

Machine Learning for Anomaly Detection in Industrial Environments [†]

Denitsa Grunova, Vasiliki Bakratsi, Eleni Vrochidou  and George A. Papakostas ^{*} 

MLV Research Group, Department of Informatics, Democritus University of Thrace, 65404 Kavala, Greece; grunovadeni@gmail.com (D.G.); vabakra@teiemt.gr (V.B.); evrochid@cs.duth.gr (E.V.)

^{*} Correspondence: gpapak@cs.duth.gr; Tel.: +30-2510462320

[†] Presented at the International Conference on Electronics, Engineering Physics and Earth Science (EEPES'24), Kavala, Greece, 19–21 June 2024.

Abstract: In modern industry, anomaly detection is an important part of safety and productivity management. Early anomaly detection could allow for timely interventions, preventing malfunctions and reducing risks for human workers and machines. This work aims to deliver an overview of the use of machine learning for anomaly detection in industrial environments, highlight the state-of-the-art, and discuss challenges and prospects for future research. Existing approaches, methodologies, and results related to anomaly detection are summarized, focusing on the application of machine learning for different types of industrial anomalies. Research findings indicate that, despite the current advances, there is still room for improvements and developments in machine learning-based anomaly detection in industrial environments, designating an important future field of research.

Keywords: machine learning; anomaly detection; industrial environments; fault detection



Citation: Grunova, D.; Bakratsi, V.; Vrochidou, E.; Papakostas, G.A. Machine Learning for Anomaly Detection in Industrial Environments. *Eng. Proc.* **2024**, *70*, 25. <https://doi.org/10.3390/engproc2024070025>

Academic Editor: Grigor Mihaylov

Published: 7 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The constantly evolving industrial technology has resulted in changes in production processes, increasing the need for effective methods to detect anomalies. Anomaly detection is the process of finding efficient ways to discover irregular values in a data set. Accident prevention is one of the most important objectives of condition monitoring and anomaly detection in industries. Anomaly detection and prediction can ensure the reliability and safety of industrial processes by establishing hazardous conditions notification systems for shutdown or control [1], thus ensuring the safety of human workers and saving costs of repair or replacement of machines due to malfunctions. However, anomalies are usually rare events in a dataset, so most existing algorithms cannot detect anomalies with (over) confidence and may lead to false alarms or false positives [2,3].

Machine-learning techniques have proven to be effective for anomaly detection, being able to uncover unusual patterns in data [4]. In order to select and apply the appropriate machine learning-based anomaly detection technique, a number of factors, such as the type of the generated sensory data stream, the type of anomaly, and the availability of training data, should be considered [5]. The most common problem in industrial research is the detection of anomalies and abnormal activity in the network [6]. In such cases, anomaly detection systems are automated security systems used to monitor, analyze, and detect anomalous activity on a network or a host computer. Lee et al. [7] reported that there are four key elements that must be considered when building an anomaly or intrusion detection system: the resource to be protected, a model to determine the typical behavior of the resource, a technique to compare the activity of the resource with healthy behavior, and a technique to determine what is abnormal. Therefore, anomaly detection is used in various applications, e.g., for cyber intrusions, in health systems, for crime investigation, etc. [8]. Many machine-learning techniques have been used and proposed for the latter applications, and they are used separately as well as in combination [9].

To this end, this work presents a comprehensive overview of the literature on machine-learning applications in industrial anomaly detection. By examining a wide range of different approaches, methodologies, and applications, this work aims to provide a holistic view of the current state of research in this area, including the different types of anomalies, machine-learning algorithms, and their performance metrics, toward comparatively evaluating anomaly detection systems. This work also examines how various machine-learning technologies are integrated into industrial environments for anomaly detection. In addition to systematically presenting research in this area, this work attempts to contribute to the body of existing knowledge by answering the following three specific research questions (RQs):

- RQ1: What are the latest scientific publications in the field of anomaly detection in industrial environments?
- RQ2: Which machine-learning models are most effective in detecting anomalies in industrial environments?
- RQ3: How does the application of machine learning in anomaly detection impact the safety of industrial environments?

The rest of this paper is structured as follows. Section 2 provides a brief review of related works. Section 3 presents the conducted research methodology, while Section 4 provides the research findings, followed by the discussions in Section 5. Finally, Section 6 concludes this paper.

2. Related Work

In order to better highlight the contribution of this work, an overview of related review works is conducted in the following.

Chandola et al. [10] provided an extensive review of several anomaly detection techniques in multiple application domains. Sodemann et al. [11] presented a survey on anomaly detection in automated surveillance, focusing on different classification models and algorithms. Zuo [12] described the three most widely used anomaly detection techniques in geochemistry. Frank et al. [13] implemented a review of general technologies for smart manufacturing in the frame of Industry 4.0, including anomaly detection.

Umer et al. [14] focused on machine learning-based industrial control against cyber-attacks, reviewing four types of machine-learning methods. Sokolov et al. [15] analyzed the problem of anomaly detection in modern industrial control systems that are vulnerable to cyber-attacks; machine-learning models to detect anomalies were evaluated, ranging from traditional to deep-learning methods. Pang et al. [16] made a comprehensive review of deep anomaly detection with a comprehensive taxonomy, covering a wide range of application categories and methods. Erhan et al. [17] reviewed the state-of-the-art anomaly detection methods in the specific area of sensor systems.

However, it is important to highlight that the existing literature lacks a comprehensive review of machine learning-based techniques for anomaly detection that exclusively focuses on industrial applications. This review aims to establish a framework for advancing theoretical knowledge and development in this field. Additionally, it seeks to pinpoint knowledge gaps and areas within industrial machine learning-based anomaly detection that remain underexplored.

3. Methodology

The first phase was conducted according to the principles of the Preferred Reference Data for Systematic Reviews and Meta-Analysis (PRISMA) statement [18]. The PRISMA methodology appropriately defined all eligibility criteria for the collection of research material, the use of information sources, the elimination of duplication, data collection procedures, and, finally, the integration of the results (Figure 1).

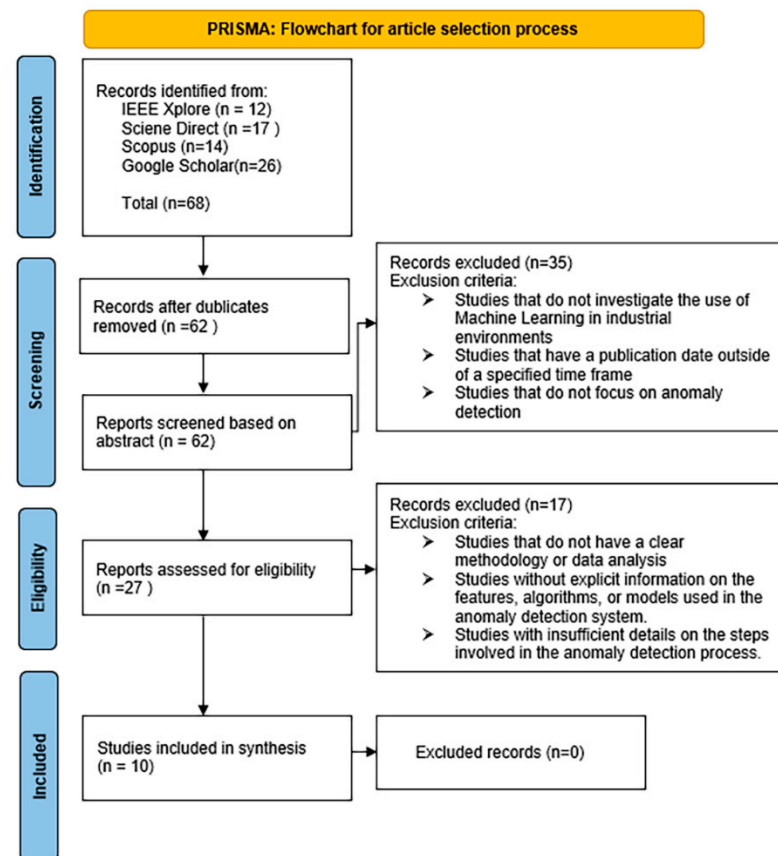


Figure 1. Research methodology.

The literature research was conducted for relevant research works in four online databases, namely, IEEE Xplore, Scopus, Science Direct, and Google Scholar. Various search terms such as “anomaly detection”, “industrial environment”, “machine learning”, and “fault detection” were used to ensure broad and saturated coverage of the relevant literature.

General inclusion and exclusion criteria for the studies included in this review were defined as follows, also considering the posed research questions:

Inclusion criteria:

- The studies were published between 2015 and 2023;
- Only studies focusing on the use of machine learning for anomaly detection in industrial environments were selected;
- Only studies that have been published in scientific journals, conferences, or books were included, thus ensuring the validity of the results;
- The studies were written in English.

Exclusion criteria:

- Lack of machine learning: studies that do not use machine learning and focus only on classical detection methods were excluded;
- Studies that did not have sufficient details about the used machine-learning algorithms, the used datasets, or reported evaluation metrics were excluded, ensuring that only studies with sufficient transparency and reproducibility were considered.
- To avoid redundancy in this review, studies that duplicated articles or presented substantially similar findings were excluded.

As a result of this review of the literature, a total of ten articles met our inclusion criteria and were further considered.

4. Results

The research findings of this review are presented in Table 1, aiming to answer the posed research questions. Table 1 includes all the information from the selected literature, the latest scientific publications, and the best-performing machine-learning algorithms, indicating a great diversity in the performance of machine-learning models in anomaly detection in industrial environments. Note that each model performs optimally under specific conditions, which reinforces the importance of adapting to the specific requirements of each industrial environment. In the context of evaluating machine-learning algorithms, there are several metrics used to compare results between different approaches, while the most common is accuracy. In particular, the time needed to create the model is of great importance, as it reflects the time needed to train the model. Table 1 also includes additional metrics for the deeper evaluation of the models' performance.

Table 1. Research results on machine learning-based anomaly detection in industrial environments.

Ref. (Year)	Dataset	Machine-Learning Model	Precision	Recall	F1-Score	Accuracy	Other
[18] (2021)	HIL ICS [19]	k-NN	97.32	97.29	97.29	97.29	AUC: 97.29%, Fitting time: 173 s, Prediction time: 104 s
		DT	99.37	99.37	99.37	99.37	AUC: 99.37%, Fitting time: 5.8 s, Prediction time: 0.0283 s
		RF	99.76	99.76	99.76	99.76	AUC: 99.76%, Fitting time: 2.21 s, Prediction time: 0.0505 s
[20] (2021)	Steel plate faults by Semeion. Research of Sciences of Communication [21]	DT	91.29	-	91.29	91.14	Sensitivity: 91.14%,
		k-NN	82.86	-	82.86	82.86	Sensitivity: 82.86%
		RF	93.86	-	92.43	93.29	Sensitivity: 93.29%
		SVM	74.57	-	79.43	86.00	Sensitivity: 86.00%
		Naïve Bayes	89.29	-	62.86	59.00	Sensitivity: 59.00%
		Logistic Regression (LR)	86.71	-	83.71	88.29	Sensitivity: 88.29%
		Multilayer Perceptron (MPL)	88.43	-	73.43	73.86	Sensitivity: 73.86%
[22] (2023)	Biz	DT	69.06	73.19	-	-	F 0.5 Score: 69.85%
		RF	87.44	70.80	-	-	F 0.5 Score: 83.51%
		k-NN	74.01	62.90	-	-	F 0.5 Score: 73.00%
		GBDT	83.03	62.06	-	-	F 0.5 Score: 77.77%
		Stacking	88.03	70.17	-	-	F 0.5 Score: 83.76%
	Mon	DT	57.07	66.60	-	-	F 0.5 Score: 58.75%
		RF	87.92	64.27	-	-	F 0.5 Score: 81.89%
		k-NN	73.23	53.73	-	-	F 0.5 Score: 68.27%
		GBDT	85.68	43.39	-	-	F 0.5 Score: 71.70%
		Stacking	87.54	69.61	-	-	F 0.5 Score: 83.25%
	Ora	DT	44.82	57.31	-	-	F 0.5 Score: 46.86%
		RF	77.99	49.01	-	-	F 0.5 Score: 69.74%
		k-NN	56.58	25.49	-	-	F 0.5 Score: 45.48%
		GBDT	61.33	21.94	-	-	F 0.5 Score: 45.13%
		Stacking	85.11	56.53	-	-	F 0.5 Score: 77.29%
	Trd	DT	41.09	58.87	-	-	F 0.5 Score: 43.73%
		RF	86.21	53.19	-	-	F 0.5 Score: 76.69%
		k-NN	81.03	33.33	-	-	F 0.5 Score: 63.00%
		GBDT	72.82	53.19	-	-	F 0.5 Score: 67.81%
		Stacking	85.42	51.90	-	-	F 0.5 Score: 75.64%
[23] (2023)	NSL-KDD [24]	IF	-	87.00	78.00	99.00	-
		SVM	-	88.00	67.00	95.00	-
[25] (2018)	Data from press-hardening processes	ANN	100	100	-	100	-
		SC-SVM	98.90	100	-	99.40	-
		IF	99.00	100	-	99.50	-

Table 1. Cont.

Ref. (Year)	Dataset	Machine-Learning Model		Precision	Recall	Performance Metrics (%)		Other
						F1-Score	Accuracy	
[26] (2020)	Pharmaceutical company data		RFA DLA	99.97 99.96	99.97 99.97	99.97 99.96	- -	- -
[27] (2019)	Modbus (D1) [28], OPC UA (D2) [29]	SVM	D1	-	-	-	92.53	Execution time: 11,712 s
			D2	-	-	-	90.81	Execution time: 0.019 s
		RF	D1	-	-	-	99.84	Execution time: 281 s
			D2	-	-	-	99.98	Execution time: 52.31 s
[30] (2017)	SWaT [31]		DNN	98.29	67.84	80.28	-	-
			SC-SVM	92.50	69.90	79.62	-	-
[32] (2021)	WSN-DS [33]		SVM	88.00	92.00	90.00	89.00	-
			Naïve Bayes	94.00	85.00	88.00	94.00	-
			DT	94.00	94.00	93.00	94.00	-
			ID-GOPA	96.00	96.00	96.00	96.00	-
			RF	94.00	85.00	88.00	94.00	-
[34] (2020)	HIL-based augmented ICS dataset		Naïve Bayes	-	-	-	54.00	Training time: 0.1 s; Prediction time: 0.1 s
			RF	-	-	-	82.93	Training time: 0.9 s; Prediction time: 0.1 s
			RF GSCN	-	-	-	82.93	Training time: 109.8 s; Prediction time: 8.2 s
			GB	-	-	-	77.58	Training time: 583.02 s; Prediction time: 0.1 s
			GB GSCV	-	-	-	83.63	Training time: 1274.2 s; Prediction time: 10 s
			ANN	-	-	-	82.79	Training time: 76 s; Prediction time: 10 s
			LSTM	-	-	-	82.81	Training time: 111 s; Prediction time: 20 s
			LSTM AE	-	-	-	82.79	Training time: 809 s; Prediction time: 10 s

Mokhtari et al. [18] focused on using the ICS dataset (HIL) for anomaly detection in industrial systems. They used k-NN, Decision Tree (DT), and Random Forest (RF) models. RF outperformed the other models, with accuracy, recall, F1 score, and precision reaching 99.76%. Fit and prediction times were relatively low, indicating efficient model performance. Overall, the RF model showed excellent anomaly detection capabilities on the ICS (HIL) dataset. Gamal et al. [19] analyzed seven machine-learning models in their study, focusing on a dataset of steel plate defects. The performance of the models was measured, and the DT and RF achieved accuracy rates of 91.14% and 93.29%, respectively. RF showed higher accuracy and repeatability compared to support vector machines (SVM) and Naive Bayes, which indicated lower accuracies. This analysis highlights the importance of the appropriate selection of machine-learning models considering the characteristics of each industry dataset.

The research of Wang et al. [20] focused on the analysis of system log files of server clusters in a financial company using various models, including DT, RF, k-NN, and gradient-based decision trees (GBDT), over four different datasets. The “stacking” method stood out with high accuracy, repeatability, and F 0.5 score values above 88%. The results indicated that methods such as “stacking” could be very effective in detecting task-related anomalies in financial IT systems.

The study conducted by Shanthi and Maruthi [21] focused on the NSL-KDD dataset and investigated the use of Isolation Forest (IF) and SVM for anomaly detection; compared to SVM, IF recorded an impressive accuracy of 99%. Results demonstrate the effectiveness of unsupervised learning approaches, such as IF, for anomaly-based intrusion detection systems in the NSL-KDD dataset. This work overall indicates that machine-learning models

constitute effective tools for anomaly detection and are able to deal with the challenge of handling complex data.

Lejon et al. [22] analyzed the performance of artificial neural networks (ANNs), single-class SVMs (SC-SVM), and IF, focusing on the use of declassified data; ANNs displayed excellent accuracy, recall, and F1-Score values, while IF and SC-SVMs showed high accuracy and recall. The results demonstrate the reliability and effectiveness of the selected models to accurately detect anomalies in the used private dataset collected from the press-hardening process, as well as the potential to implement machine learning for anomaly detection by non-experts in machine learning using specific programming libraries.

Quatrini et al. [23] focused on a pharmaceutical dataset and evaluated the performance of the Random Forest Algorithm (RFA) and the Decision Linkage Algorithm (DLA). Both algorithms showed high accuracy and recall values for anomaly detection in a private dataset of the pharmaceutical industry. Results highlight the effectiveness of both methods in anomaly detection and the need to apply them in practice in industrial settings. High-performance results reaching a precision and recall of more than 99% were reported, indicating the potential of the proposed methods to detect warnings and critical circumstances in the production phases.

Anton et al. [24] used data from two datasets, Modbus and OPC UA, in order to predict attacks and employed SVM and RF algorithms. The analysis showed that SVM outperformed the first dataset but required more runtime. RF performed better on the second dataset, with higher accuracy and faster runtime. The fact that each model behaves differently depending on the used dataset indicates the need for balance between accuracy and computational efficiency when identifying industrial anomalies. Machine learning-based anomaly detection can be highly efficient in industry since industrial settings produce plenty of data that can be used for training.

Inoue et al. [25] used the SWaT dataset and evaluated the performance of Deep Neural Networks (DNNs) and single-class SVMs in anomaly detection. The results showed that DNNs exhibited higher accuracy and repeatability than single-class SVMs, highlighting their effectiveness in anomaly detection in industrial systems, as well as the importance of selecting the appropriate model according to the application requirements.

Ifzarne et al. [26] conducted a study on anomaly detection using the WSN-DS dataset and SVM, Naïve Bayes, DT, wireless sensor network (WSN) intrusion detection model based on information gain ratio and online passive-aggressive algorithm (ID-GOPA) and RF. ID-GOPA outperformed the other models, revealing that it can work better on large datasets as its learning rate does not decrease over time. For the same reason, due to the large dataset, SVM displayed a lower accuracy since it performs better for small datasets.

Finally, Tai et al. [27] used an augmented ICS dataset and evaluated several machine-learning models, such as Gaussian Naïve Bayes, RF, RF GridSearchCV (RF GSCV), Gradient Boosting Machine (GB), GB GCSV, ANN, Long Short-Term Memory (LSTM), and LSTM autoencoder (LSTM AE). Although the results indicated that RF, GB, ANN, and LSTM classification models had profound potential for anomaly detection in industrial control systems, RF GSCV, i.e., with tuned hyperparameters, slightly outperformed the rest of the models.

5. Discussion

The analysis of the above academic papers provides an overview of machine learning-based anomaly detection in industrial environments and concludes that RFs, k-NNs, DTs, SVMs, and ANNs are the most commonly used machine-learning models that perform well. It should be noted that the choice of the appropriate model depends on the characteristics of the problem and the used dataset. The use of predictive models can improve safety in industrial settings by early detection of anomalies and, thus, timely address of the challenges inherent in the industry.

Comparing the results of the selected studies, it is clear that the choice of the dataset has a significant impact on the performance of the selected model. This review revealed

that certain algorithms can perform better on smaller and larger datasets. Model runtime is also an important factor, especially in industrial environments where the trade-off between accuracy and performance becomes apparent. Each study emphasizes the importance of selecting a model tailored to the specific characteristics of both the dataset and the problem under study. Reported results reveal that there is still potential for improvement in the field of machine learning-based anomaly detection in the industry, suggesting that this is an important scientific area of research in the future.

6. Conclusions

Anomaly detection has been an active research field for years and has attracted the interest of researchers from various fields. Detecting abnormal behaviors of systems could help reduce operational risks and prevent unseen problems and system downtime. This work presents the results of a systematic literature review regarding the use of machine-learning models in anomaly detection in industrial settings. This review provides a comprehensive overview of the developed approaches, their characteristics, and performance indicators. Consequently, summarization of the latter information is expected to guide the research community toward enhanced knowledge of the latest approaches and methodologies developed in this specific research field. Anomalies in data streams can vary widely in different domains, from healthcare to industrial environments; therefore, it is crucial to develop anomaly detection methods tailored to specific domains and environments. Researchers should explore the specific challenges posed to each particular domain and develop adaptive approaches that can effectively identify anomalies relevant to each context.

The major and fast developments in data integration technologies make machine-learning algorithms a powerful tool in industrial environments for gaining useful insights regarding the behavior of industrial systems and thus improving the efficiency of timely decision-making. Future works should also consider using the highest-performed machine-learning algorithms, as concluded from this review work, and test them to multiple industrial applications. Moreover, additional machine-learning or hybrid methods need to be developed to enhance prediction accuracies. More specifically, appropriate machine-learning algorithms need to be established to handle specific anomaly detection cases in the industry. Finally, more real-world public industrial datasets need to be developed, on which researchers could comparatively evaluate their methods.

Author Contributions: Conceptualization, G.A.P.; methodology, D.G. and V.B.; investigation, D.G., V.B. and E.V.; resources, D.G., V.B. and E.V.; data curation, D.G., V.B. and E.V.; writing—original draft preparation, D.G., V.B. and E.V.; writing—review and editing, E.V. and G.A.P.; visualization, G.A.P.; supervision, G.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this work.

Acknowledgments: This work was supported by the MPhil program “Advanced Technologies in Informatics and Computers”, which is hosted by the Department of Informatics, Democritus University of Thrace, Kavala, Greece.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Harrou, F.; Sun, Y.; Khadraoui, S. Amalgamation of Anomaly-Detection Indices for Enhanced Process Monitoring. *J. Loss Prev. Process Ind.* **2016**, *40*, 365–377. [[CrossRef](#)]
2. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [[CrossRef](#)]
3. Krawczyk, B. Learning from Imbalanced Data: Open Challenges and Future Directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [[CrossRef](#)]

4. Nassif, A.B.; Talib, M.A.; Nasir, Q.; Dakalbab, F.M. Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access* **2021**, *9*, 78658–78700. [\[CrossRef\]](#)
5. Musa, T.H.A.; Bouras, A. Anomaly Detection: A Survey. In *Lecture Notes in Networks and Systems*; ACM: New York, NY, USA, 2022; pp. 391–401.
6. Larriva-Novio, X.A.; Vega-Barbas, M.; Villagra, V.A.; Sanz Rodrigo, M. Evaluation of Cybersecurity Data Set Characteristics for Their Applicability to Neural Networks Algorithms Detecting Cybersecurity Anomalies. *IEEE Access* **2020**, *8*, 9005–9014. [\[CrossRef\]](#)
7. Lee, W.; Stolfo, S.J. Data Mining Approaches for Intrusion Detection. In Proceedings of the 7th USENIX Security Symposium, San Antonio, TX, USA, 26–29 January 1998.
8. Bauer, F.C.; Muir, D.R.; Indiveri, G. Real-Time Ultra-Low Power ECG Anomaly Detection Using an Event-Driven Neuromorphic Processor. *IEEE Trans. Biomed. Circuits Syst.* **2019**, *13*, 1575–1582. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Omar, S.; Ngadi, A.; Jebur, H. Machine Learning Techniques for Anomaly Detection: An Overview. *Int. J. Comput. Appl.* **2013**, *79*, 33–41. [\[CrossRef\]](#)
10. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection. *ACM Comput. Surv.* **2009**, *41*, 1–58. [\[CrossRef\]](#)
11. Sodemann, A.A.; Ross, M.P.; Borghetti, B.J. A Review of Anomaly Detection in Automated Surveillance. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2012**, *42*, 1257–1272. [\[CrossRef\]](#)
12. Zuo, R. Machine Learning of Mineralization-Related Geochemical Anomalies: A Review of Potential Methods. *Nat. Resour. Res.* **2017**, *26*, 457–464. [\[CrossRef\]](#)
13. Frank, A.G.; Dalenogare, L.S.; Ayala, N.F. Industry 4.0 Technologies: Implementation Patterns in Manufacturing Companies. *Int. J. Prod. Econ.* **2019**, *210*, 15–26. [\[CrossRef\]](#)
14. Umer, M.A.; Junejo, K.N.; Jilani, M.T.; Mathur, A.P. Machine Learning for Intrusion Detection in Industrial Control Systems: Applications, Challenges, and Recommendations. *Int. J. Crit. Infrastruct. Prot.* **2022**, *38*, 100516. [\[CrossRef\]](#)
15. Sokolov, A.N.; Pyatnitsky, I.A.; Alabugin, S.K. Research of Classical Machine Learning Methods and Deep Learning Models Effectiveness in Detecting Anomalies of Industrial Control System. In Proceedings of the 2018 Global Smart Industry Conference (GloSIC), Chelyabinsk, Russia, 13–15 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
16. Pang, G.; Shen, C.; Cao, L.; Hengel, A. Van Den Deep Learning for Anomaly Detection. *ACM Comput. Surv.* **2022**, *54*, 1–38. [\[CrossRef\]](#)
17. Erhan, L.; Ndubuaku, M.; Di Mauro, M.; Song, W.; Chen, M.; Fortino, G.; Bagdasar, O.; Liotta, A. Smart Anomaly Detection in Sensor Systems: A Multi-Perspective Review. *Inf. Fusion* **2021**, *67*, 64–79. [\[CrossRef\]](#)
18. Mokhtari, S.; Abbaspour, A.; Yen, K.K.; Sargolzaei, A. A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data. *Electronics* **2021**, *10*, 407. [\[CrossRef\]](#)
19. Shin, H.K.; Lee, W.; Yun, J.H.; Kim, H.C. HAI 1.0: HIL-Based Augmented ICS Security Dataset. In Proceedings of the CSET 2020—13th USENIX Workshop on Cyber Security Experimentation and Test, Co-Located with USENIX Security 2020, Online, 10 August 2020.
20. Gamal, M.; Donkol, A.; Shaban, A.; Costantino, F.; Di Gravio, G.; Patriarca, R. Anomalies Detection in Smart Manufacturing Using Machine Learning and Deep Learning Algorithms. In *Proceedings of the Proceedings of the International Conference on Industrial Engineering and Operations Management, Dhaka, Bangladesh, 26–27 December 2021*; pp. 1611–1622.
21. Steel Plates Faults Data Set, Semeion, Research Center of Sciences of Communication, Via Sersale 117, 00128, Rome, Italy. Available online: <https://archive.ics.uci.edu/ml/datasets/Steel+Plates+Faults> (accessed on 7 April 2024).
22. Wang, J.; Liu, J.; Pu, J.; Yang, Q.; Miao, Z.; Gao, J.; Song, Y. An Anomaly Prediction Framework for Financial IT Systems Using Hybrid Machine Learning Methods. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 15277–15286. [\[CrossRef\]](#)
23. Shanthi, K.; Maruthi, R. Machine Learning Approach for Anomaly-Based Intrusion Detection Systems Using Isolation Forest Model and Support Vector Machine. In Proceedings of the 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 3 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 136–139.
24. Hassan Zaib, M. NSL-KDD. Available online: <https://www.kaggle.com/datasets/hassan06/nslkdd> (accessed on 7 April 2024).
25. Lejon, E.; Kyösti, P.; Lindström, J. Machine Learning for Detection of Anomalies in Press-Hardening: Selection of Efficient Methods. *Procedia CIRP* **2018**, *72*, 1079–1083. [\[CrossRef\]](#)
26. Quatrini, E.; Costantino, F.; Di Gravio, G.; Patriarca, R. Machine Learning for Anomaly Detection and Process Phase Classification to Improve Safety and Maintenance Activities. *J. Manuf. Syst.* **2020**, *56*, 117–132. [\[CrossRef\]](#)
27. Anton, S.D.D.; Sinha, S.; Dieter Schotten, H. Anomaly-Based Intrusion Detection in Industrial Data with SVM and Random Forests. In Proceedings of the 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 19–21 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
28. Morris, T.H.; Thornton, Z.; Turnipseed, I. Industrial Control System Simulation and Data Logging for Intrusion Detection System Research. In Proceedings of the 7th Annual Southeastern Cyber Security Summit, Huntsville, AL, USA, 3–4 June 2015.
29. Antón, S.D.; Gundall, M.; Fraunholz, D.; Schotten, H.D. Implementing SCADA Scenarios and Introducing Attacks to Obtain Training Data for Intrusion Detection Methods. In Proceedings of the ICCWS 2019 14th International Conference on Cyber Warfare and Security: ICCWS 2019, Stellenbosch, South Africa, 28 February–1 March 2019.

30. Inoue, J.; Yamagata, Y.; Chen, Y.; Poskitt, C.M.; Sun, J. Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning. In Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, 18–21 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1058–1065.
31. Goh, J.; Adepu, S.; Junejo, K.N.; Mathur, A. A Dataset to Support Research in the Design of Secure Water Treatment Systems. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 88–99, ISBN 9783319713670.
32. Ifzarne, S.; Tabbaa, H.; Hafidi, I.; Lamghari, N. Anomaly Detection Using Machine Learning Techniques in Wireless Sensor Networks. *J. Phys. Conf. Ser.* **2021**, *1743*, 012021. [[CrossRef](#)]
33. Almomani, I.; Al-Kasasbeh, B.; AL-Akhras, M. WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks. *J. Sens.* **2016**, *2016*, 4731953. [[CrossRef](#)]
34. Tai, J.; Alsmadi, I.; Zhang, Y.; Qiao, F. Machine Learning Methods for Anomaly Detection in Industrial Control Systems. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 2333–2339.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.