2024

Development and Future Research Directions of AI-Based Anomaly Detection in Smart Manufacturing: A Bibliometric Analysis

Maximilian Nebel
Technische Universität Dortmund, Germany, maximilian.nebel@tu-dortmund.de

Philip Stahmann
Technische Universität Dortmund, Germany, philip.stahmann@tu-dortmund.de

Christian Janiesch
Technische Universität Dortmund, Germany, christian.janiesch@cs.tu-dortmund.de

Development and Future Research Directions of AI-Based Anomaly Detection in Smart Manufacturing: A Bibliometric Analysis

Research in Progress

Maximilian Nebel[1], Philip Stahmann[1], Christian Janiesch[1]

[1] TU Dortmund University, Chair of Enterprise Computing, Dortmund, Germany
{maximilian.nebel,philip.stahmann,christian.janiesch}@tu-dortmund.de

Abstract. Manufacturing companies face a vast increase of data. Connected sensors turn physically isolated objects into nodes in data communication networks. This development enables but also forces companies to harness their data to gain a competitive edge. In this regard, anomaly detection enables seamless processes, so that production failures can be avoided. Artificial intelligence (AI) and especially machine learning and deep learning constitute instruments to leverage statistical complexity necessary to identify anomalies in these vast amounts of data. AI-based anomaly detection has therefore been subject to an intensive academic discourse in Information Systems. This short paper provides preliminary results from a bibliometric analysis highlighting the development over time of scientific contributions in this field. Our findings show that the academic discourse has gained momentum but is still pre-mature. Additionally, we find that a technical perspective on the topic prevails in literature.

Keywords: *Artificial intelligence, anomaly detection, bibliometric analysis, manufacturing.*

1 Introduction

As is the case with numerous other sectors, the manufacturing industry is subject to considerable competitive pressure driven by the relentless pursuit of optimization. This is especially pertinent in the manufacturing sector, given that the potential for improvement is consistently being tapped into through the use of technologies such as anomaly detection to minimize flaws in the production. Therefore, there is increased pressure to optimize manufacturing processes to save costs and offer products at lower prices. Artificial Intelligence (AI)-based anomaly detection has been identified as a driver for manufacturing process optimization (Peres et al. 2018).

Anomaly detection targets the identification of unexpected events (Chandola et al. 2009). In manufacturing, the aim is to alert decision-makers regarding about the occurrence of anomalies such as unexpected machinery wear or production defects to entail initiate remediate corrective actions (Huang et al. 2021). The capability to leverage complex statistical models using machine learning supports the inference of information from data (Janiesch et al. 2021).

Despite the importance of AI-based anomaly detection in manufacturing, related literature is scattered into isolated publications. Aggregating literature that characterizes the development of the topic in science is scarce. However, publications in adjacent fields in Information Systems (IS) show that an aggregated understanding of scientific pathways of topics opens potentials for further development in academia and practice (e.g., Schöbel et al. 2021; Wanner et al. 2023; Zschech et al. 2019). On this basis, we formulate the following research question (RQ):

RQ: *How has research on AI-based anomaly detection in manufacturing developed over time?*

To answer the RQ, we conducted a bibliometric study based on 161 publications. More specifically, we analyzed types of outlets, citations, and keywords over the past six years. This short paper includes preliminary answers to the RQ. Further bibliometric analyses that build on an extended literature database to address the RQ in greater detail are still ongoing.

2 Research Methodology

To find answers to both research questions, we structured our research into the three phases of data acquisition, data preparation, and data analysis (Donthu et al. 2021). Data acquisition marks the *first phase*, the focus was on ensuring that required information is included in the data. Due to their vast diffusion in IS, we adhered to the guidelines established by vom Brocke et al. (2009). We first defined the scope on what we look for in the gathered search results. Until now, in our search we did not focus on the results or methods presented in the publications, rather than on the meta data, such as citations or keywords. In this short paper, we searched the databases IEEE Xplore, Scopus, Web of Science, EBSCOhost and AISEL as these cover a broad variety of high-quality outlets on relevant topic.

Our search term consists of three connected blocks of terms and their abbreviations. Each block represents one of the three core topics that define the AI-based anomaly detection in manufacturing domain. The first block contains terms regarding AI. The second one focuses on anomalies. Lastly, the third block refers to manufacturing terminology. The final search term is the following: (AI-based OR "Artificial Intelligence-based" OR ML-based OR "machine learning-based" OR DL-based OR "deep learning-based") AND (anomaly OR outlier OR error) AND ("Industrie 4.0" OR "industry 4.0" OR "intelligent manufacturing" OR "smart factory" OR "smart manufacturing" OR "IIoT" OR "industrial Internet of things").

In total, we found 717 publications. After deduplication and filtering of the titles, 161 sources remain. The oldest results are from 2016 with one Outlier in 2006. Independent of this, we identified the most cited authors in Google Scholar.

Subsequently, the *second phase* focuses on data preparation for data analysis in the third phase. The gathered data from the literature databases was structurally heterogeneous. To enable analysis, we aligned the data formats. The meta data includes the corresponding databases, authors, titles, publication years, countries of the authors institutions, keywords, and whether the publication is from a journal or proceeding.

The design of data and data formats is based on other bibliometric analyses in the IS domain (e.g., Honey et al. 2022; Smacchia and Za 2022; Schöbel et al. 2023). Besides structural homogenization, data preparation also included semantic alignment of keywords. For example, “industrial Internet of things (IIoT)” was reduced to “industrial Internet of things”. We hereby aimed to do as little alterations as necessary.

Phase three covers data analysis. Rather than a static, predefined procedure, we used an explorative, iterative approach as intermediate results may open avenues for further research. To determine the most cited publications according to Google Scholar, we used a Python script provided by Wittmann (2017).

3 Theoretical Background

Anomaly detection is a method of identifying patterns in data that deviate from the norm (Chandola et al. 2009). This approach is employed in manufacturing, because the time-series sensor data can be analyzed by computers to identify faults during production (Stahmann and Rieger 2021). AI has the potential to be a significant factor in this process. By employing sophisticated statistical machine learning techniques, AI can discern the attributes of patterns and indications of anomalies. The advantage of AI is that it enables the use of a variety of statistical methods, each of which may be suited best to different types of data (Russell and Norvig 2022).

Classification algorithms delineate data into normal and anomalous classes (e.g., Saci et al. 2021). It is a supervised technique, which requires a training phase before algorithm deployment. As an unsupervised technique, clustering algorithms constitute a more explorative approach, where data are assigned to homogeneous clusters (e.g., Mucchielli et al. 2021). Data points that are not assigned to expected clusters are considered anomalous. Furthermore, predictive analysis enables forecasting future states in production processes (Stefani and Zschech 2018).

Companies can take advantage of this by selecting the most appropriate machine learning methods for their needs. Machine learning can be used to predict known anomalies at an early stage, enabling intervention even before critical failures occur for preventive actions ensuring seamless processes and to minimize downtimes (Corallo et al. 2018). With the decision making behavior of AI, it has the potential to expand the automation of manufacturing processes (Arinez et al. 2020; Yang et al. 2021). For instance, an AI detects an anomaly and triggers the corresponding reaction automatically.

4 Preliminary Results and Contributions

In the following, preliminary analysis results are structured into citation analysis, outlet analysis, and keyword analysis.

Citation analysis. First, we identified the most cited publications to understand which publications constitute the most relevant basis for further scholars. To get a reliable picture, we also considered the average citations per year. These support an evaluation whether a publication constitutes a lasting foundation for future research. Figure 1 displays the most cited publications in AI-based anomaly detection in smart manufacturing. Noteworthy, the top four most cited publications in total are also the most cited publications on average over the last years. The oldest publication is from 2016 (Wang et al.), the others are from 2020 (Zonta et al; Çınar et al.), 2021 (Sarker; Zhang and Lu) or 2022 (Ahmed et al.).

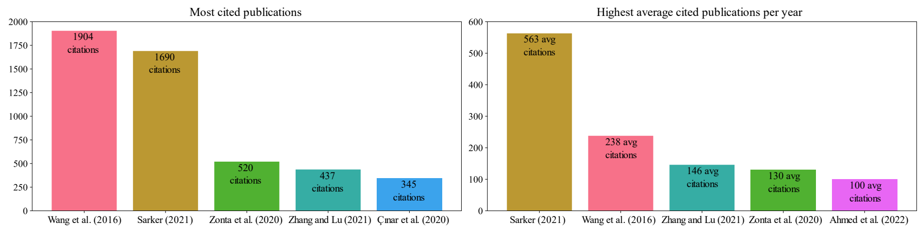

Figure 1. Most cited publications

Outlet analysis. To further identify the development of the field of the academic discourse over the last years we investigated the distribution of outlet types over time. whether a publication distributed as part of a journal or conference. Figure 2 shows the absolute number of publications in conference proceedings and journals.

Beginning from 2019, the number of publications increases with an average rate of 162% per year (185% in conference proceedings and 154% in journal publications). It is worth noting that the proportion of journal publications is gradually increasing in comparison to conference proceedings, with 25% of the publications in 2023 appearing in journals.

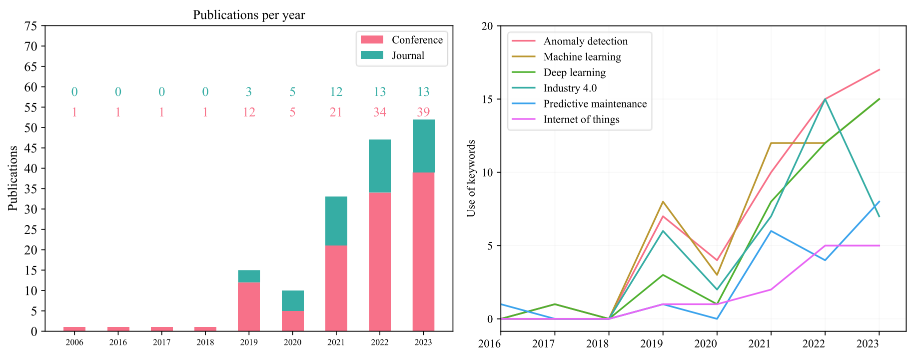

Figure 2. Annual published articles (left) and top six most used keywords (right)

Keyword analysis. The analysis of keywords aims to provide an understanding of the most relevant topics in AI-based anomaly detection in manufacturing and how they developed over time. Figure 2 shows on the right side the development of the six most frequently used keywords as chosen by the authors of the identified publications.

The most prominent keywords are "anomaly detection", "machine learning", and "deep learning". "Machine learning" dominated until 2022 and has then been overtaken by "deep learning". Deep learning is usually used for more complex problem-solving that may be due to data variety in manufacturing (Wang et al. 2018). Four out of the six most frequently used keywords relate to manufacturing as application domain. Out of these "predictive maintenance" constitute the majority in 2023. (Industrial) "Internet of things" and "industry 4.0" derive from national initiatives towards the digitization of manufacturing. The industrial Internet of things is the US American term, whereas industry 4.0 has its roots in Germany.

Preliminary practical and theoretical contributions. Our research aims for a contribution to both practice and academia. Academics can use the results as starting points for future research and by newcomers to the field to gain an overview and familiarize themselves with the various aspects of the field. Additionally, our work supports researchers in identifying relevant topics and keywords for their own literature search in their work and thus creates an incentive to explore AI-based anomaly detection in manufacturing. The results of our research offer practitioners benefits as well. They can use the findings to more accurately assess past developments and identify potential issues caused by outdated technologies. The insights into future trends enable companies to anticipate changes and capitalize on them.

5 Conclusion, Limitations, and Outlook

In our preliminary study, we conducted a bibliometric analysis following Donthu et al. (2021) to observe the development and future directions of AI-based anomaly detection. In particular, for this research-in-progress paper, we focused on the timely developments regarding the citations and keywords used in the data.

To answer our RQ: The findings suggest that the field is transitioning from a young and fast changing area towards a more mature research field. This is seen in the development of the conference to journal ratio towards journals from 2021 and onwards. We also showed the most cited works, that are journal articles that predominantly describe a state of the art or fundamentals relevant to the field.

With the keyword analysis, we find indications of a development towards more sophisticated statistical models for anomaly detection as deep learning characterizes by more complex model building than machine learning. The prominence of "Internet of things" and "industry 4.0" indicates that AI-based anomaly detection has a significant share in digitization initiatives from the perspective of the academic discourse. Furthermore, we find that the academic discourse seems to be considered through a methodological or technical lens, as "machine learning" and "deep learning" occur more frequently than terms relating to the digitization of manufacturing. An academic discourse guided by methodological considerations indicates a focus on problem-solving, that is ensuring seamless manufacturing processes with provided data.

Despite rigorous adherence to scientific guidelines of vom Brocke et al. (2009), the literature search process has some limitations. Choices on databases, search strings, and parameters have a significant influence on results of bibliometric analyses. Additionally, the selection of publications as well as database depends on subjective content interpretation. There is no guarantee that we did not miss publications that would have been relevant to investigating the research questions. Furthermore, keywords used in bibliometric analyses to reflect content of publications and literature streams (Schöbel et al. 2021; Schöbel et al. 2023) are only proxies that address essential aspects of content, but can never fully depict the complexity of contributions.

In order to obtain a broader representation of the field, we plan to expand our research by not only increasing the number of publications as a database for bibliometric analysis, but also by expanding our methodological toolbox. We will add publications found for example in the databases ACM or Springer Link. The choice of databases is due to their broad and complementary coverage of various outlets that engage with the topics of applied computer science, AI, and smart manufacturing.

In our further research, we will detail the presented analyses based on the broader literature coverage. In terms of citations, and keyword use, we plan to conduct network analysis to identify (covert) groupings by topic and geographic distributions to further investigate local clusters and focal points. We will not only look at the quantity of publications over the years, but also on the geographical diffusion of different key aspects in the domain over certain time spans. For example, we plan on investigating whether the network shows signs of saturation in regard to expansion and networking. Additionally, this can be used by practioners to identify global trends, such as on-demand machine learning as part of artificial-intelligence as a service solutions (e.g., Lins et al. 2021). Regarding outlets, we will extend the perspective on outlet types towards analyzing which conferences and journals engage with which topics over time. This may foster an understanding of the development today and reveal future research avenues (Schöbel et al. 2023). Broader keyword analyses including their topic-wise relations will support the identification clusters suitable to project future research directions (Donthu et al. 2021). The expected contribution of our full research yields at a characterization of the development of research on AI-based anomaly detection in manufacturing (RQ). In addition, we want to expand with complementary RQs to include further perspectives and also identify future research directions.

6 Acknowledgements

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the „Zukunft der Wertschöpfung - Forschung zu Produktion, Dienstleistung und Arbeit“ (Funding No. 02K23A070) and managed by Projektträger Karlsruhe (PTKA). The authors are responsible for the contents of this publication.

References

Ahmed, I., Jeon, G. & Piccialli, F. (2022), 'From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where'. *IEEE Transactions on Industrial Informatics* **18**(8), 5031–5042. https://doi.org/10.1109/TII.2022.3146552.

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C. & Zhang, J. (2020), 'Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook'. *Journal of Manufacturing Science and Engineering* **142**(11). https://doi.org/10.1115/1.4047855.

Chandola, V., Banerjee, A. & Kumar, V. (2009), 'Anomaly detection'. *ACM Computing Surveys* **41**(3), 1–58. https://doi.org/10.1145/1541880.1541882.

Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M. & Safaei, B. (2020), 'Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0'. *Sustainability* **12**(19), 8211. https://doi.org/10.3390/su12198211.

Corallo, A., Crespino, A., Dibiccari, C., Lazoi, M. & Lezzi, M. (2018), Processing Big Data in Streaming for Fault Prediction: An Industrial Application. *in* '2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)', 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 26.11.2018 - 29.11.2018. IEEE, 730–736.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. M. (2021), 'How to conduct a bibliometric analysis: An overview and guidelines'. *Journal of Business Research* **133**, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070.

Honey, K., Robb, A. & Rohde, F. (2022), Are we contributing? The who, when, where, & what of the Blockchain research landscape. *in* 'ICIS 2022 Proceedings'.

Huang, H., Yang, L., Wang, Y., Xu, X. & Lu, Y. (2021), 'Digital Twin-driven online anomaly detection for an automation system based on edge intelligence'. *Journal of Manufacturing Systems* **59**, 138–150. https://doi.org/10.1016/j.jmsy.2021.02.010.

Janiesch, C., Zschech, P. & Heinrich, K. (2021), 'Machine learning and deep learning'. *Electronic Markets* **31**(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2.

Mucchielli, P., Bhowmik, B., Ghosh, B. & Pakrashi, V. (2021), 'Real-time accurate detection of wind turbine downtime - An Irish perspective'. *Renewable Energy* **179**, 1969–1989. https://doi.org/10.1016/j.renene.2021.07.139.

Peres, R. S., Dionisio Rocha, A., Leitao, P. & Barata, J. (2018), 'IDARTS – Towards intelligent data analysis and real-time supervision for industry 4.0'. *Computers in Industry* **101**, 138–146. https://doi.org/10.1016/j.compind.2018.07.004.

Russell, S. & Norvig, P. (2022), 'Artificial intelligence'. A modern approach. 4th ed. Harlow, Pearson Education Limited.

Saci, A., Al-Dweik, A. & Shami, A. (2021), 'Autocorrelation Integrated Gaussian Based Anomaly Detection using Sensory Data in Industrial Manufacturing'. *IEEE Sensors Journal* **21**(7), 9231–9241. https://doi.org/10.1109/JSEN.2021.3053039.

Sarker, I. H. (2021), 'Machine Learning: Algorithms, Real-World Applications and Research Directions'. *SN computer science* **2**(3), 160. https://doi.org/10.1007/s42979-021-00592-x.

Schöbel, S., Saqr, M. & Janson, A. (2021), 'Two decades of game concepts in digital learning environments – A bibliometric study and research agenda'. *Computers & Education* **173**, 104296. https://doi.org/10.1016/j.compedu.2021.104296.

Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A. & Leimeister, J. M. (2023), 'Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers'. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-023-10375-9.

Smacchia, M. & Za, S. (2022), Artificial Intelligence in Organisation and Managerial Studies: A Computational Literature Review. *in* 'ICIS 2022 Proceedings'.

Stahmann, P. & Rieger, B. (2021), Requirements Identification for Real-Time Anomaly Detection in Industrie 4.0 Machine Groups: A Structured Literature Review. *in* 'Proceedings of the Annual Hawaii International Conference on System Sciences'. Hawaii International Conference on System Sciences.

Stefani, K. & Zschech, P. (2018), Constituent elements for prescriptive analytics systems. *in* 'ECIS 2018 Proceedings'.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. & Cleven, A. (2009), Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. *in* 'ECIS 2009 Proceedings'.

Wang, J., Ma, Y., Zhang, L., Gao, R. X. & Wu, D. (2018), 'Deep learning for smart manufacturing: Methods and applications'. *Journal of Manufacturing Systems* **48**, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003.

Wang, S., Wan, J., Di Li & Zhang, C. (2016), 'Implementing Smart Factory of Industrie 4.0: An Outlook'. *International Journal of Distributed Sensor Networks* **12**(1), 3159805. https://doi.org/10.1155/2016/3159805.

Wanner, J., Wissuchek, C., Welsch, G. & Janiesch, C. (2023), 'A Taxonomy and Archetypes of Business Analytics in Smart Manufacturing'. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* **54**(1), 11–45. https://doi.org/10.1145/3583581.3583584.

Wittmann, F. M. (2017), 'Optimization applied to residential non-intrusive load monitoring'. Dissertation (Masters). University of Campinas, School of Electrical and Computer Engineering, Campinas, SP.

Yang, T., Yi, X., Lu, S., Johansson, K. H. & Chai, T. (2021), 'Intelligent Manufacturing for the Process Industry Driven by Industrial Artificial Intelligence'. *Engineering* **7**(9), 1224–1230. https://doi.org/10.1016/j.eng.2021.04.023.

Zhang, C. & Lu, Y. (2021), 'Study on artificial intelligence: The state of the art and future prospects'. *Journal of Industrial Information Integration* **23**, 100224. https://doi.org/10.1016/j.jii.2021.100224.

Zonta, T., Da Costa, C. A., Da Rosa Righi, R., Lima, M. J. de, Da Trindade, E. S. & Li, G. P. (2020), 'Predictive maintenance in the Industry 4.0: A systematic literature review'. *Computers & Industrial Engineering* **150**, 106889. https://doi.org/10.1016/j.cie.2020.106889.

Zschech, P., Bernien, J. & Heinrich, K. (2019), Towards a Taxonomic Benchmarking Framework for Predictive Maintenance: The Case of NASA's Turbofan Degradation. *in* 'ICIS 2019 Proceedings'.