Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 253 (2025) 942–951

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Industry 4.0 and Smart Manufacturing

Anomaly detection using different types of machine learning models in the context of "smart maintenance technologies in the manufacturing industry"

San Giliyana[a,b,*]

[a]Mälardalen University, Universitetsplan 1, Västerås, Sweden
[b]Mälardalen Industrial Technology Center, John Engellaus Gata 1, Eskilstuna 633 61, Sweden

Abstract

Industry 4.0 presents nine technologies, including the Industrial Internet of Things (IIoT), Big Data and Analytics, Cloud Computing, etc. The progress of Industry 4.0 places a new demand for maintenance. Some of the nine technologies of Industry 4.0, such as IIoT, Big Data and Analytics, Cloud Computing and Augmented Reality (AR), as well as Cyber-Physical System (CPS) and machine learning, play an important in the development of smart maintenance technologies. In smart maintenance research, it is presented how IIoT can be used for machine connection and maintenance data collection, machine learning models for maintenance data analysis and failure prediction, AR for maintenance instructions, etc. Although previous smart maintenance research presents many technologies for smart maintenance, the manufacturing companies still face many implementation challenges when implementing and using to add benefits to maintenance organizations in line with companies main goals. Many manufacturing companies still utilize reactive maintenance and are experiencing too much downtime. In this paper, I have shown how two machine learning models, Isolation Forest and Regression Learner, and the statistical technique, Interquartile range (IQR), can be applied in order to detect anomalies in an unsupervised dataset, consisting of travel time for a linear guide, and temperature, in a drill station, which is part of a Cyber-Physical production system, located at a smart production laboratory in Sweden.

© 2025 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 6th International Conference on Industry 4.0 and Smart Manufacturing

Keywords: Smart maintenance technologies; Industry 4.0 technologies; Machine learning; Anomaly detection

* San Giliyana. Tel.: +46763136135
E-mail address: san.giliyana@mitc.se

1877-0509 © 2025 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 6th International Conference on Industry 4.0 and Smart Manufacturing
10.1016/j.procs.2025.01.156

1. Introduction

In 2011, at the Hanover Fair, Industry 4.0 was introduced [1]. In line with Liu and Xu [2], Industry 4.0 deals with Information and Communication Technologies (ICT), intelligent factories and the development of Internet and embedded system technologies.

According to many researchers, such as Vaidya et al. [3] and Alcácer and Cruz-Machado [4], Industry 4.0 presents nine technologies: 1) Industrial Internet of Things (IIoT), 2) Big Data and Analytics, 3) Augmented Reality (AR), 4) Simulation, 5) Autonomous Robots, 6) Additive Manufacturing (AM), 7) Cyber Security, 8) Cloud Computing and 9) System Integration. In line with Silvestri et al. [5], the nine technologies of Industry 4.0, Cyber-Physical System (CPS) and machine learning, play an important role in developing smart maintenance technologies.

For instance, IIoT is used for connecting machine's components to the Internet, and thereby, collecting maintenance data, such as vibration and temperature data, through different types of communication protocols, such as Open Platform Communications-Unified Architecture (OPC-UA) and Message Queuing Telemetry Transport (MQTT), which is related to Big Data and Analytics, there machine learning models are used for the Analytics and maintenance data-driven decision-making [5]. Machine learning, a part of Artificial Intelligence (AI), is about creating a model that can learn by itself based on data [6]. There are three types of machine learning: 1) Supervised learning, forecasting events based on labeled data, 2) Unsupervised learning, which is based on unlabeled data and the final one is 3) Reinforcement learning, the capability of a machine to predict its own behavior to maximize its own performance [6, 7].

The definition of maintenance, according to [15, p.8], is "combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function". Note 1 for this definition is "Technical maintenance actions include observation and analyses of the item state (e.g. inspection, monitoring, testing, diagnosis, prognosis, etc.) and active maintenance actions (e.g. repair, refurbishment)." [15, p.8]. Based on Note 1, machine learning models can be used for analysis, diagnosis and prognosis to improve traditional maintenance, such as Predetermined Maintenance, Corrective Maintenance and Condition Based Maintenance (CBM).

In this paper, Regression Learner and Isolation Forest, which are two different machine learning models, Interquartile range (IQR), which is a statistical technique to find anomalies in datasets and to calculate thresholds, are applied in order to find anomalies in an unsupervised dataset and to calculate thresholds. The data is collected from a drill station, which is a part of a Cyber-Physical production system, located at Mälardalen Industrial Technology Center (MITC), which is a smart production laboratory, a part of Mälardalen University, in Sweden. Section 2 shows the literature review in order to show how IQR, Isolation Forest and Regression Learner can be applied to support the readers. Section 3 presents the system I used in this research paper, the dataset, and input variables. Sections 4, 5 and 6 present how IQR, Isolation Forest and Regression Learner are applied in this paper. The result is presented in Section 7, and conclusions and recommendations for further research in Section 8.

2. Literature review

In [8], different techniques for anomaly detection are studied. One of them is IQR, which is, as mentioned above, a statistical technique to find anomalies in datasets and it is suitable for skewed distributed datasets [19]. IQR is the difference between the third quartile, the largest entry median, and first quartile, the smallest entry median, in a dataset, Q3 and Q1, respectively [16]. In previous research, this technique is applied for the detection of anomalies in Network Intrusion Detection System [16], to threshold anomalies scores [8] and anomalies detection in the skewed datasets [19]. The following formulas are used to calculate IQR and threshold minimum and maximum. The user decides the value of c, which is usually set to 1.5. This means anomalies occur when the value exceeds the threshold maximum or falls below the threshold minimum [8].

$$IQR = Q3 - Q1 \tag{1}$$

$$Threshold\ min = Q1 - c \cdot IQR \tag{2}$$

$$Threshold\ max = Q3 + c \cdot IQR \tag{3}$$

In order to determine the relationship between the dependent and independent variables, the machine learning model Regression Learner can be applied. Moreover, Regression Learner can also be applied in the frame of value prediction analysis, which is of high interest for maintenance development [18]. Furthermore, when applying Regression Learner, residuals larger than two can be used in order to detect anomalies in a dataset [17]. In previous research, Regression Learner is used for predictive maintenance decisions [20], predicting plant maintenance costs [21], road traffic anomalies detection [22], and online anomaly detection for smart grid data [23].

According to Liu et al. [10], anomalies are few data points that differ from the normal points in a dataset. Additionally, Liu et al. [10] state that anomalies are data with different data characteristics from normal data. Fig. 1 shows normal point X_i, which generally requires more isolated partitions than anomaly point X_0, as shown in Fig. 2 [10]. In [10], the machine learning model Isolation Forest is proposed to detect anomalies by isolating points by taking advantage of the fact that anomalies are few data points and have different data characteristics from normal points. Isolation Forest is applied for unsupervised anomaly detection for gas turbines [24], predictive maintenance [25], and fault detection [26].

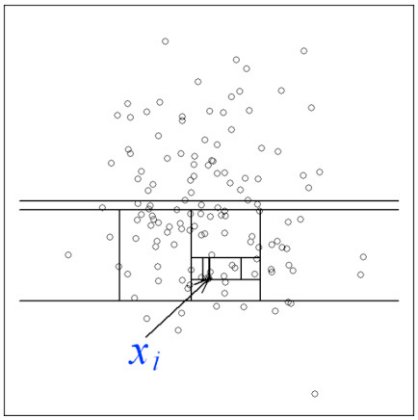

Fig. 1. Normal point X_i [10].

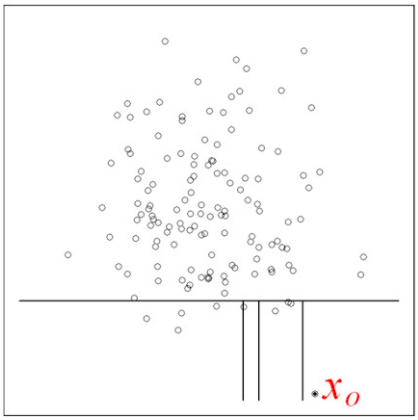

Fig. 2. Anomaly point X_0 [10].

3. System description

The system is a Cyber-Physical production system consisting of two lines, a factory line and a lab line, see Fig. 3. The robot in the middle, named as Robotino, moves producing products between the factory and lab. The lab line consists of eight stations. One of the stations is a drill station, which is used in this research paper. The drill itself moves on a linear guide between A and B, as the left side of Fig 4. shows. Furthermore, the right side of Fig. 4 shows that an external sensor placed in the drill station to collect temperature data in Celsius (C). The sensor is from Maintmaster Systems AB, a Swedish company that develops maintenance software and hardware. This sensor can collect vibration, temperature, pressure, and humidity data. Through a Raspberry Pi as a gateway, the temperature data from the sensor is sent to Maintmaster, a Computerized Maintenance Management System (CMMS), also from Maintmaster Systems AB. Fig. 5 shows the system for the temperature data collection.

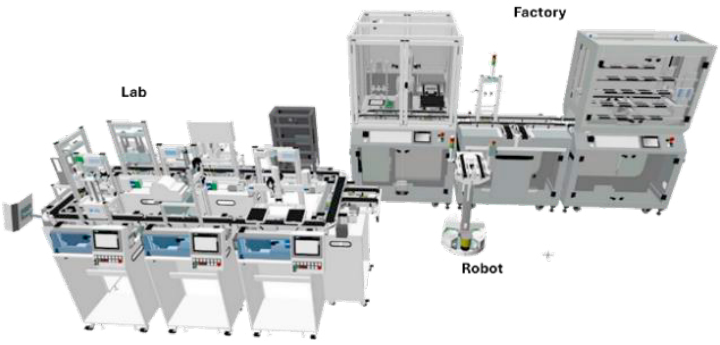

Fig. 3. Cyber-Physical production system.

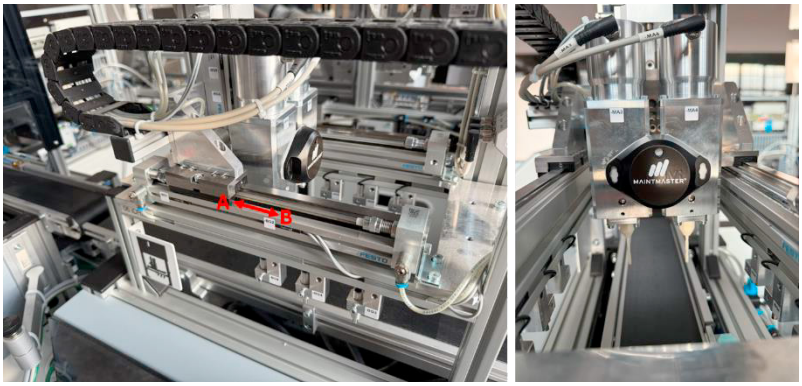

Fig. 4. The drill station. The left side shows that the drill itself moves between A and B on the linear guide. The right side shows an external sensor placed in the drill station to measure temperature.

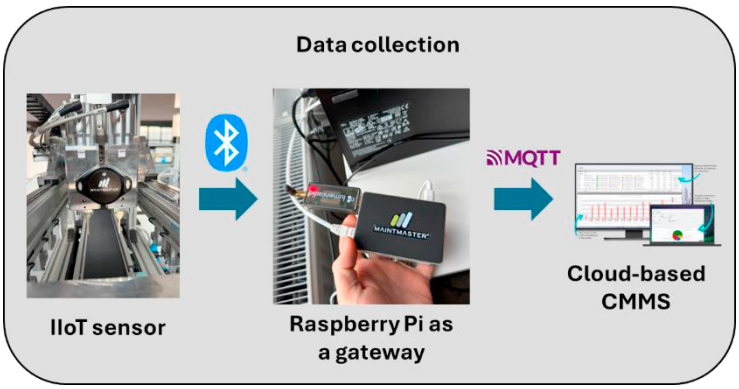

Fig. 5. The temperature data is collected using an IIoT sensor. The data is sent to the CMMS through a Raspberry Pi as a gateway. The used communications protocols are Bluetooth and MQTT.

3.1. *Dataset and input variables*

Unsupervised data is used in this research paper. The first input variable is the travel time in seconds (s) for the linear guide since, as mentioned above, the drill moves on a linear guide between A and B, as Fig. 4 shows. The travel time data is collected from the PLC using OPC UA. The second input variable is the temperature from the drill station, which is measured by the external temperature sensor. In the context of maintenance, when the travel time between A and B is longer than normal, the linear guide must be lubricated, cleaned, or replaced if broken. Both variables are

visualized on the figures below using Python library Matplotlib. Fig. 6 shows the travel time data for the drill to move on the linear guide, and Fig. 7 shows the temperature in the drill station. Furthermore, Fig. 6 shows Time To Failure (TTF) and Time Between Failure (TBF). The tops are anomaly 1, 2, 3, etc. Since the production system is new and no database of historical anomalies is available, we used Python library Numpy to simulate anomalies randomly. The healthy travel time for the linear guide is between 0.5 and 0.52 s, which we used in our simulation model to generate anomalies randomly.

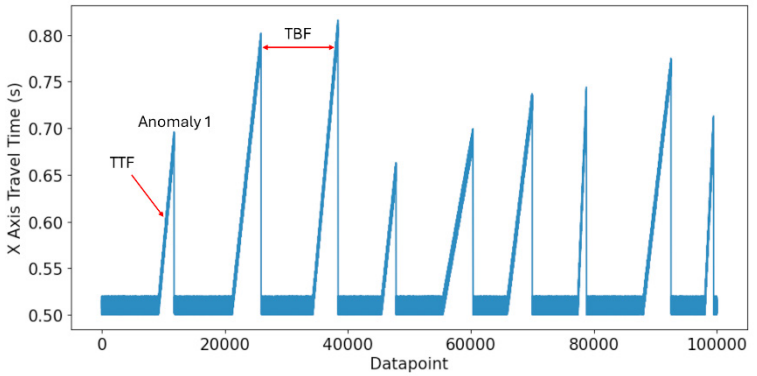

Fig. 6. Travel time data for X axis.

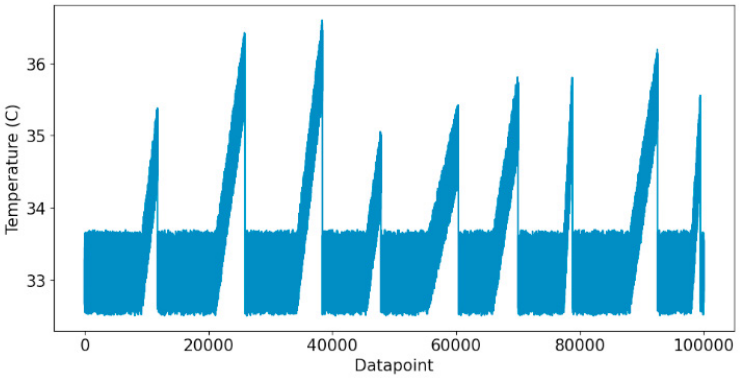

Fig. 7. Temperature data.

4. Interquartile range (IQR)

The data used in this research paper is skewed distributed, therefore IQR is used [19].

Q1 for the travel time was calculated to 0.507 C and Q3 to 0.546 C using Python library Numpy.

$$IQR = 0{,}546 - 0{,}507 = 0{,}039\ s$$

$$Threshold\ min = 0{,}507 - 1{,}5 \cdot 0{,}039 = 0{,}448\ s$$

$$Threshold\ max = 0{,}546 + 1{,}5 \cdot 0{,}039 = 0{,}605\ s$$

Q1 for the temperature was calculated to 32,947 C and Q3 to 33,595 using Python library Numpy.

$$IQR = 33{,}595 - 32{,}947 = 0{,}648\ C$$

$$Threshold\ min = 32{,}947 - 1{,}5 \cdot 0{,}648 = 31{,}975\ C$$

$$Threshold\ max = 33{,}595 + 1{,}5 \cdot 0{,}648 = 34{,}567\ C$$

Fig. 8, visualized using Python library Matplotlib, shows the maximum threshold for the travel time for the linear guide, which is 0.605 s. Fig. 9, also visualized using Python library Matplotlib, shows the maximum threshold for the temperature, which is 34.567 C. This means that anomalies occur when the travel time for the linear guide exceeds 0.605 s, and the temperature exceeds 34.567 C.

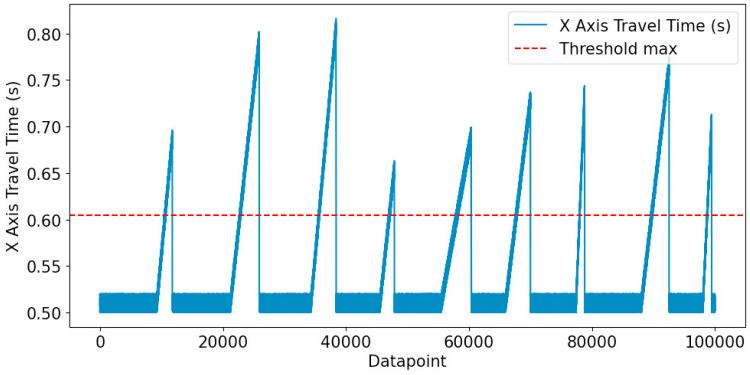

Fig. 8. The threshold for the travel time for the linear guide shows that anomalies occur when the travel time exceeds 0.6045 s.

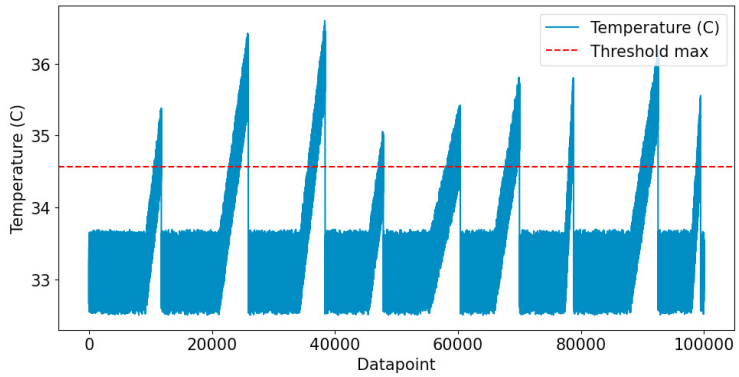

Fig. 9. The threshold for the temperature shows that anomalies occur when the temperature exceeds 34.567 C.

5. Regression Learner

This section applies Regression Learner to detect anomalies in the dataset used in this research paper. The figure below presents the residuals of the Regression Learner model, which show that most residuals are close to zero and indicate no clear signs of anomalies in the dataset. According to [17], residuals larger than two can be used to detect anomalies. Therefore, a regression learner model is unsuitable for detecting anomalies in this dataset [9].

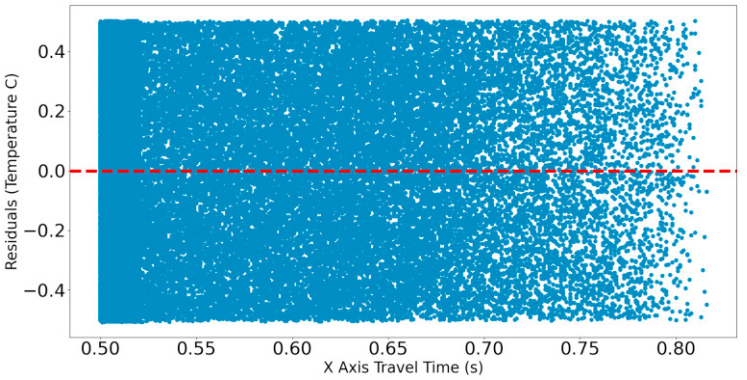

Fig. 10. The residuals of the regression model show that most residuals are close to zero and have no clear signs of anomalies in the dataset.

6. Isolation Forest

In this step, Isolation Forest has been studied to see the capability of this model to identify anomalies in the dataset used in this research paper. Fig. 11 shows the visualization based on Isolation Forest. The red area is the anomalies detected by the Isolation Forest.

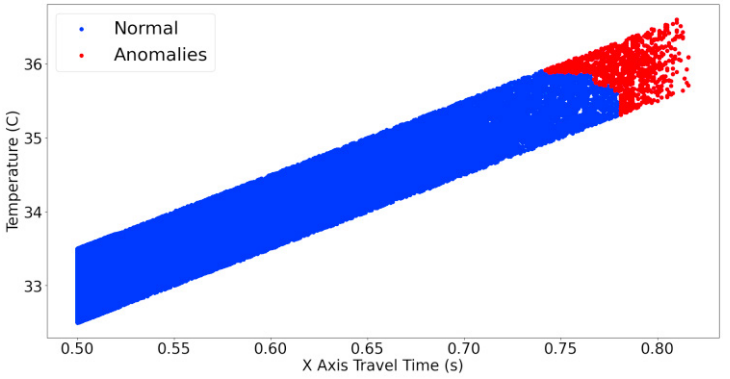

Fig.11. Isolation Forest. The red area is the detected anomalies.

The table below shows some of the anomalies detected by Isolation Forest.

Table 1. Anomalies detected by Isolation Forest.

Datapoint	Travel time (s)	Temperature (C)
24928	0.745	35.881
24943	0.745	35.913
24983	0.746	35.957
25010	0.746	35.952
25027	0.744	35.881

7. Result

IQR, Regression Learner and Isolation Forest have been used to detect anomalies in the dataset. As presented above, the residuals of the regression model show that most residuals are close to zero and indicate no clear signs of anomalies in the dataset used in this paper. Previous research presents that residuals larger than two can be used in

order to detect anomalies in a dataset [17]. Therefore, models based on Regression Learner are unsuitable for detecting anomalies in the dataset. One reason could be that the dataset has a linear relationship.

The dataset used in this research paper is skewed distributed, and previous research [19] shows that IQR can be applied to detect anomalies in these types of datasets and to calculate thresholds. IQR shows that anomalies occur when the travel time for the linear guide exceeds 0,605 s, and the temperature exceeds 34,567 C, which is in line with what Isolation Forest model shows, see Table 1.

Maintenance can be divided into Preventive Maintenance and Corrective Maintenance. There are two types of Preventive Maintenance, which are Predetermined maintenance and CBM. Predetermined Maintenance activities are performed time-based, while CBM activities are performed based on the condition of a component. One suggestion to optimize the maintenance in the drill station is to implement Predetermined Maintenance activities as rounds in the CMMS. One example is the daily cleaning and lubricating of the linear guide by operators as a part of Autonomous Maintenance activities [15].

Moreover, this paper can be of benefit to many manufacturing companies, since linear guides exist in many industrial machines.

8. Conclusion and further research

In this research paper, I have shown the power of IQR, Regression Learner and Isolation Forest, to detect anomalies in data collected from a drill station, part of a Cyber-Physical production system located at MITC, a smart production laboratory, a part of Mälardalen University, in Sweden. For further research regarding using machine learning models and other technologies in smart maintenance, I want to present some important points based on previous smart maintenance research.

Researching technologies for smart maintenance is a popular research topic. According to Lundgren et al. [11], many smart maintenance researchers focus on elaborating how Industry 4.0 technologies, CPS as well as machine learning, can be used for condition monitoring of equipment, remote services using AR, modeling wear of components, Remaining Useful Life (RUL) and prediction of failure. But one problem is that the technology focus does not consider humans and their interactions with the technologies [12]. To achieve success, the focus should not be only on technologies but also on humans using the technologies and support from organizations, such as maintenance, production and IT departments.

Industrial maintenance can significantly improve with Industry 4.0 technologies, CPS and machine learning models. But, according to Bengtsson et al. [13], to achieve this improvement, being careful when deciding which technologies to use for a given application is very important. Otherwise, the solutions might end up being cost-driven.

Furthermore, Lee et al. [14] have written about domain knowledge when working with industrial AI, and have identified four aspects:

1. Understanding the problem and focusing on industrial AI's power to solve it.
2. Understanding the system so that the right data with the right quality can be collected.
3. Understanding the physical meanings of the parameters and how they are associated with the physical characteristics of a system or process.
4. Understanding how these parameters vary from machine to machine.

Moreover, Bengtsson et al. [13] have written about domain knowledge when using AI in maintenance. Lack of domain knowledge will increase the risk of implementing technologies ineffectively [13].

The limitation of this study is that the anomalies are simulated using Python, because the Cyber-Physical production system is new and is located in a laboratory, not a real manufacturing facility. Therefore, no database of historical anomalies is available. For further research, it is recommended to use a dataset from real manufacturing to achieve better results when comparing the applied machine learning models Regression Learner and Isolation Forest, and the statistical technique IQR. One suggestion is to use time series data to enhance the results. However, it is also possible for new production systems in real manufacturing to lack a database of historical anomalies. In this case, the Simulation of potential anomalies and machine learning models could be suitable as support tools for maintenance planning. Especially when planning Predetermined Maintenance activities. Furthermore, the dataset used in this paper

has linear relationship. One suggestion is to test the applied machine learning models and IQR with a dataset that does not have linear relationship to compare the capability and result.

Moreover, [27] have mentioned several challenges to consider when implementing new technologies, such as IIoT for maintenance data collection and machine learning models for maintenance data analysis and decision-making, 1) Poor data quality, 2) Too large datasets to manage and store, 3) Competence, 4) Lack of inability to extract knowledge from data, 5) Lack of manufacturing system integration, 6) Lack of internet connectivity and 7) Security challenges related to connectivity and communication protocols. Furthermore, [28] have mentioned the challenge related to the reliability of AI and the potential lack of complete wireless connections, which is of high interest for this research paper.

Acknowledgments

The research work is a part of Industrial Technology Graduate School (INDTECH) and the AI course "AI-Based Modeling and Optimization for Industrial Systems - FOES019", at Mälardalen University in Sweden.

References

[1] Matt, Dominik, Vladimír Modrák, and Helmut Zsifkovits. (2021) "Implementing Industry 4.0 in SMEs: Concepts, Examples and Applications." *Springer Nature*.

[2] Liu, Yongkui, and Xun Xu. (2017) "Industry 4.0 and Cloud Manufacturing: A Comparative Analysis." *Journal of Manufacturing Science and Engineering* **139** (**3**): 1-8.

[3] Vaidya, Saurabh, Prashant Ambad, and Santosh Bhosle. (2018) "Industry 4.0 – A Glimpse." *Procedia Manufacturing* **20**: 233-238.

[4] Alcácer, V., and V. Cruz-Machado. (2019) "Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems." *Engineering Science and Technology, an International Journal* **22** (**3**): 899-919.

[5] Silvestri, Luca, Antonio Forcina, Vito Introna, Annalisa Santolamazza, and Vittorio Cesarotti. (2020) "Maintenance transformation through Industry 4.0 technologies: A systematic literature review." *Computers in Industry* **123**: 1-16.

[6] Kour, Herleen, and Naveen Kumar Gondhi. (2020) "Machine Learning Techniques: A Survey." In Jennifer S. Raj, Abul Bashar, S. R. Jino Ramson (eds) *Innovative Data Communication Technologies and Application*, Springer Cham, 266-275.

[7] Janiesch, Christian, Patrick Zschech, and Kai Heinrich. (2021) "Machine learning and deep learning." *Electronic Markets* **31**: 685–695.

[8] Yang, Jiawei, Susanto Rahardja, and Pasi Fränti. (2019) "Outlier detection: how to threshold outlier scores?" In João Manuel R. S. Tavares and Zeshui Xu (eds) *Proceedings of the international conference on artificial intelligence, information processing and cloud computing*, New York, Association for Computing Machinery, 1-6.

[9] Weisberg, Sanford. (2005) "Applied Linear Regression." Hoboken, New Jersey, John Wiley & Sons, Inc.

[10] Liu, Fei Tony, Kai Ting, and Zhi-Hua Zhou. (2012) "Isolation-Based Anomaly Detection." *ACM Transactions on Knowledge Discovery From Data* **6**: 1-39.

[11] Lundgren, Camilla, Cecilia Berlin, Anders Skoogh, and Anders Källström. (2023) "How industrial maintenance managers perceive socio-technical changes in leadership in the Industry 4.0 context." *International Journal of Production Research* **61** (**15**): 5282-52301.

[12] Bokrantz, Jon, Anders Skoogh, Cecilia Berlin, Thorsten Wuest, and Johan Stahre. (2020) "Smart Maintenance: an empirically grounded conceptualization." *International Journal of Production Economics* **223**: 1-17.

[13] Bengtsson, Marcus, Robert Pettersson, San Giliyana, and Antti Salonen. (2023) "The Importance of Using Domain Knowledge When Designing and Implementing Data-Driven Decision Models for Maintenance: Insights from Industrial Cases." In Uday Kumar, Ramin Karim, Diego Galar, Ravdeep Kour (eds) *International Congress and Workshop on Industrial AI*, Luleå, Springer Cham, 601-614.

[14] Lee, Jay, Hossein Davari, Jaskaran Singh, and Vibhor Pandhare. (2018) "Industrial Artificial Intelligence for industry 4.0-based manufacturing systems." *Manufacturing Letters* **18**: 20-3.

[15] Swedish Standards Institute. (2010) "Maintenance – Maintenance terminology".

[16] Vinutha, H. P., B. Poornima, and B. M. Sagar. (2018) "Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset." In Suresh Chandra Satapathy, Joao Manuel R.S. Tavares, Vikrant Bhateja, J. R. Mohanty (eds) *Information and decision sciences: Proceedings of the 6th international conference on ficta*, Singapore, Springer, 511-518.

[17] Kim, Hae-Young (2019). "Statistical notes for clinical researchers: simple linear regression 3–residual analysis." *Restorative dentistry & endodontics* **44 (1)**.

[18] Prakosa, Hendri Kurniawan and Rokhman, Nur (2021) "Anomaly Detection in Hospital Claims Using K-Means and Linear Regression." *Indonesian Journal of Computing and Cybernetics Systems* **15 (4)**: 391-402.

[19] Hubert, Mia and Van der Veeken, Stephan (2008). "Outlier detection for skewed data." *Journal of Chemometrics: A Journal of the Chemometrics Society* **22 (3-4)**: 235-246.

[20] Le, Tung, Ming Luo, Junhong Zhou and Hian L. Chan (2014) “Predictive maintenance decision using statistical linear regression and kernel methods.” In Antoni Grau, Herminio Martínez (eds) *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, Barcelona, IEEE, 1-6.

[21] Edwards, David J., Gary D. Holt and Frank C. Harris (2000) “A model for predicting plant maintenance costs.” *Construction Management & Economics* **18(1):** 65-75.

[22] Mondal, M. A. and Zeenat Regena (2020) “Road traffic outlier detection technique based on linear regression.” *Procedia Computer Science* **171**: 2547-2555.

[23] Liu, Xiufeng and Per Sieverts Nielsen (2016) “Regression-based online anomaly detection for smart grid data.” *arXiv preprint:1606.05781*.

[24] Zhong, Shisheng, Song Fu, Lin, Lin, Xuyun Fu, Zhiquan Cui and Rui Wang (2019) “A novel unsupervised anomaly detection for gas turbine using isolation forest.” In *International Conference on Prognostics and Health Management (ICPHM)*, Francisco, CA, IEEE, 1-6.

[25] Putrada, Aji Gautama, Nur Alamsyah, Syafrial Fachri Pane, Mohamad Nurkamal Fauzan and Doan Perdana (2023) “Predictive Maintenance Application on Machine Overstrain Failure with Node-Red and Isolation Forest Anomaly Detection.” In *International Conference on Communication, Networks and Satellite (COMNETSAT)*, Malang, IEEE, 64-69.

[26] Santis, Rodrigo and Marcelo Costa (2020). “Extended isolation forests for fault detection in small hydroelectric plants.” *Sustainability* **12(16)**: 6421.

[27] Bajic, Bojana, Aleksandar Rikalovic, Nikola Suzic, and Vincenzo Piuri. (2021) "Industry 4.0 Implementation Challenges and Opportunities: A Managerial Perspective." *IEEE Systems Journal* 15: 546-559.

[28] Forcina, Antonio, Vito Introna, and Alessandro Silvestri. (2021) "Enabling technology for maintenance in a smart factory: A literature review." *Procedia Computer Science* **180**: 430-435.