R. M. Forkan, F. Montori, D. Georgakopoulos, P. P. Jayaraman, A. Yavari, and A. Morshed, "An industrial IoT solution for evaluating workers' performance via activity recognition," in *39th International Conference on Distributed Computing Systems (ICDCS)*, 2019, pp. 1393-1403.
- [55] J. Vater, L. Harscheidt, and A. Knoll, "A reference architecture based on edge and cloud computing for smart manufacturing," in *Proceedings - International Conference on Computer Communications and Networks, ICCCN*, 2019, pp. 1-7.
- [56] E. O. Reséndiz-Flores, J. A. Navarro-Acosta, C. G. Mota-Gutiérrez, and Y. I. Reyes-Carlos, "Fault detection and optimal feature selection in automobile motor-head machining process," *Int. J. Adv. Manuf. Technol.*, vol. 94, pp. 2613-2622, 2018.
- [57] D. Abbott, *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*, 1st ed. Wiley, 2014.
- [58] O. Doğan, "A Roadmap for Data Driven Decision Making to Improve Quality A Roadmap for Data Driven Decision Making to Improve Quality," no. May, pp. 0-8, 2017.
- [59] H. Wiemer, L. Drowatzky, and S. Ihlenfeldt, "Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model," *Appl. Sci.*, vol. 79, pp. 403-408, 2019.
- [60] C. F. Peng, L. H. Ho, S. B. Tsai, Y. C. Hsiao, Y. Zhai, Q. Chen, L. C. Chang, and Shang, Z., "Applying the Mahalanobis-Taguchi System to improve tablet PC production processes," *Sustain.*, vol. 9, no. 9, pp. 1-17, 2017.