



# The Potential of Big Data Research in HealthCare for Medical Doctors' Learning

Manuel Au-Yong-Oliveira<sup>1</sup> · Antonio Pesqueira<sup>2</sup> · Maria José Sousa<sup>3</sup> · Francesca Dal Mas<sup>4</sup> · Mohammad Soliman<sup>5,6</sup>

Received: 29 September 2020 / Accepted: 7 December 2020

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

The main goal of this article is to identify the main dimensions of a model proposal for increasing the potential of big data research in Healthcare for medical doctors' (MDs') learning, which appears as a major issue in continuous medical education and learning. The paper employs a systematic literature review of main scientific databases (PubMed and Google Scholar), using the VOSviewer software tool, which enables the visualization of scientific landscapes. The analysis includes a co-authorship data analysis as well as the co-occurrence of terms and keywords. The results lead to the construction of the learning model proposed, which includes four health big data key areas for MDs' learning: 1) data transformation is related to the learning that occurs through medical systems; 2) health intelligence includes the learning regarding health innovation based on predictions and forecasting processes; 3) data leveraging regards the learning about patient information; and 4) the learning process is related to clinical decision-making, focused on disease diagnosis and methods to improve treatments. Practical models gathered from the scientific databases can boost the learning process and revolutionise the medical industry, as they store the most recent knowledge and innovative research.

**Keywords** Big data · Systematic literature review · MDs · Learning · VosViewer analysis

## Introduction and conceptualisation

The new technologies and the consequent increasing scale of data being generated in the healthcare sector are forcing healthcare professionals, and especially medical doctors (MDs), to understand how to connect data for new growth opportunities, enabling greater efficiency in diagnosis and research activities, resulting in better addressing the patients' needs [1].

The current literature is highlighting how healthcare analytics and data decision processes are becoming more and more critical [2]. Data allows healthcare organisations like hospitals and clinics, to maximise innovation and transformation in several different ways, affecting the decision-making process of healthcare practitioners, including MDs. Technologies like Big Data analytics and Artificial Intelligence (AI) are supporting the management of complex situations, like new treatment options for chronic diseases [3,

---

This article is part of the Topical Collection on *Education & Training*

✉ Manuel Au-Yong-Oliveira  
mao@ua.pt

Antonio Pesqueira  
antonio.pesqueira@live.com

Maria José Sousa  
maria.jose.sousa@iscte-iul.pt

Francesca Dal Mas  
fdalmas@lincoln.ac.uk

Mohammad Soliman  
msoliman.sal@cas.edu.om

<sup>1</sup> INESC TEC, GOVCOPP, Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro, Aveiro, Portugal

<sup>2</sup> Bavarian Nordic A/S, Hellerup, Denmark

<sup>3</sup> ISCTE Instituto Universitário de Lisboa, Lisbon, Portugal

<sup>4</sup> Lincoln International Business School, University of Lincoln, Lincoln, UK

<sup>5</sup> University of Technology and Applied Sciences, Salalah CAS, Salalah, Oman

<sup>6</sup> Faculty of Tourism & Hotels, Fayoum University, Fayoum, Egypt

4], less available human resources in hospitals [5] and a growing set of uncertain outcomes from viruses and bacterial phenomena, including the recent COVID-19 pandemic [6–9].

Even before the current pandemic crisis, the system was witnessing an escalating pressure on hospitals, healthcare facilities, and practitioners to be successively cost-effective [10], with increasing difficulty to justify any additional spending or investment in medical education or MD training [8, 11].

While the industry still tries to understand the best investments in medical education and clinical training, there are growing investments in new skill development around Big Data, Data Science, AI, and Machine Learning [12, 13]. Without exception, a growing need is emerging from healthcare professionals, especially MDs, to be better prepared for new systems, challenges, and demands from different spectrums of healthcare organisations (HCOs) and all their related key stakeholders.

Even though HCO decision-makers, administrators, and department heads recognise the importance of appropriate education in terms of technology and data management for MDs [14], continuing education and skills development in Big Data is often forgotten. While some professions call for continued education to meet Big Data needs and requirements, MDs are still not in the spotlight in terms of receiving that advanced training that can play a decisive role in earlier disease intervention, reduced probability of adverse reactions to drugs, less medical errors, determination of causalities, understanding of co-morbidity, and many other challenges [9]. Still, new complex situations call for new competencies and tools to support those in charge in the management of everyday medical practice. The literature has identified several issues and opportunities related to Big Data and learning for MDs and other healthcare professionals. While theoretical contributions may offer pioneering insights, they often fail in suggesting implications for practice.

Following these premises, there is a call to understand the state-of-the-art of the literature, but also how MDs learning about Big Data can be practically fostered and promoted to lead to better outcomes from the whole medical system. Therefore, starting from the results of a systematic literature review, our Research Question (RQ) is:

*RQ: Which are the main dimensions of a model proposal for increasing the potential of Big Data research in Healthcare for MDs' learning?*

## Big data research in healthcare

Big Data refers to the mass of structured and unstructured data generated worldwide [15]. In healthcare, this encompasses everything from electronic medical records, to internet-

connected (IoT) devices like medical wearables or augmented reality diagnostic tools [16, 17].

In general concepts and according to the broader literature, Big Data spans four dimensions that usually are called the 4 V's. The 4 V's are as follows: Volume, Veracity, Velocity, and Variety [18].

For any HCO, several different processes are being built on data and analytics to maintain a competitive and profitable edge in the face of all of the new regulations and governmental pressures for the maximisation of costs and results, but also given a growing need to improve performance, reduce errors in medical practice and discover new pathways for improved health outcomes [19].

Medical technologies and clinical activities are extremely complex and diverse. Big Data algorithms and data collection models are becoming more challenging than ever, where new wearables, data sources, and mobile apps that are being used by patients are increasing tremendously [17, 20]. That exponential growth of data will allow MDs to improve diagnoses, prognoses, and therapy through a better understanding of their patients' condition [11]. The impressive signs of progress of Big Data in the medical field [21] will lead to a reassessing of the worldwide view of healthcare systems and organisations [22]. Although there is still an open debate on the cost of health, it seems that only human intelligence can understand today's dispersal of data.

## Continuous medical education and learning

Nowadays MDs are also being required to support healthcare businesses to learn and quickly distil new information from masses of data [23] and transform them into actionable insights regarding cost reduction, maximisation of treatment options, diagnostic pathways, or new solutions to various problems impacting financial and human resources. The ability to extract these insights will empower the future of health and become a differentiator for companies and organisations to thrive and stay ahead of emerging competitors [24–26]. Continuous medical education allows for the improving of MDs' knowledge, preparing healthcare professionals to care for and diagnose in a very demanding and highly technological, multidisciplinary environment [23, 27].

In attempts to provide better Big Data knowledge and technical expertise to operate with different data sets, databases and insight query to MDs both efficiently and cost-effectively, hospitals and HCOs often implement online modules, hire external third-party vendors or sponsor specific academic courses. For example, most MDs are familiar with diagnostic tools and electronic medical record systems, but advanced systems that use AI or data science algorithms are still an area open to much improvement [28–31]. Although these concrete needs require knowledge and data, MDs are often too busy and too tired to go through the material and retain information

consciously. Moreover, the lack of clinical professionals does not allow much time for MDs to concentrate on training.

Not only do professionals need a constantly updated curriculum, but they also need to have a focus on inter-professional learning in the context of continuing professional development. The entry-level healthcare curriculum should introduce to all employees the concept of lifelong learning [32]. Healthcare professionals at every stage of their career must continue to learn about advances in research, treatments, knowledge, and skills required to provide safe and effective patient care [33]. It is the responsibility of the professional to keep abreast of technological changes and relevant medical knowledge [16, 22]. In this perspective, more and more e-learning options are available, with an increasing number of attendees, especially during the COVID pandemic crisis [14, 34].

Most of the worldwide HCOs are adopting robust data strategy [35] relying on Big Data techniques and concepts [36–39], where MDs are playing a decisive role to manage internal procedures and lead advanced diagnostics, treatments, medical prescriptions and consultations as well as hospital discharges/admissions more effectively. According to the literature, some of those activities and key topics can be divided into four major areas: (1) Data transformation, (2) health intelligence, (3) data leveraging, and (4) decision making, as reported in Table 1.

After a first approach to identify the main concepts and issues related to Big Data and MDs’ learning, the paper employs a systematic literature review, conducted in prestigious

scientific databases, to identify the main topics and research avenues, to later define a tentative model as a practical implication of the study.

The paper develops as follows: the following section explains the methodology, followed by the results gathered in the literature; a discussion of our findings leads to a tentative practical model to implement MD learning taking into account the Big Data issues and potential. A conclusion section ends the paper.

## Methodology

A systematic literature review, in academia, is done primarily in order to identify where we currently stand on a specific academic topic or regarding a specific academic research question – identifying what has been discovered (trends, patterns, lines of research, new theory, prominent researchers) and what remains to be discovered (including gaps in the literature). For this purpose we used VOSviewer software [49], which is good for “visualizing scientific landscapes” [50]. VOSviewer represents an evidence-based technique that uses a reliable and structured method to perform the identification, analysis, and interpretation of several relevant academic papers related to the selected topic. The topic herein is Big Data research in the healthcare context and applied to MDs’ learning and training [51]. On the VOSviewer software tool one may add the following:

**Table 1** Four Big Data key areas for MDs

Area	Description
Data Transformation [40, 41]	MDs are turning data into compelling stories and insights communicating with different internal and external partners. Traditionally consent exists that guides patients towards making better healthcare decisions and to be more aware of treatment options; MDs are helping in bridging the gap between engagement and actions.
Health Intelligence [2, 42, 43]	Currently, capable MDs with the right skills to gather insights from predictive health intelligence systems are playing a decisive role in the success of new solutions and medical research. Integration of real-time and historical health data to better analyse a patient’s history, medical record, and to create personalised and anticipatory experiences are crucial nowadays for several HCOs.
Data leveraging [40, 44, 45]	With the growing barriers presented to medical practice, MDs will need to better equip their decision capabilities with more patient- and service-focused decisions, wherein nowadays patients are possessing more information capacity and better decision judgments that are bringing new complexities in doctor-patient discussions.
Decision making [33, 46–48]	The democratisation of data and generalised data access to medicine specialists and general practitioners is bringing new real capacity to treatment, diagnostics, and medication selection to healthcare professionals. Currently, a number of different organisations are already investing exponentially in data centralisation and new rules for data access and security that can bring better outcomes and insights to any HCO department as well as professional decisions, where new technical, as well as non-technical skills, will be required from MDs to understand, interpret and expedite on-demand decision-making processes.

“VOSviewer is a software tool for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature.” [52].

The literature review aims to identify possible missing connections between the study of Big Data processes, technology, and procedures with continuous medical education, learning, and training in the healthcare context.

Examples of Big Data in healthcare and MDs’ education were identified by an initial literature review, after which the added value of 215 items was evaluated, based on the assessment of the added value and the quality of the pieces of evidence.

The selected search terms were derived from the keywords from our research question and also from all the previous literature review exercise where we were able to identify the key terms for the systematic literature review exercise. To better understand some of the key concepts and future network of terms, we also performed a preliminary literature analysis of the most relevant search terms, which will be further explained in the following paragraphs.

Additionally, to minimize the probability of missing relevant articles and publications, a combined search strategy was used with the preliminary literature review, involving the free resource PubMed search engine (accessing biomedical literature primarily from the bibliographic database MEDLINE) and the free Google Scholar search engine. The searches involved snowballing (backward and forwards). The snowballing approach was initially performed on the set of papers identified through the database searches. The approach included using the necessary basic criteria in which the selection applied and without major limitations (e.g. year, content) to identify the title and abstract of the papers, citing our set of selected papers.

After the preliminary and basic literature review of the most relevant publications and articles from an original initial selection of 758 items, we defined a search string used to search databases containing scientific papers in the context of Big Data research applied to medical education and learning. After applying the essential criteria exclusion, the resulting papers were defined as the starting set for the snowballing process. After executing the snowballing iterations, we used the advanced criteria exclusion, which is related to the actual data extraction and quality assessment. The whole process is reported in Fig. 1.

The search terms were based on the research questions using synonyms and related terms. The following keywords were used to formulate the search string: a) Big

Data Research; b) HealthCare learning and education c) continuous Medical Education; d) decision making MDS; e) Medical Doctors.

The above-selected search terms are derived from the keywords from our research question and also from all the previous literature review exercise where we were able to identify the key terms for the systematic literature review exercise. Additionally, the previously presented search terms were based on the areas that we want to further investigate the connections but avoiding terms and relationships that might go outside the keywords selected.

Performing the necessary searches with a restricted list of keywords also allowed a better quality and control of the selected relevant publications and more targeted content in terms of the necessary advanced criteria exclusions based on the VOSViewer requirements.

Our target population was all relevant papers in a healthcare context with medical, clinical, or biology-related activities or processes, where papers outside the healthcare industry were automatically removed as we intended to capture papers in that context regardless of the type of research performed.

Regarding the data sources, the goal was to cover the literature published about Big Data research applied to MDS learning, education and training, so the PubMed and Google Scholar digital databases were selected for data retrieval.

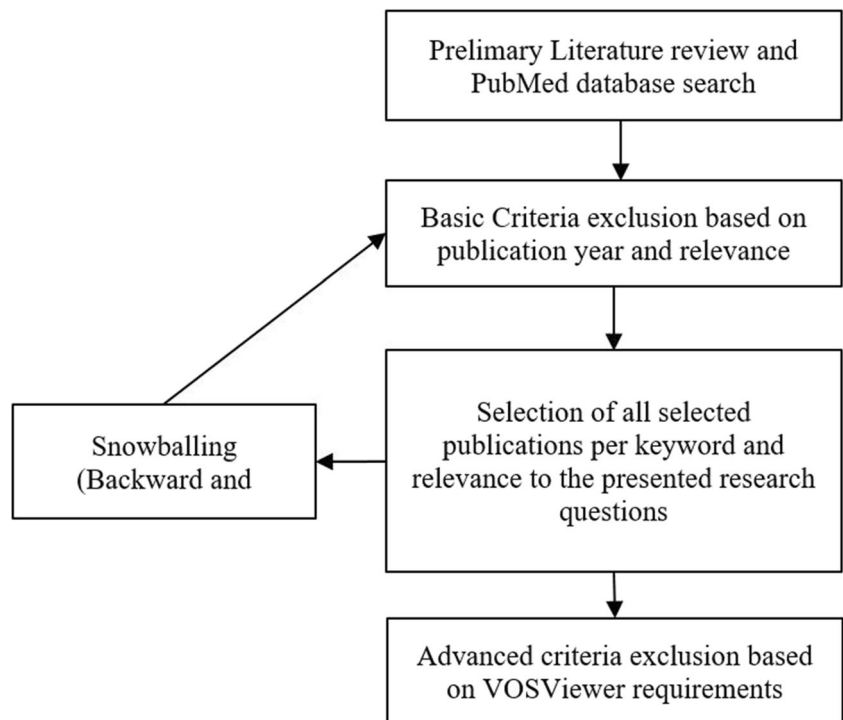
Before applying the selection criteria given the topic of the review, generic exclusion criteria were defined: a) published in non-peer-reviewed publication channels such as books, thesis or dissertations, tutorials, keynotes, and others. OR; b) Not available in English OR; c) A duplicate OR; d) published before 2010. The first two criteria were implemented in the search strings executed in the selected digital library, and the remaining papers were evaluated through two sets of selection criteria: a) basic and b) advanced. Figure 2 summarises the process.

## Search and selection

The first criteria were based on the titles, content, abstract, language, year, and presented results of the papers, articles, or publications. These criteria were applied to the articles that passed a generic exclusion criterion and were identified through the database search or snowballing as described in the section above. In this context, the items included were related to big data research, healthcare learning, continuous medical education, decision making MDS, big data healthcare, and medical doctors.

Following the procedure presented [15], the papers were classified as: a) Relevant; b) Irrelevant; or c) Uncertain (in this case, the available information in the title and abstract is inconclusive).

**Fig. 1** Overview of the applied systematic literature review

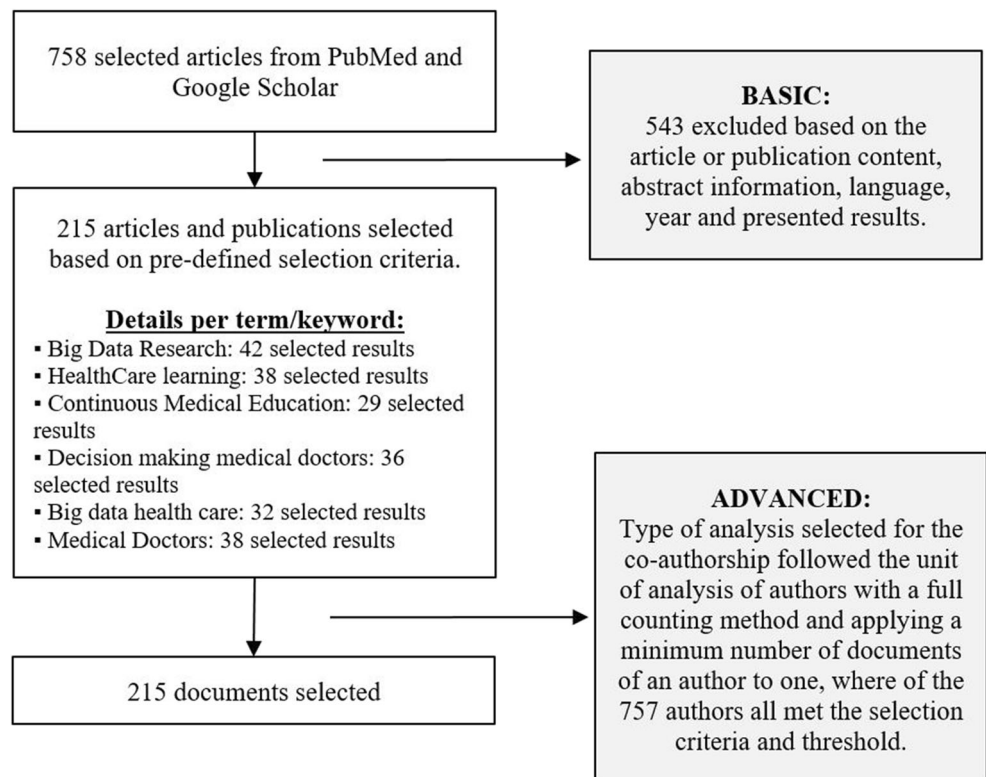


For a profound understanding of all the results, we selected two types of analysis: (1) co-authorship analysis and (2) co-occurrence.

(1) This type of analysis presented the relatedness of all selected authors and co-authors and was determined

based on their number of documents and networks of relationships. In this paper and terms of the advanced search criteria, the type of analysis selected for the co-authorship followed the unit of analysis of authors with a full counting method and applying a minimum number of documents of an author to one,

**Fig. 2** Overview of the database search and selection criteria



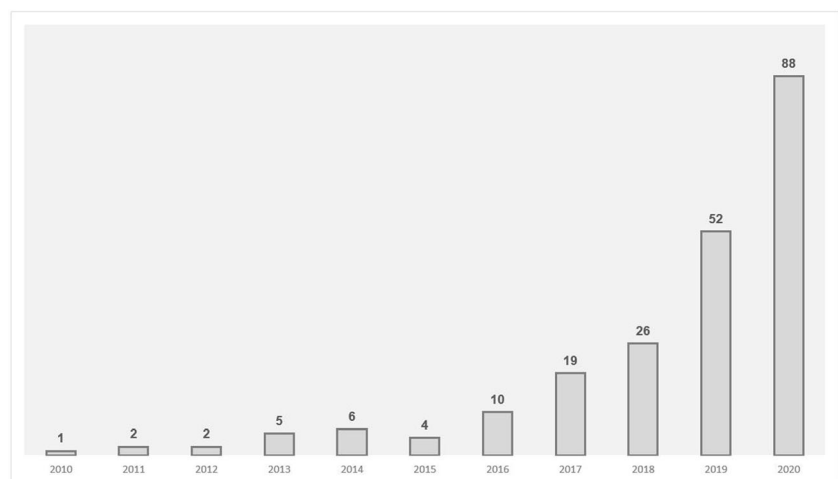
whereof the 757 authors all met the selection criteria and threshold. For each of the 757 authors from the 215 publications, the total strength of the co-authorship links with other authors was also calculated, and the authors with the greatest total link of strength were selected.

- (2) Co-occurrence analysis: the relationship and relatedness of all presented items and selection have a main determination of the number of documents and all related search terms and keywords in which they occur together.

In most of the presented analysis, the counting of the terms and keywords were presented as a full count and instead of fractional count analysis, where the full counting represents that each of the co-authorship or co-occurrence link has the same weight. In the fractional counting, we should understand the counting method as the weight of a link, where for instance if an author co-authors a document with 10 other authors, each of the 20-co-authorship links weigh 0.10 or 1/10.

The advanced criteria were also related to the actual data extraction from the Mendeley platform and upload into VOSViewer, in which some of the additional criteria for the data analysis were thoroughly applied. Some of those additional and final options were: (1) minimum occurrence of a keyword from the total of keywords; (2) total of the strength of the co-occurrence with other keywords calculations followed a pre-defined selection of the number of keywords to be selected for all of the presented analysis (3) weight of occurrences for an effective data visualisation; (4) analysis and validation of all of the selected cluster of analysed items from the VOSViewer algorithm; (5) usage of default values in association with the normalisation for data layout visualisation (6) merging of small clusters for an easy to understand visualisation of all the presented figures in this paper.

**Fig. 3** distribution per year.



## Results

In this section, the findings of the systematic review process are presented. All presented results had as the main basis the bibliometric indicators from the 215 selected articles. As explained in the previous point, the whereof regarding the statistical analysis allowed us to better understand the results with concrete data and measurable indicators from the scientific network of all of the co-authorship analysis as well as of the co-occurrence analysis.

### Publication year

Figure 3 visualises the distribution of all 215 items since 2010 and until today. From the total number of the studied items, 88 were from the current year, showing a clear reflection of the importance of the research question to the entire community and research world.

Since 2015, a growing trend of the number of publications related to the selected keywords can be observed, while in the last 3 years the growth appears as exponential.

### Co-authorship data analysis

VosViewer is used by the academic community to create co-authorship, keyword co-occurrences, citations, bibliographic coupling or co-citation maps based on bibliographic data. In this section, we will be presenting co-authorship analysis based on network graphs where the association strength network has been applied while constructing the network graph.

This type of analysis presents the relatedness of all selected authors and co-authors and was determined based on their number of documents and networks of relationships. In this paper and terms of the advanced search criteria, the type of analysis selected for the co-authorship followed the unit of analysis of authors with a full counting method and applying a minimum number of documents of an author to one,

whereof the 757 authors all met the selection criteria and threshold. For each of the 757 authors from the 215 publications, the total strength of the co-authorship links with other authors was also calculated, and the authors with the greatest total link of strength were selected.

The strength of a link may, for example, indicate the number of cited references two publications have in common, the number of publications two researchers have co-authored, or the number of publications in which two terms occur together. For example, in the case of co-authorship links between researchers and authors, the links attribute indicates the number of co-authorship links of a given author, researcher, or publication with others. Therefore the Total link strength attribute indicates the total strength of the co-authorship links of a given researcher with other researchers.

From the previously explained overview of the database search and selection criteria, the final number of the verified selected authors from the 215 items was

757. As some of the 757 authors in the selected network were not connected as should also be expected, the largest set of connected authors consisted of 15 authors that showed connections between all of them. For the benefit of our readers, we also kept the analysis of all sets of networks and showing the different amounts of clusters from the selected analysis. While running the layout algorithm, we understood that from all 757 authors or full datasets, we needed to make 209 clusters from all the different connections of authors. As explained in the above paragraph, only 15 authors from the 757 had connections between them, and as we can see in fig. 5. For each of the authors from the 215 publications, the total strength of the co-authorship links with other authors was also calculated, and the authors with the greatest total link of strength were selected. Figs. 4 and 5 report the results of our analysis in terms of coauthorships and networks.

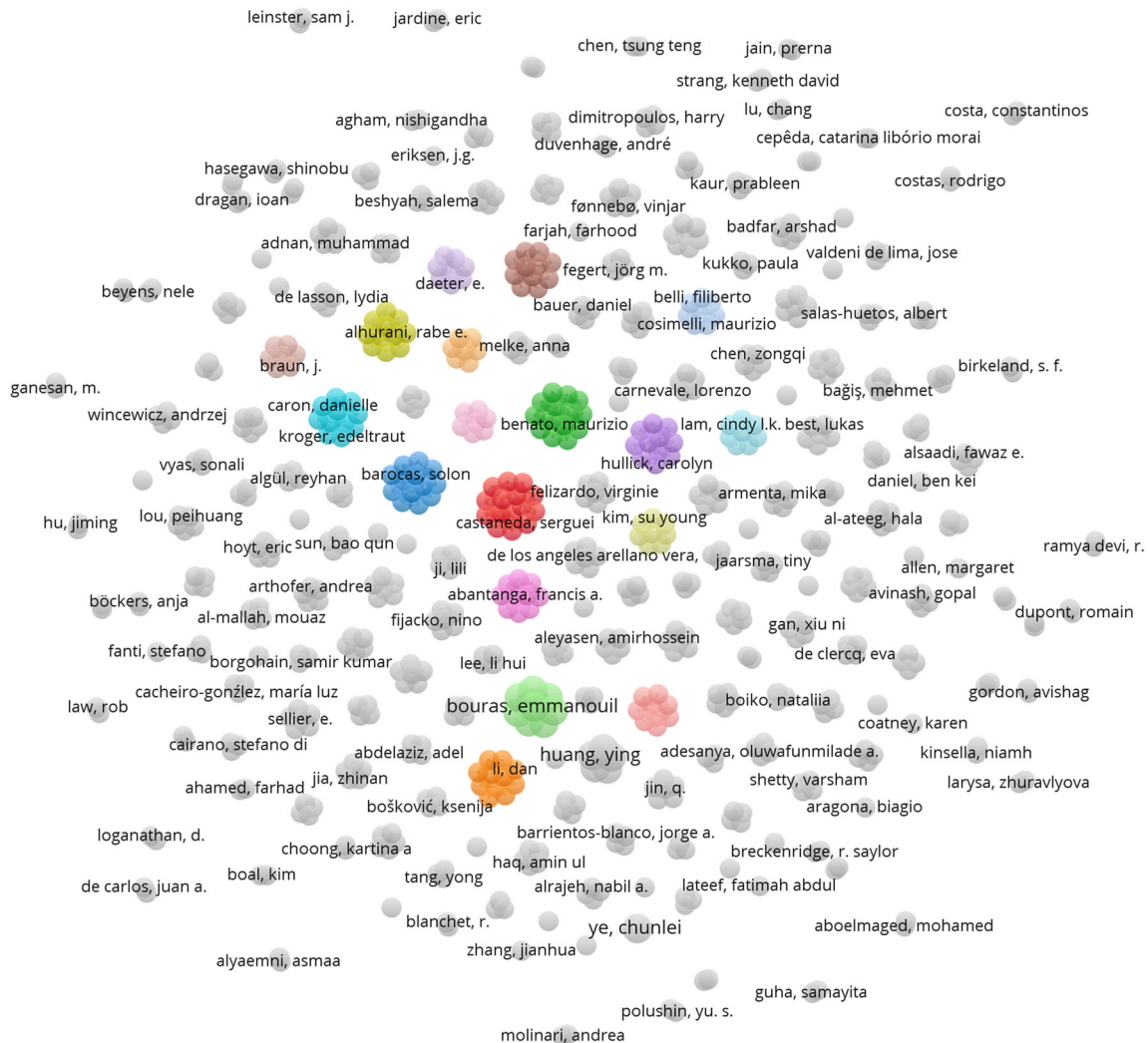
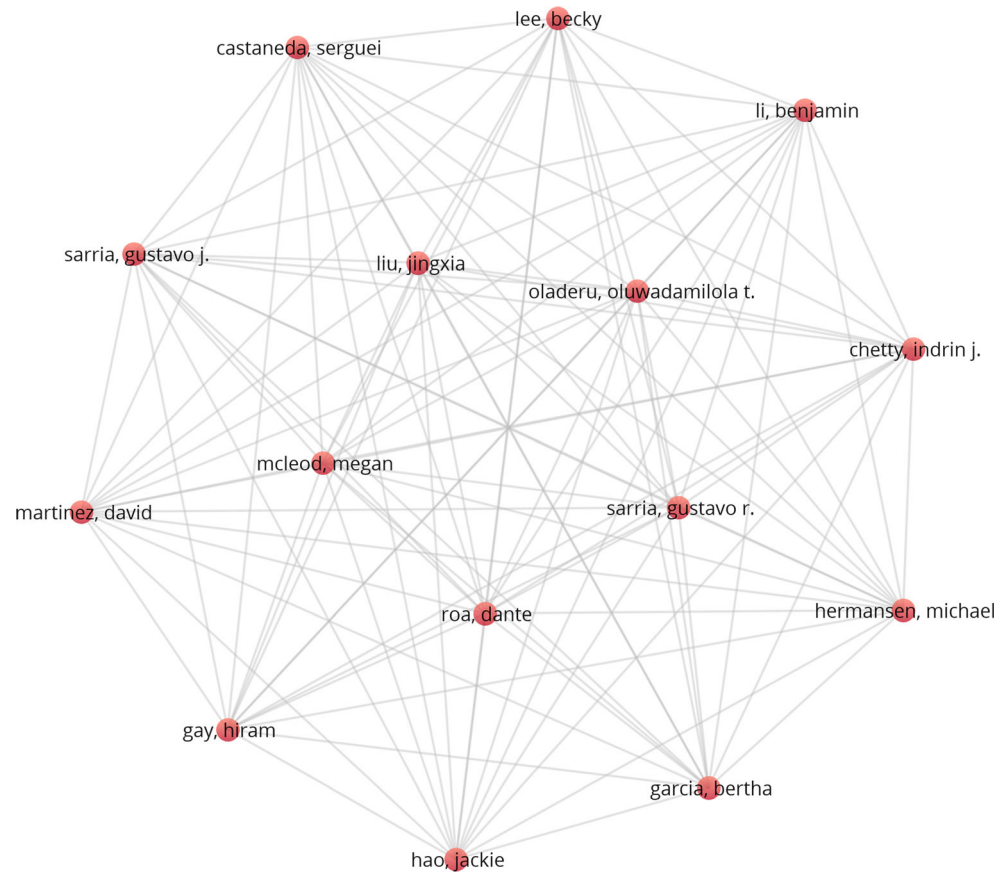


Fig. 4: a high-level overview of all co-authorship relationship networks from the 575 verified authors.

**Fig. 5** Fifteen authors with a relationship



### Co-occurrence of terms and keywords - data analysis

The type of analysis for the co-occurrence had as the main basis the keywords as the unit of analysis and the full counting as the selected method. From the 569 keywords, only 379 of the keywords were part of the largest set of the connected items, according to Fig. 6.

The mapping technique from the VOSViewer software of the keywords allowed us to understand that Big Data, continuous medical education, MDs, have a close relationship with a high number of incidence and good relationship quality.

Figure 7 is also part of the co-occurrence analysis. 53 keywords present a concrete strength of the co-occurrence links with other keywords and allowing us to better understand the total link strength from all 569 keywords and where the minimum of occurrences were 2 keywords as the threshold requirement for the analysis. From the verified selected keywords, the ones with more occurrences were big data ( $n = 30$ , total link of strength = 47), healthcare ( $n = 17$ , total link of strength = 36), and machine learning ( $n = 14$ , total link of strength = 20).

From the current literature, we can also conclude that the connections and linkages between big data, continuous medical education, and medical doctors' learning and training are not clear.

### Model proposal for MDs' learning based on big data research in health

In trying to address our research question: "Which are the main dimensions of a model proposal for increasing the potential of Big Data research in Healthcare for MDs' learning?," we tried to develop a learning model proposal based on Big Data research in healthcare for MDs. The proposed model is supported by the results of the systematic literature review. It includes the Big Data key-areas in health identified in our research as being fundamental for MDs' learning processes: data transformation; health intelligence, data leveraging, and decision-making; associated to each one of these key-areas emerge the learning categories as presented in Table 2.

Each of the model dimensions allows MDs to develop new skills and competencies to address better the challenges posed by the modern healthcare ecosystem [34, 53], that needs to review its global strategies because of the technological shifts.

The first dimension is connected to data transformation. Four learning categories have been identified. Health information services allow MDs to use the new technologies in assessing the patients' data and history, including lab results, X-rays, clinical information, demographic information, and notes. The automatic management of such data can offer immediate insights and practical working support to MDs. Still,





**Table 2** Model Proposal for MDs' Learning based on Big Data key-areas

Health Big Data key-areas	MDs' learning categories			
Data Transformation	Health information Services	New health applications	Medical Images	AI systems applied to health procedures
Health Intelligence	Epidemic Outbreak Forecasting	Drug Discovery	Big data analytics to predict diseases	Genome data
Data leveraging	Patient information	Clinical records	Operational data	Public health data
Decision-making	Disease Diagnosis	Real-time monitoring of patients	Methods to improve treatments	Patient-centred care

Last but not least, AI represents one of the most promising technologies for healthcare applications, able to disrupt the system in terms of diagnosis, treatment, and follow-ups of several clinical disciplines, including oncology, radiology, and surgery [28, 31, 58]. Still, AI and its related technologies like machine learning and deep learning are not included in medical curricula, both undergraduate or postgraduate.

The second dimension concerns health intelligence, which includes the learning regarding epidemic outbreak forecasting, drug discovery, big data analytics to predict diseases, and genome data. The main sources of data are health systems, communities, media and numerous other informal outlets gather epidemiological and other data to predict diseases and to discover new treatments and drugs. The goal is to learn how to gather and to make the interpretation, review, processing and generation of useful information for public health professionals, clinicians, decision-makers and policymakers to achieve a higher level of knowledge and define the appropriate health strategies [59–61].

The third dimension, data leveraging regards the learning about patient information, clinical records, operational data, and public health data, covering information specific to patients, such as diagnosis notes, doctor's prescription, medical photographs, pharmacy records, most of which are supported in research data. This process of learning can also help to improve the decision-making process, including disease diagnosis, real-time monitoring patient methodologies, methods to improve treatments and patient-centred care. It can also be a framework for enhancing healthcare efficiency and productivity using big data analytics, reflecting the development of new clinical models to face problems such as how to determine the most appropriate disease treatment or which medication is the best to apply to a particular case.

A structured learning model can help to improve the quality of care, to avoid risk factors in the early diagnosis of diseases and to better manage the hospital information system. Healthcare big data helps to define emerging sources of data such as social media platforms, connected devices, and others,

where MDs can study new medical situation or find a solution already discovered by other colleagues and accessible by technology which links several sources including patient medical history, data from diagnostic and clinical trials, and index of drug efficacy. It offers a valuable source of knowledge for MDs to achieve innovative healthcare solutions when the mixture of these data sources and analytics are combined together.

The main areas of learning for MDs are researched in a very significant way and are present in scientific databases, where the most recent studies are stored and constitute a knowledge bank to develop specific competencies. The scientific databases are repositories of big data in healthcare and they can be used to support the medical industry helping to provide for better health services based on cutting-edge knowledge. In this regard, the proposed model can help the MDs to develop specific and updated skills and competencies and to be up-to-date regarding health knowledge.

## Discussions and conclusions

The current paper has sought to specify the key dimensions to propose a model for strengthening the likely adaptation of healthcare big data research for medical education and MDs' learning and training. In doing so, this study employed a methodology involving the conducting of a rigorous systematic bibliometric literature analysis and review with large-scale data research within the healthcare setting (i.e., medical, clinical, or biology-related activities or processes). More specifically, the present study conducted two types of bibliometric analysis on the associated literature, namely co-authorship analysis and co-occurrence analysis. The collected data involved 215 peer-reviewed articles published in the PubMed and Google Scholar databases during the period from 2010 to 2020. VOSviewer software was used to construct and view bibliometric maps on the subject of this study.

Overall, it is evident that this area of research has experienced a considerable expansion in popularity in the last three years (from 2018 to 2020). According to co-authorship relationships network analysis, out of 757 researchers, there are

only fifteen authors who had collaborations on some joint research in this field. With respect to keyword co-occurrence analysis, the results indicated that there are 379 keywords, out of 569 keywords, that were part of the greatest collection of the associated items. In conclusion, it is apparent that, based on the selected and analysed data, no obvious associations have been substantially shown between healthcare big data and continuous medical education, learning and training.

This study has several contributions and implications for practice. Due to the increasing significance of big data among different contexts, studying, systematically reviewing and critically analysing large-scale data are attracting scholars' attention from various disciplines, including in the healthcare field. In this vein, a number of literature reviews on healthcare big data have been published. However, this fundamental work expands prior literature sources on healthcare big data by presenting novel insights and guides for scholars and healthcare stakeholders around the world to support the further production of work in this area of research, along with its implementation among medical education and MDs' learning and training. The mapping technique via VOSViewer software of the keywords allowed us to understand that Big Data, continuous medical education, MDs, have a good and close relationship. In addition, this paper adds to the existing body of knowledge concerning bibliometric studies on large-scale data in the healthcare context. Although there are some bibliometric studies that reviewed and analysed the previous publications in healthcare, learning and training aspects were not really made clear, hence this study provides valuable findings related to the potential of healthcare Big Data for MDs' learning and training: 1) data transformation is related to the learning that occurs through medical systems; 2) health intelligence includes the learning regarding health innovation based on predictions and forecasting processes; 3) data leveraging regards the learning about patient information; and 4) the learning process is related to the decision-making process, focused on disease diagnosis and methods to improve treatments. The above can facilitate a deeper understanding of the debates in this field as well as contributing to the existing literature by further mapping the connections among the retrieved publications within this domain. Furthermore, this research objectively and analytically outlined the selected literature with respect to healthcare large-scale data in order to present a crucial and valuable source for concerned scholars to better understand the overall evolution and trends of this area of research. The obvious implication of the existing study is that the systematic bibliometric literature review analysis of healthcare large-scale data by different computational methods can be practiced to obtain significant interpretations and outlines for improving the quality of medical education, learning and training.

Moreover, this paper sought to determine the missing gaps and links between studying and analysing big data processes,

technology, and procedures and ongoing MDs' learning and training within the healthcare sector including hospitals and clinics. As a result, the findings of this research provide a clear overview of the current status on the effective utilisation of healthcare big data for continuing medical education and MDs' learning and training. In addition, the findings of this work can produce outstanding academic and practical lines and paths for healthcare stakeholders and professionals, including MDs, to be quick to pick up new information and technical experience and to be eager to learn from this large-scale of data regarding healthcare. This will result in improving their knowledge, professional skills, decision-making ability, and personal qualifications in particular and supporting the healthcare sector they are involved in, in general.

Additionally, one particularly interesting fact highlighted by the results of co-authorship analysis is that collaborations between scholars are limited amongst this area of research. To this end, this paper contributes to knowledge in the healthcare setting by addressing the substantial lack and absence of a cooperative research community and considerable collaboration among authors. Thus, this paper provides explicit insights and directions to effectively enhance the collaboration between researchers from different institutions and countries. This will aid in presenting more valuable cooperation and comprehensive sources regarding healthcare big data research and utilising them for medical education and MDs' learning and training.

Finally, and in sum, professionals need a constantly updated curriculum, they need to have a focus on inter-professional learning, and the entry-level healthcare curriculum must introduce to all employees the concept of lifelong learning – as continuing education and skills development in Big Data is often forgotten.

## Limitations and future research

Even though this study provides several useful insights, a number of limitations have to be acknowledged, leading to further research avenues and opportunities. First, the current paper has conducted two patterns of bibliometric analyses which are: co-authorship analysis and co-occurrence. Thus, future studies are suggested to employ different series and metrics of bibliometric analyses on large-scale data research within the healthcare setting, including, prolific authors, journals' production trends, citation and H-index analysis, distribution of the most productive countries and institutions, countries' and/or institutions' co-authorship analysis and co-citation analysis on the published work in the topic. Additionally, future research is recommended to analyse the research methods, data collection and analysis procedures applied in previous studies of this domain. Additional in-depth

analysis may be conducted, even employing other types of literature reviews such as more structured ones [62].

This paper employed VOSviewer software to construct and visualise bibliometric networks on the topic. Further studies can therefore adopt other software programs such as CiteSpace, BibExcel, Eigenfactor Score, R package, among others, which are common and efficient software programs employed to gain quantitative and visual information and mapping in a certain domain.

Furthermore, the focus of this study was on peer-reviewed articles published in the PubMed and Google Scholar digital databases, thus focusing on more clinical-related journals or publications. Future studies may concentrate on more sources published in other databases such as Elsevier's Scopus and/or Web of Science (SCIE, SSCI, and/or ESCI), as these datasets include journals, reviews, and books of other scientific fields, which may create synergies on the topic, like Social Sciences, Business and Management, Engineering, or Education. As recent studies have reported, practical and innovative inputs and solutions about healthcare-related matters may be gathered and applied from other domains, especially Business and Management [41].

In addition, future research may also expand to other non-peer-reviewed publications, including books and book chapters, proceedings papers, reports, guidelines, book reviews, among others to provide a broader overview. Such sources may report pioneering and promising early-stage ideas and concepts, which may not yet be published in ranked journals. Relevant publications written in other languages, other than English, may also be targeted in future work.

**Funding** This study was funded by Fundação para a Ciência e Tecnologia, Grant: UIDB/00315/2020.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Dal Mas F, Piccolo D, Edvinsson L, Skrap M, D'Auria S. Strategy Innovation, Intellectual Capital Management and the Future of Healthcare. The case of Kiron by Nucleode. In: Matos F, Vairinhos V, Salavisa I, Edvinsson L, Massaro M, editors. Knowledge, People, and Digital Transformation: Approaches for a Sustainable Future. Cham: Springer; 2020. p. 119–31.
- Dal Mas F, Piccolo D, Ruzza D. Overcoming cognitive bias through intellectual capital management . The case of pediatric medicine . In: Ordonez de Pablos P, Edvinsson L, editors. Intellectual Capital in the Digital Economy. London: Routledge; 2020. p. 123–33.
- Miceli L, Bednarova R, Di Cesare M, Santori E, Spizzichino M, Di Minco L, et al. Outpatient therapeutic chronic opioid consumption in Italy: a one-year survey. *Minerva Anestesiol.* 2017;83(1):33–40.
- Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: Three decades of research. A comprehensive review. *J Clin Pharm Ther.* 2001;26(5):331–42.
- Limb M. World will lack 18 million health workers by 2030 without adequate investment, warns UN. *Br Med J.* 2016;354.
- Cobianchi L, Pugliese L, Peloso A, Dal Mas F, Angelos P. To a New Normal: Surgery and COVID-19 during the Transition Phase. *Ann Surg.* 2020;272:e49–51.
- Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA.* 2020;323(14):1341–1342.
- Jain P, Kaur A. Big Data Analysis for Prediction of Coronary Artery Disease. In: 2018 4th International Conference on Computing Sciences (ICCS). Jalandhar; 2018. p. 188–93.
- Kaur P, Sharma M, Mittal M. Big Data and Machine Learning Based Secure Healthcare Framework. *Procedia Comput Sci.* 2018;132:1049–59.
- Massaro M, Dumay J, Garlatti A. Public sector knowledge management: A structured literature review. *J Knowl Manag.* 2015;19(3):530–58.
- Kim J, Diesner J, Kim H, Aleyasen A, Kim H. Why name ambiguity resolution matters for scholarly big data research. In: 2014 IEEE International Conference on Big Data (Big Data). Washington, DC; 2014. p. 1–6.
- Carra G, Salluh JIF, da Silva Ramos FJ, Meyfroidt G. Data-driven ICU management: Using Big Data and algorithms to improve outcomes. *J Crit Care.* 2020;60:300–4.
- Pesqueira A, Sousa MJ, Rocha A, Sousa M. Data Science in Pharmaceutical Industry. In: Rocha A, Reis LP, Costanzo S, editors. Advances in Intelligent Systems and Computing. Cham: Springer; 2020.
- Garcia Vazquez A, Verde JM, Dal Mas F, Palermo M, Cobianchi L, Marescaux J, et al. Image-guided surgical e-learning in the post-COVID-19 pandemic era: what is next? *J Laparoendosc Adv Surg Tech.* 2020;30(9):993–7.
- Baro E, Degoul S, Beuscart R, Chazard E. Toward a Literature-Driven Definition of Big Data in Healthcare. *Biomed Res Int.* 2015;(639021):9.
- Li Z, Li Y, Lin P. The Security Evaluation of Big Data Research for Smart Grid. In: 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). Tangier, Morocco; 2019. p. 1055–9.
- Bednarova R, Biancuzzi H, Rizzardo A, Dal Mas F, Massaro M, Cobianchi L, et al. Cancer rehabilitation and physical activity: The “Oncology in Motion” project. *J Cancer Educ.* 2020;(In press).
- Wu X, Chen H, Wu G, Liu J, Zheng Q, He X, et al. Knowledge Engineering with Big Data. *IEEE Intell Syst.* 2015;30(5):46–55.
- Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Heal Inf Sci Syst.* 2014;2(3).
- Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Per Med.* 2018;15(5):429–448.
- Pesqueira A, Sousa M, Rocha A. Big Data Skills Sustainable Development in Healthcare and Pharmaceuticals. *J Med Syst.* 2020;(forthcoming).

22. Luo J-D, Liu J, Yang K, Fu X. Big data research guided by sociological theory: a triadic dialogue among big data analysis, theory, and predictive models. *J Chinese Sociol.* 2019;6(1).
23. Medin-Eastwood D, Podhorska IKP. Healthcare Big Data Systems, Wearable Medical Devices, and Remote Patient Care. *Am J Med Res.* 2019;6(1):48.
24. Bin Ali N, Petersen K, Wohlin C. A Systematic Literature Review on the Industrial Use of Software Process Simulation. *J Syst Softw.* 2014;97:65–85.
25. Patgiri R, Nayak S, Borgohain SK. Role of Bloom Filter in Big Data Research: A Survey. *Int J Adv Comput Sci Appl.* 2018;9(11): 655–61.
26. Radhika TV, Gouda KC, Kumar SS. Big data research in climate science. In: 2016 International Conference on Communication and Electronics Systems (ICCES). Coimbatore; 2016. p. 1–6.
27. Rajabion L, Shaltook AA, Taghikhah M, Ghasemi A, Badfar A. Healthcare big data processing mechanisms: The role of cloud computing. *Int J Inf Manage.* 2019;49:271–89.
28. Maubert A, Birtwisle L, Bernard JL, Benizri E, Bereder JM. Can machine learning predict resectability of a peritoneal carcinomatosis? *Surg Oncol* [Internet]. 2019;29:120–5. Available from: <http://www.sciencedirect.com/science/article/pii/S0960740419301057>
29. Aruni G, Amit G, Dasgupta P. New surgical robots on the horizon and the potential role of artificial intelligence. *Investig Clin Urol.* 2018;59(4):221–2.
30. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism.* 2017;69(1):S36–40.
31. Guan J. Artificial Intelligence in Healthcare and Medicine: Promises, Ethical Challenges and Governance. *Chinese Med Sci Journal.* 2019;34(2):76–83.
32. Shaughnessy AF, Slawson DC. Are we providing doctors with the training and tools for lifelong learning? *BMJ.* 1999;319(1280).
33. Sousa MJ, Dal Mas F, Garcia-Perez A, Cobianchi L. Knowledge in Transition in Healthcare. *Eur J Investig Heal Psychol Educ.* 2020;10(3):733–48.
34. Cobianchi L, Dal Mas F, Peloso A, Pugliese L, Massaro M, Bagnoli C, et al. Planning the Full Recovery Phase: An Antifragile Perspective on Surgery after COVID-19. *Ann Surg.* 2020;272(6): e296–9.
35. Pesqueira A, Sousa MJ. Pharmaceuticals and Life Sciences: Role of Competitive Intelligence in Innovation. In: Jamil GL, Ribeiro F, Malheiro da Silva A, Maravilhas Lopes S, editors. *Handbook of Research on Emerging Technologies for Effective Project Management.* Hershey, PA: IGI Global; 2020. p. 237–54.
36. Dogaru DI, Dumitrache I. Big Data and Machine Learning Framework in Healthcare. In: 2019 E-Health and Bioengineering Conference (EHB). Iasi, Romania; 2019. p. 1–4.
37. Gillespie B, Otto C, Young C. Bridging the academic-practice gap through big data research. *Int J Mark Res.* 2018;60(1):11–13.
38. Guha S, Kumar S. Emergence of Big Data Research in Operations Management, Information Systems, and Healthcare: Past Contributions and Future Roadmap. *Prod Oper Manag.* 2018;27(9):1724–35.
39. Sousa MJ, Pesqueira A, Lemos C, Sousa M, Rocha A. Decision-Making based on Big Data Analytics for People Management in Healthcare Organizations. *J Med Syst.* 2019;43(9):290.
40. Angelos P. Interventions to Improve Informed Consent Perhaps Surgeons Should Speak Less and Listen More. *JAMA Surg.* 2020;155(1):13–4.
41. Dal Mas F, Garcia-Perez A, Sousa MJ, Lopes da Costa R, Cobianchi L. Knowledge Translation in the Healthcare Sector. A Structured Literature Review. *Electron J Knowl Manag.* 2020;18(3):198–211.
42. Dal Mas F, Piccolo D, Cobianchi L, Edvinsson L, Presch G, Massaro M, et al. The effects of Artificial Intelligence, Robotics, and Industry 4.0 technologies. Insights from the Healthcare Sector. In: Proceedings of the first European Conference on the impact of Artificial Intelligence and Robotics. Academic Conferences and Publishing International Limited; 2019. p. 88–95.
43. Becker A. Artificial intelligence in medicine: What is it doing for us today? *Heal Policy Technol* [Internet]. 2019;8(2):198–205. Available from: <http://www.sciencedirect.com/science/article/pii/S2211883718301758>
44. Dal Mas F, Biancuzzi H, Massaro M, Miceli L. Adopting a knowledge translation approach in healthcare co-production. A case study. *Manag Decis.* 2020;In Press.
45. Biancuzzi H, Dal Mas F, Miceli L, Bednarova R. Post breast cancer coaching path: a co-production experience for women. In: Paoloni P, Lombardi R, editors. *Gender Studies, Entrepreneurship and Human Capital IPAZIA 2019 Springer Proceedings in Business and Economics.* Cham: Springer; 2020. p. 11–23.
46. Dal Mas F, Bagarotto EM, Cobianchi L. Soft Skills effects on Knowledge Translation in healthcare. Evidence from the field. In: Lepeley MT, Beutell N, Abarca N, Majluf N, editors. *Soft Skills for Human Centered Management and Global Sustainability.* London: Routledge; 2021.
47. Yule S, Smink DS. Non-Technical Skill Countermeasures for Pandemic Response. *Ann Surg.* 2020;272(3):e213–5.
48. Schutt S, Holloway D, Linegar D, Deman D. Using simulated digital role plays to teach healthcare “soft skills.” In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health, SeGAH 2017. 2017.
49. Saviano M, Bassano C, Picocchi P, Di Nauta P, Lettieri M. Monitoring Viability and Sustainability in Healthcare Organizations. *Sustainability.* 2018;10:3548.
50. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* [Internet]. 2010;84(2):523–38. Available from: <https://doi.org/10.1007/s11192-009-0146-3>
51. Hoyt E, Hughes K, Long D, Tran A, Ponto K. Scaled Entity Search: A method for media historiography and response to critiques of big humanities data research. In: 2014 IEEE International Conference on Big Data (Big Data). Washington, DC; 2014. p. 51–9.
52. VosViewer. Welcome to VOSviewer [Internet]. Leiden University. 2020 [cited 2020 Dec 2]. Available from: <https://www.vosviewer.com/>
53. Derbyshire J, Wright G. Preparing for the future: Development of an “antifragile” methodology that complements scenario planning by omitting causation. *Technol Forecast Soc Change* [Internet]. 2014;82(1):215–25. Available from: <https://doi.org/10.1016/j.techfore.2013.07.001>
54. Grenda TR, Whang S, Evans NR. Transitioning a Surgery Practice to Telehealth During COVID-19. *Ann Surg.* 2020;272(2):e168–9.
55. Jain S, Khera R, Lin Z, Ross JS, Krumholz HM. Availability of Telemedicine Services Across Hospitals in the United States in 2018: A Cross-sectional Study. *Ann Intern Med* [Internet]. 2020 Apr 30;173(6):503–5. Available from: <https://doi.org/10.7326/M20-1201>
56. Reed ME, Huang J, Parikh R, Millman A, Ballard DW, Barr I, et al. Patient–Provider Video Telemedicine Integrated With Clinical Care: Patient Experiences. *Ann Intern Med* [Internet]. 2019 Apr 30;171(3):222–4. Available from: <https://www.acpjournals.org/doi/abs/10.7326/M18-3081>
57. Davrieux CF, Palermo M, Cúneo T, Zanutini D, Giménez ME. What Is the Role of Image-Guided Endovascular Surgery in Postbariatric Surgery Bleeding Complications? *J Laparoendosc Adv Surg Tech A.* 2020 Sep
58. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep Learning in Radiology. *Acad Radiol* [Internet]. 2018;25(11):1472–80. Available from: <https://doi.org/10.1016/j.acra.2018.02.018>

59. Hatchett RJ, Meecher CE, Lipsitch M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc Natl Acad Sci U S A*. 2007;104(18):7582–7.
60. Trequattrini R, Shams R, Lardo A, Lombardi R. Risk of an epidemic impact when adopting the Internet of Things: the role of sector-based resistance. *Bus Process Manag J*. 2016;22(2):403–19.
61. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? *Lancet* [Internet]. 2020;395(10227):848–50. Available from: [https://doi.org/10.1016/S0140-6736\(20\)30551-1](https://doi.org/10.1016/S0140-6736(20)30551-1)
62. Massaro M, Dumay JC, Guthrie J. On the shoulders of giants: Undertaking a structured literature review in accounting. *Accounting, Audit Account J*. 2016;29(5):767–901.

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