Artificial Intelligence Applications for Defect Detection in Industrial Processes: a Bibliometric Analysis

Diogo Gomes de Sousa Bezerra [1*[0009-0009-9838-0980]], Sanderson César Macêdo Barbalho [1[0000-0003-1664-4866]], Diego Cesar Florencio de Queiroz [1[0009-0006-8646-8238]], Marcelo Becker [2[0000-0002-7508-5817]] and Leonardo Ananias de Lima Felippe [3]

[1] Universidade de Brasília, Brasília, DF, Brazil
[2] Universidade de São Paulo, São Paulo, SP, Brazil
[3] Extreme Digital Solutions, São Paulo, SP, Brazil
* diogogsb@hotmail.com

Abstract. Artificial intelligence algorithms, especially machine learning and deep learning techniques, are increasingly present in our daily lives, often presenting performance evolutions and new applications. With the advent of new technologies and the so-called industry 4.0, applications in industrial and manufacturing processes are increasingly present. This work aims to stipulate a theoretical scope and trends regarding the works that emphasize AI applications in industrial and production processes. A bibliometric analysis was carried out (using VOSViewer software) on scientific publications linked to terms such as machine and deep learning, associated with applications in failure and error detection of products and procedures in industrial and manufacturing processes. The study reviewed publications from 2015 to 2022 in Scopus and Web of Science platforms.

Keywords: Deep Learning, Machine Learning, Bibliometric analysis.

1 Introduction

Artificial intelligence (AI) can be defined as the effort to automate intellectual tasks performed by humans. AI is a broad field that encompasses machine learning (ML) techniques and deep learning (DL) techniques, but it is not limited to ML and DL. It also has approaches that do not depend on learning, such as clustering techniques [1] and anomaly detection [14].

There are countless possible applications for the various existing AI techniques, many of them present in the daily lives of the vast majority of the world's population, such as product, film, and series recommendations systems, spam recognition systems in e-mails, facial recognition systems on smartphones, and others. One of the application areas that has been gaining more and more relevance in several international scenarios (Germany, United States, China, and Japan) is focused on

intelligent manufacturing systems [2]. According to Xu and Hua (2017), these systems have as their main characteristics *high correlation and deep integration, dynamic reconfiguration, and a huge volume of data.* In terms of high correlation, all systems are highly correlated, exchanging information in different layers of the process, such as manufacturing and storage/warehouse systems exchanging information. In dynamic reconfiguration, all systems are connected to create the best strategies for different products/services automatically. Finally, in terms of the large volume of data, each existing "intelligent" system generates a huge amount of data from the most diverse possible sources to be used for monitoring, transmission, processing, and storage.

Techniques encompassed by AI can be implemented at practically all manufacturing system levels, such as in crucial parts that determine the quality and success of a product, such as fault detection [3, 4] and aiding manual assembly steps [5, 15]. Thamm et al. (2021) have proposed an augmented intelligence-based worker assistance system in manual assembly.

This work aims to perform a bibliometric analysis of publications related to terms associated with Artificial Intelligence, such as Machine Learning and Deep Learning. We associated them with other terms that show applications in failure and error detection, products and procedures and industrial and manufacturing processes. Our purpose was to stipulate a theoretical scope and discover trends regarding research being carried out, as well as identify main authors, publications, countries and all relevant data for this characterization, emphasizing AI applications in industrial processes. It considers publications from 2015 to 2022 in the Scopus and Web of Science databases. The data obtained were analyzed in terms of various characteristics with the aid of the VOSViewer software for bibliometric analysis. Possible applications, case studies, trends, and research gaps on the topics in question were identified.

2 Methodology

Bibliometrics analysis can be defined as the use of mathematical and statistical methods to show quantitative and qualitative changes surrounding a certain area of research to stipulate important aspects in this area, such as trends, main works, authors, and others [13].

The advanced bibliometric analysis was based on the steps defined and specified in this section.

2.1 Databases

Initially, the databases were selected to search for scientific articles. In this work, the databases used were Scopus and Web of Science. They are platforms with large

databases of reviewed scientific publications containing works such as books, scientific journals, and conference proceedings from the most diverse scientific areas (technology, medicine, social sciences, arts, and humanities), in addition to presenting several search tools and filters.

2.2 Research Protocol

In the form that the results obtained present the most relevant and necessary data for the objective in question, a research protocol was defined to relate keywords with Boolean operators (OR and AND) to generate search sentences. These search strategies were later applied on the platforms in their respective search systems in keywords, abstracts, and titles of publications. The search phrase used was:

((recognition AND (defect OR fail)) AND ((assemb* OR manufact*) AND (lines OR production)) AND ("deep learning" OR "machine learning" AND algorithm))*

It relates aspects of fault detection in production lines from AI techniques, and the * character is interpreted as a wildcard, accepting any combination of letters after the last term. It was done to cover more related works.

Thirty-eight publications were obtained in the Scopus database and 58 in the Web of Science database. Still, as both bases share journals in common, some samples were present in both groups of results, so it was necessary to eliminate duplicate works in the sample, obtaining then 16 works in the Scopus database and 58 in the Web of Science database, totaling 74 publications. It was also defined that only works written in English would be used in the sample, which made it necessary to eliminate one more publication present in the sample from the Scopus base, because it was written in Chinese, thus totaling 73 publications. Table 1 summarizes the research protocol used.

Table 1. Publication selection protocol.

Methodological Procedures	
Database	Scopus and Web of Science
Publishing period	2015 to 2022
Boolean operators	“AND”, “OR” and “*”
Publication types	Articles, conference papers and review
Keywords Combination	
Textual content	TITLE-ABS-KEY-AUTH
Terms	"recognition", “defect”, “fail”, “assemb*”, "manufact*", "lines", ”algorithm”, "production”, "deep learning", “machine learning”
Final combination	TITLE-ABS-KEY-AUTH ((recognition AND (defect OR fail*)) AND ((assemb* OR manufact*) AND (lines OR production))

	AND ("deep learning" OR "machine learning" AND algorithm))
Portfolio of References Selected for Quantitative Analysis	
Publications found and used in the quantitative analysis:	96 results
Elimination of duplicate publication (present in both bases)s:	22 results
Elimination of publications in languages other than English and Portuguese:	1 result
Final portfolio of publications	**73 results**

2.3 Bibliometric Data

With the list of selected publications, it is possible to export various publication information from the databases, such as the number of citations, year of publication, the congress in which it was published, and various other information. In possession of these samples, it is possible to evaluate some data that describe the characteristics of the research areas under analysis. The open-source software VOSviewer was used to build and visualize bibliometric networks. It permits us to analyze the data present in the sample, which can be formed by categories such as authors, journals, publications, and others, which can be interrelated based on citations, bibliographic coupling, co-citation, and co-authorship. These data are further explored in section 3.

3 Result and Discussion

3.1 Bibliometric Analysis

This section presents and discusses the results obtained from the procedures described above. From the 73 articles obtained with the defined research protocol applied to the Scopus and Web of Science databases, it was possible to obtain relevant data to quantify the maturity and characteristics of the research area. We identified studied themes, such as main authors, countries with more publications, more impactful works, and others.

The graph below (Fig. 1) shows the number of works published in the defined period. The growth trend in these publications over the years is clear. It can be characterized as evidence that the scientific community increasingly works on studies related to the topics in question. The year 2022 had the highest number, with 32

publications. An increase of 68.42% over 2021, while the number of publications in 2021 corresponds to an increase of 58.33% in 2020.

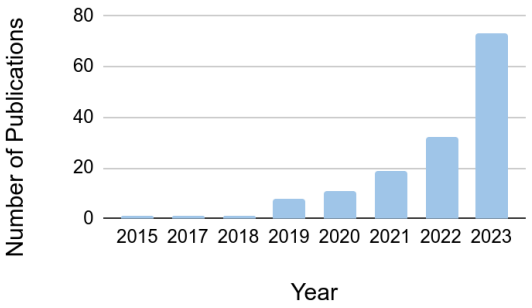

Fig. 1. Number of publications between 2015 and September 2022.

Table 2 presents the nine most cited works among the 73 works present in the sample, together with their Journal Citation Reports (JCR) index, which was created to evaluate the relevance of a journal, taking into account data such as the number of citations and papers in a journal [6].

Table 2. Most cited works, among the 74 present in the sample.

Authors	Title	Citations	Journal	JCR
Bertolini, M. et al. (2021)	Machine Learning for industrial applications: A comprehensive literature review	**58**	Expert Systems with Applications	**8.665**
Tan, SC. et al. (2015)	Evolutionary Fuzzy ARTMAP Neural Networks for Classification of Semiconductor Defects	**46**	IEEE Transactions on Neural Networks and Learning Systems	**14.255**
Wang, JL. et al. (2020)	Deformable Convolutional Networks for Efficient Mixed-Type Wafer Defect Pattern Recognition	**34**	IEEE Transactions on Semiconductor Manufacturing	**2.796**
Yang, YT. et al (2020)	A High-Performance Deep Learning Algorithm for the Automated Optical Inspection of Laser Welding	**33**	Applied Sciences	**2.838**

Gao, YP. et al (2020)	A Multilevel Information Fusion-Based Deep Learning Method for Vision-Based Defect Recognition	**26**	IEEE Transactions on Neural Networks and Learning Systems	**12.255**
O'Leary et al. (2020)	Deep Learning for Classification of the Chemical Composition of Particle Defects on semiconductor Wafers	**22**	IEEE Transactions on Semiconductor Manufacturing	**2.796**
Gao, YP et al. (2022)	A Review on Recent Advances in Vision-based Defect Recognition towards Industrial Intelligence	**20**	Journal of Manufacturing Systems	**8.498**
Xu, L. et al. (2020)	A Weakly Supervised Surface Defect Detection Based on Convolutional Neural Network	**19**	IEEE Acess	**3.476**
Bahaghighat, M; Akbari, L; Xin, Q. (2019)	A Machine Learning-Based Approach for Counting Blister Cards Within Drug Packages	**15**	IEEE Acess	**3.476**

Although the most cited publication in the sample was not the oldest, let alone the work published in the journal with the highest JCR index, this number of citations is directly linked to the fact that this work is a literature review.

Intending to identify the countries that publish the most works on the topics in question, the graph in Fig. 2 presents the number of publications per country, in which publications in 25 countries were identified. China holds the largest number of publications, about 35.71% of publications in the stipulated period, followed by Taiwan (10.71%), Germany (9.52 %), and the USA (5.95%), in the graph, "others" represent countries that contain only one publication.

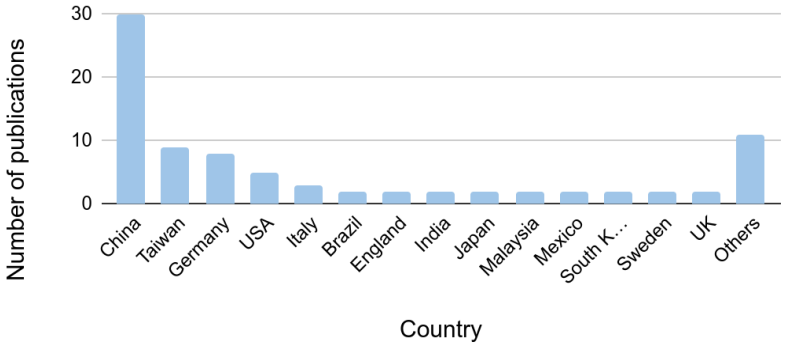

Fig. 2. Number of publications per country.

The keyword analysis presents us with possible trends for future studies by connecting with other unused keywords in our search protocol. It was possible to correlate some data present in the publications in the sample, allowing us to identify the main references, authors, and trends. The keyword co-occurrence map in the sample of 73 papers can be seen in Fig. 3, in which keywords that occur in at least two papers were selected, totaling 79. Table 3 presents the most frequent among them, with the keyword *deep learning* at the top of the list, representing about 34.17% of the total. It also highlights *convolutional neural networks* (and their similarities like *CNN, convolutional neural network,* etc.) representing 27.84% and the keyword *classification* (20.25%).

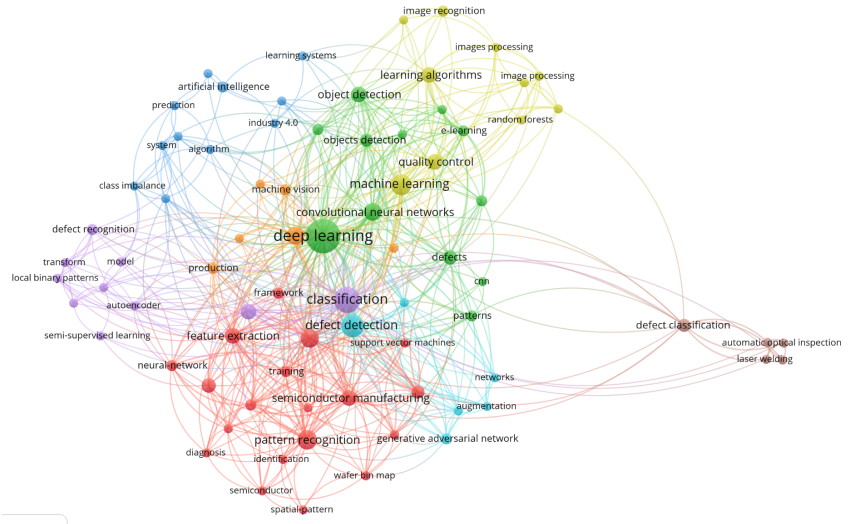

Fig. 3. Obtained keyword occurrence map.

Table 3. Most frequent keywords in the sample.

Keyword	Occurrence
Deep learning	27
Convolutional neural networks	22
Classification	16
Defect detection	12
Machine Learning	10
Pattern recognition	9
Inspection	8
Recognition	8
Feature extraction	6

The keywords co-occurrence map is composed of 6 main clusters, with the main keywords *deep learning* (green cluster), *cla*ssification (purple cluster), *defect detection* (blue cluster), *pattern recognition* (red cluster), *machine learning* (yellow cluster), and *defect detection* (brown cluster). By analyzing these clusters, we can identify the most used terms and how they relate to each other to identify the main inspirations and trends for the areas that are being analyzed, such as the keywords *augmentation*, *generative adversarial network* (GAN), and *extraction features* present in the blue cluster, linked to terms such as *clustering algorithms*, *data augmentation,* and *training* that can demonstrate the use of deep learning architectures such as GAN to generate data for supervised AI algorithms, as well as the use of clustering techniques for pattern recognition. Another case that can be observed is the presence of the keywords *surface defect recognition* and *semi-supervised learning* linked to the terms *classification*, *defect detection,* and *quality control*, which can be seen as possible application trends.

Other maps that can be created from VOSVIewer are the bibliographic coupling and co-citation maps. Bibliographic coupling establishes associations between articles based on bibliographic references present in both articles, which may take publications or authors into account. On the other hand, co-citation associates two (or more) documents or authors based on the presence of these same documents/authors in the same list of references of a publication [8].

The minimum number of documents present simultaneously in two or more works in our sample to be considered a coupling was defined according to the Pareto principle. So that 20% of the 73 publications obtained were represented. Fig. 4 shows the bibliographic coupling map obtained, in which 15 (approximately 20% of our sample) linked documents were obtained. A point to be noted is that all the most cited publications in our sample are among these 13 documents, even just three being literature reviews on the studied themes. This analysis allows us to assimilate the proximity of publications from their references on the theme, theoretical,

methodological nature, or another similar shared aspect [8,9]. It is also possible to note that the works by Bertolini, M et al. (2021) and Gao YP et al. (2022) are linked to the main clusters on the map, covering almost the entire map, which can be good evidence that these works contain good references for the topics in question.

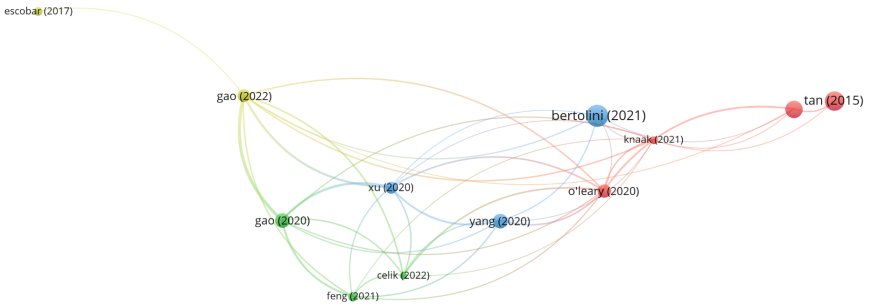

Fig. 4. Obtained bibliographic coupling map.

For creating the co-citation map of authors, it was defined that to be considered a co-citation, two or more authors must be cited simultaneously in about 20% (using the Pareto principle) of the publications in our sample. We then obtained seven co-cited authors. The map obtained can be seen in Fig. 5, where the main authors cited in our sample and their relationships are present, while Table 4 shows the most co-cited authors.

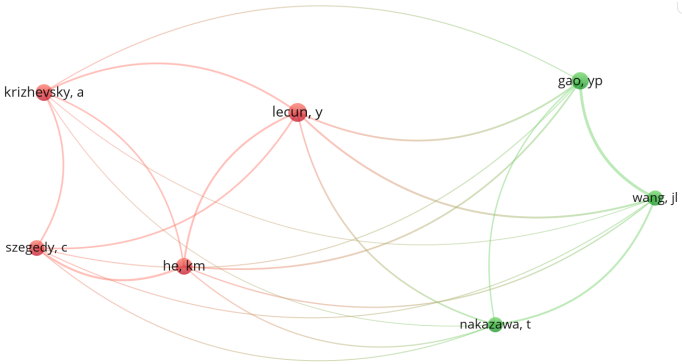

Fig. 5. Author co-citation map.

Among the most cited authors, it is possible to observe famous names in the academic area on the subjects, especially in the machine and deep learning. That is the case of Lecun, Y., who is responsible for pioneering works in deep learning, such as the

creation of CNN (Convolutional Neural Network) named LeNet [10]. He is the chief AI scientist on Facebook and Professor at New York University nowadays.

Table 4. Most co-cited authors are present in the co-citation analysis.

Author	Citations
Lecun, Y.	18
Gao, Yp.	15
He, KM.	14
Krizhevsky, A.	14
Szegedy, C.	13
Nakazawa, T.	12
Redmon, J.	11

At the same time, Redmon, J. (researcher at the University of Washington) is known for creating an object detection architecture optimized concerning inference time, popularly known as YOLO [11]. This architecture continues to be studied and evolved year after year next to state the art for object detection tasks. Based on these data, it is possible to state that these authors are pioneers in the algorithms used by most of the works present in the sample, addressing the most diverse aspects, such as new architectures for different tasks, optimization methods, and the creation of datasets.

4 Conclusions

The study identified that the number of publications about the applications of techniques and artificial intelligence algorithms to detect failures in manufacturing and industrial processes is small, even with the research protocol being strictly specific and applied to the two largest and most frequently used platforms. However, it is possible to notice a growing trend in the observed period, showing signs of increasing research opportunities and gaps that are of interest.

Bibliometric analysis also reveals relevant information, such as major countries investing in research of this subject, i.e., China and the United States. It is important to emphasize that the low number of publications in South America may reflect the industry crisis that occurred in the region during the time period. in which all works published in this period are of Brazilian origin, making the national research and development of works related to the theme of this paper even more relevant.

It is also important to consider data analyzed through VOSViewer, which allows significant analysis for future works since the most relevant researchers were found through the number of citations and co-citations of their published documents. Moreover, we identified the main keywords and their relationships, which present us

with the main areas and subareas on the subject. This can be observed when we analyze the most cited keywords in the sample along with the main authors. That analysis highlighted Deep Learning and Convolutional Neural Networks as standpoints for future studies. It also allow us to identify that the keywords related to recognition and detection are the most present in a general context, which are directly linked to the main authors presented in this work.

Inspired by the results of this research, the authors are investigating the development of a system capable of detecting failures in manual assembly processes in industrial environment. In addition, a qualitative analysis can be performed on the publications investigated in this research to further explore aspects of the main works on which these themes can be further investigated, generating even more research insights and the development of an analytical framework for evaluating new technologies in this field.

Finally, it is possible to state that this work validates the relevance and growth of future studies on these topics, facilitating research directions, providing quality references, and fostering relevant insights in academia, in order to present the main techniques, algorithms and paradigms applied to tasks in question.

References

1. Chollet, F.: Deep Learning with Python. [S.l.]: Manning Shelter Island (2018).
2. Xu, X.. Hua, Q.: Industrial big data analysis in smart factory: Current status and research strategies, in IEEE Access, vol. 5, pp. 17 543–17 551, (2017).
3. Sassi, P., Tripicchio, P., and Avizzano, A. C.: A smart monitoring system for automatic welding defect detection, IEEE Transactions on Industrial Electronics, vol. 66, no. 12, pp. 9641–9650, (2019).
4. Du, W., Shen, H., Fu, J., Zhang, G., He, Q.: Approaches for improvement of the x-ray image defect detection of automobile casting aluminum parts based on deep learning, NDT E International, vol. 107, p.p 102144, (2019).
5. Lai, Z.-H., Tao, W., Leu, M. C., Yin, Z.: Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing," Journal of Manufacturing Systems, vol. 55, pp. 69–81, (2020).
6. JOURNAL CITATION REPORTS. Cartão de referência rápida: suportado pelo Isi Web of Knowledge: Thomson Reuters, (2009) Disponível em: <https://bibliotecaquimicaufmg2010.files.wordpress.com/2013/03/jcr_qrc_pt_20101026.pdf>. last accessed 2023/01/08.
7. Escobar, C.A., Morales-Menendez, R.: Machine Learning and Pattern Recognition Techniques for Information Extraction to Improve Production Control and Design Decisions. In: Perner, P. (eds) Advances in Data Mining. Applications and Theoretical Aspects. ICDM 2017. Lecture Notes in Computer Science(), vol 10357. Springer, Cham (2017).

8. Grácio, M. C. C.: Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 21, n. 47, pp. 82-99, (2016).
9. Lucas, E. O.: Garcia-Zorita, J. C.: Sanz-Casado, E.: Evolução histórica de investigação em informetria: ponto de vista espanhol. Linc em Revista, v. 9, n. 1, pp. 255-270, (2013).
10. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition, in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, (1998).
11. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.; You Only Look Once: Unified, Real-Time Object Detection,in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788 (2016).
12. Lihao, W. and Yanni, D.: A Fault Diagnosis Method of Tread Production Line Based on Convolutional Neural Network, 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), pp. 987-990, Beijing, China (2018).
13. De Bakker, F. G. A., Groenewegen, P., & Den Hond, F.: A bibliometric analysis of 30 years of research and theory on corporate social responsibility and corporate social performance. Business and Society, vol 44(3), pp 283-317 (2005).
14. Breunig, M. M., Kriegel, H-P., Raymond T. Ng, Sander, J. 2000. LOF: identifying density-based local outliers. SIGMOD Rec. 29, 2 ,pp. 93–104 (2000).
15. Thamm, S., Huebser, L., Adam, T., Hellebrandt, T., Heine, I., Barbalho, S., Knorr V. S., Becker, M., Bagnato, S. V., Schmitt, H. R.: Concept for an augmented intelligence-based quality assurance of assembly tasks in global value networks, Procedia CIRP, vol 97, pp 423-428 (2021).
16. Bertolini, M., Mezzogori, D., Neroni, M., Zammori, F.: Machine Learning for industrial applications: A comprehensive literature review, Expert Systems with Applications, vol 175, 114820 (2021).
17. Gao, Y., Li, X., Wang, X. V., Wang, L., Gao, L.: A Review on Recent Advances in Vision-based Defect Recognition towards Industrial Intelligence, Journal of Manufacturing Systems, vol 62, pp 753-766, (2022).
18. van Eck, N. J.; Waltman, L. (2010) VOSViewer: Visualizing Scientific Landscapes [Software]. Available from https://www.vosviewer.com