



(REVIEW ARTICLE)



Enhancing predictive accuracy in manufacturing through AI/ML-driven data labeling

Srinivas Vikram *

Independent Researcher, USA.

International Journal of Science and Research Archive, 2025, 16(03), 871-879

Publication history: Received on 16 May 2025; revised on 01 July 2025; accepted on 04 July 2025

Article DOI: <https://doi.org/10.30574/ijrsra.2025.16.3.1694>

Abstract

This research focuses on data-labeling techniques using AI and ML that are overcoming challenges in manufacturing. Data labeling that supports supervised learning plays a key role in allowing machines to spot patterns and get their predictions correct. The study themes give importance to gathering data in real time during manufacturing, developing advanced ways to improve labels are added and upgraded, and achieving great results, such as making decisions more efficiently, having instant insights into operations, and building systems that can learn independently. The automatic data labeling process is important for reaching intelligent, predictive, and autonomous manufacturing.

Keywords: AI In Manufacturing; Machine Learning; Data Labeling; Predictive Insights; Real-Time Analytics; Automated Labeling; Intelligent Decision-Making; Industry 4.0

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) in the manufacturing industry play a vital role. Using these digital technologies, the companies can work more efficiently, control expenses, and encourage new ideas. AI or ML has proved that it can help transform the manufacturing industry by offering timely and useful advice using data. The effectiveness of their models mainly depends on the quality and quantity of the labelled data that is used for training.

Even though AI or ML offer great opportunities for manufacturing, there is a major barrier in the data labeling process. Most traditional ways of annotating data require time, effort, and plenty of resources. Manual organizing of such various data is not possible [1]. This can prevent the use of large-scale data and slow the process of developing models. AI or ML models cannot handle the demands of important manufacturing cases without clear and correct labels on the data.

Data labeling is indeed important in manufacturing, but the industry is not the only one facing those issues because the same situation occurs in the medical and technology industries also. The crucial aspect of accuracy in healthcare is patient safety, and autonomous vehicles need a lot of labeled image databases in a variety of conditions [2]. In comparison with these industries, manufacturing will have to deal with high-volume real-time operational data, whereas scalability and integration are crucial factors.

Aim and Objectives of the Research

The aim of this research is to examine the AI or ML-driven data labeling process that can enhance predictive insights in the manufacturing sector.

- To analyze the impact of accurate and automated data labeling on the performance of AI or ML models in manufacturing.

* Corresponding author: Srinivas Vikram

- To explore current techniques and advancements in real-time and AI-assisted data labeling that support intelligent, data-driven decision-making in manufacturing environments.

This study explains that AI and ML can be used in their final applications as well as to boost their process of data labeling. The science deals with intelligent tools, real-time data solutions, and algorithms that update themselves, making annotation less complex for all types of data, such as images, sensor information, logs, and issues. Secondary qualitative analysis will show that making data labeling more accurate leads to better predictive results and helps businesses decide faster. This also helps to create systems that can act independently in failure management [2]. AI or ML is used to label data at scale; manufacturing companies can make the most of predictive analytics and direct their decisions based on data.

2. Literature review

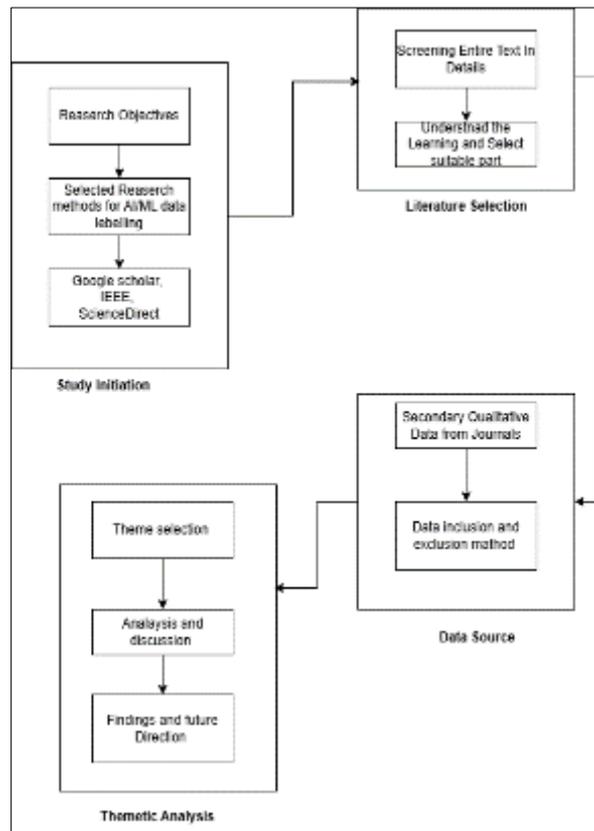


Figure 1 Flow of the Research

In this section, a comprehensive literature review has been compiled based on the following steps:

- **Identification of Key Concepts:** Key themes such as artificial intelligence, machine learning, data labeling, predictive analytics, and their applications in manufacturing are identified to guide the research scope.
- **Selection of Relevant Studies:** Scholarly articles, review papers, case studies, and technical reports relevant to AI/ML in manufacturing and data labeling are selected through keyword-based searches.
- **Critical Analysis and Synthesis:** The selected literature is analyzed to understand current trends, gaps, and contributions related to AI-driven data labeling and its impact on predictive capabilities in the manufacturing domain.

The literature review articles have been searched from the following reputable and well-recognized scientific publishers and organizations:

- **IEEE Xplore Digital Library:** The technical papers and conference proceedings on AI/ML and industrial automation are collected from this source.

- **ScienceDirect:** The peer-reviewed journals related to manufacturing, data science, and intelligent systems are collected from this source.
- **Google Scholar:** The academic articles focusing on AI applications, industrial informatics, and data engineering in smart manufacturing systems are collected from this source.

2.1. Searching Study

Relevant literature was identified using keywords such as “AI in manufacturing”, “machine learning”, “data labeling”, and “predictive insights” across databases like IEEE Xplore, ScienceDirect, and Google Scholar.

2.2. Selection of Journal Articles

The secondary analysis is made to select journal articles that reviewed the link between AI and ML in manufacturing, mainly focusing on data labeling that is applied there. The articles selected had a connection to predictive analytics, working with fresh data, and automation in plants. Resulting studies are mainly from the last 5 to 7 years, so that both existing and innovative techniques will be included. Research is limited to papers from prestigious publishers, IEEE, Elsevier, and Scholar. All these papers are reliable and related to the research objectives.

2.3. The Goal of the Review

The main purpose of this literature review is to study and combine previous studies on adding AI and ML to manufacturing by emphasizing the key role of data labeling for correct predictions. It plans to find out the AI/ML model performance that is affected by data labeling and discuss recent advancements in real-time and automated labeling. It also points out some of the issues and opportunities in using them for industry. This part gives an initial explanation of data quality, and good labels are important for making smart and efficient decisions in manufacturing.

2.4. Study of Previous Literature

2.4.1. AI and ML Integration in Manufacturing



Figure 2 Importance of AI and ML in Manufacturing

Various journals show that AI and ML have progressively come into use in manufacturing businesses. As stated by [3], AI methods have improved the way automation, finding defects, quality checks, and predictive maintenance are done. Apps in this area depend on good data to work and have improved their operations. According to [4], applying machine learning allows predicting problems with equipment and adjusting maintenance to enhance product quality, cost, time saving and customer trust. The efficacy of the outcomes mainly depends on the accuracy of the labeling of the training data.

2.4.2. The Role of Data Labelling in Predictive Analytics

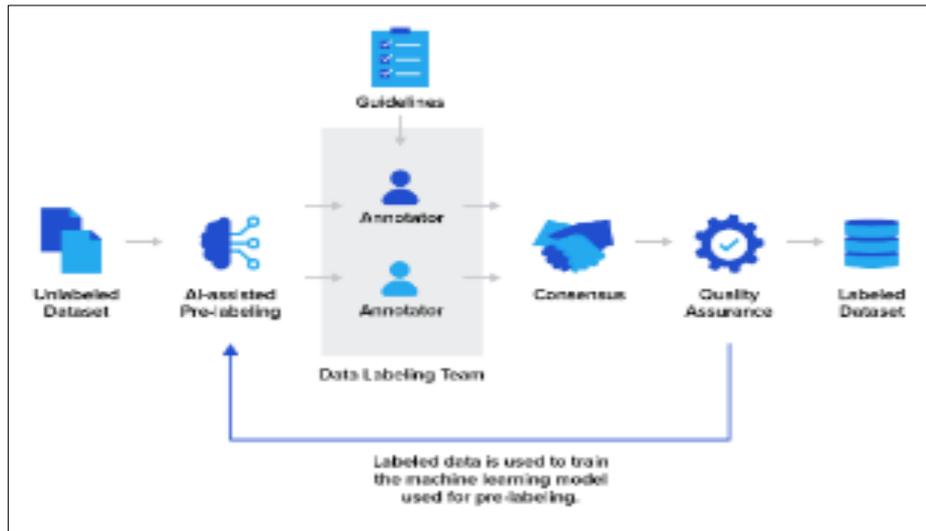


Figure 3 Data labeling process

A recent study by [6] stated that having better and more labelled data improves supervised machine learning. Some of the typical data labeling steps in manufacturing are to sort sensor results, identify flaws in images, and include tags on machine logs. Inaccurate labeling lowers the quality of predictions obtained by the model [7]. Correct and steady labeling helps the system handle quick decisions, discover problems at once, and make better forecasts. Evidence indicates that context in manufacturing, such as configurations, test results, and solutions to fixing problems, should be logged for use when the training data is gathered.

2.4.3. Automation and AI-Augmented Labelling

Traditional labeling techniques, being labor-intensive, present scalability challenges. This research has explored AI-assisted labeling to address this. According to [7], using semi-supervised learning and active learning can take care of a big portion of the labeling work, thus reducing manual effort. As a result, machines can recommend probable labels that can be confirmed by human experts [8]. This system can work both quickly and consistently. Weak supervision and the production of synthetic data are being used by industries that do not have access to fully labelled datasets.

2.4.4. Real-Time Data Labelling for Manufacturing Intelligence

According to [9], getting accurate insights by using neural networks requires a steady flow of data labeling in real time. Images taken for review, and operational values monitored, real-time labeling means that ML models always respond appropriately to new situations, with sensors recording data. According to [10], joining labeling systems to Manufacturing execution systems (MES) and IoT platforms helps supply chain managers make the best recommendations. This integration, predictive outcomes keep being improved with every new set of labelled data.

2.4.5. Addressing Challenges in Automated Data Labeling

Although the automated data labeling has many advantages in terms of improved speed, scalability, and accuracy, it also introduces certain challenges in case of the manufacturing setting. Also, the initial set-up can be quite difficult, needing more specific configurations, depending on the specific production processes. Another issue that creates a risk of data privacy is working with sensitive manufacturing or customer-related data [11]. Moreover, legacy systems integration may not be easy since they are not easily compatible with newer AI or ML technologies. Companies are also using the modular, flexible solutions of labeling systems, enforcing a high data governance policy to resolve these issues. Also, the companies employ middleware to fill the infrastructural gaps between the old system and the emerging new sources of technologies.

2.5. Literature gap

There is a lot of research showing AI/ML and data labeling are important for manufacturing, but there are very few studies that investigate the influence of real-time AI labeling on predictive decision-making [10]. It is important to

understand the process of data labeling. This research will help to analyze various secondary resources to understand the process of data labeling.

3. Methodology

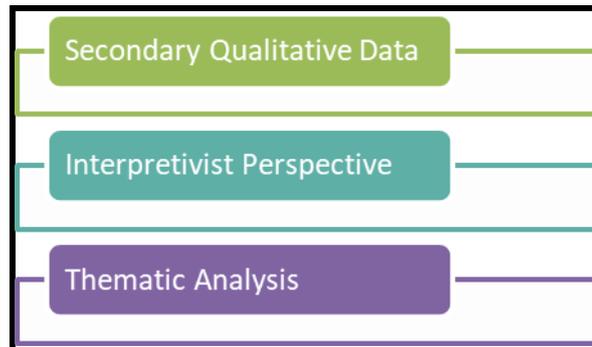


Figure 4 Method used

This research adopts a secondary qualitative approach that is guided by the interpretivist perspective that can help to explain and interpret human experiences, actions, and technology. This style of studying was chosen since its goal is to understand AI/ML that supported data labeling works in factories, not to measure data or perform experiments.

A qualitative approach enables to look deeply into current literature, cases of technology use, whitepapers, and academic articles on the topic of data labeling that helps to making predictions [12]. The fact that the work covers new technologies and developing fields makes qualitative research better suited than quantitative research to note trends, difficulties, and what experts are saying.

Information about the topic is found by accessing secondary data from IEEE Xplore, ScienceDirect, and Google Scholar. These sources are chosen because they are applicable, trustworthy, and have access to all peer-reviewed articles [13]. The paper concentrated on works from the last 4 years to incorporate the most current information about AI, ML, and data tagging in manufacturing.

Thematic analysis is carried out on the data that are selected from these mentioned sources. Keywords such as labeling datasets, labeling data using machine learning, integrating labeling in real time, and the results for making predictions are gathered from reading important literature [14].

4. Data analysis

4.1. Thematic Analysis

4.1.1. Theme 1: Using correct and automated labeling boosts the performance of predictive models

Precise and automated labeling makes ML models used in manufacturing work with greater accuracy and speed. The use of AI and ML in predictive maintenance, quality inspection, and improving processes counts on having a lot of accurate data. It is easy for labels created manually to be inconsistent or wrong and difficult to increase in large amounts [15]. The help of active learning, semi-supervised learning, and natural language processing, automated data labeling fixes the issues mentioned above [16]. Using good labels helps predictive models to notice patterns, lessen distractions, and give more accurate warnings for problems such as failure of machines, poor product quality, or breaking of procedures.

This theme addresses the gap by demonstrating that automating data labeling does more than simple labeling. It also improves the need for producing top ML models. Based on the correct labeling labels, the model's decisions are easier to understand and have more confidence that is very important in critical manufacturing areas [17]. Even though it is very important, there are not many studies that connect well a label is applied to actual results in real industries. This study states that good labeling practices must be included in the core workflow of any predictive system that is being adopted in production.

The data labeling process in manufacturing involves collecting real-time data from sensors, machines, and quality control systems, followed by annotating this data with meaningful labels such as “defective,” “passed,” or “requires maintenance”. This labeled data trains machine learning models to detect anomalies, predict equipment failures, and optimize production processes [18]. The data labeling process supports smarter decision-making, reduces downtime, and enhances product quality, making it a critical enabler for Industry 4.0 and digital transformation in manufacturing.

4.1.2. Theme 2: Advancements in Real-Time and AI-Assisted Labelling for Intelligent Manufacturing

Introducing AI into data labeling helps a lot in developing intelligent manufacturing systems [19]. The person working on a shop can see data storing in nonstop and in many forms, such as results from sensors, records of inspections, and reports on problems and maintenance [20]. Sensors, IoT equipment, PLCs, and MES pick up in real time temperature, pressure, machine performance rates, and changes in voltages. Data circulates through the networks using Ethernet/IP, Modbus to the historians and MES. Data is handled on-site by edge computing, while cloud platforms offer the option to check and analyze information remotely. Timely labeling of data helps factories respond straight away to unexpected situations and base their decisions on reliable information. AI technologies such as machine vision, robotic process automation, and cognitive computing are equipped to interpret hard-to-understand patterns and give correct labels mostly without help from humans.

4.2. Real-Life Examples

The data labeling effect of AI/ML is observed in real life through the following cases in manufacturing. **Siemens** is one such company that adopts AI to conduct predictive maintenance by tagging patterns of machine behaviors, thereby lowering downtimes and working at a higher efficiency level [21]. **Tesla** uses an automated process of labeling images to make quality checks, and it allows optimizing the process based on detecting defects in real-time [22]. Implementation of intelligence labeling to increase productivity, precision, and decision-making capabilities indicates that integration of AI into modern manufacturing systems will be beneficial.

This theme brings the concepts together by showing the real-time labeling that helps put predictive systems and autonomous thinking into practice. Traditional batch-processing cannot adjust model predictions constantly, unlike real-time labeling, that makes the feedback loop happen continuously. Failure patterns are tagged as they appear, AI models can respond very quickly with suggestions to overcome the issue or put steps in place to prevent it. There has not been much study on AI that can make it possible for independent labeling in various changing workplaces. This research has proven that having intelligent labels in IoT systems encourages independent automated insight creation that helps achieve better flexibility and secure operations in manufacturing.

5. Result

The themes are analyzed that AI/ML-driven data labeling has a major effect on forecasting in manufacturing. Using the themes of precise and automated data annotation along with instant AI-backed labeling. It is explained in this research that the ability to correctly and speedily label data makes a big difference in the performance of models. Many studies have shown that ML models use special data well and are better at detecting issues related to manufacturing failures, unexpected changes in quality, and when maintenance is required [23]. Tools such as active learning and weak supervision now replace much of the manual work, so labeling can help larger amounts of data. They suggest that automation not only helps with more labels but also guarantees accuracy that makes AI models stronger. As a result, organizations can make better decisions faster and avoid making many mistakes during their work [24]. It reveals that better data labeling plays a major role in making predictive modelling in smart manufacturing successful.

The second theme highlights the importance of quick and AI-supported ways of labeling data for smarter and more flexible manufacturing systems. Moving to Industry 4.0 it will be easy to execute the labeling process of data. Emerging AI-driven labeling systems such as machine vision systems, real-time sensors, and automatic logging equipment with day-to-day procedures raises the adaptability of these systems at the time of addressing troublesome situations [25]. Bringing artificial intelligence to Forex means that predictions can change constantly and fast, so minimal human input is needed. As a result, companies can operate more proactively, improving the way things are done and making them much safer and more efficient.

Having accurate data labels while training AI/ML models benefits the manufacturing sector in areas of prediction. The main point is that these findings focus on a little-studied link between unsupervised data labeling and the quality of a model's predictions [26]. Using AI-powered self-labeling allows manufacturing to rely on smart and flexible decision-making.

6. Discussion

This research adds to the discussion about AI/ML in the field of manufacturing by examining in detail the effect of data labeling on better predictions. While the use of AI/ML in operations and prediction of failures is acknowledged, most literature has missed the key fact that accurate data labeling in real time is vital for everything to work smoothly. The research investigates automation and real-time AI assistance for labeling models that can greatly affect the performance of manufacturing systems in predicting data.

The analysis for this study relies on two main insights into its main themes. The data labeling process has been discussed in the theme one to make it understandable [27]. It starts by proving that high-quality predictive models are achieved with good and accurate labels for datasets. Previously, considered labeling only as a preparation for later procedures, but the current study views it as an actual working approach. Through combining several points, this analysis makes it clear that using automation for labeling increases speed, consistency of the data, and the model's reliability [28]. This view is different and fits well in factories that need to make large and diverse products, because manual labeling becomes very difficult. Programming languages such as Python provide robust tools and frameworks for implementing automated labeling processes efficiently.

The study goes further by making real-time automated labeling the main tool for flexible decision-making. Many studies talk about predictive maintenance and quality checks as two different categories. The study shows that it is now possible to label issues, anomalies, or failures in real time so that the data can help automated systems predict future events instantly [29]. Although it is not widely used, this style of control loop offers a lot of value in the Manufacturing setting. It helps develop systems that can spot, study, and react to repeating patterns all by themselves. As a result, people can view AI deployment in manufacturing as connected to everything from data creation and labeling to prediction and decision-making, not only as a tool for predicting outcomes. It proposes a way for labeling to be integrated into AI-based manufacturing. This solves the problem mentioned in the literature and forms a valid base for further research and use in factories.

6.1. Future directions

The research should focus on bringing together self-supervised and federated learning to improve automated data labeling in manufacturing sectors. As a result, these advanced strategies require less labelled data, but they do not lose in performance or privacy within manufacturing units. It is important to constantly review data provided by sensors, cameras, and machines to develop labeling systems that can change with the equipment's status. Developing automation in labeling that takes corrective feedback from users could lead to progress in improving models' performance [30]. It is possible to configure these platforms so that they handle many types of data, including texts, images, and different forms of text, that helps them cover many areas of manufacturing.

Using explainable AI helps increase trust at the time it comes to decision-making in highly important manufacturing settings. Considering that production lines and systems include real-time, AI-driven labeling, industries gain the power to make instant, autonomous choices that aids in developing productive, quick-response, predictive manufacturing solutions, ultimately advancing Industry 4.0.

7. Conclusion

The research investigated the major role labeling of data plays in releasing the power of AI and ML for the manufacturing sector. It is found, from a review of secondary material, that reliable, automatic, and ongoing data labeling improves quality control, maintenance forecasting, and operational efficiency. According to the review of topics, labeling forms a key aspect in the creation of accurate and smart AI systems. By focusing attention on the topic's existing knowledge gap, this research clearly shows that real time automatic data labeling process is needed and should be included in every system with the rise of automation and data-driven activities in industries. Data labeling is essential for building smart manufacturing, since it supports quick and accurate decision-making and brings constant advancements to production.

References

- [1] Vatansever, S., Schlessinger, A., Wacker, D., Kaniskan, H. Ü., Jin, J., Zhou, M. M., and Zhang, B. (2021). Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Medicinal research reviews*, 41(3), 1427-1473.

- [2] Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., and Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, 13(12), 1216.
- [3] Plathottam, S. J., Rzonca, A., Lakhnori, R., and Iloeje, C. O. (2023). A review of artificial intelligence applications in manufacturing operations. *Journal of Advanced Manufacturing and Processing*, 5(3), e10159.
- [4] Agrawal, R., Majumdar, A., Kumar, A., and Luthra, S. (2023). Integration of artificial intelligence in sustainable manufacturing: current status and future opportunities. *Operations Management Research*, 16(4), 1720-1741.
- [5] Munappy, A. R., Bosch, J., Olsson, H. H., Arpteg, A., and Brinne, B. (2022). Data management for production quality deep learning models: Challenges and solutions. *Journal of Systems and Software*, 191, 111359.
- [6] Xu, J., Kovatsch, M., Mattern, D., Mazza, F., Harasic, M., Paschke, A., and Lucia, S. (2022). A review on AI for smart manufacturing: Deep learning challenges and solutions. *Applied Sciences*, 12(16), 8239.
- [7] Zhou, L., Zhang, L., and Konz, N. (2022). Computer vision techniques in manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(1), 105-117.
- [8] Desmond, M., Duesterwald, E., Brimijoin, K., Brachman, M., and Pan, Q. (2021, August). Semi-automated data labeling. In *NeurIPS 2020 competition and demonstration track* (pp. 156-169). PMLR.
- [9] Zha, D., Bhat, Z. P., Lai, K. H., Yang, F., and Hu, X. (2023). Data-centric ai: Perspectives and challenges. In *Proceedings of the 2023 SIAM international conference on data mining (SDM)* (pp. 945-948). *Society for Industrial and Applied Mathematics*.
- [10] Linaza, M. T., Posada, J., Bund, J., Eisert, P., Quartulli, M., Döllner, J., ... and Lucat, L. (2021). Data-driven artificial intelligence applications for sustainable precision agriculture. *Agronomy*, 11(6), 1227.
- [11] Fredriksson, T. (2023). *Opportunities, Challenges and Solutions for Automatic Labeling of Data Using Machine Learning*. Chalmers Tekniska Högskola (Sweden).
- [12] Mazzei, D., and Ramjattan, R. (2022). Machine learning for industry 4.0: a systematic review using deep learning-based topic modelling. *Sensors*, 22(22), 8641.
- [13] Akbarighatar, P., Pappas, I., and Vassilakopoulou, P. (2023). A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. *International Journal of Information Management Data Insights*, 3(2), 100193.
- [14] Gladysz, B., Tran, T. A., Romero, D., van Erp, T., Abonyi, J., and Ruppert, T. (2023). Current development on the Operator 4.0 and transition towards the Operator 5.0: A systematic literature review in light of Industry 5.0. *Journal of Manufacturing Systems*, 70, 160-185.
- [15] Cerquitelli, T., Pagliari, D. J., Calimera, A., Bottaccioli, L., Patti, E., Acquaviva, A., and Poncino, M. (2021). Manufacturing as a data-driven practice: methodologies, technologies, and tools. *Proceedings of the IEEE*, 109(4), 399-422.
- [16] Tomar, B., Kumar, N., and Sreejeth, M. (2023). Real Time Automation and Ratio Control Using PLC & SCADA in Industry 4.0. *Computer Systems Science & Engineering*, 45(2).
- [17] Engin, D., Engin, M., Pinar, S. and Candan, M., (2021). PLC and SCADA Based Real-Time Monitoring and Control of Networked Processes. *J. Multidiscip. Eng. Sci. Technol*, 8, pp.14070-14080.
- [18] Enemosah, A., and Chukwunweike, J. (2022). Next-Generation SCADA Architectures for Enhanced Field Automation and Real-Time Remote Control in Oil and Gas Fields. *Int J Comput Appl Technol Res*, 11(12), 514-29.
- [19] Kumar, V. (2023). Digital enablers. In *The economic value of digital disruption: A holistic assessment for CXOs* (pp. 1-110). *Singapore: Springer Nature Singapore*.
- [20] Bhaskaran, P. E., Maheswari, C., Thangavel, S., Ponnibala, M., Kalavathidevi, T., and Sivakumar, N. S. (2021). IoT Based monitoring and control of fluid transportation using machine learning. *Computers & Electrical Engineering*, 89, 106899.
- [21] Demir, I. B. (2023). Artificial Intelligence for predictive maintenance.
- [22] Weiss, E., Caplan, S., Horn, K., and Sharabi, M. (2024). Real-time defect detection in electronic components during assembly through deep learning. *Electronics*, 13(8), 1551.
- [23] Whang, S. E., Roh, Y., Song, H., and Lee, J. G. (2023). Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, 32(4), 791-813.

- [24] Sivathapandi, P. K. P. (2022). Advanced AI Algorithms for Automating Data Preprocessing in Healthcare: *Optimizing Data Quality and Reducing Processing Time*.
- [25] Mathew, D., Brintha, N. C., and Jappes, J. W. (2023). Artificial intelligence powered automation for industry 4.0. In *New horizons for Industry 4.0 in modern business* (pp. 1-28). Cham: Springer International Publishing.
- [26] Ko, H., Witherell, P., Lu, Y., Kim, S., and Rosen, D. W. (2021). Machine learning and knowledge graph based design rule construction for additive manufacturing. *Additive Manufacturing*, 37, 101620.
- [27] Plathottam, S. J., Rzonca, A., Lakhnori, R., and Iloeje, C. O. (2023). A review of artificial intelligence applications in manufacturing operations. *Journal of Advanced Manufacturing and Processing*, 5(3), e10159.
- [28] Khalil, R. A., Saeed, N., Masood, M., Fard, Y. M., Alouini, M. S., and Al-Naffouri, T. Y. (2021). Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet of Things Journal*, 8(14), 11016-11040.
- [29] Adams, D., and Krulicky, T. (2021). Artificial Intelligence-driven Big Data Analytics, Real-Time Sensor Networks, and Product Decision-Making Information Systems in Sustainable Manufacturing Internet of Things. *Economics, Management & Financial Markets*, 16(3).
- [30] Gärtler, M., Khaydarov, V., Klöpffer, B., and Urbas, L. (2021). The machine learning life cycle in chemical operations—status and open challenges. *Chemie Ingenieur Technik*, 93(12), 2063-2080.