



Artificial intelligence applied in pulmonary hypertension: a bibliometric analysis

Germaine Tchuente Foguem¹ · Aurelien Teguede Keleko²

Received: 7 December 2022 / Accepted: 7 February 2023 / Published online: 7 March 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Introduction Advances in Artificial Intelligence (AI) offer new Information Technology (IT) opportunities in various applications and fields (industry, health, etc.). The medical informatics scientific community expends tremendous effort on the management of diseases affecting vital organs making it a complex disease (lungs, heart, brain, kidneys, pancreas, and liver). Scientific research becomes more complex when several organs are simultaneously affected, as is the case with Pulmonary Hypertension (PH), which affects both the lungs and the heart. Therefore, early detection and diagnosis of PH are essential to monitor the disease's progression and prevent associated mortality.

Method The issue addressed relates to knowledge of recent developments in AI approaches applied to PH. The aim is to provide a systematic review through a quantitative analysis of the scientific production concerning PH and the analysis of the networks of this production. This bibliometric approach is based on various statistical, data mining, and data visualization methods to assess research performance using scientific publications and various indicators (e.g., direct indicators of scientific production and scientific impact).

Results The main sources used to obtain citation data are the Web of Science Core Collection and Google Scholar. The results indicate a diversity of journals (e.g., IEEE Access, Computers in Biology and Medicine, Biology Signal Processing and Control, Frontiers in Cardiovascular Medicine, Sensors) at the top of publications. The most relevant affiliations are universities from United States of America (Boston Univ, Harvard Med Sch, Univ Oxford, Stanford Univ) and United Kingdom (Imperial Coll London). The most cited keywords are “Classification”, “Diagnosis”, “Disease”, “Prediction”, and “Risk”.

Conclusion This bibliometric study is a crucial part of the review of the scientific literature on PH. It can be viewed as a guideline or tool that helps researchers and practitioners to understand the main scientific issues and challenges of AI modeling applied to PH. On the one hand, it makes it possible to increase the visibility of the progress made or the limits observed. Consequently, it promotes their wide dissemination. Furthermore, it offers valuable assistance in understanding the evolution of scientific AI activities applied to managing the diagnosis, treatment, and prognosis of PH. Finally, ethical considerations are described in each activity of data collection, treatment, and exploitation to preserve patients' legitimate rights.

Keywords Bibliometrics · Medical informatics · Pulmonary hypertension · Artificial intelligence · Machine learning · Deep learning · Ethics · GDPR

Abbreviations

AI Artificial intelligence
ML Machine learning

DL Deep learning
WoS Web of science
IT Information technology
DS Data science
DNNs Deep neural networks
GDPR General data protection regulation
TF-IDF Term frequency inverse document frequency
ECG Electrocardiography
PH Pulmonary hypertension
ERC European Research Council
ND Number of documents
NP Number of publications

✉ Aurelien Teguede Keleko
kelekoaurelien@yahoo.fr
Germaine Tchuente Foguem
tchuente@gmail.com

¹ Faculté des Sciences, Université de Yaoundé I, PO Box 812, Yaoundé, Cameroon

² Ecole Nationale d'Ingénieurs de Tarbes (ENIT), 47 Avenue Azereix, BP 1629, 65016 Tarbes, France

IEEE	Institute of Electrical and Electronics Engineers
NN	Neural networks
CVD	Cardiovascular disease
MRI	Magnetic resonance imaging
RI	Resonance imaging
CT	Computer tomography
US	Ultrasound
CNN	Convolutional neural network
UK	United Kingdom
cIMT	Carotid intima-media thickness
PA	Plaque area
PAH	Pulmonary arterial hypertension
mPAP	Mean pulmonary arterial pressure
CAR	C-reactive protein to albumin
CRP	C-reactive protein
COPD	Chronic obstructive pulmonary disease
IPH	Idiopathic pulmonary hypertension
CTEPH	Chronic thromboembolic pulmonary hypertension
BPA	Balloon pulmonary angioplasty
RVLS	Right-ventricular longitudinal strain
RVFAC	Right-ventricular fractional area change
RVESRI	Index of right-ventricular systolic remodeling
TAPSE	Tricuspid annular plane systolic excursion
SSVM	Survival support vector machine
minSUP	Minimum SUPport
minCONF	Minimum CONFidence
minLIFT	Minimum LIFT
GGT	γ -Glutamyl transferase
AST	Spartate aminotransferase
ALT	Alanine aminotransferase
NLR	Neutrophil/lymphocyte ratio
USA	United States of America

1 Introduction

1.1 Data sciences

Artificial Intelligence (AI) can be defined as the ability to reproduce human intelligence (e.g., vision, hearing, language comprehension, object grasping, adaptation, and reasoning) with machines. In its revival, Data Science exploits considerable masses of data (big data) to perform increasingly impressive processing. This was not the case in the 1970s and 1990s when we saw AI as the ability to make logic. Now, we have realized that at the heart of AI, there is, first of all, learning from data, and it is this ability to take these data and extract important information or features that generates the knowledge [1]. This knowledge is defined as an insight that can be expressed using rules or constraints characterizing a reflection on the information analyzed from

the contextualization of data. In other words, knowledge is the fact of a particular medical field extract from data which is accessed from a variety of different sources/systems to take decisions and learn from the resulting patterns for the diagnosis and treatment of patients.

Scientific research on “Data Sciences” calls for the aggregation of several disciplines, mainly computer science, mathematics, and statistics. This involves several branches of mathematics, because we obviously see geometry appearing, analysis, probability, and statistics, since one has data and must predict the results with some confidence. The mixture of scientific fields is essential for modeling and analyzing large amounts of heterogeneous data. The theoretical component highlighted underlying bridges between domains beyond their particularities. The experimental component fostered the application of insightful insights whose tremendous problem-solving power led to improved runtime performance. We understand the operation of the algorithms used, which are in a certain way mastered, but their capacity for generalization may seem considerable, but it is essentially based on a form of regularity [2] of the function linking the input data and the predicted response. The most remarkable applications in medicine concern the ability to generate new data, the ability to aid in the diagnosis, and the ability to make precision and personalized medicine [3] from examinations or the genome and, therefore, to adapt to any disease (e.g., oncology and cardiology).

1.2 Machine learning and deep learning

AI is a general concept consisting of using a computer to model intelligent behavior with minimal human involvement. In AI applied to big data, we have two main approaches, such as Deep Learning (DL) and Machine Learning (ML). Learning by Deep Neural Networks (DNNs) reflects the fact that AI, in a certain way, tries to reproduce the functioning of multilayer neurons in the human brain. Indeed, it was a surprise to see that using artificial neurons, which are structurally considerable simplifications of a biological neuron. We can develop algorithms which have quite remarkable and performing learning capacities or abilities. The organization of these artificial neurons with algorithms such as they are implemented in computers has, however, made it possible to obtain predictive capabilities in a wide variety of problems ranging from vision to hearing and language processing [4].

There are two reasons for the explosion of AI in recent years [5]: (i) a technological reason linked to the accumulation of masses of data via the Internet or connected objects and the prodigious increase in the speed of computers; (ii) a conceptual reason with the understanding of the contribution of the integration of statistical elements in digital learning in addition to symbolic reasoning based on logic. The specificity of AI algorithms is the

possession of many parameters fixed by examples. We give examples of the algorithm, and it is from these examples that it optimizes its parameters and configures itself. Once the algorithm is configured, it can provide answers about data it does not know in advance. Deep Learning (DL) is a particular architecture or a particular class of algorithms with multiple learning layers [6]. This whole field generally focuses on the data to learn the patterns or structures. This learning process allows understanding the nature of a phenomenon by extracting relevant knowledge. This Learning is made possible, because the phenomenon has a form of regularity, and therefore, it can be generalized. The fact that these algorithms can generalize from examples means that, in some way, they have discovered the deep structures of the problem.

We, therefore, try to understand the sources of regularity in a phenomenon and from a few examples which are new by interpolation. The DL approach is geometric rather than logical, because it can predict a response to a question from examples we have already seen. Then, we have to establish similarities between these examples via the definition of distances, like in a geometric space. Understanding these analogies and similarities makes it possible to find the answer in a case where we do not know it by relating it to a point where we know the answer. It is, therefore, more of the order of analogy than the order of logic. Indeed, hierarchical organization and symmetries both play a fundamental role in understanding the regularity of most complex learning functions. The considered data can be in a structured, semi-structured, and non-structured formats presenting a challenge for the computer machine that currently includes notions that are difficult to encapsulate from a mathematical point of view. With analytics comes the ability to generate new data. For example, one can envision generating extraordinarily complex images or videos [7]. Overall, in the DS approach, current developments are exciting fields of knowledge, but with systematic adaptations and regulation. We will want it, especially in the interface with work.

Certain types of actions based on AI will undoubtedly require more explanations [8]. In addition, we will have to design medical systems that are ethical and explainable [9]. We will have to see how to establish collaboration, particularly for robotics and human services. There is also access to private data, and obviously, it is required to consider the General Data Protection Regulation (GDPR). The GDPR Requirements, such as data collection and management methods, lead to reflections at the continental level, such as for the European Union [10, 11]. In general, when you have technologies as powerful as this, it is essential to regulate their use according to society's adaptation processes.

1.3 Contributions and research objectives of the paper

This article aims to provide an overview of scientific research on the subject of AI applied to medical problems like Pulmonary Hypertension (PH). The proposed method is based on bibliometric analysis, which applies a data mining approach. Bibliometric is a research methodology applied in information sciences. This analysis approach uses statistical and mathematical methods to measure, evaluate some indicators, and describe patterns of publications. These indicators can be the production and distribution of publications, the most productive sources, articles, authors, and institutes. Furthermore, it offers effective processing of scientific information to characterize the evolution and the current state of work about research considered. In other words, bibliometric analysis can be viewed as a decision-making tool that provides valuable guidelines (e.g., ethical questions relating to the trust of the data collected and the understanding of their processing). This can help researchers and practitioners to understand the main scientific challenges and most relevant questions related to the field of study [12]. Medical society is also impacted by recent progress in AI, and it is, therefore, wise to shed light on the use of its new computer methods concerning a health problem [13]. In this case, we demonstrate that ML and DL techniques are currently used in managing cardiovascular issues, including PH. The development of powerful DL algorithms offers new opportunities in cardiology using data and signal processing methods, clinical risk prioritization, or image processing [14].

In these circumstances, the main contributions produced relate to the challenges of analyzing bibliometric information on a bi-disciplinary subject, and they are described below:

- The first challenge relates to understanding the problems of medical informatics specific to pulmonary hypertension. Our first contribution concerns the consideration of the specificities of numerical reasoning in the field of AI for using indicators of bibliometric evaluation in cardiology.
- The second challenge is supporting the analysis of bibliometric data with an in-depth examination of the articles. Our contribution is inspired by a solution recommended by the European Research Council (ERC), examining the five or ten most cited articles to assess the work's quality, degree of originality, and influence.
- The third challenge relates to using indicators to evaluate AI for the progress of medical science. Our contribution is based on functional information on the medical devices used to provide an argument with a visual representation of the data of the different strategic axes of using AI for pulmonary hypertension.

- The fourth challenge concerns the respect of ethics regarding AI approaches and the right to explanation in the General Data Protection Regulation (GDPR). We will highlight the articles that deal with these topics.

The rest of the document is organized into five sections: Sect. 2 presents the research methodology and bibliometrics engaged in the paper. Section 3 describes the analysis made and the associated results. Section 4 concerns the discussion of this bibliometrics. Section 5 is about the conclusion and perspectives.

The following section proposes a methodology to describe the steps of the considered way for analyzing bibliometric data.

2 Research methodology

2.1 Bibliometric analysis

The space for university research has grown considerably over the past half-century. Throughout the world, in both developed and developing countries, we regularly see the creation of new university establishments and research organizations. This results in increased human, material, and software resources with various digital solutions for scientific production dissemination (e.g., conferences and journals). This is accentuated by the establishment of multiple, more intense collaborations with more mobility or virtual exchanges between researchers on a national and international scale.

The profusion of scientific production coupled with the perpetual growth of digital channels for the distribution of research documents complicates the synthesis of the information to distinguish the most significant. Knowledge of the most important advances is critical regarding research that could impact human health. This is particularly the case of research on AI applied to diseases affecting the vital organs protected by the rib cage and the cranium.

Bibliometrics is an approach to evaluate, monitor, and visualize how scientific fields are structured [15]. In addition, it helps to define the information about the publications and to determine the productivity and quality (e.g., issues with bibliometrics relating to the representativeness of the publication sources and the relevance of their selection criteria) of researchers and countries, and organizations including their relative impact. Bibliometrics can be viewed as a method of evaluation in the scientific research community. In scientific research, productivity does not always equal to quality. In particular, individual evaluation through bibliometrics can have adverse effects on the quality of research. In fact, a person could be preoccupied with increasing his/her index. As an example, he can use the same idea in

several papers. Also, a researcher can split the same idea into several slices, each idea being published separately in different journals. Despite these different arguments, some organizations use this method and productivity indicator as a criterion for allocating the financial support to the most productive or innovative researchers [16].

This approach can be applied in several fields, such as malware study [17], Finnish schizophrenia research [15], global Parkinson's disease research trends [18], global scientific research on COVID-19 [19], and trends in dry eye disease research [20]. This article is a bibliometric study about the methods of AI applied to PH disease. Although this disease is not quite common, its characteristic element is the rise in blood pressure in the vessels departing from the heart's right ventricle and supplying the lungs. The difficulty of establishing the diagnosis could delay the discovery of the disease by generating aggravating factors with potentially harmful consequences on the lives of patients and families who would be victims.

Given these considerations, AI can significantly help with advanced data processing and image analysis methods to aid reasoning in the medical management of this disease. In this context, bibliometrics becomes an essential means for the overall understanding of the evolution of research work about applying AI techniques to PH disease. This global vision offers the opportunity to know the journals most involved in disseminating this scientific information while highlighting the universities with the most prolific authors. In addition, highlighting the main keywords allows a more efficient selection of the information sought according to the centers of interest of the actors and organizations concerned. In addition, we will be identifying the most used technologies and their respective trends. We will also present the main research challenges related to the context of the study.

2.2 Recommended workflow for science mapping

We describe in this section the methodology adopted to conduct the bibliometric research process. Figure 1 represents the workflow that describes the best practice or method for performing this analysis. The methodology framework for this bibliometric investigation comprises multiple steps: research definition, data collection, data processing, data analysis, and visualization. In the following paragraphs, we describe the five main steps of this methodology.

- *The definition of the research*

This step consists of defining the keywords and expressions, especially a report with the whole of the topics which are approached or developed in the topic of research considered. These keywords will be used as input in the next step to extract the documents.

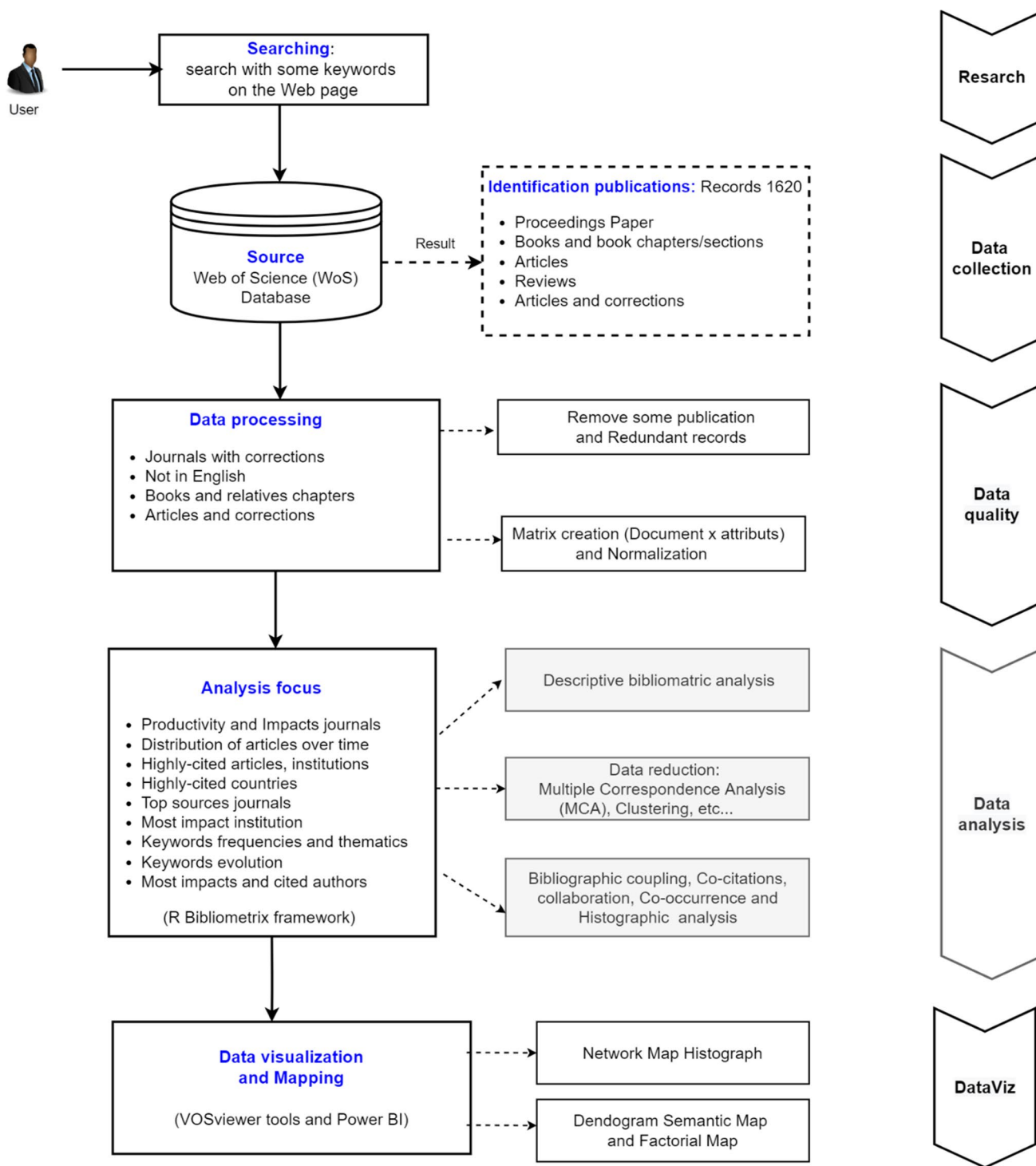


Fig. 1 Methodology adopted for bibliometric analysis

• *Data collection*

This second step is done by identifying publications selected from conference proceedings, books and book

chapters, and articles (original papers or reviews) from journals published or finalized. These documents must meet other requirements, such as the author's writing and the year of publication.

- *Data processing*

This step defines the techniques we will use to process the collected data. The most common operations are, respectively, data cleaning to eliminate redundancies in the collected data, data transformation to create a square matrix (documents \times attributes), and, finally, data normalization to remove the size effect on the data.

- *Data analysis*

This is the most important step of the bibliometric study. Data analysis includes a descriptive analysis of bibliometrics, data reduction via various methods [e.g., clustering, Principal component analysis (PCA), Multiple Correspondence Analysis (MCA), Multidimensional Scaling (MDS)], and bibliometric coupling (co-citations, collaboration, co-occurrence) with histographic analysis.

- *Data visualization*

The last step involves visualizing the results of the analysis. It provides a set of methods for graphically summarizing data to show the relationships between data sets. The graphical representation of data can take the form of maps, graphs, diagrams, or tables, to facilitate the understanding of analytical data (e.g., dendrogram, clustering, and semantic and factorial maps).

2.3 Web of science and data collection

The collection of bibliometric data can be done via several platforms accessible online on the Internet or databases [21]. These databases are Web of Science (WoS) from Clarivate Analytics, Scopus from Elsevier, Google Scholar, Science Direct, and IEEE. WoS provides access to more high-impact, multidisciplinary journals in each subject (e.g., medicine, mathematics, physics, health sciences, biology, etc.) and the research is more exhaustive. WoS and Scopus resources are large multidisciplinary bibliographic databases containing citations and abstracts from peer-reviewed journals, scholarly journals, books, patent records, conference proceedings, etc. In addition, the searches can be performed according to several criteria, including the subject, the author, the title of the article, the years of publication, etc. These databases also provide tools for tracking, analyzing, and visualizing research results. They also offer bibliometric modules and constitute the reference data for several international university rankings, such as Times Higher Education or Shanghai.

Considering the connectivity opportunities via our professional environment's Virtual Private Network (VPN), our study focused on the bibliographic and citational database

WoS. WoS is a bibliographic database platform that allows the generation of bibliometric indicators (statistical analysis of publications). This scientific database has the following features:

- *A multidisciplinary* of data concerning all exact and applied sciences.
- *A remote accessibility* via the connection identifiers of university establishments or research organizations.
- *A non-exhaustiveness*, because all the titles are therefore not present, particularly journals that are not in English or journals with low impact.

In addition, WoS provides an operational analysis function on a set of references and an interesting “citation mapping” function to visually see the publication path. While Scopus is content to directly provide macro-analysis of the results on the title of periodicals, authors, years, types of documents, and major fields, institutions have no analysis.

2.4 Scanning and keywords search

The search for scientific documents is done by specifying “keywords”. The research process uses the Term Frequency Inverse Document Frequency (TF-IDF) method of records [22]. This method can be defined as calculating the relevance of a word of a series or a corpus concerning a text. The meaning increases proportionally to the number of times a word appears in the text but is compensated by the frequency of the word in the corpus (data set).

TF-IDF analysis has the advantage of emphasizing relevance and uniqueness as the decisive criteria for frequency weighting while combining document-specific analysis and general analysis disciplines. Also, the results are smoothed using logarithms to get more relevant data. The disadvantage of the TF-IDF analysis concerns the permanent examination of the complete editorial content of a document and the inadequacy of short texts containing few words.

In this case, it is necessary to study the semantic scope of the identified keywords to obtain the most relevant documentary elements based on the TF-IDF method. The approach can be iterative with modifications and iterations, so it is important to start by refining the list of keywords using common sense.

In Web of Science, there are two types of searches that can be done:

- *General search* allows you to search for one or more terms in different fields. You can choose from various search fields: topic, title, author, group author, publication name, year published, address, language, document type, etc.

- *Advanced search* allows you to create complex queries using search symbols and perform combinations. The use of advanced search requires a good command of search equations.

For both types of searches, the query is launched by default on the four databases that make up the Web of Science: SCI (Science Citation Index), SSCI (Social Science Citation Index), AHCI (Arts & Humanities Citation Index), and CPCI (Conference Proceedings Citation Index). It is possible to limit the search for one or more databases by selecting them in “Citation databases” and to limit by dates by clicking on the “Timespan” button.

In this study, we performed the advanced search on August 24, 2022, in the Web of Science database. We, therefore, carried out a progressive definition of keywords and iteration with corrections based on the observation of the results obtained. In addition to the years of publication, we also consider the articles' abstract, author, or indexing keywords. The combination of keywords we used generates a query. This query is described as follows:

(Pulmonary Hypertension **and** Machine Learning **OR** Deep Learning and Heart Disease) **and** (Timespan: 1997–2023)

3 Analysis and results

In this section, we present the analysis of the most important results of the bibliometric study.

3.1 Main information about the collection

3.1.1 Distribution of the documents

Exploring the distribution of documents is quite informative, as seen with the *TreeMap chart distribution* (Fig. 2). We note that the problem of the study is correlated or connected with several other areas.

It is observed that the topics covered are of great diversity, ranging from biomedical engineering to cardiovascular systems, passing to computer science methods with AI. In the first position, biomedical engineering appears thanks to the development of devices used for diagnosing and treating patients [e.g., electroencephalograph (EEG)]. We find the cardiovascular heart systems naturally because they are impacted, and their examination can make it possible to perceive evocative anomalies of pulmonary hypertension. We also found AI, the potential for improving patient care mentioned, particularly for predicting cardiovascular risk [23]. Our work endeavored to report on the observed diverse thematic areas; however, in this section, we will only describe the first five themes:



Fig. 2 TreeMap chart distribution

- In the first position, we find “*Biomedical engineering*”, which explicitly concerns high-tech materials and equipment for the hospital sector, care establishments, and health professionals. The biomedical engineer has a perfect knowledge of diagnostic devices, treatment, and assistance with the most advanced techniques.
- The second position is “*Engineering Electrical Electronic*”, because recording the heart's electrical activity with electrocardiography (ECG) is an essential cardiology examination that makes it possible to follow the heartbeats as they occur. Depending on the results, we can guide the diagnosis and request, if necessary, the performance of additional examinations such as a stress ECG.
- In the third position, we find “Radiology Nuclear Medicine Medical imaging,” a sub-specialty of radiology involving using radioactive drugs (radiopharmaceuticals) to diagnose and treat disease. These radioactive materials are usually injected into a vein but are sometimes swallowed or inhaled. Non-invasive imaging plays an important role in assessing and managing pulmonary hypertension [24].
- In the fourth position, the “Cardiac, and cardiovascular systems” appear naturally because they are mainly impacted by pulmonary hypertension. Indeed, examining these systems is a determining factor in the perception of possible abnormalities suggestive of pulmonary hypertension. Further, the progression of pulmonary hypertension in the lungs’ arteries can be categorized (e.g., normal, mild pulmonary hypertension, and hypertension).
- The fifth position is “Computer Science Artificial Intelligence”, since artificial intelligence mainly has a computer component and a mathematical component. We can note the sub-themes of deep learning, machine learning, computer vision, intelligent agents, and game theory for artificial intelligence. By way of illustration, AI plays a crucial role in pursuit of improving patient’s care through prediction of possible risk of cardiovascular [23].

Documents retrieved from queries include book chapters, articles, conference proceedings, editorial letters, and reviews of conference proceedings and books (Fig. 3). Moreover, all the collected documents are analyzed without filters (years, type of documents, and languages).

3.1.2 Information about data

We use a search query to retrieve scientific documents (papers, books, authors, etc.) published between 1997 and 2023 on from the scientific database WoS, however, this must be balanced against the confirmation bias [25] which could lead to explaining the information by consistent data and neglecting the data which contradicts them. In point of

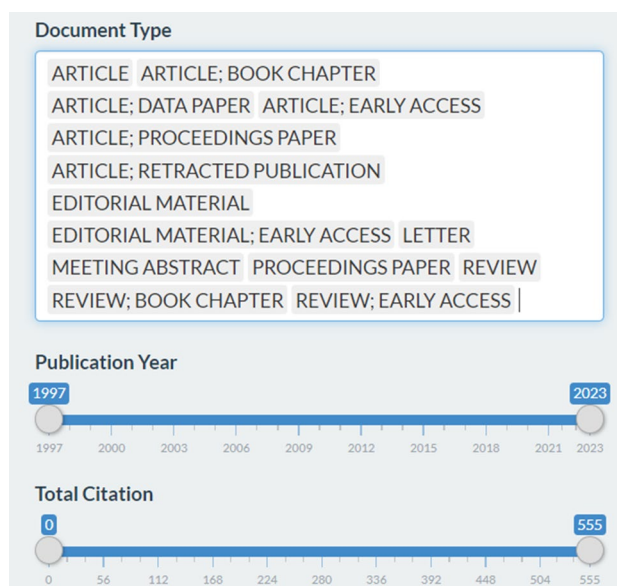


Fig. 3 The type of documents collected during the period 1997–2023

fact, ethical concerns with bibliometric models arise from their inclination to unintentionally generate unfair information. This neglects some sensitive factors such as the level of expertise on the subject of the authors. Table 1 presents the primary information and statistics regarding the publications collected according to the request. The number of documents is 1620, with the main articles numbering 1068. The average citation per document is 13.69, while the number of citations per year per document is 3.67. The number of authors who wrote a single document is quite negligible (30) compared to the number of those who wrote several (8439). Regarding collaboration, a tiny minority of documents are written by a single author (32), while the average number of co-authors per document is relatively high (7). The Keywords Plus (ID) generated algorithmically [26] from the words or phrases frequently appearing in the titles of the references of an article are almost three times less numerous (2489) than the Author's Keywords (DE) described explicitly by the authors (6153) of an article. However, Keywords Plus helps improve the power of cited reference searching by searching across all relevant disciplines in all articles with commonly cited references. Furthermore, some research has shown that the Keywords Plus parameters accurately capture the scientific content and concepts presented in the articles [27].

3.2 Annual scientific publication trend

3.2.1 Annual scientific production

The publications are relatively sparse, with documents published close together and spaced between 1995 and 2015. Figures 4 and 5 represent the annual

Table 1 Information on documents' types, contents, authors, and collaboration

Description	Results
<i>Main information about data</i>	
Timespan	1997:2023
Sources (journals, books, etc.)	731
Documents	1620
Average years of publication	1.96
Average citations per documents	13.69
Average citations per year per doc	3.678
References	1
<i>Document types</i>	
Article	1068
Article; book chapter	7
Article; data paper	1
Article; early access	53
Article; proceedings paper	4
Article; retracted publication	1
Editorial material	14
Editorial material; early access	1
Letter	4
Meeting abstract	21
Proceedings paper	266
Review	175
Review; book chapter	1
Review; early access	4
<i>Document contents</i>	
Keywords plus (ID)	2489
Author's keywords (DE)	6153
<i>Authors</i>	
Authors	8469
Author appearances	11.910
Authors of single-authored documents	30
Authors of multi-authored documents	8439
<i>Authors collaboration</i>	
Single-authored documents	32
Documents per author	0.191
Authors per document	5.23
Co-authors per documents	7.35
Collaboration index	5.31

scientific production published in WoS journals over the last 30 years. This time interval can therefore be considered negligible and not considered in the analysis. Just after this period, the volume of publications becomes significant. We noticed an exponential evolution of publications in the field of study. Especially, a successively strictly increasing curve of publications with a peak in 220 out of 465, then strictly decreasing with only one publication in 2023. The analysis will, therefore, focus on the documents published between 2015 and 2022.

Table 2 shows the evolution of published papers and their respective percentages. In particular, the number of documents reaches its peak in 2021 with more than 465 published articles (30%). However, during the first semester of the year 2022, we register approximately 307 indexed documents, which is about 20% of the collected documents.

3.2.2 Average citation per year

In this subsection, we produce a graph (Fig. 6) showing the metric “Average number of citations per paper” and the “Average number of citations per year” over the period 2015–2022. The average number of citations per article is determined by dividing the total number of citations by the total number of articles. This value can be a useful metric to assess a journal's or author's average impact. The average number of citations per year is determined by dividing the total number of citations by the number of years the author or journal has published articles. This can provide a significant measure for assessing a journal's or author's annual impact. In particular, the year 2017 provides the highest pair of values with an “Average number of citations per paper (MeanTCperArt)” of 57 and an “Average number of citations per year (MeanTCperYear)” of 11.

3.3 Sources analysis

3.3.1 Most relevant sources

The journals are distinguished by their productivity values characterizing the annual occurrences of their Number of Publications (NP) on the topic studied. Figure 7 and Table 3 represent the most relevant sources. We note that the indicator NP represents the number of articles published in the journal. For example, the top five relevant scientific journals are listed below:

- *IEEE Access* is an open-access scientific journal published by the Institute of Electrical and Electronics Engineers (IEEE). The journal emphasizes interdisciplinary topics. The fields of applications are electrical engineering, electronics, and computer technology. There are also publications on the use of digital technologies in cardiovascular informatics and vascular biology.
- *Computers in Biology and Medicine* are concerned with the application of the computer in the fields of bioscience and medicine. The topics of interest in this journal include computer aids to biocontrol-system engineering, automatic computer analysis of pictures of biological and medical importance, accumulating and recalling individual medical records, and interfaces to patient monitors.
- *Biology Signal Processing and Control* deliver interdisciplinary research in the measurement and analysis

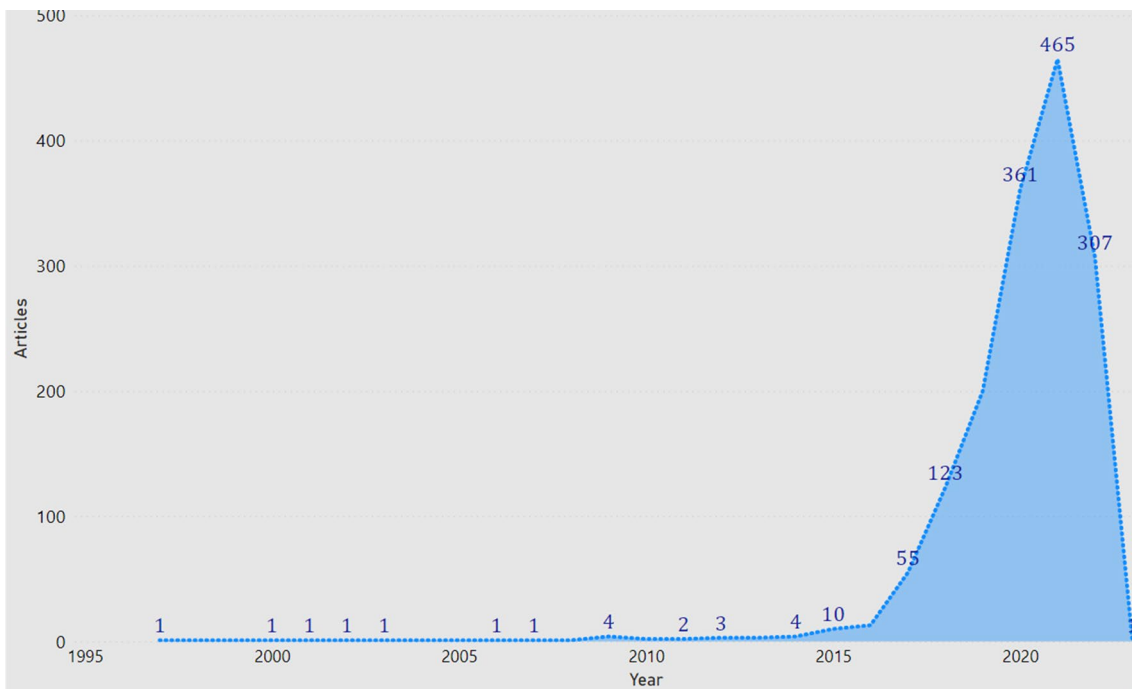


Fig. 4 The evolution of the volume of publications from 1995 to 2023. Between 1970 and 2015, the number of documents is not representative, and in 2023, we only have one document. We will delete

the documents concerned. The analysis will focus on the documents collected between 2015 and 2022 (Fig. 5). We have reduced about 85 papers (1660–1535)

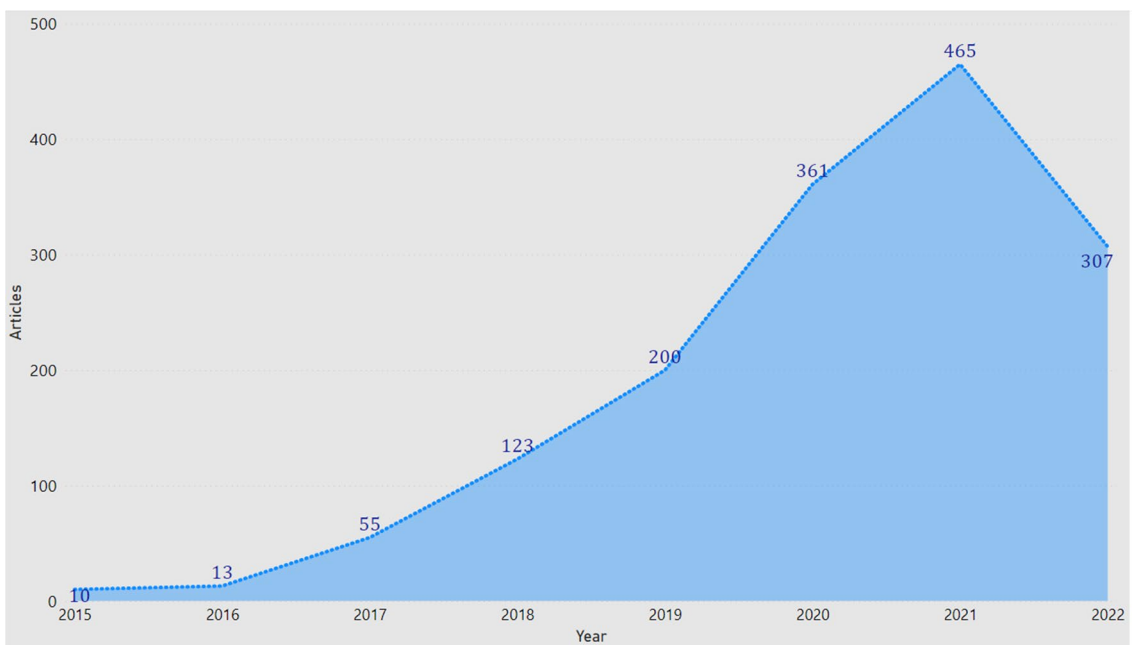


Fig. 5 The evolution of the volume of publications from 2015 to 2022

Table 2 Annual number of published between 2015 and 2022 on WoS

Years	ND	ND (%)
2015	10	0.65
2016	13	0.85
2017	55	3.59
2018	123	8
2019	200	13
2020	361	23.52
2021	465	30.30
2022	307	20

The metrics ND and ND (%) are, respectively, the numbers of documents and the number of documents in percent

of signals and images in clinical medicine and the biological sciences. The focus is on contributions concerning the practical, applications-led research on selecting methods and devices in clinical diagnosis, patient monitoring, and management.

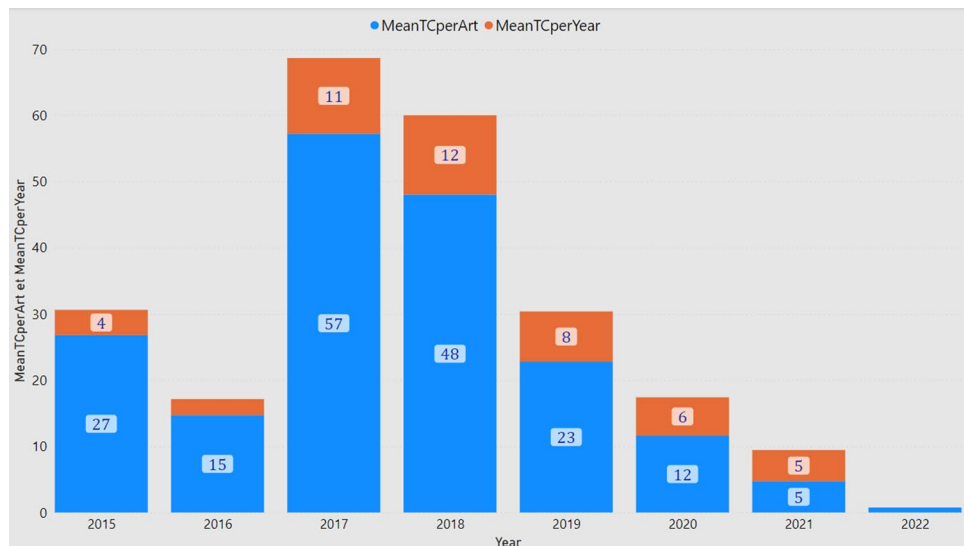
- *Frontiers in Cardiovascular Medicine* provide research articles across basic, translational, and clinical cardiovascular medicine. The topic of the journal concerns cardiovascular medicine with a focus on studies that verify treatments and practices in cardiovascular care or facilitate the translation of scientific advances into the clinic as new therapies or diagnostic tools.
- *Sensors* offer information on the technological developments and scientific research in the vast area of physical, chemical, and biochemical sensors, with remote sensing and sensor networks. The topics cover Smart/Intelligent sensors, Sensing principles, Signal processing, data fusion and deep learning in sensor systems, Sensing and imaging, Communications, and signal processing.

3.3.2 Sources impact

As a preliminary comment, it can easily be observed that the most productive sources are not necessarily the most relevant or cited ones. The main criteria for assessing the relevance of digital document information are the provenance of the document, the reliability of the content, and the purpose pursued by the author [28]. To assess the document's relevance, it is not necessary to read the entire document. A quick exploration is enough to understand the document's content. To assess this relevance, we can once again ask ourselves a few questions that will relate to the content of the papers and the level (scientific, professional, general interest, popular, etc.) of the information conveyed. To assess the level of information, we will consider the following elements:

- *The nature of the document:* It must be remembered that it is a matter of carrying out work at the university level and knowing how to recognize the nature of the journals consulted.
- *The quality of information:* review, if necessary, the criteria for assessing data quality that can be assessed relative to certain criteria such availability (e.g., search for missing data, in particular those that are not striking, not memorable or not recent) or representativeness (e.g., weighting of observations according to the size and relevance of the selected sample). In addition, we mentioned these data quality issues in the preceding subsection (see Sect. 2.2).
- *The introduction and conclusion:* consult the document's introduction and conclusion. A quick reading makes it possible to evaluate the level of information, to know what the starting question is and what conclusions the author draws from it.

Fig. 6 The “average number of citations per paper” and the “average number of citations per year” over 2015–2022



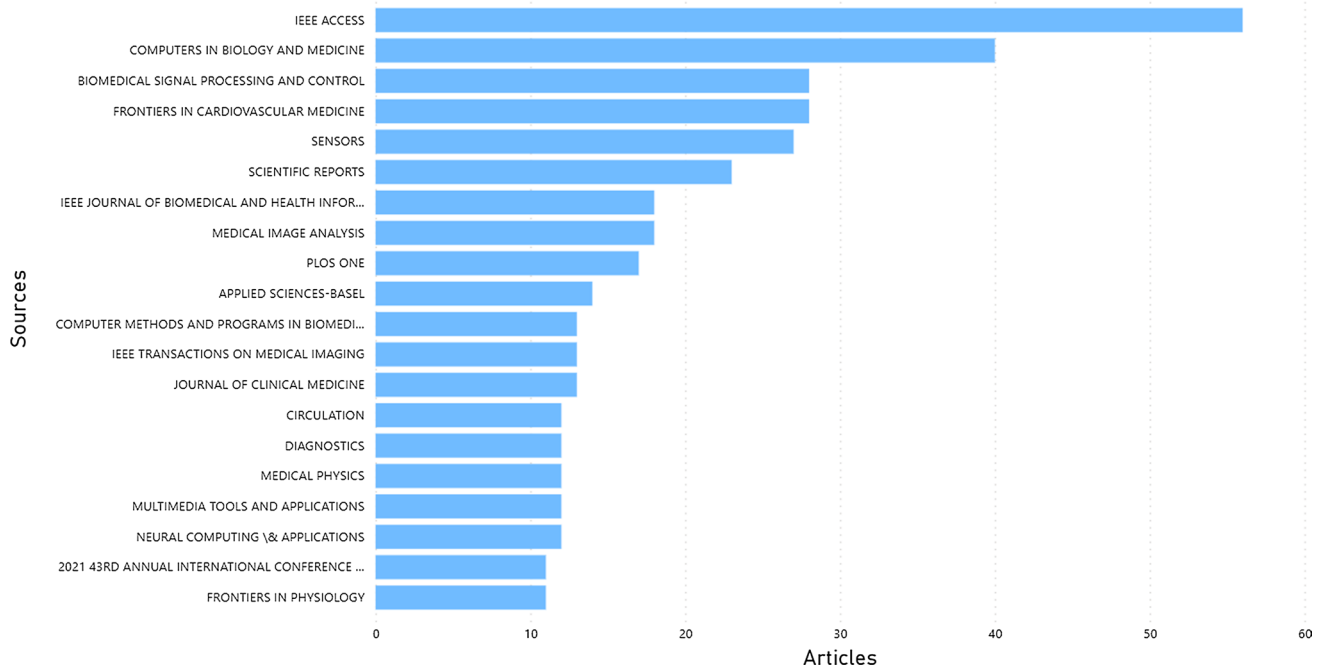


Fig. 7 Most relevant sources from the numbers of documents

Table 3 Top 20 sources' documents ordered by the number of published articles

Sources	NP
<i>IEEE Access</i>	56
<i>Computers in Biology and Medicine</i>	40
<i>Biomedical Signal Processing and Control</i>	28
<i>Frontiers in Cardiovascular Medicine</i>	28
<i>Sensors</i>	27
<i>Scientific Reports</i>	23
<i>IEEE Journal of Biomedical and Health Informatics</i>	18
<i>Medical Image Analysis</i>	18
<i>PLoS One</i>	16
<i>Applied Sciences-Basel</i>	14
<i>Computer Methods and Programs in Biomedicine</i>	13
<i>IEEE Transactions on Medical Imaging</i>	13
<i>Journal of Clinical Medicine</i>	13
<i>Diagnostics</i>	12
2021 43rd Annual international conference of the IEEE engineering in medicine & biology society (EMBC)	11
<i>Circulation</i>	11
<i>Frontiers in Physiology</i>	11
<i>CMC-Computers Materials & Continua</i>	10
<i>Journal of the American College of Cardiology</i>	10
<i>Medical Physics</i>	10

- *The vocabulary specialization*: note the level of specialization of the vocabulary used.

The sources can also be sorted according to their impact to better understand their respective influence in a field considered or, more specifically, on the theme studied. This sorting can be done on a criterion such as the *h*-index (or *h*-factor), which assumes the number of publications, their sources, and their number of citations (see Fig. 8 and Table 4).

According to Fig. 9, the evolution curve of the IEEE Access journal has a bell shape describing a phase of growth (2015–2020) followed by a phase of decline. The form of evolution is, therefore, of the parabolic type with a peak of maximum value between the years 2020 and 2021. The evolution curves of the journals “Computers in Biology and Medicine” and “Biomedical Signal Processing and Control” have a polynomial appearance. The form of evolution has a strictly increasing trend, at least over the observation period of 2015–2022. The journals “Frontiers in Cardiovascular Medicine” and “Sensors” evolution curves have a logarithmic pace. The form of evolution has a strictly increasing trend, at least over the observation period of 2015–2022.

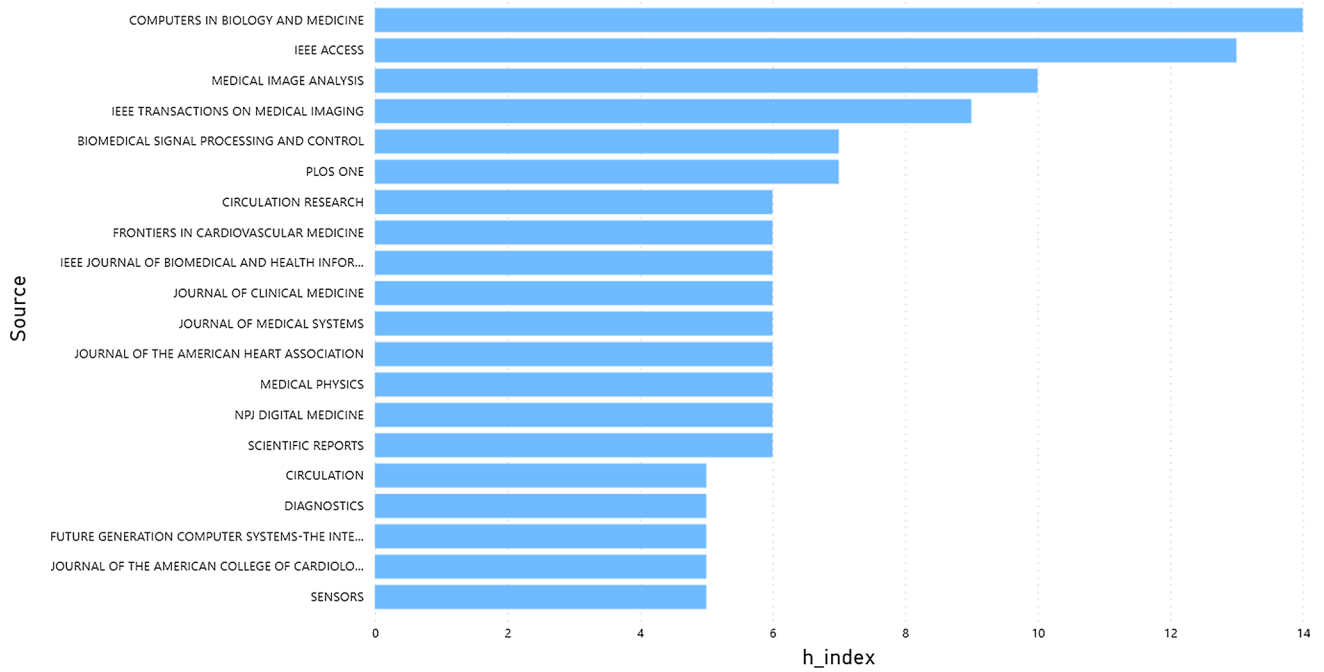


Fig. 8 Source impact based on h-index

Table 4 Most relevant sources are ordered by h-index value. According to TC value, the most highly cited sources are not necessarily the most relevant

Source	h_index	g_index	TC	PY_start
<i>Computers in Biology and Medicine</i>	14	37	1403	2017
<i>IEEE Access</i>	13	25	729	2017
<i>Medical Image Analysis</i>	10	18	456	2018
<i>IEEE Transactions on Medical Imaging</i>	9	13	379	2017
<i>Biomedical Signal Processing and Control</i>	7	14	217	2018
<i>PLoS One</i>	6	12	157	2017
<i>Frontiers in Cardiovascular Medicine</i>	6	11	131	2020
<i>Scientific Reports</i>	6	13	193	2019
<i>IEEE Journal of Biomedical and Health Informatics</i>	6	13	184	2018
<i>Journal of Clinical Medicine</i>	6	11	136	2018
<i>Medical Physics</i>	6	10	178	2019
<i>Circulation Research</i>	6	9	417	2020
<i>NPJ Digital Medicine</i>	6	8	279	2018
<i>Journal of the American Heart Association</i>	6	7	157	2019
<i>Journal of Medical Systems</i>	6	8	122	2017
<i>Sensors</i>	6	6	143	2019
<i>Diagnostics</i>	5	7	57	2020
<i>Circulation</i>	5	11	349	2017
<i>Journal of the American College of Cardiology</i>	5	10	852	2017
<i>Future Generation Computer Systems-The International Journal of eScience</i>	5	8	310	2019

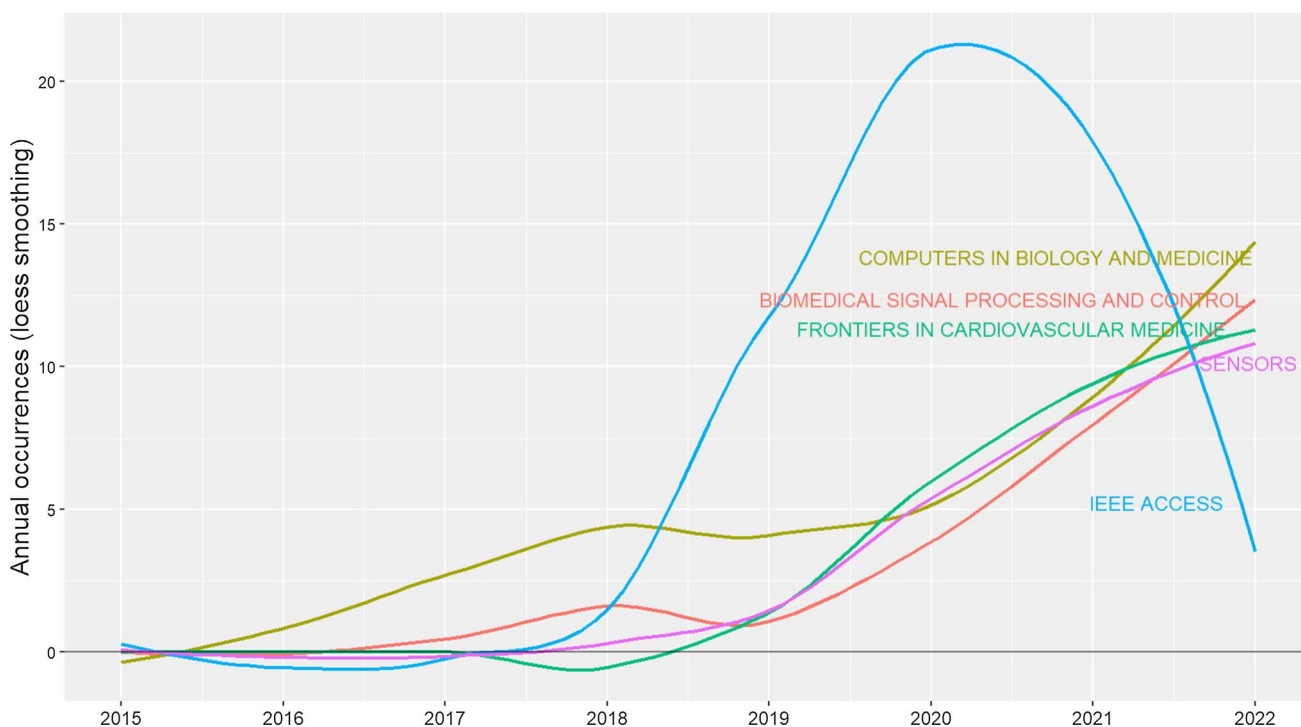


Fig. 9 Source curves according to their annual occurrences

Table 5 Most frequent words and their occurrences

Words	Occurrences
Classification	207
Diagnosis	135
Disease	130
Prediction	88
Risk	84
Segmentation	82
Heart	78
Model	76
Heart failure	56
Heart disease	55
System	54
Features	53
Association	50
Mortality	50
Recognition	50
Validation	48
Identification	46
Algorithm	44
Neural networks	44

3.3.3 Most frequent words

High-frequency words are very important words for readers in any search field. In AI applied to health, as in other application domains, there can be a relatively small group

of words that occur very often and a much larger group of words that occur less often. The small group of occurring words is often known as high-frequency vocabulary. Table 5 and Fig. 10 show the most frequent word according to the Keywords Plus parameters. In this case, the five most frequent words are:

- *Classification*: this word is found in the first position, because research work considers that it is often essential to classify the type of disease considered before any subsequent reasoning.
- *Diagnosis*: this word is found in the second position, because research activities consider that it is fundamental to identify the type of disease before any conservative or curative action.
- *Disease*: this word is in the third position, because although it is central to the medical field, scientific research activity no longer focuses on its relatively already well-established definition.
- *Risks*: this word is found in the fourth position, because it is more than crucial to be able to establish the characteristics of the root causes leading to a pathological situation.
- *Prediction*: this word is found in the fifth position, because one of AI’s significant contributions is the desire to provide decision support in the projection of potential future scenarios.

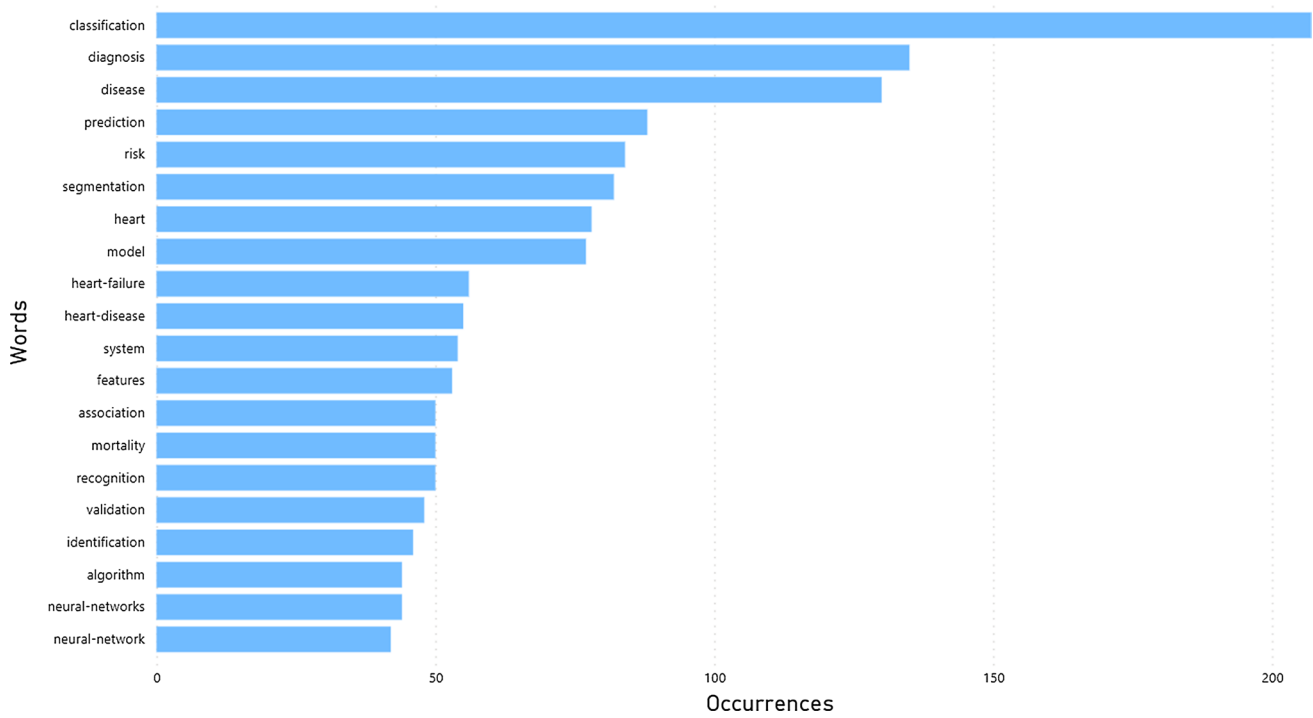


Fig. 10 Most relevant words based on occurrences

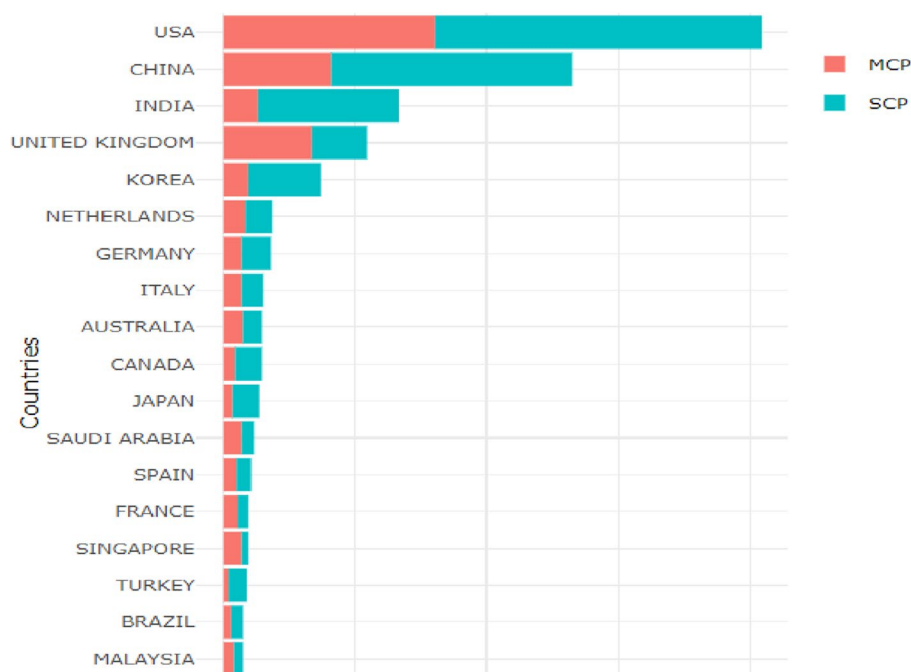
Table 6 Countries and criteria from the production of their corresponding authors

Country	Articles	Freq	SCP	MCP	MCP_Ratio
USA	409	0.26979	248	161	0.394
China	265	0.1748	183	82	0.309
India	133	0.08773	107	26	0.195
United kingdom	109	0.0719	42	67	0.615
Korea	74	0.04881	55	19	0.257
The Netherlands	37	0.02441	20	17	0.459
Germany	36	0.02375	22	14	0.389
Italy	30	0.01979	16	14	0.467
Australia	29	0.01913	14	15	0.517
Canada	29	0.01913	20	9	0.31
Japan	27	0.01781	20	7	0.259
Saudi Arabia	23	0.01517	9	14	0.609
Spain	21	0.01385	11	10	0.476
France	19	0.01253	8	11	0.579
Singapore	19	0.01253	5	14	0.737
Turkey	18	0.01187	14	4	0.222
Brazil	15	0.00989	9	6	0.4
Malaysia	15	0.00989	7	8	0.533
Iran	14	0.00923	9	5	0.357
Pakistan	13	0.00858	6	7	0.538

Table 7 Countries and their scientific production

Country	Frequency
USA	3023
China	1489
UK	891
India	588
South Korea	446
Germany	417
The Netherlands	341
Australia	306
Canada	306
Italy	269
Japan	210
Spain	183
Singapore	160
France	156
Greece	119
Belgium	108
Saudi Arabia	105
Pakistan	91
Iran	80
Turkey	76

Fig. 11 Number of documents from the corresponding author's country



3.3.4 Corresponding author's country and country scientific production

The joint analysis of the graph of the countries of origin of the authors of correspondence of the articles and the graph of the global scientific production of the countries shows in descending order from the first to the fifth the following countries: the United States of America, China, India, the UK, and South Korea (Tables 6 and 7; Fig. 11).

3.4 Most productive and cited sources

3.4.1 Most productive authors

The top 5 most productive authors are SURI JS (USA), SABA L (Italy), KHANNA NN (India), ACHARYA UR (Singapore), and LAIRD JR (USA), respectively, whose profile information is as follows:

- SURI JS belongs to the Stroke Monitoring and Diagnostic Division, Roseville, United States of America. His *h*-index is 49, and his citation count is 10,928. His research interests are Diabetes Mellitus, Logistic Regression Analysis, Carotid Artery, Risk Stratification, Student *T* Test, Atherosclerosis, Cardiovascular Risk, logistic regression, Intima-Media Thickness, support vector machine, Hemoglobin A1c, and Classification Accuracy. In particular, in 2021, he worked on AI techniques to characterize acute respiratory distress syndrome in COVID-19-infected lungs. The developed AI tools assist pulmonologists in detecting early the presence of a virus. In addition, this tool can classify the different types of pneumonia and measure their severity in patients affected by COVID-19 [29].
- SABA L belongs to the affiliation Department of Radiology, University of Cagliari, Italy. His *h*-index is 44, and his number of citations is 9196. His research topics are Common Carotid Artery, Echography, Speckle Noise, Atherosclerotic Plaques, Carotid Stenosis, Magnetic Resonance Imaging, Arterial Wall Thickness, Carotid Artery Diseases, and Carotid Intima-Media Thickness. For example, this author recently published a narrative review on stroke and Cardiovascular disease (CVD). Their study highlights AI-based tools (machine learning, deep learning, and transfer learning) applied to non-invasive imaging technologies such as Magnetic Resonance Imaging (MRI), Computer Tomography (CT), and Ultrasound (US). These models help to characterize and classify the carotid plaque tissue in a multimodal manner. This plaque represents a significant cause of these diseases [30].
- KHANNA NN belongs to Indraprastha Apollo Hospitals, New Delhi, India. His *h*-index is 24, and his citation count is 1744. His research interests include Arterial Wall Thickness, Carotid Diseases, Carotid Intima-Media Thickness, Common Carotid Artery, Echography, Speckle Noise, Radiological Findings, Clinical Features, and COVID-19. The author reviewed the chronic autoimmune disease called rheumatoid arthritis (RA). This disease can cause cardiovascular (CV) and stroke risks. For improved understanding and control of patients suffering from

RA, he exploited a non-invasive method of imaging the carotid artery using 2D ultrasound. This technique provides a tissue-specific image of the atherosclerotic plaque. In addition, it helps to characterize the type of plaque morphologically, and the distinct phenotypes (wall thickness and wall variability) are measured. Thus, ML and DL techniques are exploited to automate these risk characterization processes and to facilitate early and accurate CV risk stratification [31].

- ACHARYA UR is affiliated with Ngee Ann Polytechnic, Singapore City, Singapore. Its *h*-index is 105, and its citation count is 40,344. His research topics are Heart Arrhythmia, Electrocardiograph, Convolutional Neural Networks, Support Vector Machines, Seizures, Bonn, Real Image, Ophthalmology, and Diabetic Retinopathy. The authors analyze the information (signals) from an electrocardiogram (ECG) that monitors the activities of the heart. Arrhythmia diagnosis is identified by each heartbeat (normal or abnormal). In addition, correct classification is performed under the ECG morphology. Their case study subdivided these heartbeats into five groups: non-ectopic, supraventricular ectopic, ventricular ectopic, fusion, and unknown beats. The authors proposed a deep Convolutional Neural Network (CNN) of nine hidden layers to automatically identify the different types and frequencies of arrhythmic heartbeats from ECGs. When properly trained, the model has an accuracy rate close to 94.03%. This result is valid if we apply the technique of artificial data augmentation to address the problem of unbalanced classes [32].
- LAIRD JR belongs to Adventist Health, Roseville, United States of America. His *h*-index is 46, and his citation count is 7734. His research interests include Restenosis, Cardiovascular Risk, Kaplan Meier Method, Intima-Media Thickness, Target Lesion Revascularization, Atherosclerosis, Hemoglobin A1c, Diabetes Mellitus, Carotid Artery, Percutaneous Transluminal Angioplasty, Risk Stratification, and Rank Sum Test. The author presents a detailed AI framework for examining joint measurement of carotid intima-media thickness and plaque area by ultrasound for cardiovascular/stroke risk monitoring. The study exploits modern AI methods for the detection and segmentation of carotid Intima-Media Thickness (cIMT) and Plaque Area (PA) from carotid vascular images [33] (Table 8).

Table 8 Authors and their production of articles

Authors	Articles	Articles fractionalized
Suri JS	33	2.37
Saba L	31	2.35
Khanna NN	23	1.44
Acharya UR	21	4.05
Laird JR	21	1.38
Wang J	21	4.03
Wang X	21	3.33
Zhang H	21	3.43
Zhang Y	21	2.67
Johri AM	19	1.23
Li Y	18	2.57
Wang Y	18	2.98
Zhang X	18	3.49
Liu X	17	3.01
Miner M	17	0.77
Kitas GD	16	0.87
Li H	16	3.26
Chen X	15	1.77
Pareek G	15	0.66
Wang SH	15	3.76

Table 9 Affiliations and their production

Affiliations	Articles
Boston Univ	127
Harvard Med Sch	117
Univ Oxford	101
Stanford Univ	98
Imperial Coll London	82
Hallym Univ	66
Univ Med Ctr Utrecht	66
Univ Calif Los Angeles	65
Emory Univ	59
Univ Leicester	57
Yonsei Univ	53
Massachusetts Gen Hosp	45
Case western Reserve Univ	44
Univ Amsterdam	43
Univ Virginia	42
Icahn Sch Med Mt Sinai	41
Brigham and Womens Hosp	39
Johns Hopkins Univ	39
Univ Calif San Francisco	39
Univ IOWA	39

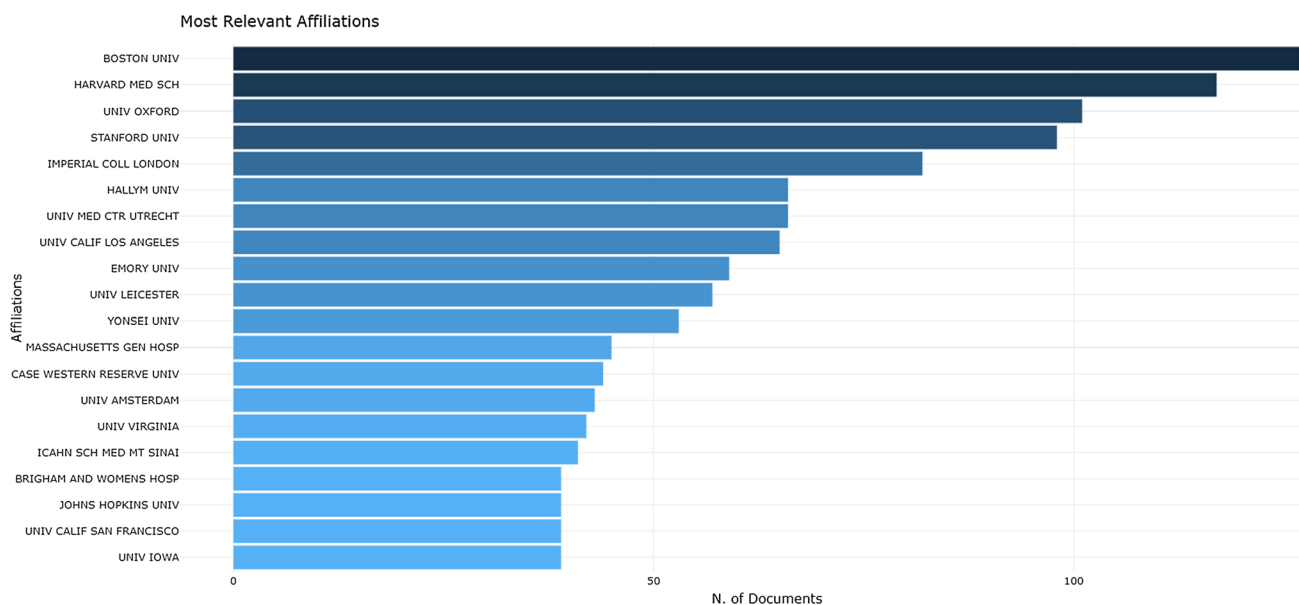


Fig. 12 Most relevant affiliations from their number of documents

Table 10 Country scientific production

Country	Total citations	Average article citations
USA	8837	21.606
China	1833	6.917
United Kingdom	1435	13.165
Korea	1176	15.892
Singapore	1046	55.053
India	771	5.797
Netherlands	551	14.892
Japan	515	19.074
Germany	492	13.667
Australia	450	15.517
Turkey	437	24.278
France	364	19.158
Saudi Arabia	317	13.783
Italy	296	9.867
Canada	284	9.793
Spain	238	11.333
Poland	147	21
New Zealand	140	14
Belgium	126	11.455
Iran	99	7.071

3.4.2 Most productive and cited affiliations

We find among the first five only universities from United States of America and United Kingdom to evaluate the most productive and most cited institutions. Thus, the most

productive and quoted institutions are from first to fifth: Boston University, Harvard Medical School, University of Oxford, Stanford University, and Imperial College of London (Table 9 and Fig. 12).

3.4.3 Most productive and cited country

As for the most cited countries, we see Singapore in the fourth position and the demotion of India in the sixth position. Thus, the most cited countries from the first to the fifth are the following: the United States of America, China, the United Kingdom, South Korea, and Singapore (Table 10; Fig. 13).

3.4.4 Most global cited papers and references

In this section, we will briefly present the five most cited articles by describing each of them with the general idea, the strong points, and the weak points:

- [34] Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. 10. 1038/s41551-018-0195-0
- *General idea:* This article uses DL models to predict cardiovascular risk factors from retinal fundus images. The objective is to help the scientific community understand cardiovascular disease processes and risk factors affecting patients' retinal vasculature or optic disc.
- *Strengths:* Retinal images are used as inputs to Neural Networks (NN) for varying degrees of predicting several cardiovascular risk factors. These models are trained on

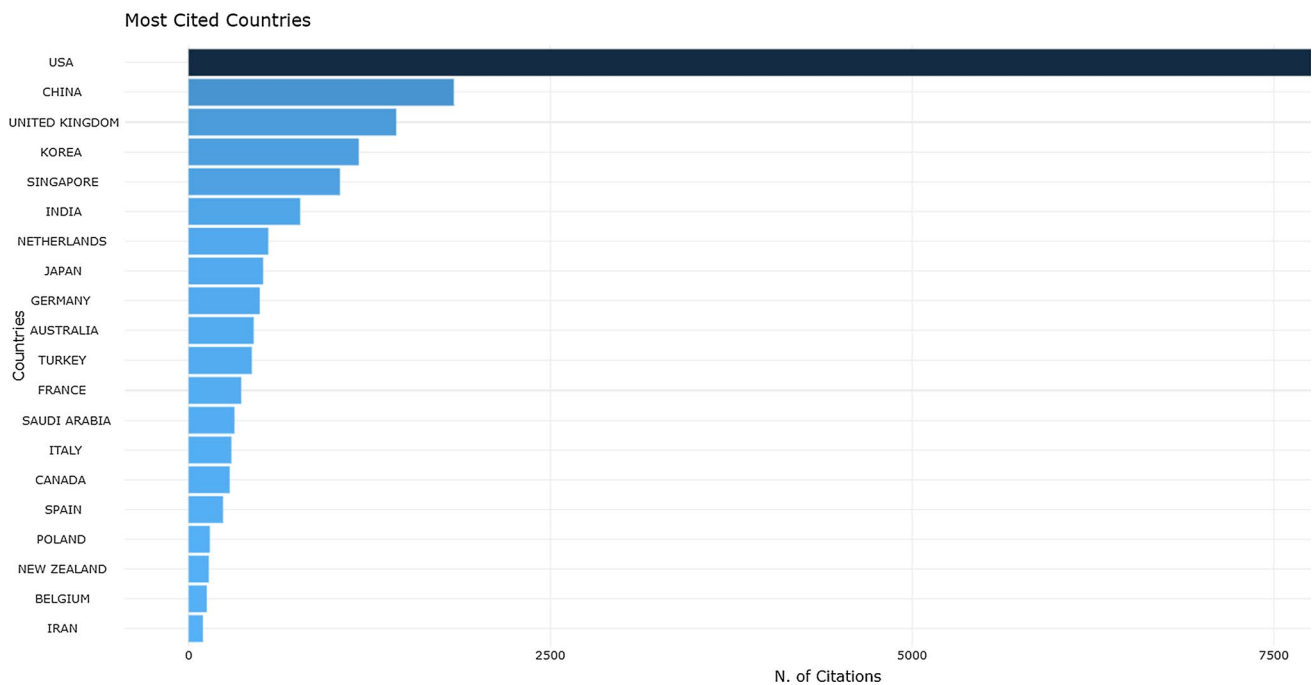


Fig. 13 Most cited countries

nearly 300,000 patients and validated on two independent databases of 12,000 and 1000 patients. The authors demonstrate that anatomical characteristics, such as the optic disc or blood vessels are sufficient to predict risks, such as age, gender, smoking status, systolic blood pressure, and major adverse cardiac events.

- **Weaknesses:** These predictions had a low coefficient of determination values, because the algorithm used does not allow us to predict these parameters with high accuracy. For example, in 78% of the cases, the age prediction had a margin of error of plus or minus five years, while the basic predictions had a margin of minus 5 years in only 44% of the cases.
- [32] A deep convolutional neural network model to classify heartbeats. 10. 1016/j.compbimed.2017.08.022
- **General idea:** This article describes the use of deep neural networks to distinguish heartbeats which can be subdivided into five categories, namely non-ectopic, ectopic supraventricular, ectopic ventricular, fusion, and unknown beats.
- **Strengths:** A methodology for automatically identifying the five categories of heartbeats in electrocardiogram (ECG) signals is proposed. A method includes pre-processing (noise suppression and ECG heartbeat segmentation), synthetic data generation (to overcome data imbalance in the five classes), and a convolutional neural network (rotational and translational invariance).
- **Weaknesses:** Training the neural network requires long hours and specialized hardware (GPU) to be effective, which is computationally expensive. Moreover, a vast number of images is necessary for this model training for the reliable recognition of several patterns.
- [32] Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. 10. 1016/j.ins.2017.06.027
- **General idea:** This article proposes a new approach to automatically detect myocardial infarction (MI) using electrocardiogram (ECG) signals, since the visual interpretation is difficult due to their low amplitude and duration.
- **Strengths:** In this study, a convolutional neural network (CNN) algorithm is implemented for the automated detection of ordinary and MI ECG (with noise and without noise). This indicates that the proposed CNN model can understand the underlying structure and recognize the class of a possibly noisy ECG beat signal.
- **Weaknesses:** The performance and reliability of the proposed system could be improved by applying the bagging technique (bootstrap + aggregation) and obtaining more open-source data concerning other cardiovascular diseases, such as heart failure, hypertensive heart disease, and cardiomyopathy.
- [35] Artificial Intelligence in Precision Cardiovascular Medicine, 10. 1016/j.jacc.2017.03.571
- **General idea:** This article provides an overview of the application of AI in cardiovascular clinical care and discusses its potential role in facilitating precision cardiovascular medicine. Computing is highlighted involving

- self-learning with machine learning, pattern recognition, and natural language processing to simulate the operation of human thought processes.
- *Strengths*: The authors present the clinical implications of the application of artificial intelligence in cardiology. Five categories of use are mentioned: (i) exploration of new risk factors, (ii) classification of new phenotypes and genotypes, (iii) prediction of the risk of bleeding and stroke, (iv) identification of other stroke risk factors, and (v) prediction of left-ventricular ejection fraction.
 - *Weaknesses*: The challenges of using AI for physicians to generate reasoning hypotheses, perform extensive data analysis, and optimize AI applications in clinical practice are not explicitly described. However, ignoring these challenges of AI could overshadow its strengths and be a hindrance to genuinely reaching the era of precision cardiovascular medicine.
 - [36] Artificial Intelligence in Cardiology. 10.1016/j.jacc.2018.03.521
 - *General idea*: This article provides a selection of the relevant applications of AI methods in cardiology by identifying avenues for a pertinent integration of these methods in cardiovascular medicine later. The paper also reviews predictive modeling concepts relevant to cardiology, such as feature selection.

- *Strengths*: The document presents the workflow of machine learning in cardiology starting from a variety of sources (experimental, biological, environmental, portable, and clinical data) to perform four activities: (i) feature selection, (ii) the selection of the machine learning algorithm, (iii) the development of the model, and (iv) the evaluation of the model and its prediction.
- *Weaknesses*: There is a lack of formal description of the reasoning process that can combine deep learning numerical reasoning with medical expertise (e.g., knowledge of pathophysiology and clinical presentation). The avenues for interpreting the models and explaining the results are not detailed (Table 11; Fig. 14).

4 Research trends in AI applied to pulmonary hypertension

4.1 Keywords dynamics analysis and trend topics

The Keywords Plus WordCloud offers greater diversity, frequency, and density of keywords than the Author and Document Title Wordclouds. In addition, the WordCloud of Keywords plus highlights the keyword classification, while the WordClouds of keywords and titles highlight the keyword of deep learning.

Table 11 Most cited authors with the reference of their documents

Paper	DOI	Total citations	TC per year	Normalized TC
Poplin R, 2018, Nat Biomed Eng	10.1038/s41551-018-0195-0	555	111	11.5645
Acharya UR, 2017, Comput Biol Med	10.1016/j.compbio.2017.08.022	487	81.167	8.514
Acharya UR, 2017, Inf Sci	10.1016/j.ins.2017.06.027	363	60.5	6.3462
Krittanawong C, 2017, J Am Coll Cardiol	10.1016/j.jacc.2017.03.571	317	52.833	5.542
Johnson KW, 2018, J Am Coll Cardiol	10.1016/j.jacc.2018.03.521	305	61	6.3552
Choi E, 2017, J Am Med Inf Assoc	10.1093/jamia/ocw112	298	49.667	5.2098
Yildirim O, 2018, Comput Biol Med	10.1016/j.compbio.2018.09.009	291	58.2	6.0635
Saussele S, 2018, Lancet Oncol	10.1016/S1470-2045(18)30,192-X	282	56.4	5.876
Zhang J, 2018, Circulation	10.1161/CIRCULATIONAHA.118.034338	269	53.8	5.6051
Bai W, 2018, J Cardiovasc Magn Reson	10.1186/s12968-018-0471-x	249	49.8	5.1884
Han SS, 2018, J Invest Dermatol	10.1016/j.jid.2018.01.028	222	44.4	4.6258
Madani A, 2018, NPJ Digit Med	10.1038/s41746-018-0065-x	206	41.2	4.2924
Ambale-Venkatesh B, 2017, Circ Res	10.1161/CIRCRESAHA.117.311312	204	34	3.5664
Xiao C, 2018, J Am Med Inf Assoc	10.1093/jamia/ocy068	194	38.8	4.0424
Zeng X, 2019, Bioinformatics	10.1093/bioinformatics/btz418	190	47.5	8.337
Tan JH, 2018, Comput Biol Med	10.1016/j.compbio.2017.12.023	185	37	3.8548
Shameer K, 2018, HEART	10.1136/heartjnl-2017-311,198	173	34.6	3.6048
dey D, 2019, J Am Coll Cardiol	10.1016/j.jacc.2018.12.054	161	40.25	7.0645
COENEN A, 2018, Circ Cardiovasc Imaging	10.1161/CIRCIMAGING.117.007217	159	31.8	3.3131

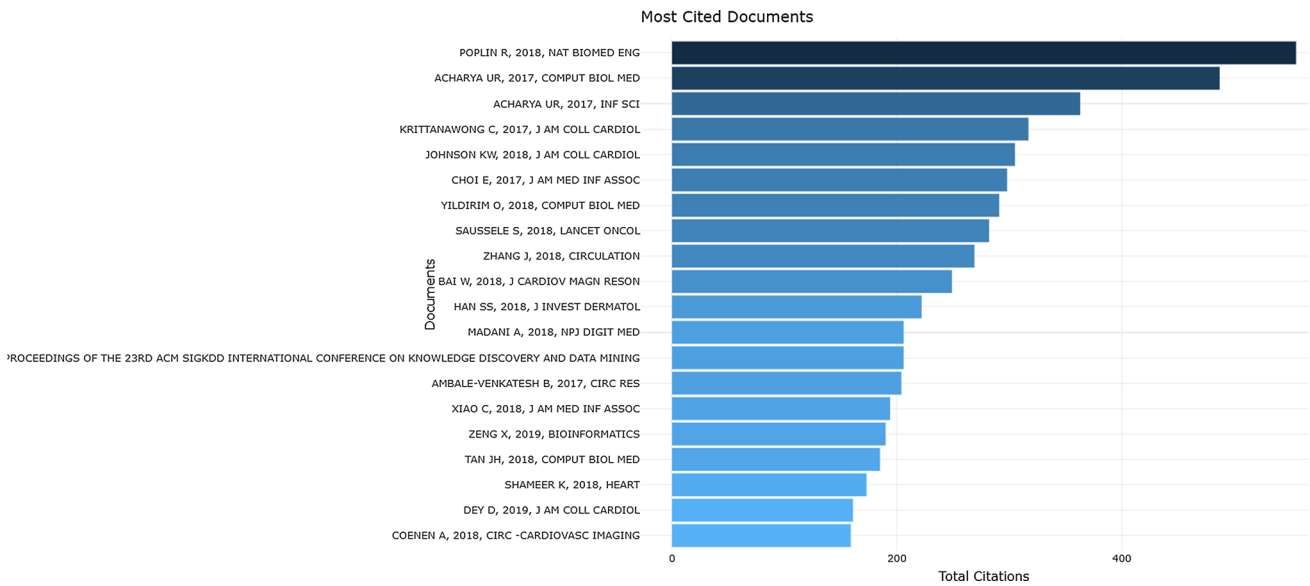


Fig. 14 Most cited documents based on their total citations

Fig. 15 WordCloud by keyword plus



Fig. 16 WordCloud by author's keyword



- *WordCloud by keyword plus* (see Fig. 15) in the center, we find the keyword classification, and in the first circle, the keywords disease and diagnosis appear. The keywords prediction, model, risk, heart, and segmentation are in the second circle. In the third circle, we find various keywords, including heart disease, neural networks, coronary artery disease, identification, features, validation, recognition, system, mortality, algorithm, association, and artificial intelligence.
- *WordCloud by author's keyword* (see Fig. 16) in the center, we find the keyword deep learning, and in the first circle, the keyword machine learning appears. The keywords convolution neural network and artificial intelligence are in the second circle. In the third cycle, there are various keywords, including electrocardiogram, segmentation, classification, learning, heart disease, and cardiovascular disease.
- *WordCloud by Titles* (see Fig. 17) in the center, we find the keywords deep and learning, and in the first circle, the words disease, neural, and heart appear. In the second circle are the keywords classification, network, segmentation, prediction, detection, cardiac, and neural. In the third cycle, there are various keywords, including analysis, images, risk, ECG, diagnosis, artificial, coronary, intelligence, convolutional, and cardiovascular.

In sum, these graphs (wordClouds) highlight the use of AI techniques such as ML and DL. In particular, we can note that supervised learning (precisely the classification case) is extensively applied to predict patient states in medicine.

4.1.1 Evolution curves and logarithmic frequency scale of words

The evolution curves (Fig. 18) of the five most relevant keywords (classification, diagnosis, disease, prediction, and risk) all have a parabolic shape with peaks that almost all appear around the year 2021. The bell shape of these curves is very marked, with peaks at high values relative to their neighborhoods. However, these five peaks have decreasing values from the first most relevant classification keyword to the fifth most relevant risk keyword.

The evolution curves of the following keywords have a less-marked bell shape with notably lower peaks close to neighboring values. We note, in this case, the presence of the five most frequent keywords in descending order: segmentation, heart, model, heart failure, and heart disease. Overall, there is a clear dominance of the first three keywords classification, diagnosis, and disease. On the other hand, the plots of curves of the different keywords are much closer to several intersects.

It can be observed that these themes almost all reach their maximum values in the interval between the years 2018 and 2022. This is described in more detail in Table 12.

Logarithmic scales have some advantages (e.g., useful for plotting rates of change, a wider range of data can be displayed and increased data for smaller values) but also disadvantages (e.g., easy to make errors plotting, difficult to analyze, zero cannot be plotted, and negative and positive values cannot be displayed on the same graph). The logarithmic function then makes it possible to obtain a representation of the evolution of the data with a graph whose values on the ordinate axis are less distant. The graph of the logarithm of frequency shows five major groups (Fig. 19):

Fig. 17 WordCloud by title



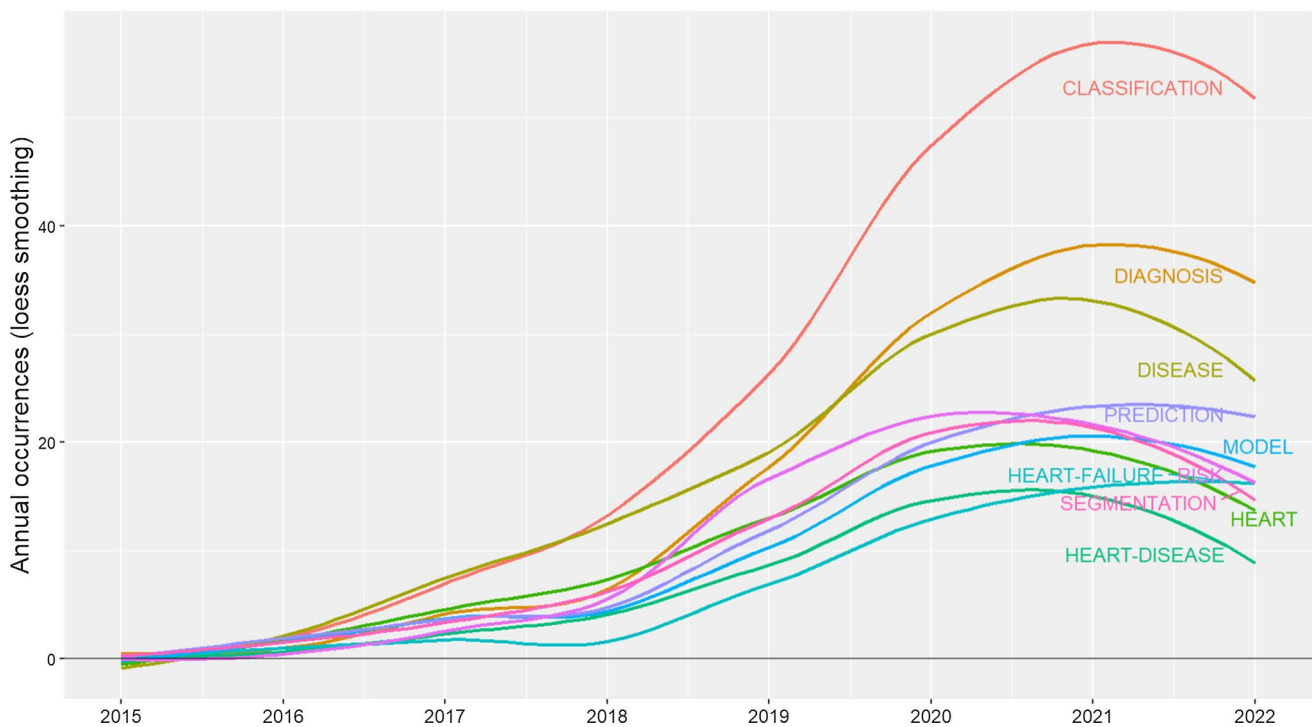


Fig. 18 Evolution curves of most cited words

- The *first group* formed around 2018 includes the keywords optical coherence, tomography, lumen diameter, level set, mutations, architectures, pressure, and meta-analysis.
- A *second group* formed around 2020 includes the keywords disease, prediction, risk, segmentation, and heart.
- A *third group* was formed around the year 2021, which includes the keywords classification and diagnosis.
- A *fourth group* was formed around the year 2021, which includes the keywords heart failure, systems, and features.
- A *fifth group* was formed around the year 2022, which includes the keywords feature extraction, cardiomyopathies, dysfunction, hypertension, brain, abnormalities, and time.
- *Driving themes* these are well-developed and important structuring themes in the field of research. In this case, validation is the main driving theme.
- *Niche themes* these are peripheral and highly specialized themes. In this case, segmentation is the main niche theme.
- *Emerging or declining themes* low centrality and low-density themes represent reducing or emerging themes. In this case, artificial intelligence is the leading emerging theme.
- *Basic themes* these are important themes without specific development but are fundamental, general, and transversal to the field of research. In this case, classification and disease are the two main basic themes.

4.1.2 Thematic map and conceptual structure

The thematic map is a graphic representation having as the axis of abscissas the concept of centrality and as axes of ordinates the idea of density. Centrality measures how closely a keyword network interacts with a keyword network. Density measures the degree of internal connection of keywords in a network. Themes can be categorized into four groups according to their centrality and density values [37] (Fig. 20):

To understand and interpret the themes and their relationships, we exploited word networks based on the Louvain algorithm [38] and Jaccard distance. Using this approach, we can extract the structure of the themes on a large network, as well as show that a given topic is common with other topics. Moreover, we are able to visualize the hierarchical relationship structure between different groups of themes. It is important to note that the size of the nodes is a function of the causal effects, and the thickness of the links indicates the degree of relationships between the themes. The relationships between words are stronger when the link thickness is higher. The number of words is 4500 and the minimum frequency is 50.

Table 12 Items and the year when the maximum value is reached

Item	Maximum value of frequency	Year
Classification	207	2021
Diagnosis	135	2021
Disease	130	2020
Prediction	88	2020
Risk	84	2020
Segmentation	82	2020
Heart	78	2020
Heart failure	56	2021
System	54	2021
Features	53	2021
Cardiomyopathy	9	2022
Optical coherence tomography	9	2019
Brain	8	2022
Dysfunction	8	2022
Feature- extraction	8	2022
Level set	8	2019
Abnormalities	7	2022
Hypertension	7	2022
Lumen diameter	7	2019
Mutations	7	2019
Time	7	2022
Architectures	6	2018
Deep-vein thrombosis	6	2019
Gene expression	6	2019
Multicenter	6	2019
Pressure	6	2018
Pulmonary embolism	6	2019
Radiographs	6	2019
Thrombosis	6	2019
Metaanalysis	5	2018

Figure 21 represents the word networks found on the documents collected between 1997 and 2023. In particular, it shows that the topics can be divided into three main groups. Therefore, the central node or engine of each network is, respectively, “classification”, “disease”, and “risk”. The densest network (blue) is composed of several topics, such as “classification”, “diagnosis”, “prediction”, “neural network”, and “feature”. The thickness of the links between topics shows that there is a strong relationship between these topics. Indeed, neural network models are often used to solve classification and diagnosis tasks. Mostly, it is used to predict the health status of the patient such as the variation of the heart rate.

The second network (green) is formed by “disease”, “segmentation”, “heart”, “artificial intelligence”, and “performance”. Finally, the last cluster (red) is composed of “heart failure”, “risk”, “validation”, “association”, and “mortality”.

By grouping words, we can reduce the number of topics that are similar or strongly related to each other.

The conceptual structure aims to provide a graphical representation of the analysis of the co-occurrence of keywords observed in a bibliographic collection. This analysis can be done using reduction techniques such as Multiple Correspondence Analysis (MCA) (Fig. 22).

The conceptual design consists of clusters in which each color describes the outlines of a cluster. In the case considered, two groups are identified:

- *The first cluster of red color* has inside the words MRI, diagnosis, prediction, classification, or risk. The word signals, ECG, coronary heart disease, management, association, big data, models, and convolutional neural network are included on its borders.
- *The second cluster of blue color* has inside the word’s performance, coronary artery disease, quantification, heart, and CT angiography, and includes on its borders the terms artificial intelligence, angiography, computed tomography, low fractional reserve, left ventricle, and segmentation.

4.2 AI models and ethical issues

4.2.1 Ethical considerations of artificial intelligence in cardiology

In the health sector, many achievements or initiatives have taken place intending to implement and develop AI technology in the sector, including cardiology. These different projects were based on several objectives. For these purposes, we mention the data collection, data analysis, and data mining with the ethical considerations comprising inclusion, biases, data protection, and proper implementation [39]. The exploitation of AI in healthcare has many ethical implications. In the past, decision-making was done by humans. However, intelligent machines assist or help healthcare providers make decisions [40]. These machines raise questions of accountability, transparency, authorization, and confidentiality. The ethical issues [41] related to the application of AI to healthcare and arose for each of these activities are:

- *Data collection using connected objects (IoT)* Indeed, and as is the case in most sectors of activity, connected objects are excellent means of collecting data in a short time. Thanks to the detection algorithms and the various sensors contained therein, it is possible to gather many parameters on a patient, his habits, or the symptoms of a disease. The main ethical consideration that must be considered relates to the need to obtain the patient’s informed consent, since any access to private medical

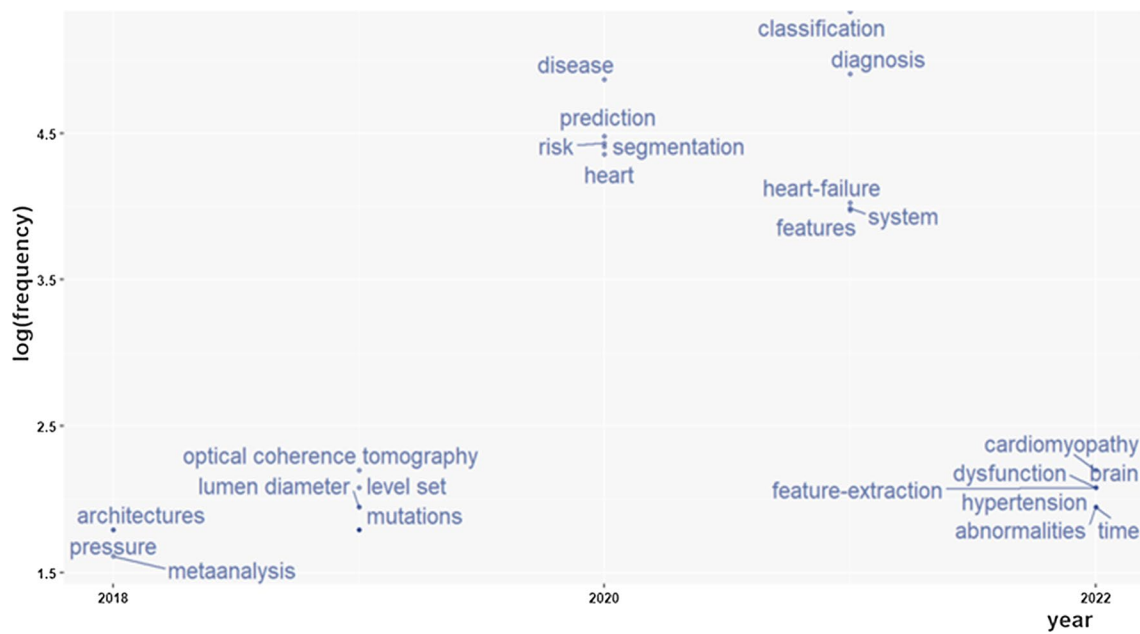


Fig. 19 Logarithms of word frequencies

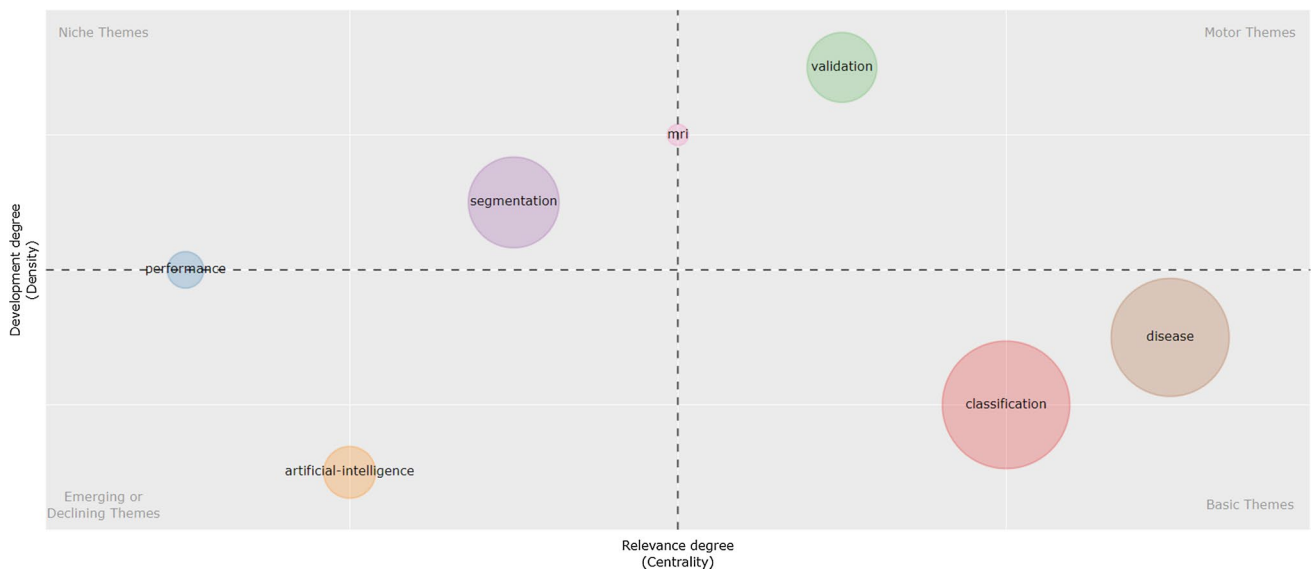


Fig. 20 Thematic map using centrality and density

data collected and transferred requires the permission of the person concerned or his legal representative.

- *Data analysis using neural networks* Indeed, neural networks are one of the ideal technologies for processing massive data. Thanks to the machine learning algorithms in place hospitals can carry out tests, precise and preventive analyses, and more efficient treatment methods to treat their patients and treat serious illnesses. In addition, thanks to these same neural networks, they can also

perform predictive analyses of the latter to anticipate the consequences of applying a product or treatment method. The main ethical consideration that must be considered concerns the accuracy of the algorithmic treatments used by limiting potential biases, such as those related to the gender, age, and ethnicity of the patient.

- *The exploitation of data thanks to machine learning* Indeed, the machine learning of neural networks is an excellent way to exploit the analyzes carried out on medi-

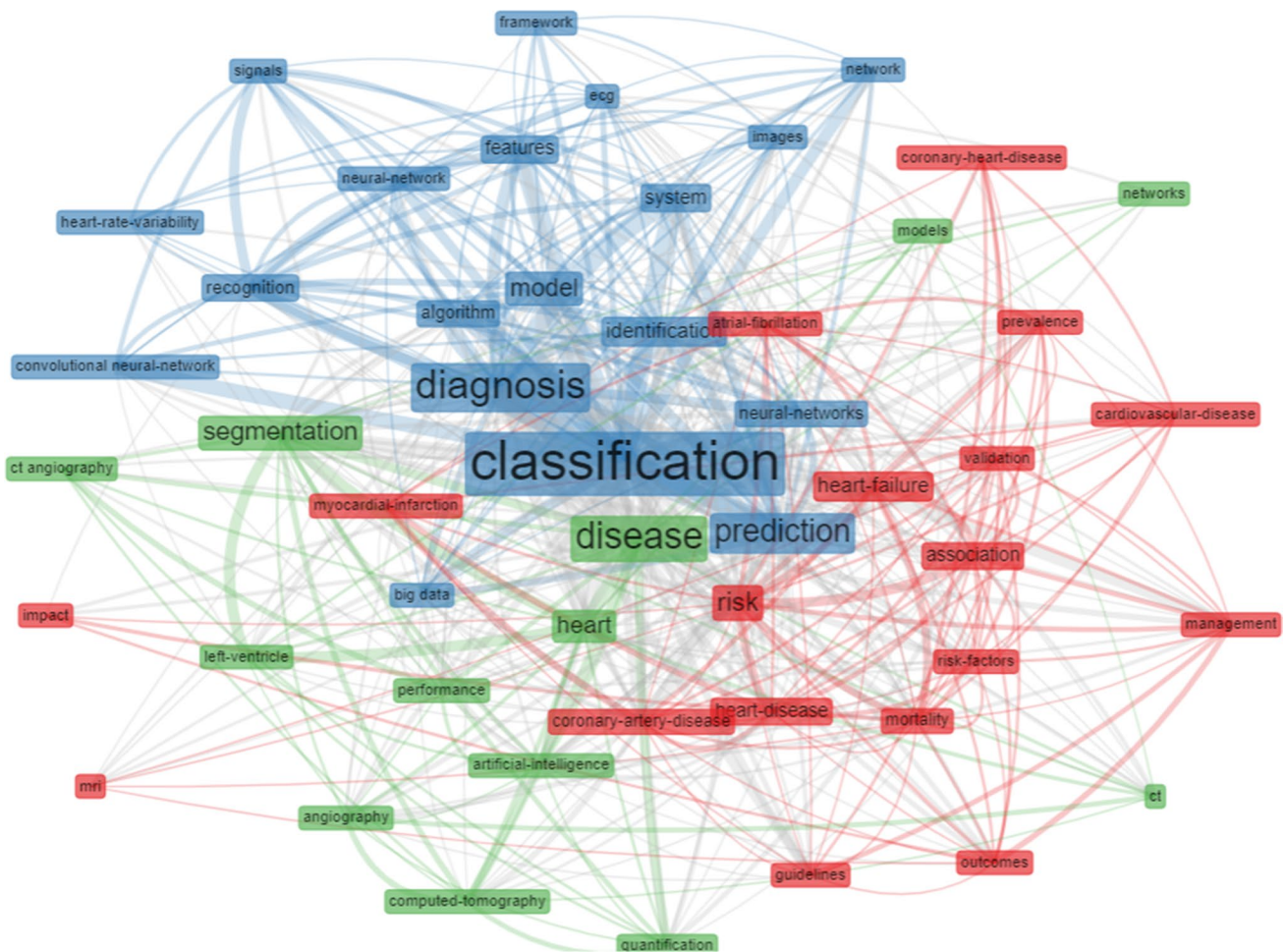


Fig. 21 Graphical representation of the keywords network. The size of each node is proportional to its degree and the thickness of the links represents the strength of the link

cal data. Thanks to the different learning models and the different methods of using the algorithms of neural networks, organizations in the health sector (health centers and hospitals) are now able to perform precision and data prediction analyses to optimize the disease treatment process and anticipate the appearance of critical situations. The main ethical consideration that must be considered concerns the communication to the patient of any use of his private data, explaining to him beforehand how this data will be used to be able to guarantee the strictly medical nature of the target use.

- *Mistakes resulting from IA algorithms* These AI algorithms can be subject to errors in predicting patient states. In this case, it is difficult to establish or attribute responsibility. In addition, systems based on artificial intelligence tools can be biased by the data. For example, they could predict a high probability of disease based on data that are not causal factors such as the gender or

race of the patient. Authorities or governments should establish laws to address AI's ethical and responsibility issues. By way of illustration, the French government issued decree no. 2017–330 of March 14, 2017, relating to the rights of persons subject to individual decisions taken on the basis of algorithmic processing [42]. In this decree, article R. 311-3-1-1 explicitly mentions the right to obtain communication of the rules defining this algorithmic processing and the main characteristics of its implementation, as well as the procedures for exercising this right to communication and referral to the commission for access to administrative documents. This will help to better regulate the key problems and governance mechanisms to limit these AI systems' negative implications.

- *Use of medical AI and anti-discrimination laws* Besides the challenge mentioned above, medical AI also faces issues of equality and discrimination [43]. Particularly

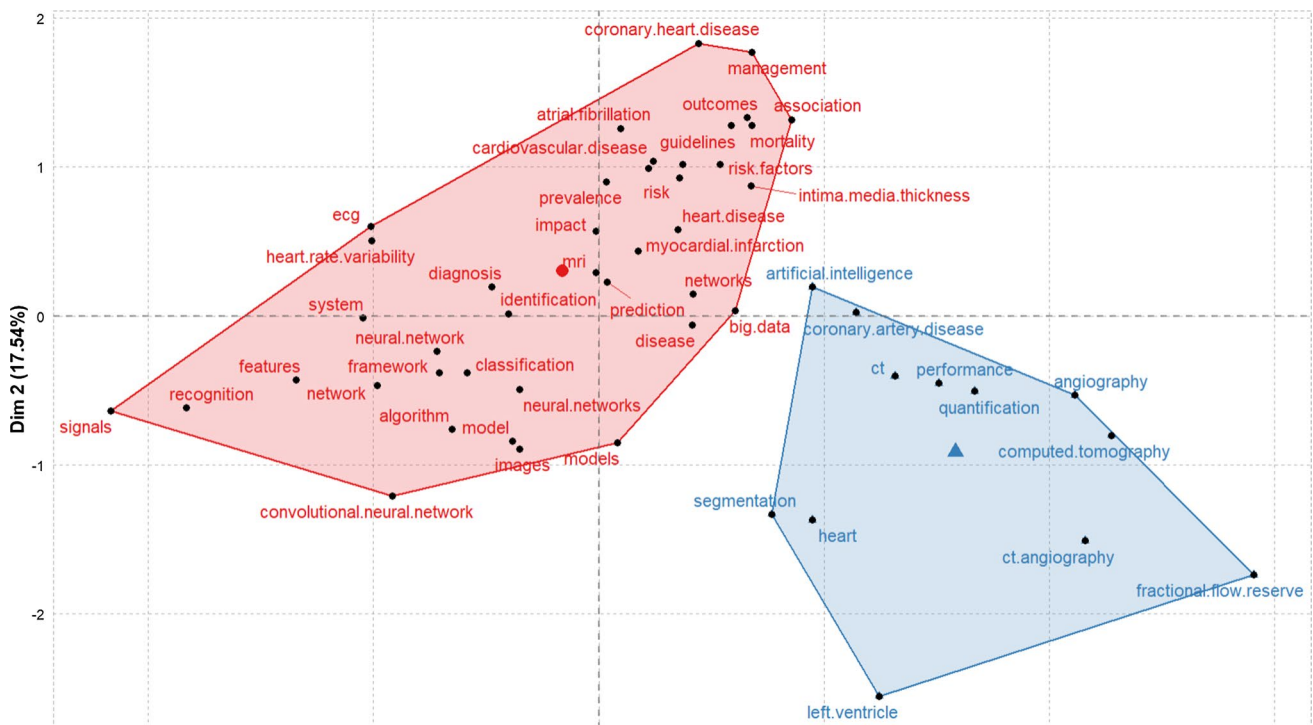


Fig. 22 Conceptual structure of co-word analysis

with respect to gender, race, children, the elderly, and people with disabilities. In addition, the data exploited should be carefully reviewed to detect any form of bias [10]. AI-based models must be autonomous, accessible, and monitored for potential forms of unfairness (equity). Therefore, to avoid misinterpreting AI model results, we need to take into account the diversity of people. The validity of the results must be especially critically questioned when medical AI is applied to specific population groups that are poorly represented in the training data [44].

4.2.2 Ethical issues related to GDPR in medical AI

The positioning of the notion of ethics seems to be relatively changing according to the evolution of societal considerations on the exploitation of data. In this case, the protection of personal data (GDPR) has introduced the obligation to develop the so-called transparent processing. In terms of processing and algorithms, this notion is associated with the concept of explainability [45]. This fundamental notion consists of being able to trace the decision-making process of the automated processing from the input data to the output result. However, transparency takes on other aspects, on which the regulation does not yet provide a framework, such as transparency in the choices made by data scientists in

terms of data transformation. As a result, many other factors should be considered to measure the risks and ethical problems generated by algorithmic processing that can support decisions made automatically. Although some stakeholders do not understand the technical details of IA approaches, we have to ensure that all information related to the data is simple and easy to access and understand.

Besides, current health data privacy regulations in the United States of America (USA) address only partially the critical right of patients to have access to digital health data. To guarantee ethical guidelines, patients must have unrestricted access to the raw data that is collected via any computing system. [46]. Thus, it is necessary to assess the impact of the models on the respect of data protection before their release (e.g., rights and freedoms of individuals, risk of infringement of physical or mental health, and privacy). Thus, it is necessary to assess the impact of the models on the respect of data protection before their release (e.g., rights and freedoms of individuals, risk of infringement of physical or mental health, and privacy).

4.2.3 Difference between the DL and ML in the context of study

In computational medicine and data science, ML and DL are used to address the various prediction tasks. Machine

learning is a subset of AI, this approach has been around for a long time, as a set of techniques that enables programs from datasets or correction signals. ML are arousing great interest in some application such as medical data analysis. These models are mostly used, because their mathematical law and decision rule can be known by users. However, ML is subject to degradation and inefficiency (e.g., missing data). Moreover, it is inefficient for processing unstructured data such as blurry or masked images. Furthermore, ML suffers from over-fitting and stability during the learning process; thus, the predictions can be distorted. The recent scientific and technological progress has made deep neural network learning appear as a real advance over the traditional ML.

Deep Learning is a particularly recent ML inspired by the neural networks of human brains. DL uses sophisticated computer programming and training to learn about hidden patterns in larger datasets. Besides, it is more efficient to train unstructured and it can also process high-dimensional data more quickly than ML methods. However, these methods are also called “black box” opaque models. In this context, the decision-making rules are not systematically known by the practitioners. Furthermore, these models generate results that are often incomprehensible or difficult to explain. Thus, it is important to understand the process or mathematical rules produced by the model to ensure their reliability and improve their performance and potential ethical issues.

At the ethical level, the capacities for explaining and interpreting the decisions suggested by a machine learning system are relative to the architectures used. These being more complex in deep learning, we are often faced with the dilemma between efficiency and fairness of the reasoning resulting from the interweaving of a multitude of learning layers. This is particularly noticeable when sensitive information with limited interpretations or explanations can be derived from the merging of protected and unprotected features.

5 Discussion

5.1 AI interest of the bibliometric analysis

The realization of this bibliometric study of the applications of AI to the problem of Pulmonary Hypertension. However, the observations made are to be moderated in view of the potential biases influencing the results and arising from technological developments and trending or sensational subjects. These biases can be induced by the scientific database used, the construction of the query to retrieve the data, and the time span defined. Despite these reservations, we have made

some observations deserving particular attention and which are described below:

- The first observation is that the collected documents do not specifically mention this disease, but the cardiovascular system more regularly. This can be explained by the fact that cardiovascular diseases are often analyzed in their entirety before studying conditions affecting their elements, including the heart and vessels (arteries and veins).
- The second observation is linked to the learning considered. We note a clear dominance of documents on supervised learning, which concerns both classification problems with discrete output and regression problems with continuous output. However, works with unsupervised learning are few. This could be due to the higher level of difficulty presented by unsupervised learning problems.
- The third observation concerns the five most cited publications consisting of three specific studies and two literature review articles. The first most cited article proposes an original and surprising approach to predicting cardiovascular risk factors from photographs of the retinal fundus by deep learning. At the same time, the second most cited paper is about deep learning to classify heartbeats. On the other hand, the third most cited paper proposes deep learning for automated detection of myocardial infarction using ECG signals.
- The fourth observation relates to the limitations of this study, which mainly used the Web of Sciences database and, peripherally, the Scopus database. It would be interesting to conduct other additional studies where the roles would be reversed or even others, including the Google Scholar database.
- The fifth observation highlights that despite the success of AI models in medical applications, some questions arise about the ethics and responsibility of these models. These ethical principles impact the dimension and nature of the data and information used to extract new knowledge and support decision-making systems.

5.2 AI models and ethical issues

Despite the previously mentioned advantages, this approach has some issues.

In the context of the acceptability of AI applied to disease prediction in medicine, there is a need for collaboration between all stakeholders (e.g., data scientists, medical personnel, etc.). This supports enforcement and compliance with legislation that addresses the impact of AI-based autonomous machines and sensitive issues related to societal, ethical, and privacy factors. The perspectives adopted in bibliometric analysis need the verification of the consistency

of indicators and the use of data sources that can generate biases. In fact, the results obtained can be dependent on the scientific databases (Scopus and Web of Science) and any filters performed (languages of the writing of the document, year of publication construction of the query). According to the databases and the types of indicators, a researcher can receive different measures of impact. Moreover, the results can be impacted by database quality (incorrectly indexed articles, duplicate articles, and synonym problem of the produced authors). We need to perform a careful pre-processing of the data prior to beginning the analyses.

To make the indicators more robust, we can combine several databases and remove redundant publications. In addition, the “impact factor”, which is used to evaluate the quality of an article and the importance of a given journal for the scientific community, can lead to a distorted interpretation or conclusion (e.g., articles published in journals with a high-impact factor tend to be of higher quality than those in journals with a low impact factor) [47]. Besides, the exploitation of this factor tends to indicate that a paper published in a high-impact journal receives the most citations. In contrast, in their study, the authors show that the most productive article is not automatically the most cited [48].

Besides, the nature of the bibliometric methodology could be a potential limitation. To address this issue, we can exploit a combination of bibliometric approaches. Although the results of the analysis are quantitative in nature, the relationship between quantitative and qualitative results can be complicated [49]. Therefore, we must be more careful when concluding findings. Finally, bibliometric studies can only provide short-term predictions for research areas; thus, researchers are cautioned against making overambitious statements about research areas and their long-term implications.

6 Conclusion

The use of artificial intelligence via the management of large amounts of data will allow assistance on several levels: (i) a help with risk stratification to more accurately estimate a patient's level of risk; (ii) an aid in imaging to precisely recognize anomalies; (iii) help to detect or screen for heart rhythm problems; (iv) help in the use of algorithms in the follow-up of specific patients.

In this document, we have been guided by a desire to provide an overview of research on AI's applications to the Pulmonary Hypertension problem.

- The first conclusion is at the level of the group of five journals that stand out for a high number of publications on the topics of AI applied to cardiology: IEEE Access, Computers in Biology and Medicine, Biology Signal

Processing and Control, *Frontiers in Cardiovascular Medicine*, *Sensors*.

- The second conclusion concerns scientific production, with the dominance of five countries: United States of America (e.g., Boston, Harvard, and Stanford), India (e.g., Visvesvaraya National Institute of Technology Nagpur), United Kingdom (e.g., Imperial College London and the University of Oxford), and South Korea (e.g., Hallym University and Yonsei University) and Singapore (e.g., Ngee Ann Polytechnic). In addition, collaborations between these affiliations generate several joint publications.
- The third conclusion is related to the top ten keywords that appear most frequently. The first group of the five most relevant keywords (classification, diagnosis, disease, prediction, and risk) has a parabola-like plot with high peaks around 2021. The second group of the other five most frequent keywords (segmentation, heart, model, heart failure, and heart disease) also have a parabola-like pattern with lower peaks around mid-2020.
- The fourth conclusion: AI models applied to the medical field is highly related to questions of GDPR, human rights, privacy, and autonomy. It is thus an area of AI with legislative developments in progress.

AI approaches correlate with massive data or big data through connected medical objects, and we need to analyze and monitor these data over time. In the context of personalized medicine, the data from these adapted analyses will make it possible to refine diagnoses and treatments. The breakthroughs and innovations of this AI began with precision medicine initiatives, which are tangible, in particular on the use of portable devices (e.g., connected watches) that make it possible to detect cardiovascular abnormalities such as arrhythmias which were not captured during too sequential recordings.

Declarations

Conflict of interest There is no conflict of interest that might bias the described work: no financial, no personal, and no other relationships with other people or organizations within 3 years of beginning the work submitted. We verify that the data contained in the manuscript being submitted have not been previously published, have not been submitted elsewhere, and will not be submitted elsewhere while under consideration at *AI and Ethics—Journal*. We verify that all authors have reviewed the contents of the manuscript being submitted, approve of its contents, and validate the accuracy of the data. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Waring, J., Lindvall, C., Umeton, R.: Automated machine learning: review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* **104**, 101822 (2020). <https://doi.org/10.1016/j.artmed.2020.101822>
- Tian, Y., Zhao, X., Huang, W.: Meta-learning approaches for learning-to-learn in deep learning: a survey. *Neurocomputing* **494**, 203–223 (2022). <https://doi.org/10.1016/j.neucom.2022.04.078>
- Thirunavukarasu, R., George Priya Doss, C., Gnanasambandan, R., Gopikrishnan, M., Palanisamy, V.: Towards computational solutions for precision medicine based extensive data healthcare system using deep learning models: a review. *Comput. Biol. Med.* **149**, 106020 (2022). <https://doi.org/10.1016/j.compbiomed.2022.106020>
- Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., Uchibe, E., Morimoto, J.: Deep learning, reinforcement learning, and world models. *Neural Netw.* **152**, 267–275 (2022). <https://doi.org/10.1016/j.neunet.2022.03.037>
- Dong, S., Wang, P., Abbas, K.: A survey on deep learning and its applications. *Comput. Sci. Rev.* **40**, 100379 (2021). <https://doi.org/10.1016/j.cosrev.2021.100379>
- LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436–444 (2015). <https://doi.org/10.1038/nature14539>
- Mehta, N., Pandit, A., Shukla, S.: Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. *J. Biomed. Inform.* **100**, 103311 (2019). <https://doi.org/10.1016/j.jbi.2019.103311>
- Tjoa, E., Guan, C.: A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(11), 4793–4813 (2020)
- Caroprese, L., Vocaturro, E., Zumpano, E.: Argumentation approaches for explainable AI in medical informatics. *Intell. Syst. Appl.* **16**, 200109 (2022)
- Schneeberger, D., Stöger, K., Holzinger, A.: The European legal framework for medical AI. In: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pp. 209–226. Springer, Cham (2020)
- Schneeberger, D., Stöger, K., Holzinger, A.: The European legal framework for medical AI. In: *Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4*, pp. 209–226. Springer International Publishing, (2020)
- Guo, Y., Hao, Z., Zhao, S., Gong, J., Yang, F.: Artificial intelligence in health care: bibliometric analysis. *J. Med. Internet Res.* **22**(7), e18228 (2020)
- Martinez-Millana, A., Saez-Saez, A., Tornero-Costa, R., Azzopardi-Muscat, N., Traver, V., Novillo-Ortiz, D.: Artificial intelligence and its impact on the domains of universal health coverage, health emergencies and health promotion: an overview of systematic reviews. *Int. J. Med. Inform.* **166**, 104855 (2022). <https://doi.org/10.1016/j.ijmedinf.2022.104855>
- Panayides, A.S., Amini, A., Filipovic, N.D., Sharma, A., Tsafaris, S.A., Young, A., et al.: AI in medical imaging informatics: current challenges and future directions. *IEEE J. Biomed. Health Inform.* **24**(7), 1837–1857 (2020)
- Koskinen, J., Isohanni, M., Paajala, H., Jääskeläinen, E., Nieminen, P., Koponen, H., et al.: How to use bibliometric methods in evaluation of scientific research? An example from Finnish schizophrenia research. *Nord. J. Psychiatry* **62**(2), 136–143 (2008)
- Dhanalakshmi, Y., Babu, I.R.: Intrusion detection using data mining along fuzzy logic and genetic algorithms. *Int. J. Comput. Sci. Netw. Secur.* **8**(2), 27–32 (2008)
- Ab Razak, M.F., Anuar, N.B., Salleh, R., Firdaus, A.: The rise of “malware”: bibliometric analysis of malware study. *J. Netw. Comput. Appl.* **75**, 58–76 (2016)
- Li, T., Ho, Y.S., Li, C.Y.: Bibliometric analysis on global Parkinson’s disease research trends during 1991–2006. *Neurosci. Lett.* **441**(3), 248–252 (2008)
- Dehghanbanadaki, H., Seif, F., Vahidi, Y., Razi, F., Hashemi, E., Khoshmirisafa, M., Aazami, H.: Bibliometric analysis of global scientific research on coronavirus (COVID-19). *Med. J. Islam Repub. Iran* **34**, 51 (2020)
- Boudry, C., Baudouin, C., Mouriaux, F.: International publication trends in dry eye disease research: a bibliometric analysis. *Ocul. Surf.* **16**(1), 173–179 (2018)
- Gorraiz, J., Schloegl, C.: A bibliometric analysis of pharmacology and pharmacy journals: scopus versus web of science. *J. Inf. Sci.* **34**(5), 715–725 (2008)
- Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **24**(5), 513–523 (1988)
- Roussel, A., Dellamonica, D., Menuet, R., Pineda Armando, L., Sabatine, M.S., Giugliano, R.P., Trichelair, P., Zaslavskiy, M., Ricci, L.: Can machine learning bring cardiovascular risk assessment to the next level? A methodological study using Fourier trial data. *Eur. Heart J. Digit. Health* **3**(1), 38–48 (2022). <https://doi.org/10.1093/ehjdh/ztab093>
- Remy-Jardin, M., Ryerson, C.J., Schiebler, M.L., et al.: Imaging of pulmonary hypertension in adults: a position paper from the Fleischner Society. *Eur. Respir. J.* **57**, 2004455 (2021). <https://doi.org/10.1183/13993003.04455-2020>
- Nickerson, R.S.: Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2**(2), 175–220 (1998)
- Garfield, E., Sher, I.H.: Keywords plus algorithmic derivative indexing. *J. Am. Soc. Inf. Sci.* **44**(5), 298–299 (1993)
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., Duan, Z.: Comparing keywords plus of WOS and author keywords: a case study of patient adherence research. *J. Am. Soc. Inf. Sci.* **67**(4), 967–972 (2016)
- Huang, C.-H., Liu, J.S., Ho, M.H.-C., Chou, T.-C.: Towards more convergent main paths: a relevance-based approach. *J. Inform.* **16**(3), 101317 (2022). <https://doi.org/10.1016/j.joi.2022.101317>
- Suri, J.S., Agarwal, S., Carriero, A., Paschè, A., Danna, P.S., Columbu, M., et al.: COVLIAS 1.0 vs. MedSeg: artificial intelligence-based comparative study for automated COVID-19 computed tomography lung segmentation in Italian and Croatian Cohorts. *Diagnostics* **11**(12), 2367 (2021)
- Saba, L., Sanagala, S.S., Gupta, S.K., Koppala, V.K., Johri, A.M., Khanna, N.N., et al.: Multimodality carotid plaque tissue characterization and classification in the artificial intelligence paradigm: a narrative review for stroke application. *Ann. Transl. Med.* **9**(14), 1206 (2021)
- Khanna, N.N., Jamthikar, A.D., Gupta, D., Piga, M., Saba, L., Carcassi, C., et al.: Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning-based tissue characterization. *Curr. Atheroscler. Rep.* **21**(2), 1–14 (2019)
- Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San Tan, R.: A deep convolutional neural network model to classify heartbeats. *Comput. Biol. Med.* **89**, 389–396 (2017)
- Biswas, M., Saba, L., Omerzu, T., Johri, A.M., Khanna, N.N., Viskovic, K., et al.: A review on joint carotid intima-media thickness and plaque area measurement in ultrasound for cardiovascular/stroke risk monitoring: artificial intelligence framework. *J. Digit. Imaging* **34**(3), 581–604 (2021)
- Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R.: Prediction of

- cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**(3), 158–164 (2018)
35. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., Kitai, T.: Artificial intelligence in precision cardiovascular medicine. *J. Am. Coll. Cardiol.* **69**(21), 2657–2664 (2017)
 36. Johnson, K.W., Torres Soto, J., Glicksberg, B.S., Shameer, K., Miotto, R., Ali, M., Ashley, E., Dudley, J.T.: Artificial intelligence in cardiology. *J. Am. Coll. Cardiol.* **71**(23), 2668–2679 (2018)
 37. Cobo, M.J., Martínez, M.A., Gutiérrez-Salcedo, M., Fujita, H., Herrera-Viedma, E.: 25 years at knowledge-based systems: a bibliometric analysis. *Knowl. Based Syst.* **80**, 3–13 (2015)
 38. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**(10), P10008 (2008)
 39. Martin, C., DeStefano, K., Haran, H., Zink, S., Dai, J., Ahmed, D., Razzak, A., Lin, K., Kogler, A., Waller, J., Kazmi, K., Umair, M.: The ethical considerations including inclusion and biases, data protection, and proper implementation among AI in radiology and potential implications. *Intell. Based Med.* (2022). <https://doi.org/10.1016/j.ibmed.2022.100073>
 40. Davenport, T., Kalakota, R.: The potential for artificial intelligence in healthcare. *Future Healthc. J.* **6**(2), 94 (2019)
 41. Safdar, N.M., Banja, J.D., Meltzer, C.C.: Ethical considerations in artificial intelligence. *Eur. J. Radiol.* **122**, 108768 (2020). <https://doi.org/10.1016/j.ejrad.2019.108768>
 42. Desmoulin-Canselier, S., Le Métayer, D.: Algorithmic decision systems in the health and justice sectors: certification and explanations for algorithms in European and French law. *Eur. J. Law Technol.* **9**(3) (2018)
 43. Hoffman, S., Podgurski, A.: Artificial intelligence and discrimination in health care. *Yale J. Health Pol'y L. Ethics* **19**, 1 (2019)
 44. FRA, D.: Quality and artificial intelligence—mitigating bias and error to protect fundamental rights. European Union Agency for Fundamental Rights (2019)
 45. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020)
 46. Cohen, I.G., Gerke, S., Kramer, D.B.: Ethical and legal implications of remote monitoring of medical devices. *Milbank Q.* **98**(4), 1257–1289 (2020)
 47. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., Lim, W.M.: How to conduct a bibliometric analysis: an overview and guidelines. *J. Bus. Res.* **133**, 285–296 (2021)
 48. Keleko, A.T., Kamsu-Foguem, B., Ngouna, R.H., Tongne, A.: Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis. *AI Ethic.* **2**, 553–577 (2022)
 49. Wallin, J.A.: Bibliometric methods: pitfalls and possibilities. *Basic Clin. Pharmacol. Toxicol.* **97**(5), 261–275 (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.