The Journal of Finance and Data Science 9 (2023) 100102

Contents lists available at ScienceDirect

The Journal of Finance and Data Science

journal homepage: www.keaipublishing.com/en/journals/jfds

Review Article

The applications of big data in the insurance industry: A bibliometric and systematic review of relevant literature

Nejla Ellili [a], Haitham Nobanee [a,b,c,*], Lama Alsaiari [a], Hiba Shanti [a], Bettylucille Hillebrand [a], Nadeen Hassanain [a], Leen Elfout [a]

[a] College of Business, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates
[b] Oxford Centre for Islamic Studies, University of Oxford, Marston Rd, Headington, Oxford OX3 0EE, UK
[c] The University of Liverpool Management School, The University of Liverpool, Liverpool, Lancashire, United Kingdom

ARTICLE INFO

Keywords:
Bibliometric analysis
Big data
Insurance
Insurance industry
Risk

ABSTRACT

The insurance industry has changed rapidly over the last few decades. One factor in this change is the continuous growth of massive amounts of data that need to be processed properly to be optimally utilized. This has led to a strong wave of advanced processing technologies that can systematically manage big datasets, such as machine learning and artificial intelligence. This study analyzes the current state of research on big data and insurance. Bibliometric analysis and a systematic review were conducted to analyze the patterns and trends of the subject area, with the main focus on citations as a key measurement unit. This analysis is important to fill the existing gap in the examined area because no other bibliometric analysis has been conducted previously on the same subject; it will also help in establishing a scientific background for future research. The research findings verify that the United States is the most popular and cited country in the research area of big data and insurance at both the single authorship and co-authorship levels. Finally, the major impact of the relationship between big data and the insurance sector was marked by human-related aspects.

1. Introduction

In 2022, the insurance sector around the world generated revenues of more than USD 6 trillion. A total that exceeds the entire Gross Domestic Product of the world's largest economies, like Japan, Germany, UK, India, Italy, France, Canada, and draws in double the intake of the oil industry (Hassani et al., 2020). That is what makes actuarial science nowadays of great importance and versatile in various contexts. In other words, the insurance sector depends mainly on risk analysis through the use of mathematical and statistical methods (Frees and Huang, 2023). The increasing dependence of the insurance sector on the application of advanced technologies, such as big data, artificial intelligence, and machine learning is a strong indication of the importance of this research topic. Big data can help insurers increase the accuracy level of policy pricing due to the higher availability of data and improve the efficiency of their operations by providing more effective and efficient insight into their customers. This can lead to lower costs and improve the customer satisfaction (Hassani et al., 2020). No one knows exactly when the idea of insurance started; however, it is predicted that it dates back to antiquity. However, actuarial science that we know today was founded in the late 1600s when requirements for long-term insurance packages such

* Corresponding author. College of Business, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates.
E-mail addresses: haitham.nobanee@adu.ac.ae, nobanee@gmail.com, haitham.nobanee@liverpool.ac.uk, haitham.nobanee@oxcis.ac.uk (H. Nobanee).

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.jfds.2023.100102
Received 12 February 2023; Received in revised form 16 July 2023; Accepted 16 July 2023
Available online 20 July 2023

as life insurance, annuities, and burial were growing fast; therefore, a new mathematical and statistical discipline was needed to manage these variables (Landsman and Sherris, 2001).

Big data is the main information resource for insurers. It is defined by Reference (Sagiroglu and Sinanc, 2013) as a massive amount of data sets that cannot be easily stored, processed, analyzed, or visualized because of the high complexity of its structure. The application of big data in insurance can lead to various benefits, such as better risk assessments, claims management, underwriting, retention, and customer satisfaction. Machine learning, and artificial intelligence are commonly used in the insurance industry utilize and analyze big data. These technologies can help insurance companies improve the quality of their services by automating the data processing, identifying trends and patterns, and improving the data quality. The conceptual background of the application of big data in the insurance industry is strongly explained by the increase in the availability and accessibility of large and complex datasets as well as the need of insurers to use these datasets to extract valuable insights and create sustainable business value. The data that insurance companies utilize include personal and non-personal information coming from internal or external sources as well as information that might be structured or unstructured (Keller et al., 2018.). An analysis conducted by SNS Telecom and Information Technology (IT) in 2018 confirms that the utilization of big data is growing rapidly, forecasting a predicted $3.6 billion worth of investments to take place by 2021 (Telecom SNC & IT, 2018). The research analysis also shows how the embracing of big data leads to 40–70% cost savings, 30% healthier access to insurance packages, and 60% improvement in identifying fraud. Moreover, another study conducted in the same year, conducted only on 4% of insurance firms around the world, claimed that around 74% of insurance companies from their sample confirmed that the utilization of big data and analytics has resulted in competitive gains for their businesses (Corbett et al., 2018). There are several examples of the important role that big data plays in the insurance sector, such as automobile insurance, life insurance (using the so-called mortality modeling), health insurance, harvest risk, catastrophe risk (such as hurricanes, tornadoes, geomagnetic events, earthquakes, floods, and fires), climate risks, and cyber risks (which is the weakest one regarding security and safety).

Big data has revolutionized the insurance sector, improving customer experiences, risk assessment, operations, and fraud detection. Insurers use big data analytics to gather and analyze structured and unstructured data from various sources, enabling more accurate risk pricing and personalized offerings. By understanding customer behavior patterns, insurers can offer tailored policies and marketing campaigns, enhancing customer relationships and retention. Big data plays a crucial role in fraud detection, identifying suspicious patterns for quicker investigations and reduced losses. Predictive analytics enables better customer service by anticipating needs and providing proactive support. It streamlines claims processing, improving response times and customer satisfaction. Additionally, big data predicts future trends, optimizing insurance processes and reducing operational costs. In health and life insurance, it monitors and analyzes customer health data, leading to personalized plans and healthier lifestyles. Reinsurance companies also benefit from big data, accurately assessing risks for long-term sustainability. Overall, big data continues to transform the insurance industry, empowering companies with valuable insights and driving innovation.

The application of big data in the insurance industry has garnered significant attention, leading to a comprehensive bibliometric and systematic review of relevant literature. This study aims to explore and analyze the various ways big data is utilized in the insurance sector. Through a meticulous examination of existing research, the review highlights the impact of big data on customer experiences, risk assessment, operational efficiency, fraud detection, and claims processing in the insurance domain. The findings of this review shed light on the emerging trends and potential future directions for leveraging big data in insurance. By synthesizing and analyzing the current body of literature, this study offers valuable insights for researchers, practitioners, and policymakers in the insurance industry.

A bibliometric and systematic review study is crucial for understanding big data applications in the insurance industry. The vast and evolving nature of this domain requires comprehensive analysis. A bibliometric review enables systematic analysis of numerous studies, identifying key trends and gaps in the literature. It uncovers emerging trends, ensuring the latest advancements are captured. The review evaluates research methodologies, enhancing understanding of evidence reliability. By identifying underexplored areas, the study pinpoints topics requiring further investigation, guiding future research. Synthesizing diverse findings from various domains like customer experiences, risk assessment, claims processing, and fraud detection, the systematic review provides a consolidated understanding of big data's impact on insurance. Policymakers, practitioners, and insurers benefit from evidence-based insights to adopt big data strategies and address challenges. The study advances knowledge by synthesizing existing research, identifying gaps, and formulating new research questions. Ultimately, the bibliometric and systematic review serves as an indispensable tool, providing a solid foundation for further research and guiding practical implications in the insurance industry.

The applications of big data in the insurance industry have been extensively explored in the existing literature, revealing its significant impact on various aspects of the sector. However, despite the wealth of research available, there still exists a notable gap in understanding the long-term effects and sustainability of big data implementation in insurance companies. While many studies have focused on short-term benefits such as improved risk assessment and customer experiences, there is limited research on the long-term implications and challenges faced by insurers in maintaining big data initiatives. Additionally, the ethical and privacy concerns arising from the collection and analysis of massive amounts of personal data in the insurance context require further investigation. Closing this literature gap is essential for insurers and policymakers to make informed decisions on the responsible and effective use of big data in the insurance industry, ensuring its continued success and relevance in the future.

This study aims to provide a comprehensive analysis of the research growth and trends in the subject area of big data and the insurance industry. The significance of this analysis is to provide researchers and practitioners with thorough knowledge of the current situation in the studied research area and what areas of improvement they need to focus on. Considering the rapid evolution of the insurance sector and the increasing prevalence of big data, a systematic review of studies on its application in the insurance industry is necessary. This study had seven research questions, as follows: (1) what's the growth in the research on the application of big data in the insurance industry? (2) what are the most cited documents on this research topic? (3) what are the most productive sources, authors, organizations, and countries? (4) what are the most co-cited references, sources, and authors? (5) what are the different collaborations

between authors, organizations, and countries? (6) what are the most occurrent keywords in the research on the application of big data in the insurance industry? (7) what are the suggestions for future research on this field?

This study aims to achieve two main objectives. Firstly, this review will provide an up-date analysis of the current state of the application of big data in the insurance industry. Since this sector constantly changes, new research studies and innovations have emerged. A systematic review will include the most recent studies and developments and provide insights into the opportunities and challenges that big data can present in the insurance industry. Secondly, a systematic review will highlight gaps in the current research regarding the application of big data in the insurance industry and identify future research ideas on this topic to guide the development of more efficient and effective big data applications within the insurance sector.

The remainder of this study is structured as follows. First, a literature review of previous research papers on this topic is presented. Second, the methodology developed for this research is explained, including the analytical methods used, scientific mapping, and data network tools. Next, the results of the bibliometric analysis are presented in terms of publication growth, authors' contributions, most cited documents, most popular countries, most active sources, organizations of authors, and keywords' occurrence. Finally, the conclusions of this study are critically discussed. Additionally, suggestions for future research are provided with the purpose of helping interested researchers improve the subject area, in addition to the limitations of this research study.

2. Literature review

Big data refers to the large and complex datasets that can't be processed using standard tools. In the past few years, the big data has gained widespread significance in various sectors, including the insurance. The application of such data has allowed insurance companies to improve their customer engagement and retention, risk assessment and the efficiency of their fraud prevention as well as mitigate their costs (Banu, 2022; Rana et al., 2022). Big data is often associated with the development of machine learning (ML) and artificial intelligence (AI). In fact, the AI and ML technologies require big data in order to be successfully implemented.

Despite the increasing interest in the application of big data in the insurance sector, this research topic remains underexplored. This review aims to provide an in-depth analysis of the literature related to the potential application of this technology in the insurance industry. In addition, this review aims to identify the various gaps in the literature and provide recommendations for future research. This will help policymakers, researchers, and practitioners in the insurance sector by understanding how big data can be effectively used to improve their insurance operations.

According to Reference (Hussain et al., 2016), big data usage in the finance and insurance sector has a great impact on the level of performance of the company itself. These advantages can be summarized in two main points. First, it enhances customer understanding, engagement, and involvement. This is because having the financial services and products digitalized and increasing the trend of clients engaging with groups and/or brands in the digitalized space will pave the way for financial services companies, including insurance, to improve their level of consumer involvement, thus enhancing the clients' overall experience. Second, the use of Big Data in insurance and financial institutions will help improve fraud recognition and prevention facilities. This is critical because financial service institutions are highly vulnerable to fraud in every process in which they engage. Previously, without the presence of big data, banks analyzed only a few samples of transactions to identify fraud. In fact, this would result in several fraudulent actions falling through the net and other "false positives." Therefore, the application of big data has allowed these corporations to utilize larger datasets to detect trends that signal fraud to aid in the process of minimizing exposure to this type of risk (Hussain et al., 2016). Moreover, reference (Hussain et al., 2016) describes several main potential uses of big data in the financial and insurance sector that will help in further enhancing the financial institutions' operations. First, the process of what is called "Reputational risk management" which simply provides an evaluation of exposure to reputational risk related to consulting services provided by banks to consumers. A pessimistic assessment can negatively impact a bank's power to sustain existing or create more business connections and access funding sources. Thus, the rise in the possibility of default, which is often referred to as credit risk, the increase in price volatility, and the complications of exchanging specific financial products in restricted markets have led to the growth of operational and reputational risks related to advisory and brokerage services. Many banks and financial institutions provide third-party financial services. This suggests that the poor performance of a third-party product has substantial effects on the relationship between the bank and its clients. The second is what is known as a retail brokerage. The use of big data will help determine topic trends, identify events, or assist in portfolio optimization and asset distribution. This attention is not based on figures built on quantitative historical data. However, investors seek indicators that have a certain predictive component and are easy to understand. On the other hand, Reference (Senousy et al., 2020) focused on the use of big data in social insurance. Social insurance simply refers to a person's guard against risks, such as disability, retirement, or death. Utilizing big data mining and analytics paves the way for insurers and brokers to develop ideal solutions for individuals seeking social insurance. Reference (Senousy et al., 2020) further focused on Egyptian society, specifically social insurance, and performed several experiments, implementing some mining techniques, such as clustering and classification algorithms on the Egyptian social insurance dataset to prove the usefulness of big data in the insurance sector.

Reference (Ho et al., 2020) proposes that a strong regulatory and ethical environment may be similar for big data analytics regarding health insurance. Furthermore, we discuss several approaches to precautions and participating mechanisms that should be recognized. First, it is vital to develop a transparent and efficient data-governance framework. Permitted standards must be passed, and insurers must be urged and offered incentives to acquire human-centered strategies in the creation and usage of artificial intelligence and big data analytics. Subsequently, an accountable and strong process is required to clarify what information could be utilized and how it could be utilized. Thus, individuals whose information would be employed must be supported through their active involvement in understanding how their personal information can be handled and regulated. Finally, insurers and governance corporations that consist of officials and policymakers are required to act collectively to guarantee that the established big data analytics based on artificial

intelligence are clear and precise. If no appropriate ethical environment is implemented, the management of such analytics will result in the spread of unconnected data systems, destroying current inequalities, credibility, and trust (Ho et al., 2020).

Additionally, reference (Manoj Kumar et al., 2016) supports the rising interest in the big data field and insurance. At the beginning, they clearly stated that big data is a developing technology utilized in nearly all life aspects, such as e-commerce, E−healthcare, and the banking industry. Furthermore, it highlights how insurance corporations are currently demonstrating great interest in analyzing their enormous datasets that contain hospital and patient records, paving the way for insurance companies to extract valuable information. These companies focus on the success and failure percentages and responses provided by patients. Insurance companies will receive hospital invoices, discharge summaries, and medical reports from their patients. Subsequently, these patients' reports, symptoms, and responses were further analyzed using big data technologies such as Infinispan and map-reduce concepts for separation and data extraction in E-health insurance (Manoj Kumar et al., 2016).

However, Reference (Hanafy and Ming, 2021) took the topic of big data and insurance to the next level, in which they discussed the relation of big data and machine learning to a specific type of insurance, that is, auto insurance. They first stated that there is an urgent need for creative techniques to handle the increasing trend in the amount and severity of auto insurance claims effectively. Machine learning (ML), which uses big data to understand future trends, is a technique that is helpful in resolving this problem. Car insurers who frequently work to enhance their customer service began acquiring and utilizing machine learning techniques to improve their knowledge and understanding of their records more efficiently; therefore, they are able to enhance their customer service through a good understanding of their clients' needs. This study focuses on how automotive insurance corporations integrate machinery learning in their businesses and investigates how machine learning models can be applied to insurance big data. XGBoost, Logistic regression, decision trees, random forest naïve Bayes, and K-Nearest Neighbors (K-NN) algorithm were used to forecast claim occurrence. Likewise, the study also evaluated and compared the performance of these models (Hanafy and Ming, 2021).

Thus, after analyzing key papers related to big data and insurance, it can be seen that there is a need for a comprehensive bibliometric analysis of the subject area, as no previous paper has provided such an analysis. The paper will therefore serve as a guide for future researchers since it will help them identify remarkable documents, journals, countries, and organizations; along with helping them identify the link between different authors, countries, keywords, and citations. This paper also proposes cluster analysis and suggestions for future research to enable potential researchers to conduct their studies effectively.

3. Methodology

To identify the role of big data in the development of the insurance industry and the extent to which this research area has grown, an analysis method called "Bibliometric analysis" was employed. It is a quantitative method that aims to systematically analyze, connect, map, and visualize publications and contributions related to a certain research topic (Nobanee, 2021).

To map the research field on big data and insurance multidimensionally, four types of bibliographic analysis were applied: bibliographic coupling, co-citation, co-authorship, and co-occurrence analysis, along with the utilization of their analysis properties to explore the datasets of this research.

First, bibliographic coupling is a type of similarity measure that applies a similarity measure as a means of categorizing documents into specific, similar relationships (Khatib et al., 2022). This type of analysis considers the author's research areas, the citation network, and the paper content among other things (Lim and Buntine, 2016). Co-citation is not any different, as it is also considered to be a similarity measure, albeit the semantic type, for documents cited simultaneously (Boyack and Klavans, 2010). Reference (Small, 1973) defines co-citation analysis as a sort of measurement that evaluates the frequency of two cited documents together to assist structure the similarity of a topic area. The third conducted analysis is co-authorship, which refers to a sort of collaboration between two or more authors who share the authorship of a publication, as well as the effort necessary to publish it (Glänzel and Schubert, 2005). In co-authorship, a journal article will represent any form of individual contribution; this will allow each author to hold and share fair responsibility for the results (Ponomariov and Boardman, 2016). Lastly, when dealing with a variety of data mining techniques when faced with both numeric and textual data, co-occurrence analysis tends to be applied as the basis for and framework of such data (Buzydlowski, 2015). Co-occurrence analysis, which is also known as coincidence or concurrence, is the process of quantitively counting a pair of word-based data with some mutual correlation (Muppidi and Reddy, 2020).

For this research, Scopus was used as the primary database tool to build the search query and collect the required information for the conducted analysis. Scopus is the most leading abstract and citation database of extremely substantial peer-reviewed publications, including journal articles, conference proceedings, books, and book chapters (Nobanee et al., 2021). Using such a database helped in accessing a structured data set of documents and articles centered primarily on how insurance sectors utilize Big Data. A query was built and used to explore the Scopus database on April 30, 2023.

As shown in Table 1 and Figure 1, the built query was refined three times to narrow the final dataset to the most relevant documents only. This was done by including the conditions that satisfied our report objective. The refining process is as follows: The set search criteria for the first query included the appearance of the two main keywords "big data" AND "insurance" in the titles, abstracts, or keywords of the database documents. The search criteria resulted in the generation of 1214 documents. Then, the search criteria of the query were modified to include all types of documents and exclude books, erratum, trade journals, and undefined documents. In addition, the query was further refined by limiting the search results to documents written in English only, which led to a reduction in the number of documents from an initial set of 1214 to 1078. Furthermore, the final built query considered all subject areas, in addition to inclusive access to both open- and non-open-access documents. The final refined dataset included documents published within the last five years (2012–2023). The data was extracted on April 30, 2023.

Table 1
SEARCH QUERY

Description	Conditions	Documents
Search query	TITLE-ABS-KEY (“big data” AND “insurance”)	1214 Documents
Search query after refine	TITLE-ABS-KEY (“big data” AND “insurance”) AND (EXCLUDE (SRCTYPE, “b") OR EXCLUDE (SRCTYPE, “d") OR EXCLUDE (SRCTYPE, “Undefined”))AND (EXCLUDE (DOCTYPE, “bk”))AND (EXCLUDE (DOCTYPE, “er”)) AND (LIMIT-TO (LANGUAGE, “English”)) AND (EXCLUDE (LANGUAGE, “French”) OR EXCLUDE (LANGUAGE, “German”))	1078 Documents
Access	We included both open access and non-open access documents.	
Search date	April 30, 2023	
Years	All years (2012–2023 (April))	
Subject area	We have excluded books, Erratum, trade journals, and undefined documents	
Source type		
Language	We limited our search to the English language.	

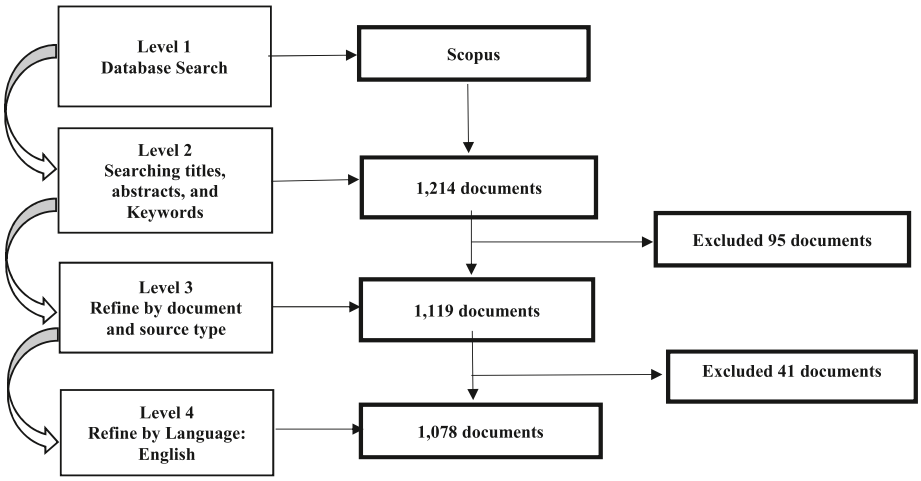

Fig. 1. The process applied for delimiting literature (PRISMA chart).

VOSviewer was used to process and analyze the dataset collected from Scopus. The software is a scientific mapping and data network tool that tends to systematically visualize a large amount of bibliometric data. The initials of the software's name “VOS” stand for “Visualization of Similarities,” which is represented in terms of the analyzed items and their distances (Vanhala et al., 2020). VOSviewer is a bibliometric analysis software that enables researchers to explore large bibliographic datasets and identify patterns, connections, and trends in a specific research topic (van Eck and Waltman, 2010). VOSviewer is one of the most used tools in several bibliometric reviews (Baker et al., 2020; Khandelwal et al., 2022; Ellili, 2022; Nobanee and Ellili, 2023). VOSviewer has several advantages in bibliometric analysis (van Eck and Waltman, 2010). These advantages include the free accessibility to VOSviewer, the simple design of VOSviewer's interface, flexible data input (Scopus, Web of Science, and other databases), customizable analysis (keywords analysis, co-citation analysis, and bibliographic coupling analysis), and several visualization options (cluster maps and network maps).

4. Results

4.1. Publication growth

Data retrieved from the Scopus database on big data and insurance included 1078 different documents. Published documents were published from 2012 April 30, 2023. Fig. 2 illustrates the overall trends in the topic. In 2012, only one document was published, and started to increase on a yearly basis and reached a maximum number of 180 documents in 2021. Thus, the overall trend is upward and is expected to grow even further because big data and insurance are emerging topics that are attracting scholars from all around the world.

Fig. 3 Presents the distribution of documents by type, while Fig. 4 shows the distribution of the same documents per subject. Both figures indicate various trends across the published documents. More particularly, Fig. 3 indicates that journal articles dominate the

A. PUBLICATION GROWTH

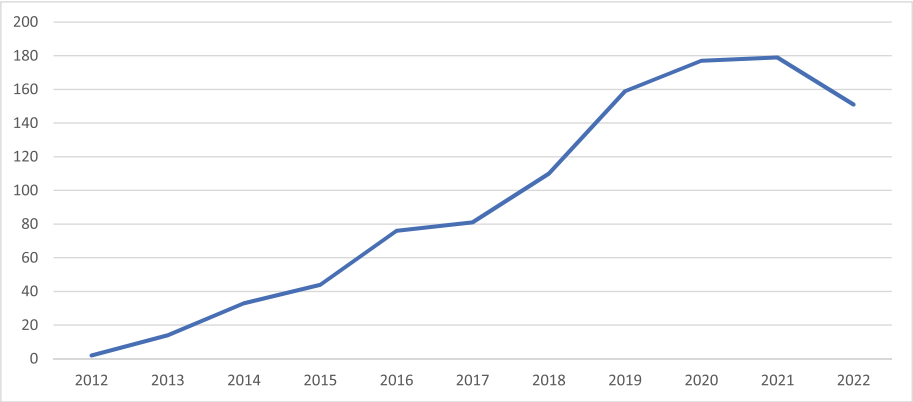

Fig. 2. Annual number of publications on big data and insurance.

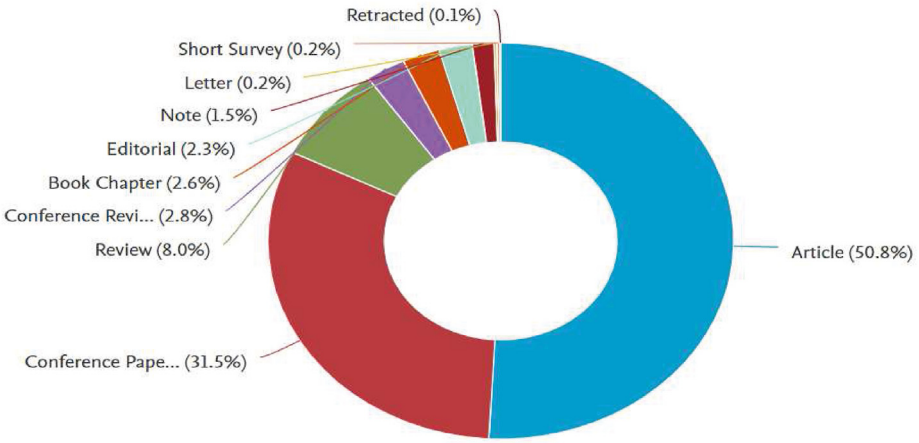

Fig. 3. Documents by type.

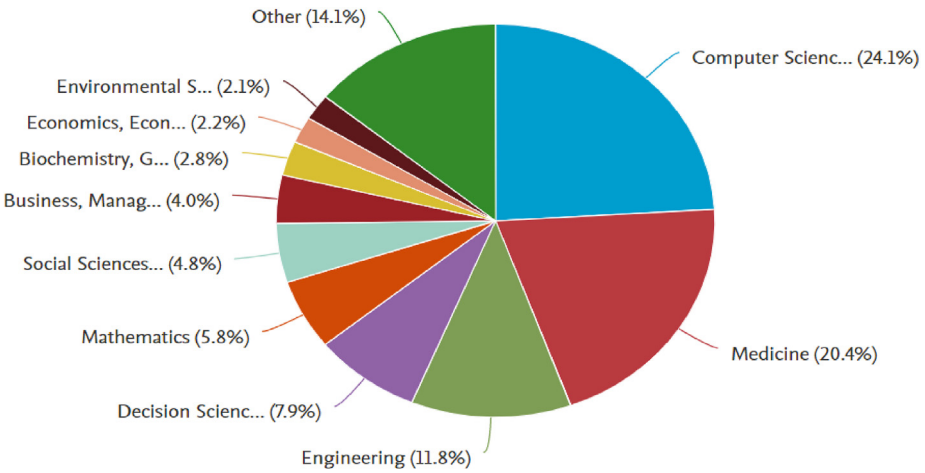

Fig. 4. Documents by subject area.

publication on big data in insurance research (50.8%), followed by conference papers (31.5%). In addition, Fig. 4 shows that most documents on big data in insurance research were conducted in computer science field (24.1%), while only 4% of the documents were related to business, management and accounting and only 2.2% of published documents were related to economics, econometrics, and finance.

4.2. Citation analysis

4.2.1. Top documents

Utilizing VOSviewer, the top documents published in journals with their number of citations were identified. The ranking of the documents was based on the number of citations. This type of analysis will provide future scholars with the top documents with the most relevant data regarding the topic to gain a deeper insight into big data and insurance and know exactly where to start (see Table 2).

Several papers with a high number of citations have been published. In the first place came the document "Big data analytics in healthcare: Promise and potential" with 1779 citations which is considered a high number as the paper has been published in 2014. In addition, this document has the highest total link strength indicating its strong connection with other documents. Moreover, the paper has around 3 times the citation of the paper placed second which is "Taiwan's national health insurance research database: Past and future" that has 525 citations (see Fig. 5). In general, the papers have a relatively high number of citations since the topic started recently in 2012; thus, long time has passed for each paper to be well recognized, and because the topic is significantly increasing over time.

Table 2
Top 10 cited documents on "big data" and "insurance"

	Document	Citations	Total Link Strength
1	Raghupathi and Raghupathi (2014)	1779	18
2	Hsieh et al. (2019)	525	1
3	Citron and Pasquale (2014)	370	2
4	Price and Cohen (2019)	358	5
5	Xu et al. (2021)	232	0
6	Massie et al. (2014)	232	1
7	Lehrer et al. (2018)	193	1
8	Fröhlich et al. (2018)	178	0
9	Patil and Seshadri (2014)	167	0
10	Lin et al. (2017)	160	2

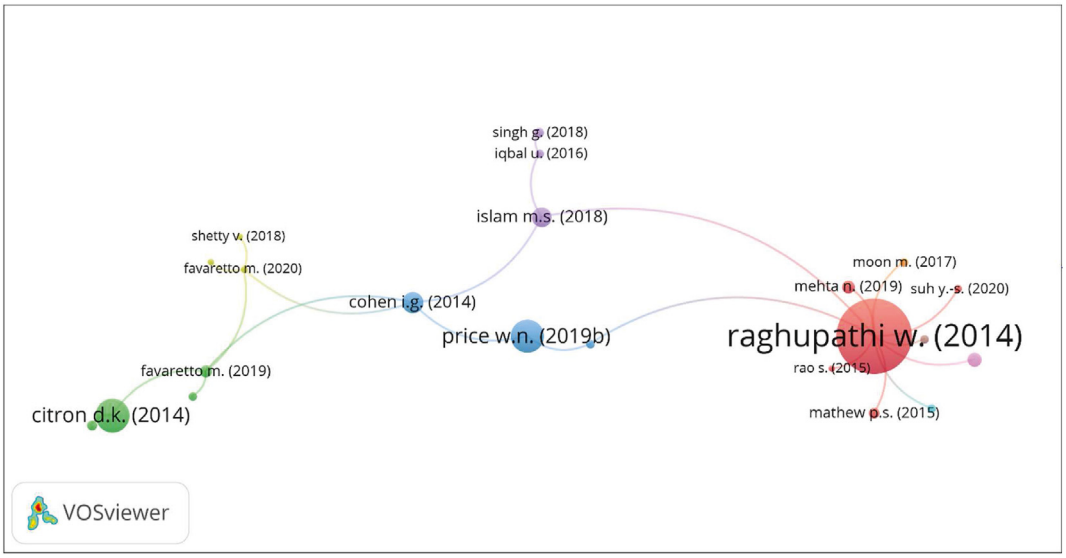

Fig. 5. Top documents map. Source: Own elaboration using VOSviewer.

4.2.2. Most productive sources

Table 3 and Figure 6 lists the top ten sources based on the number of citations. First, it came to *Health Information Science and Systems* with a total number of 1779 citations and the highest total link strength (18), although it has only one published document in this regard. However, it could be understood because of the document "Big data analytics in healthcare: Promise and potential" is the top document in this field and is published in *Health Information Science and Systems*. The same goes for the second-place journal *Clinical Epidemiology* with a total of 528 citations as it has only published two documents including the second top citing document which is "Taiwan's national health insurance research database: Past and future".

Table 3
Top 20 most productive sources by number of citations.

	Source	Documents	Citations	Total Link Strength
1	Health Information Science and Systems	1	1779	18
2	Clinical Epidemiology	2	528	1
3	Washington Law Review	1	370	2
4	Nature Medicine	1	358	6
5	Journal of The American College of Cardiology	3	238	0
6	American Journal of Transplantation	1	232	1
7	Journal of Healthcare Informatics Research	1	232	0
8	Journal of Medical Internet Research	10	225	1
9	IEEE Access	5	211	2
10	Health Affairs	3	209	5
11	Journal of Management Information Systems	2	198	1
12	Bmc Medicine	2	184	0
13	IEEE International Congress on Big Data, 2014	2	177	0
14	Journal of Big Data	7	174	11
15	2nd IEEE International Conference on Big Data Security on Cloud, 2016	2	148	0
16	Procedia Computer Science	2	139	4
17	Ageing Research Reviews	1	131	0
18	Bulletin of the World Health Organization	2	131	4
19	IT Professional	2	130	3
20	Healthcare	2	129	4

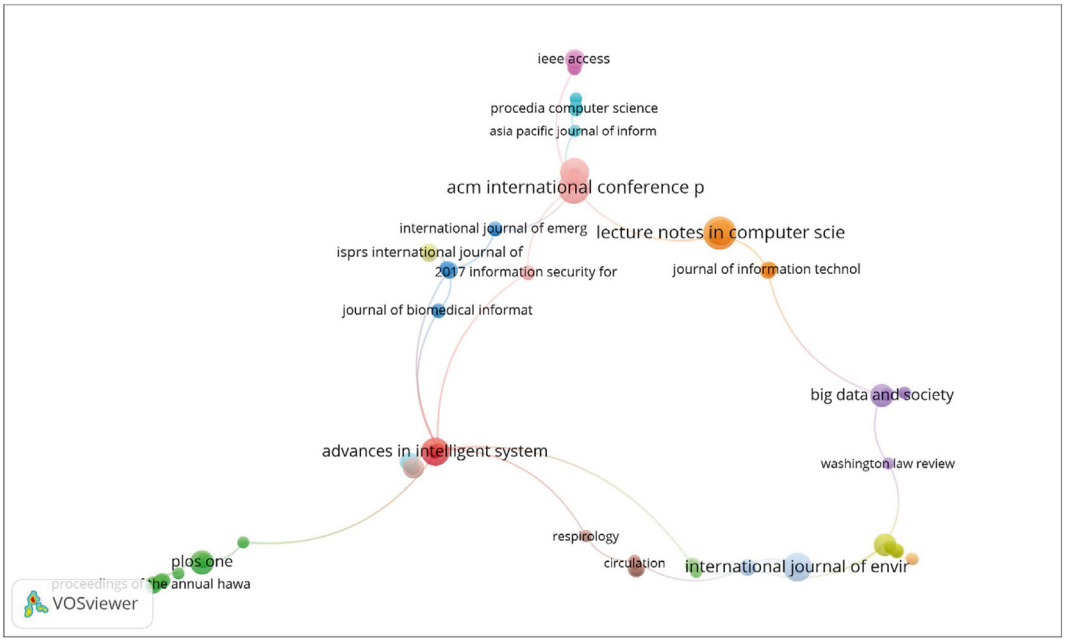

Fig. 6. Mapping of most productive sources map. Source: Own elaboration using VOSviewer.

4.2.3. Most influential authors

Examining influential authors in big data and insurance will pave the way for future scholars willing to publish in this regard to communicate and collaborate with top authors in this field. These scholars could also benefit from the vast knowledge that top authors possess in a particular field as well as in other fields. Therefore, this paper focuses on the leading authors based on the number of citations because the topic is yet to arise, and all authors have only one to three documents published in this regard except Khoshgoftaar T.M. Based on the analysis generated by the VOSviewer, in the first place came all authors who contributed to writing the top document of big data and insurance titled "Big data analytics in healthcare: Promise and potential" which are Raghupathi V. and Raghupathi W. Both authors have 1779 citations. Similarly, Hsieh C.-Y., Lai E.C.-C., Lin S.-J., Shao S., –C, Su C.-C., Sung S.-F., and Yang Y.-H.K. came in second place as the authors for the second top document each with 525 citations (Table 4 and Fig. 7).

Table 4
Top 20 influential authors by number of citations.

	Author	Documents	Citations	Total Link Strength
1	Raghupathi V.	1	1779	5
2	Raghupathi W.	1	1779	5
3	Hsieh C.-Y.	1	525	0
4	Lai E.C.-C.	1	525	0
5	Lin S.-J.	1	525	0
6	Shao S.-C.	1	525	0
7	Su C.-C.	1	525	0
8	Sung S.-F.	1	525	0
9	Yang Y.-H.K.	1	525	0
10	Price W.N.	3	400	7
11	Citron D.K.	1	370	0
12	Pasquale F.	1	370	0
13	Ii, Cohen I.G.	1	358	7
14	Khoshgoftaar T.M.	28	285	17
15	Wang F.	2	259	0
16	Lin L.	3	244	0
17	Bian J.	1	232	0
18	Glicksberg B.S.	1	232	0
19	Kuricka L.M.	1	232	0
20	Massie A.B.	1	232	0

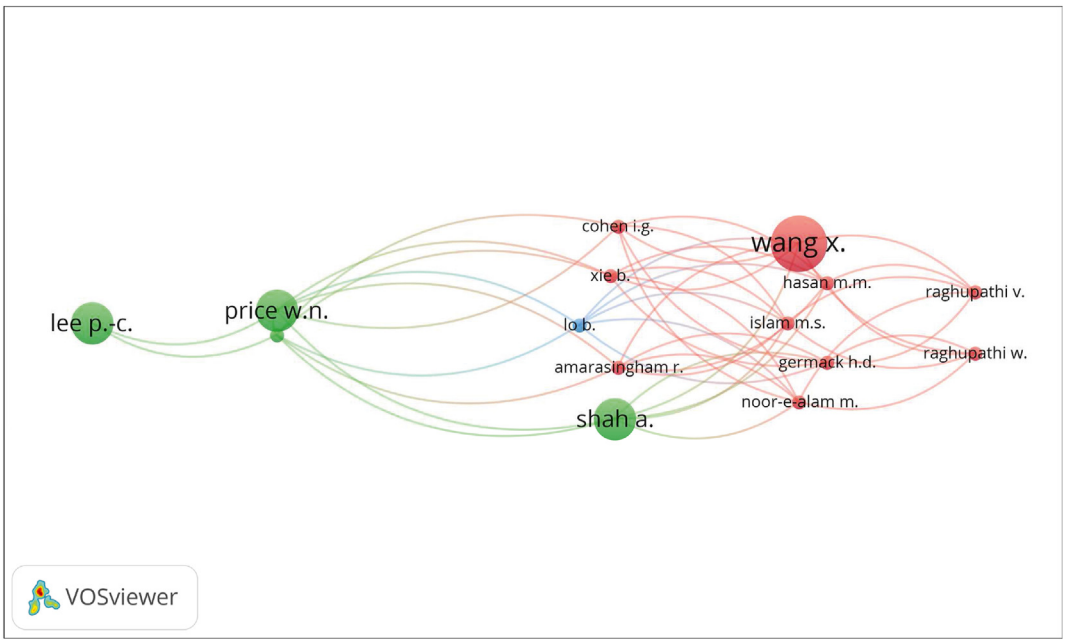

Fig. 7. Mapping of most influential authors. Source: Own elaboration using VOSviewer.

4.2.4. Top contributed organizations

Defining top organizations will help future scholars to identify entities who are interested in the topic. Thus, researchers could try to receive help from these entities when they have a great idea but require a lot of professional and financial support. Additionally, postgraduate students could seek those organizations if they would like to work and specialize in related fields.

Using VOSviewer, the top organizations were identified based on the number of citations. Came in the first place are the two organizations where the top authors of the top document are affiliated to. These organizations are Brooklyn College, City University of New York, and Graduate School of Business, Fordham University and both are located in the United States. These organizations have 1779 citations from the top documents that they published. Similarly, the second highest organizations were National Chung Cheng University, Tainan Sin Lau Hospital, Chang Gung Memorial Hospital, National Cheng Kung University Hospital, Ditmanson Medical Foundation Chiayi Christian Hospital, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University and all of them are located in Taiwan in addition to University of Illinois at Chicago located in the United States (see Fig. 8).All these organizations have 525 citations since they published the second top document in the field of big data in insurance (see Table 5).

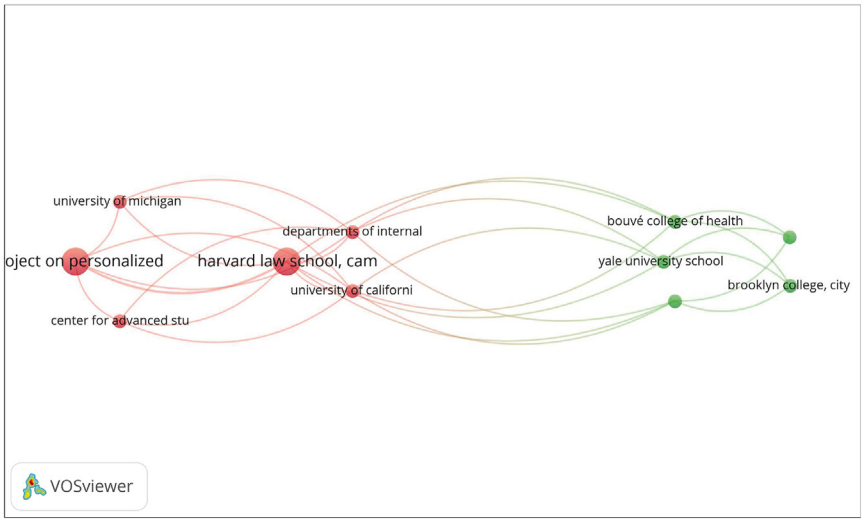

Fig. 8. Mapping of most contributed organizations. Source: Own elaboration using VOSviewer.

Table 5
Top 20 contributed organizations by number of citations.

	Organization	Documents	Citations	Total Link Strength
1	Brooklyn College, City University of New York, United States	1	1779	3
2	Graduate School of Business, Fordham University, United States	1	1779	3
3	Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Taiwan	1	525	0
4	Department of Neurology, Tainan Sin Lau Hospital, Taiwan	1	525	0
5	Department of Pharmacy Systems, Outcomes & Policy, University of Illinois at Chicago, United States	1	525	0
6	Department of Pharmacy, Chang Gung Memorial Hospital, Taiwan	1	525	0
7	Department of Pharmacy, National Cheng Kung University Hospital, Taiwan	1	525	0
8	Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chiayi Christian Hospital, Taiwan	1	525	0
9	School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Taiwan	1	525	0
10	Harvard Law School, Cambridge, United States	2	504	9
11	University of Maryland, United States	1	370	0
12	Project on Personalized Medicine, Artificial Intelligence, & Law, Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, United States	2	363	6
13	Center for Advanced Studies in Biomedical Innovation Law, University of Copenhagen, Denmark	1	358	4
14	University of Michigan Law School, United States	1	358	4
15	Department of Epidemiology, Johns Hopkins School of Public Health, United States	1	232	0
16	Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, United States	1	232	0
17	Department of Population Health Sciences, Weill Cornell Medicine, New York, United States	1	232	0
18	Department of Surgery, Johns Hopkins University, School of Medicine, United States	1	232	0
19	Institute for Digital Health, Icahn School of Medicine at Mount Sinai, United States	1	232	0
20	U.S. Department of Defense Joint Artificial Intelligence Centre, United States	1	232	0

4.2.5. Most popular countries

Identifying the top countries publishing on the topic of big data and insurance will help researchers find countries that are currently interested in and working in this field. Additionally, it could pave the way for new graduates who would like to continue their studies in this field to look for opportunities in these countries as they already have knowledge in this emerging field.

Using VOSviewer, papers were ranked based on both the highest number of documents and the highest number of citations. Based on the number of documents, the United States came in first place with 286 different documents (30.36% of total publications), followed by China with 149 documents (around 15.81% of total publications), and finally South Korea with 148 documents (around 15.71% of total publications). Hence, it could be seen that the top three countries have developed and emerging economies. Thus, demonstrating their interest in adopting the latest technologies to maintain and enhance their economies even further.

Similarly, the United States was placed first on the citation level with a total of 8216 citations. This could be understood as the top document published in this field by American affiliations, thus contributing significantly to the citation score of the United States. This was followed by China with a total number of 1395 citations which is around half of the United States' score (see Fig. 9). However, these numbers and ranks are subject to change easily in the coming five to ten years, as the topic is still rising, and much research is yet to be done (see Tables 6 and 7).

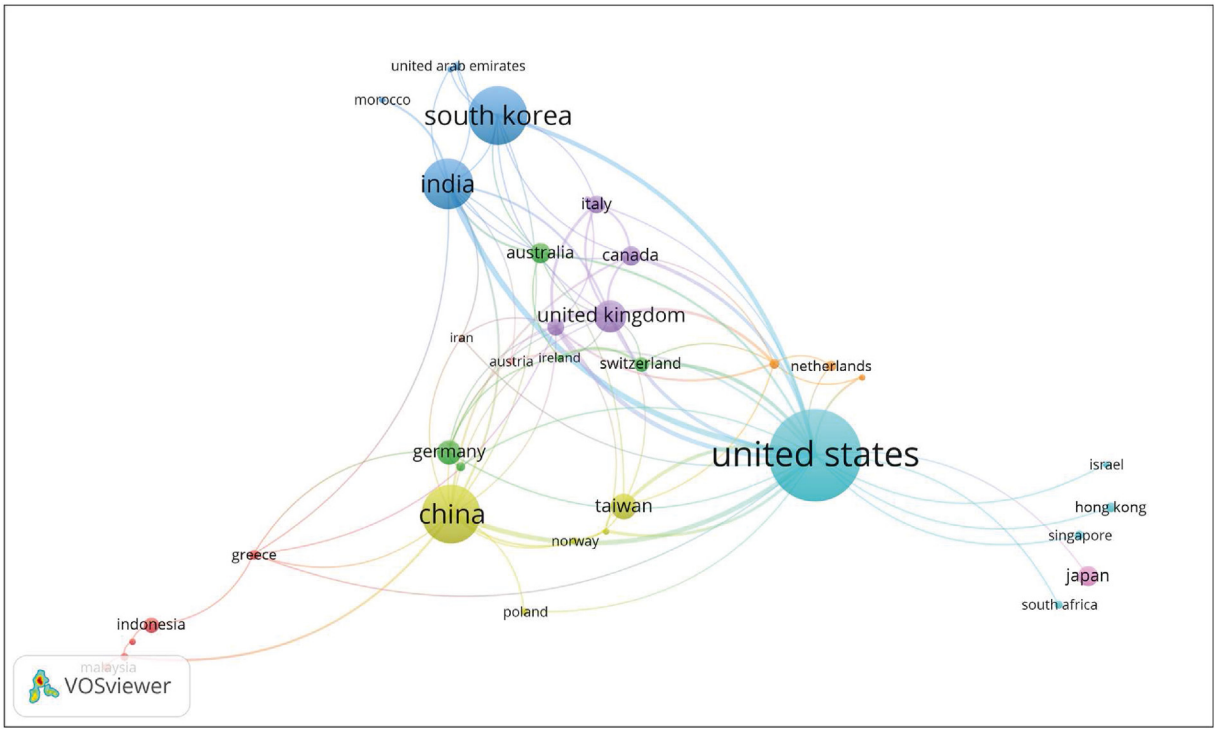

Fig. 9. Mapping of most popular countries. Source: Own elaboration using VOSviewer.

Table 6
Top 10 countries by number of documents.

	Country	Documents	Citations	Total Link Strength
1	United States	286	8216	65
2	China	149	1395	23
3	South Korea	148	1021	14
4	India	119	823	25
5	United Kingdom	61	1139	25
6	Taiwan	45	955	9
7	Germany	41	566	6
8	Australia	32	554	7
9	Japan	32	169	1
10	Canada	29	475	13

Table 7
Top 10 countries by number of citations.

	Country	Documents	Citations	Total Link Strength
1	United States	286	8216	65
2	China	149	1395	23
3	United Kingdom	61	1139	25
4	South Korea	148	1021	14
5	Taiwan	45	955	9
6	India	119	823	25
7	Germany	41	566	6
8	Switzerland	20	558	8
9	Australia	32	554	7
10	Denmark	6	490	7

4.3. CO-CITATION ANALYSIS.

4.3.1. TOP CO-CITED REFERENCES

Co-citation analysis is a type of analysis that helps to structure the similarity of a particular subject area by evaluating the frequency of a cited reference. Two documents are called co-cited references when they both are included in the reference list of another third document (Khatib et al., 2022). As shown in Fig. 10, the co-citation analysis found that there are 10 documents out of 1779 that are co-cited at least five times. The analysis resulted in assigning all the co-cited references to only one cluster, which means that all the co-cited references in the network map are closely related to each other.

The relatedness strength of the analysis items, such as cited references, is represented by the distance between items appearing on the network map. Thus, the closer the cited references are to each other, the stronger their relationship.

In this co-citation analysis, the total link strength, which is provided by VOSviewer, was used to measure the strength of a pair of connected references. Based on Table 8 (Dutta and Ang, 2016), is the most co-cited reference with a total of 40 citations but it has the lowest link strength. All other co-cited references were cited at least five times with total link strength ranging between 2 and 57. These co-cited references were published between 1949 (Tukey, 1949) and 2018 (Herland et al., 2018). Most of these references are publications in computer sciences field that focus on Big Data. Relatively, the highest connected references-with has 57 total link strengths have five citations and are on cloud computing and algorithm (Gai et al., 2016; Qiu et al., 2015).

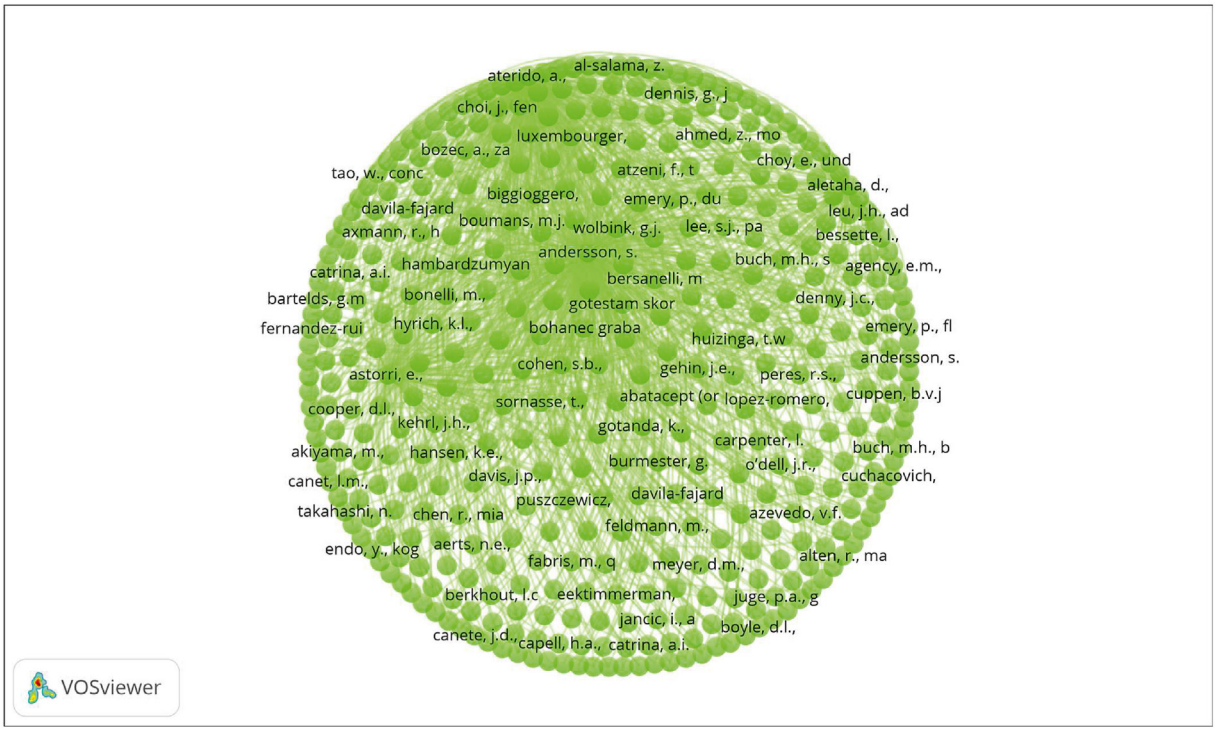

Fig. 10. Mapping of top co-cited references. Source: Own elaboration using VOSviewer.

Table 8
Top 10 co-cited references.

	Cited reference	Citations	Total Link Strength
1	Dutta and Ang (2016)	40	0
2	Breiman (2001)	11	2
3	Branting et al. (2016)	6	14
4	Bauder and Khoshgoftaar (2016)	6	13
5	Tukey (1949)	6	11
6	Gai et al. (2016)	5	57
7	Qiu et al. (2015)	5	57
8	Herland et al. (2018)	5	12
9	Chandola et al. (2013)	5	10
10	Gai (2014)	5	48

4.3.2. *Source co-citation analysis*

Co-citation analysis in terms of sources provides a more insightful overview of the most significant subject areas where relevant documents were co-cited. The threshold set for this analysis was a minimum of 25 citations per source. As shown in Fig. 11 shows 13 sources out of a total of 160.

The sorted sources are mostly related to medical, natural, and computer sciences. They are classified into five clusters. The purple and yellow clusters are related to natural sciences and led by Plos one and Nature, respectively. The green and blue clusters are related to medical sciences and led by JAMA (Journal of the American Medical Association) and BMJ (British Medical Journal), respectively. While the greenish-yellow cluster is related to computer sciences and led by IEEE Access. The co-citation links of sources included in purple, yellow, green, and blue clusters have a higher level of relatedness than other sources which indicates that most of the publications on big data in insurance combine between natural and medical sciences. In addition, most of medical sciences are related to respiratory and rheumatic diseases.

As listed in Table 9, Plos One recorded 228 citations, which is more that the double of the average citation count (101 citations). This journal was the second strongest connected source with a total of link strength of 3486. JAMA is the second highest cited journal with 209 citations and a total link strength of 2614. The Lancet comes in the third place with 174 citations and a total link strength of 2632. This finding indicates that the strongest co-cited documents are those that discuss the relationship between Big Data and insurance from the natural and medical aspects.

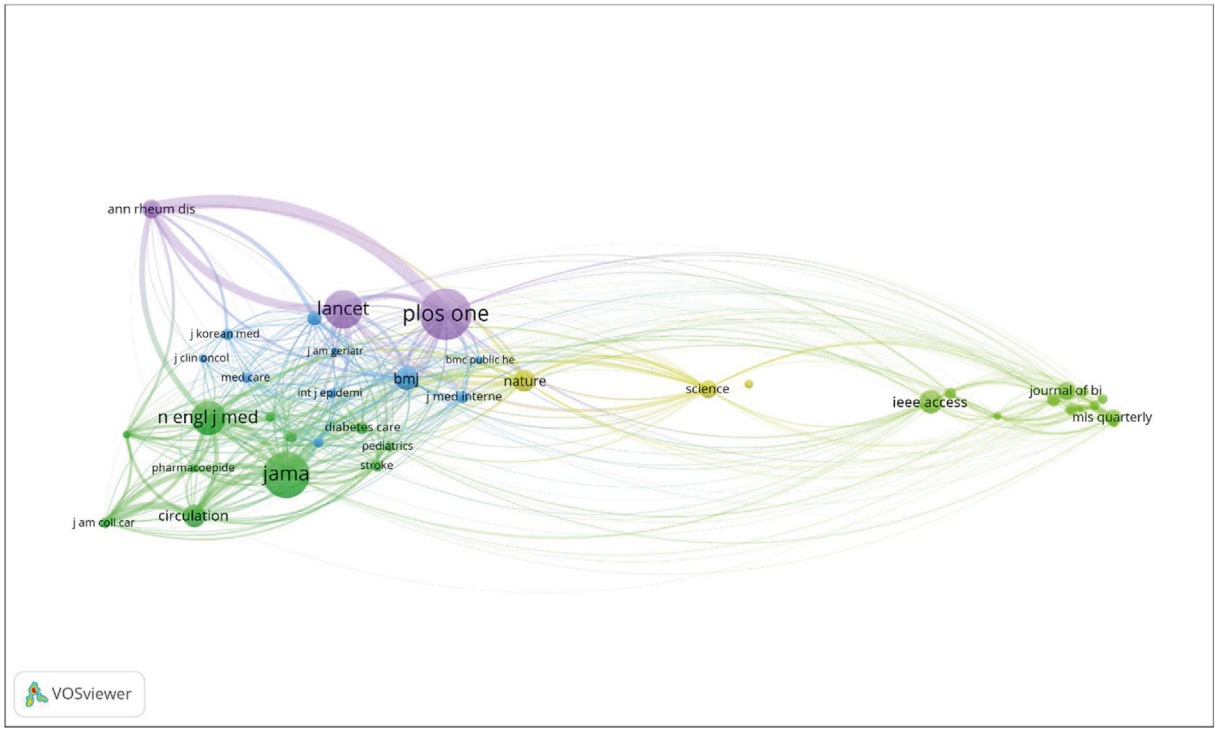

Fig. 11. Sources co-citation map. Source: Own elaboration using VOSviewer.

Table 9
Top co-citation sources.

	Source	Citations	Total Link Strength
1	Plos One	228	3486
2	JAMA	209	2614
3	Lancet	174	2632
4	The New England Journal of Medicine	156	2419
5	BMJ	106	1259
6	Circulation	103	2056
7	Nature	98	1131
8	Annals of the Rheumatic Diseases	82	6016
9	Journal of the American College of Cardiology	49	1445
10	Arthritis Research & Therapy	31	3461
11	The Journal of Rheumatology	30	2337
12	The Annals of Thoracic Surgery	30	1133
13	Arthritis & Rheumatology	26	2308

4.3.3. Author co-citation analysis

A third co-citation analysis was conducted by the authors. The author's co-citation analysis aims to analyze the intellectual origins of the studied topic, Big Data, and insurance. The analysis criterion was set to be 20 citations or above; out of 160 authors mentioned in the generated dataset, only 20 authors were cited at least twenty times. Furthermore, the results of the analysis formed only two clusters. As shown in Fig. 12, the two clusters are completely differentiated because of the distinct distance, which indicates no significant inter-relationship between them. Unlike the nature of the relationships between clusters, the relationships among the items of each cluster are highly interrelated.

Based on Table 10, Khoshgoftaar, T.M. is the most frequently cited author, with a total of 265 citations, and he is also the fourth strongly connected to author to others (6014 total link strength). His publications focused mainly on exploring big data in medicare and were published within a period of 5 years (2018–2023). In terms of citations, Qiu, M. comes in the second place with 177 citations and has the highest total link strength of 16,338. Qiu, M. is interested in cloud computing for Big Data, and this interest is reflected in the topics of his five cited articles (2016–2018). Gai, K. is the third most cited author with 129 citations and the second highest total link strength of 12,980. Qiu, M. and Gai, K. shared a common interest in cloud computing, and their four co-authored articles on this research topic were published between 2016 and 2017. In comparison to Khoshgoftaar, T.M., there is a very large difference in the total link strength, as the link strength for Khoshgoftaar, T.M. represents 36.81% and 46.33% of Qiu, M.'s and Gai, K's link strengths, respectively.

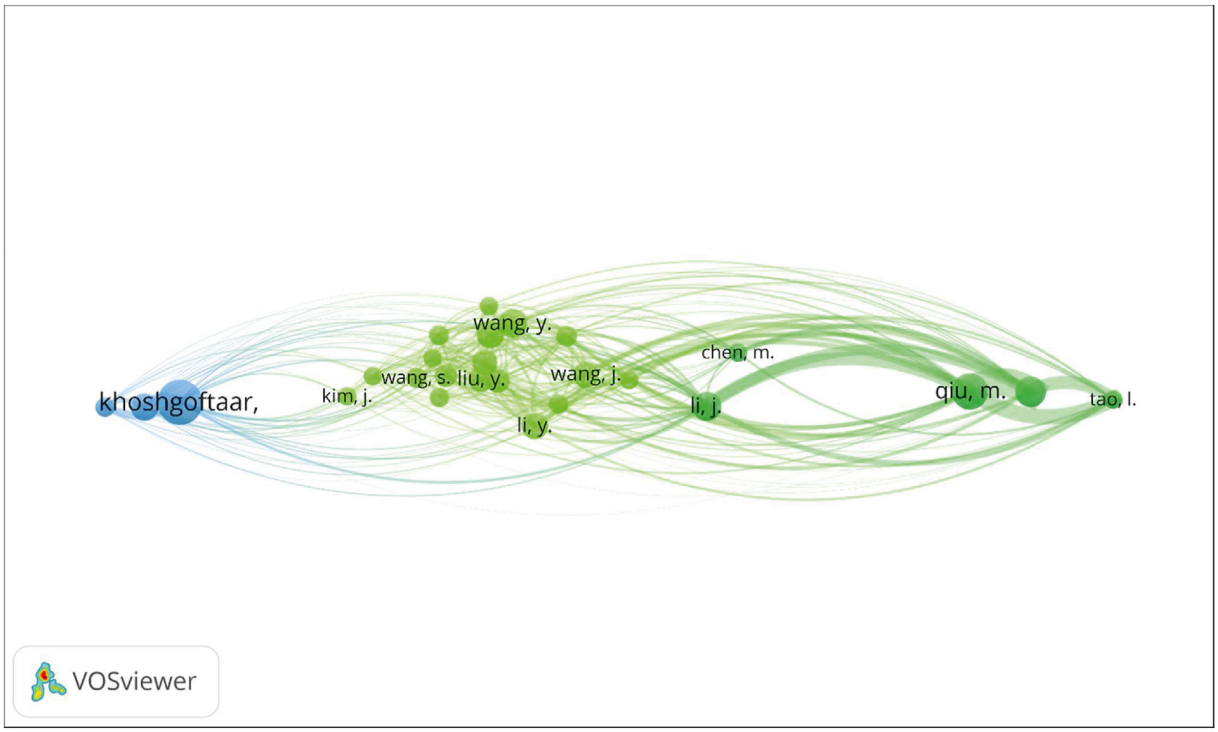

Fig. 12. Authors co-citation map. Source: Own elaboration using VOSviewer.

Table 10
Top co-citation authors.

	Author	Citations	Total Link Strength
1	Khoshgoftaar, T.M.	285	6014
2	Qiu, M.	177	16,338
3	Gai, K.	129	12,980
4	Li, J.	120	7237
5	Wang, Y.	117	3382
6	Zhang, Y.	111	3646
7	Bauder, R.A.	108	3070
8	Li, Y.	99	3595
9	Wang, J.	92	3512
10	Liu, Y.	89	2815
11	Wang, H.	87	2706
12	Li, X.	67	3294
13	Tao, L.	51	5505
14	Chen, Z.	51	3016
15	Zhang, L.	48	2353
16	Ming, Z.	47	5661
17	Zhao, H.	40	4032
18	Liu, M.	36	3135
19	Woo, J.	25	4411
20	Qin, X.	23	2653

The other listed 17 authors,' relevant publications were cited between 23 and 120 times, and a total link strength ranging between 2353 and 7237.

4.4. Co-authorship analysis.

4.4.1. Top collaborated co-authors

Co-authorship analysis assesses the extent of collaboration between two or more authors who share the authorship of a publication as well as the efforts required to produce it. Such scientific collaboration plays a vital role in broadening the research scope and accelerating innovation (E Fonseca et al., 2016). Furthermore, it helps in unifying cross-country efforts and building an international collaboration in various research fields, such as Big Data and the insurance industry. The set threshold for this co-authorship analysis was the publication of at least one document in co-authorship. There were 125 identified authors in the research dataset who met this threshold of a total of 160. The developed VOSviewer network (Fig. 13) for this co-authorship analysis resulted in grouping the authors into one cluster which indicates that all the authors in the network map are closely related to each other.

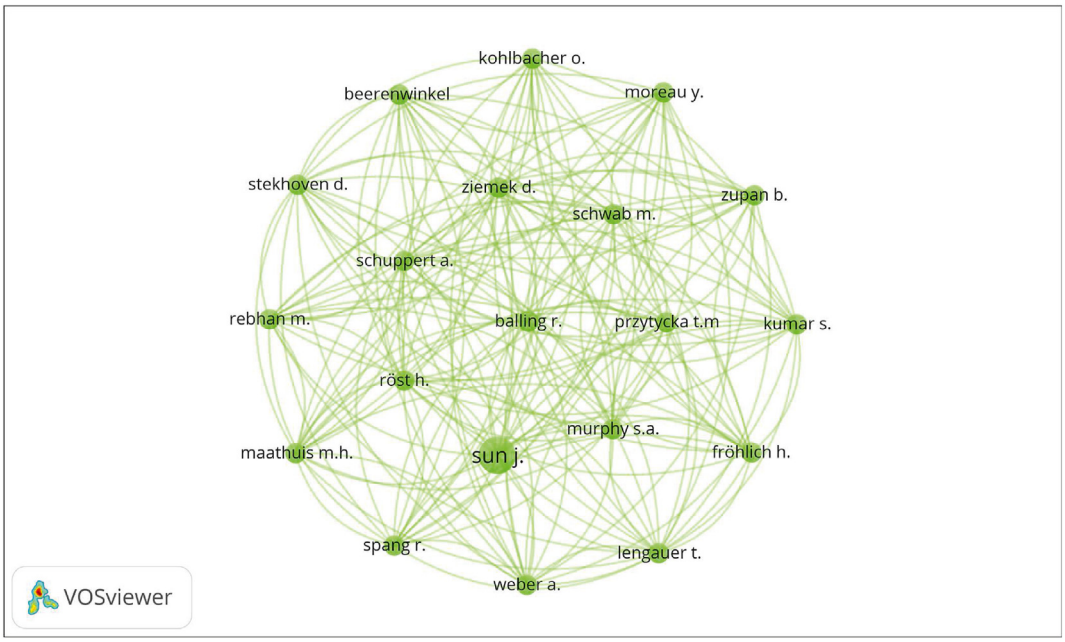

Fig. 13. Mapping of top co-authors. Source: Own elaboration using VOSviewer.

In addition to the threshold for one document, Table 11 lists 20 authors whose documents were cited at least 232 times. Interestingly, the results of the identified authors and their number of citations in this analysis are the same as those of the most influential authors identified in the citation analysis. For instance, Raghupathi V. and Raghupathi W. are the top two cited authors in both types of analysis, with 1779 citations each, which is much higher than the average citation (522 citations per author). In terms of total link strength, Khoshgoftaar T.M. had the strongest co-authorship links with other researchers (21 links each) and he is affiliated to Florida Atlantic University in the United States. Wang F. has the second highest total link strength of 17 and he is affiliated to College of Management, Yuan Ze University, Chungli, in Taiwan. These findings corroborate those of the organization analysis indicating the active research on big data in insurance in both the United States and Taiwan.

4.4.2. Top collaborated organizations

The second co-authorship analysis was conducted at the organizational level, through which organizations and affiliations that share the co-authorship of a certain document were identified. The criterion set for this VOSviewer-based analysis was publishing at least one document, and 87 organizations identified in the dataset met this criterion. As shown in Fig. 14, this analysis resulted in the formation of one cluster. Most of the organizations that appear on the map (Fig. 14) are health and medical organizations.

In terms of citations, the results are indistinguishable from those of the citation analysis of organizations. The first two listed organizations in Table 12 have one document and 1779 citations. Both organizations operate in the field of computer science. Unsurprisingly, these top two listed organizations are based in the United States, and this result is expected since the top two co-authors identified in the previous analysis were also originally from the United States. This finding proves the outstanding scientific importance that American research organizations have performed to precisely track and identify the impact of big data on the health and insurance industry. In addition, it has been found that Tai organizations have the highest strength of corporate collaboration with a total link strength of up to 12. "Department of Preventive Medicine" and "Institute of Artificial Intelligence and Big Data in Medicine" are the most connected organizations when it comes to a collaboration of authorship (Total link strength:6). Apart from this finding, it can be said that medicine-related organizations are more strongly connected than computer science organizations in terms of how they collaborate with others to improve the research field.

In general, the co-authorship and organization co-authorship analyses concluded that Tai authors and organizations have the greatest total link strength in terms of authorship collaboration. This finding is despite the fact that American publications are much more cited than Tai publications.

4.4.3. Top collaborated countries

The third co-authorship analysis focuses on a wider context than the previous two co-authorship analyses. This analysis was conducted at the country level, and the extent to which a country shares the co-authorship of publications relevant to Big Data and insurance. All identified countries in this research dataset met the criterion of publishing at least one document with co-authorship (47 countries). As shown in Fig. 15, the countries were geographically divided into four clusters. The United States dominates the research field of big data and insurance with 8216 citations, representing 39.78% of the total citations for the listed 47 countries. This finding can be explained by the results of the previous two co-authorship analyses, since both the top co-authors and organizations are from the United States. In the previously discussed citation analysis, the United States was also found to be the most popular country of origin of authors interested in investigating the role of Big Data in the insurance industry. In addition, the United States had the largest number of co-authorship citations, and the highest total link strength.

Table 11
Top 20 co-authors by number of citations.

	Author	Documents	Citations	Total Link Strength
1	Raghupathi V.	1	1779	1
2	Raghupathi W.	1	1779	1
3	Hsieh C.-Y.	1	525	6
4	Lai E.C.-C.	1	525	6
5	Lin S.-J.	1	525	6
6	Shao S.-C.	1	525	6
7	Su C.-C.	1	525	6
8	Sung S.-F.	1	525	6
9	Yang Y.-H.K.	1	525	6
10	Price W.N.	1	400	1
11	Citron D.K.	1	370	1
12	Pasquale F.	1	370	1
13	Ii, Cohen I.G.	1	358	1
14	Khoshgoftaar T.M.	28	285	21
15	Wang F.	2	259	8
16	Lin L.	3	244	7
17	Bian J.	1	232	5
18	Glicksberg B.S.	1	232	5
19	Kuricka L.M.	1	232	2
20	Massie A.B.	1	232	2

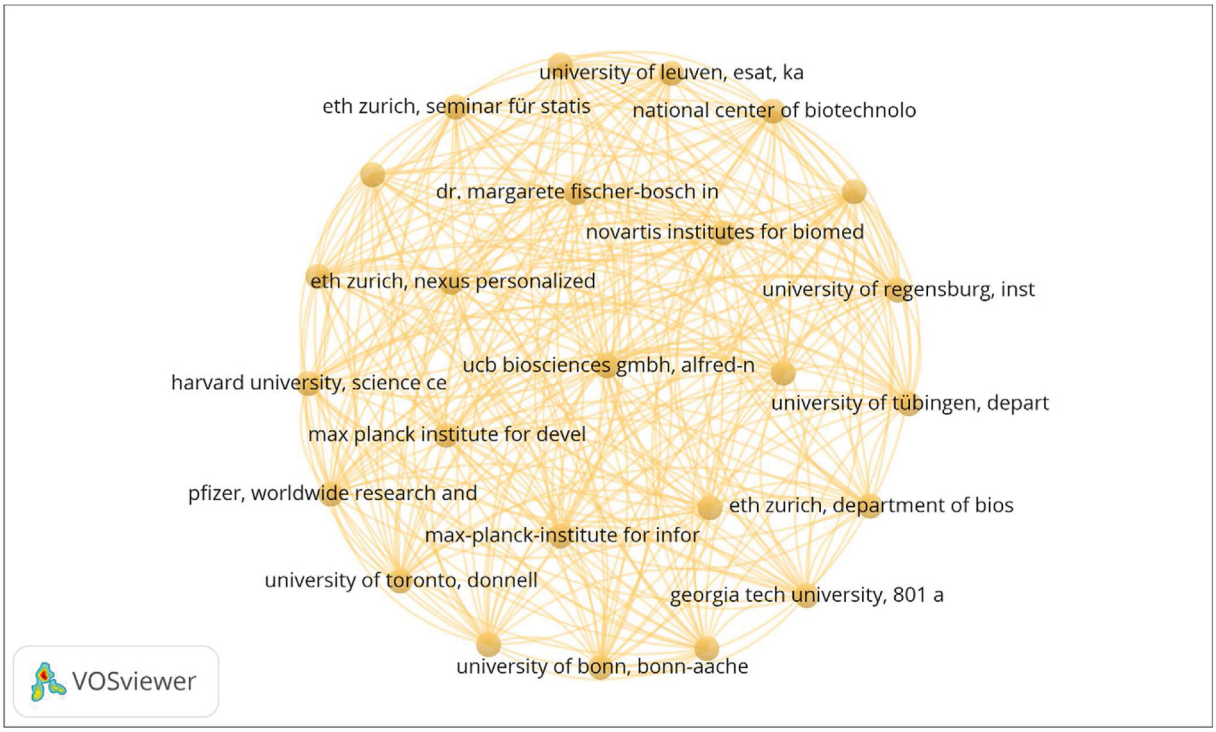

Fig. 14. Organizations co-authorship network. Source: Own elaboration using VOSviewer.

Table 12
Top 20 collaborated organizations by number of citations.

	Organization	Documents	Citations	Total Link Strength
1	Brooklyn College, City University of New York, United States	1	1779	1
2	Graduate School of Business, Fordham University, United States	1	1779	1
3	Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Taiwan	1	525	6
4	Department of Neurology, Tainan Sin Lau Hospital, Taiwan	1	525	6
5	Department of Pharmacy Systems, Outcomes & Policy, College of Pharmacy, University of Illinois At Chicago, United States	1	525	6
6	Department of Pharmacy, Chang Gung Memorial Hospital, Taiwan	1	525	6
7	Department of Pharmacy, National Cheng Kung University Hospital, Taiwan	1	525	6
8	Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chiayi Christian Hospital, Taiwan	1	525	6
9	School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Taiwan	1	525	6
10	Harvard Law School, United States	2	504	5
11	University of Maryland, Francis King Carey School of Law, United States	1	370	
12	Project on Personalized Medicine, Artificial Intelligence, & Law, Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, United States	2	363	3
13	Center for Advanced Studies in Biomedical Innovation Law, University of Copenhagen, Denmark	1	358	3
14	University of Michigan Law School, United States	1	358	3
15	Department of Epidemiology, Johns Hopkins School of Public Health, United States	1	232	3
16	Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, United States	1	232	3
17	Department of Population Health Sciences, Weill Cornell Medicine, United States	1	232	3
18	Department of Surgery, Johns Hopkins University, School of Medicine, United States	1	232	1
19	Institute for Digital Health, Icahn School of Medicine at Mount Sinai, United States	1	232	3
20	U.S. Department of Defense Joint Artificial Intelligence Centre, United States	1	232	3

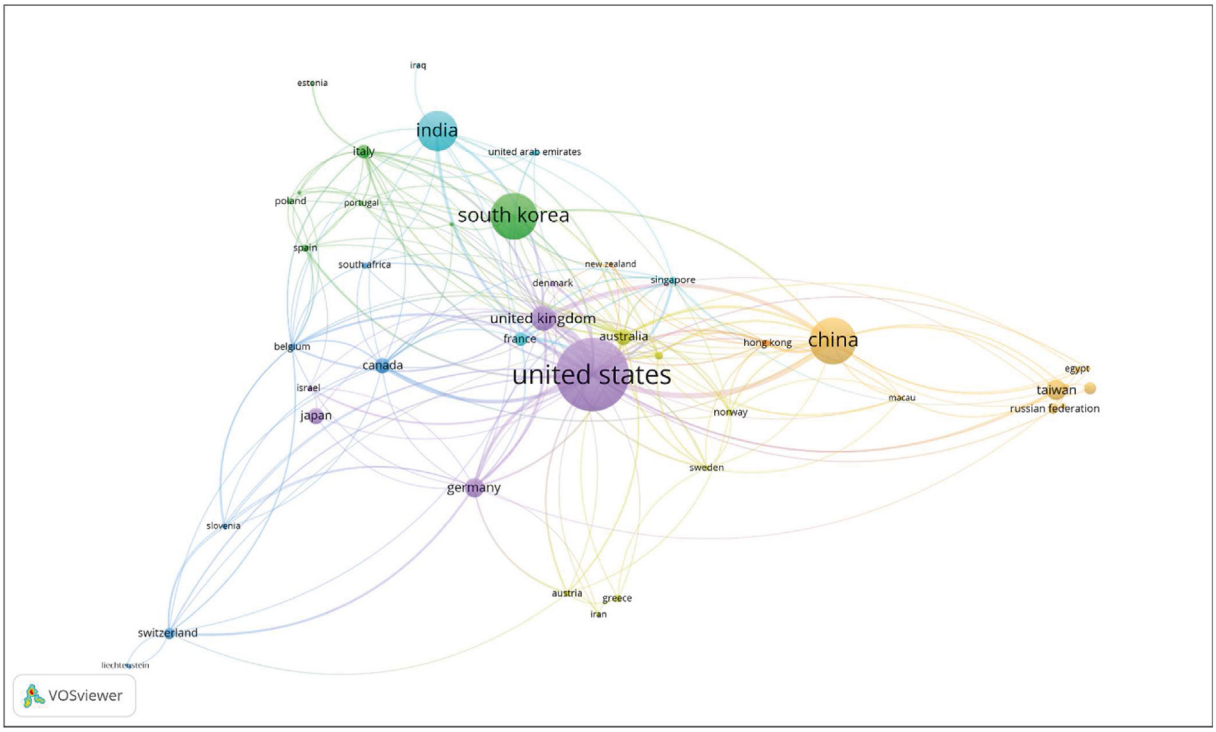

Fig. 15. Countries co-authorship network. Source: Own elaboration using VOSviewer.

Based on the size of the labels displayed in Fig. 15, China is the second most popular country after the United States, where the co-authorship documents originated from were cited 149 times. The United Kingdom is also the most interconnected country in terms of international collaboration of research authorship with 90 link strengths (Table 13).

4.5. Keywords analysis

4.5.1. Most relevant keywords

In this study, keyword co-occurrence analysis was used to identify research trends within the subject area of big data and insurance by exploring published documents. Additionally, the measurement of the total link strength was used in this analysis to determine the number of published documents in which two keywords appear together. The results of the conducted co-occurrence analysis were visualized using VOSviewer software, in which the size of each keyword's circle indicates how frequently the listed keyword is mentioned in the articles of the generated dataset. The larger the circle size, the more frequently the keyword appears in the analysis dataset.

Table 13
Top collaborated countries by number of citations.

	Country	Documents	Citations	Total Link Strength
1	United States	286	8216	144
2	China	149	1395	65
3	United Kingdom	61	1139	90
4	South Korea	148	1021	12
5	Taiwan	45	955	10
6	India	119	823	24
7	Germany	41	566	40
8	Switzerland	20	558	16
9	Australia	32	554	57
10	Denmark	6	490	11
11	Canada	29	475	28
12	Belgium	11	321	25
13	Netherlands	12	310	24
14	Sweden	6	228	15
15	Italy	26	223	28

The keyword network map (Fig. 16) shows that there are 15 critical keywords with a minimum of 100 occurrences, leading related keywords to the investigated subject. The analysis results confirm that "Big Data" has significantly dominated the research context on the role of Big Data in the insurance industry (685 occurrences, 4113 links strength). The relationship strength among all keywords under both the blue-coded and green-coded clusters is distinguishable because of the similar distances among them. Additionally, almost all listed keywords are not fundamentally related to each other, considering the moderate distance between them. The keyword "Big Data" leads the brown-coded cluster, and it can be said that it is the main linkage point between the two clusters and among the items of the cluster itself. However, the blue-coded cluster is led by the keyword "human".

The results from this analysis show that the research area of how big data can impact the insurance industry focuses mainly on humans, which occurred in 392 documents and was found as the strongest connected keyword to others— after "big data" with 4113 links (Table 14). This finding indicates that most research papers that discuss the relationship between big data and insurance examine its impact on human-related aspects. A possible justification for this finding is that individuals, families, or employees are the main targeted segments by the insurance industry; hence, any impact of big data on the insurance industry will impact humans either directly or indirectly.

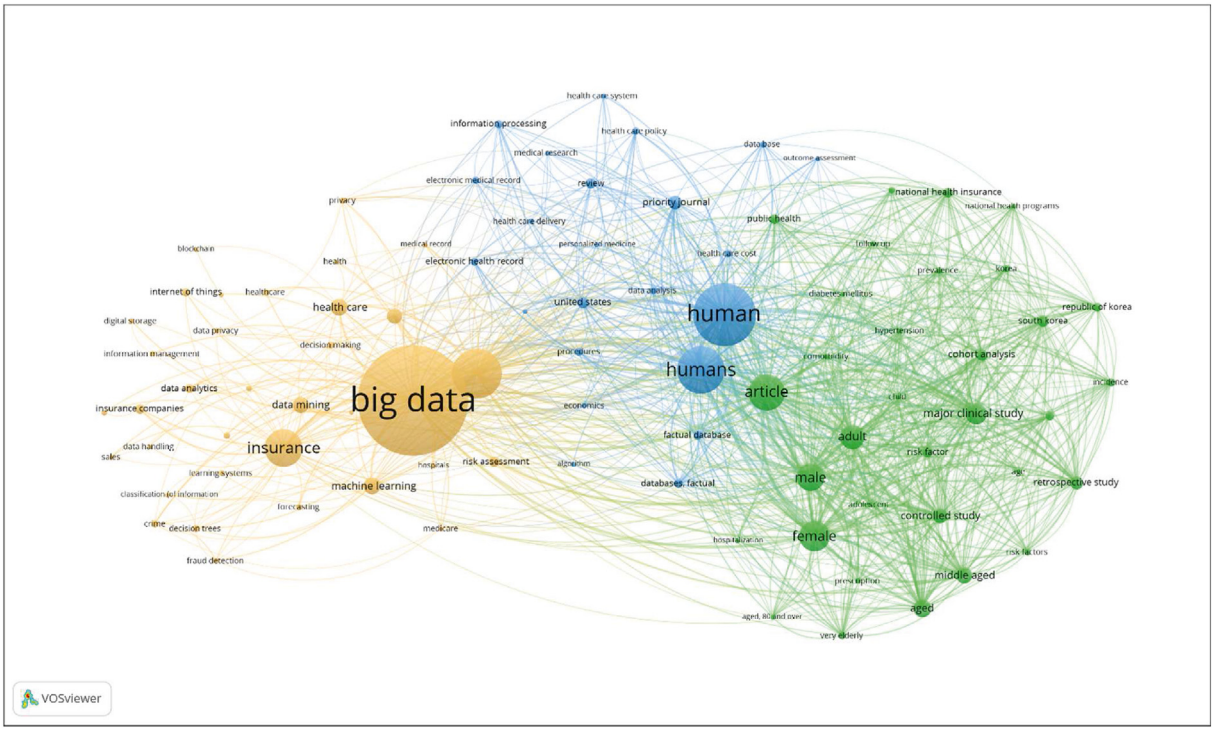

Fig. 16. Mapping of most relevant keywords. Source: Own elaboration using VOSviewer.

Table 14
Top 15 relevant keywords by number of occurrences.

	Keyword	Occurrences	Total Link Strength
1	Big Data	685	4113
2	Human	392	4384
3	Health Insurance	311	2493
4	Humans	298	3470
5	Insurance	239	1240
6	Article	226	2927
7	Female	191	2705
8	Male	174	2529
9	Adult	151	2252
10	Major Clinical Study	130	2023
11	Aged	110	1682
12	Health Care	107	761
13	Machine Learning	106	822
14	Data Mining	104	594
15	Middle Aged	100	1601

4.5.2. Top author keywords

First, an author keyword refers to a type of keyword that best reflects a document's content upon the author selection (Scopus, 2173). Through using VOSviewer, our keyword analysis resulted in identifying 20 author keywords, which are selected by minimum of 2 authors as the most relevant keywords to their research papers on big data applications in the insurance industry. This filtering process resulted in the formation of five color-coded clusters. Fig. 17 provides unarguable evidence of "Big Data" as the most important keyword chosen by authors who contributed to the investigation of Big Data in the insurance industry. Moreover, the network map shows that Big Data is the most keyword that has a co-occurrence linkage with all the other 19 author keywords. This finding indicates that Big Data is the essence and center of the studied research area, which is normal and highly expected because Big Data is a key variable in this research area. In terms of the keywords' relatedness, Fig. 17 displays that "Healthcare," "Blockchain", and "insurance"- which belong to the y brown, blue and green-coded cluster-are somehow related to each other because of their closest distances to "Big Data". This level of relatedness is also applicable for the two keywords "Big Data Analytics" and "Data Mining" that come under the brown-coded cluster. Similarly, for the yellow-coded cluster, "Fraud Detection" and "Class Imbalance" are relatively stronger than the other keywords in the other clusters because the distance between them is shorter.

Based on the figures presented in Table 15, there is a very huge gap between "Big Data" and other author keywords. Because it was mentionned by 472 documents as the primer keyword that reflects their research's contents, whereas other author keywords were chosen by only two to three authors. This is not only because of the high number of occurrences, but also because Big Data has the highest value of total link strength. To illustrate, 'Big Data' was selected almost 7 times as the best keyword that reflects the content of the same document where other listed author keywords were also selected. This is an expected finding, as all relevant publications mainly focus on big data, either as a primary or secondary research variable. In other words, there is no other keyword like "Big Data" that can describe the content of publications relevant to the studied research area except the keyword "Machine learning," which surprisingly was selected by only 65 documents. "Big Data", "machine learning", "health insurance", "healthcare", "data mining", and "insurance" appeared in the previous analysis of most used keywords. However, "Big Data" and "machine learning" appeared around 50% less than the previous analysis.

4.5.3. Top indexed keywords

Indexed keywords refer to keywords that content suppliers and providers select based on their high relevance to the content. Such keywords are standardized based on publicly accessible vocabularies. The selection of an indexed keyword takes into consideration available synonyms, plural forms, and different spellings (Scopus, 2173). The related indexed keywords to the topic of this research

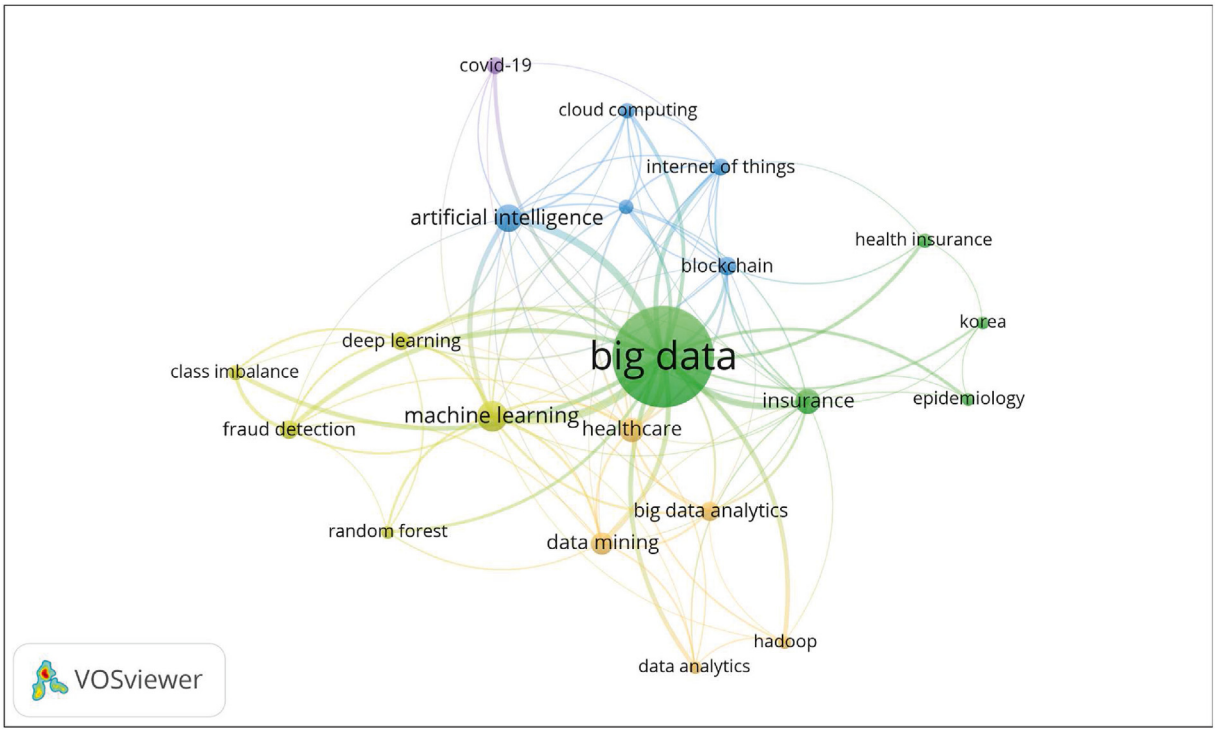

Fig. 17. Mapping network of keywords authors. Source: Own elaboration using VOSviewer.

Table 15
Top keywords authors by number of occurrences.

	Keyword	Occurrences	Total Link Strength
1	Big Data	389	262
2	Machine Learning	59	87
3	Artificial Intelligence	49	65
4	Insurance	42	47
5	Healthcare	40	57
6	Data Mining	36	41
7	Big Data Analytics	29	27
8	Fraud Detection	27	37
9	Blockchain	26	32
10	Deep Learning	26	34
11	Internet of Things	24	34
12	Covid-19	23	15
13	Cloud Computing	20	24
14	Privacy	19	29
15	Class Imbalance	18	27
16	Health Insurance	18	12
17	Data Analytics	16	23
18	Epidemiology	16	10
19	Hadoop	15	21
20	Random Forest	15	17

paper were explored by the co-occurrence's analysis on the VOSviewer. The criterion for minimum occurrences was set to 100, and 13 indexed keywords met this criterion, resulting in the formation of only three clusters, as shown in Fig. 18. The green-coded cluster is led by the keyword "Human," whereas the blue-coded cluster is led by the keywords "Big data", and the brown-coded cluster is led by the keyword "Female". Additionally, the distance among the indexed keywords, which is displayed in Fig. 18, reveals that the relatedness level among the keywords, which either belong to the same cluster or differ, is not very strong.

As presented in Table 16, the indexed keywords listed in this analysis are identical to the previous analysis of the most used keywords. Moreover, the 13 keywords that appear in this analysis have higher occurrences than in the earlier analysis. In terms of connection strength, "Human" is the keyword most linked with other relevant keywords regarding publications related to Big Data and insurance. This confirms that the two variables of this relationship have mostly been investigated from a human-based perspective. In general, it can be concluded from the overall keyword co-occurrence analysis that "Big Data" is the most frequently occurring keyword

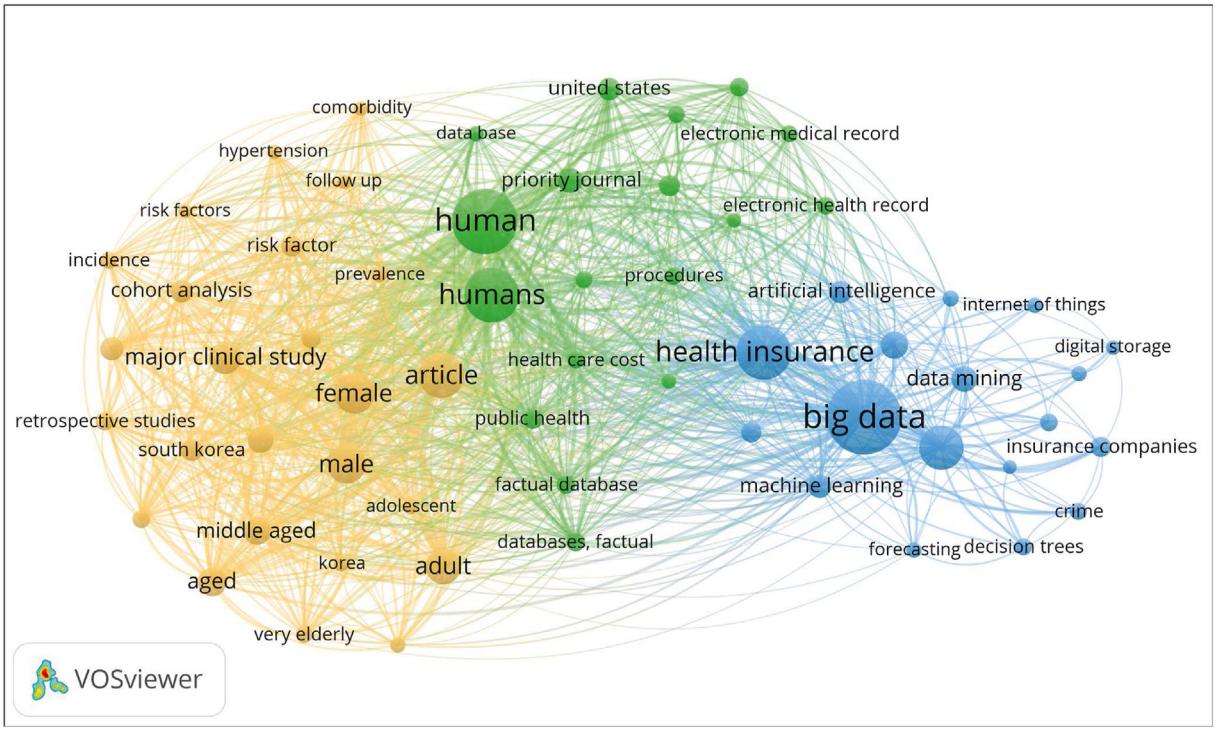

Fig. 18. Network mapping of indexed keywords. Source: Own elaboration using VOSviewer.

Table 16
Top indexed keywords by number occurrences

	Keyword	Occurrences	Total Link Strength
1	Big Data	491	2349
2	Human	392	3698
3	Health Insurance	301	1958
4	Humans	298	2957
5	Article	226	2511
6	Insurance	216	951
7	Female	191	2364
8	Male	174	2209
9	Adult	151	1966
10	Major Clinical Study	130	1770
11	Aged	110	1463
12	Health Care	100	554
13	Middle Aged	100	1397

at all three levels. This finding was expected because big data is one of the main research variables for publications discussing big data in the insurance industry.

5. Cluster analysis

The most influential papers were analyzed, resulting in four clusters. These clusters identified the keywords "big data" and "insurance" within each cluster as follows: big data in the medical field, big data and COVID 19, big data and individualized insurance, and big data application within insurance (Table 17). A comprehensive analysis was conducted on each article within the clusters to obtain insights into the purpose, findings, and recommendations for future research in the form of research questions.

Based on Table 17, the first cluster included articles that covered the application of big data technologies in the medical field. As reference (Kim et al., 2018a) explains that the use of big data technologies resulted in more accurate statistics and therefore improved research results compared to the traditional studies that relied on survey-based statistics. Similarly, in the second cluster, the research covered big data and the influence of the statistical data provided by such technologies in combatting the COVID-19 pandemic.

Cluster 3 moved away from the medical field to include research papers on big data and its applications in the insurance industry. Reference (Barry and Charpentier, 2020) suggests that the application of big data technologies to personalize insurance would not be realistic and jeopardizes the existence of insurance as it is based on the pooling of risk and demand uncertainty. In contrast, reference (Arumugam and Bhargavi, 2019) suggests that a personalized premium can be determined with the application of big data technologies.

The last cluster addresses the application of big data technologies to improve the current insurance industry practices. Reference [56]recommends a new customer profitability prediction method, while Reference (Lin et al., 2017) recommends a new classification model for insurance based on big data technologies.

Analysis of the clusters revealed two broad categories. Two of the four clusters focused on big data in the medical field, and the other two focused on big data technologies applied directly to insurance practices and methods.

6. Discussion

The purpose of this bibliometric analysis is to review the current literature on the application of big data technologies in the insurance industry, identify emerging trends, and determine gaps in this research topic. Data retrieved from the Scopus database using the keywords "big data" and "insurance" revealed 1078 documents from 2012 to 2023 (April), with a significant increase in 2021. This indicated a growing interest in this field and the relative novelty of this subject area. Bibliometric analysis provides an overview of the relevant publications and a foundation for understanding this topic. According to Reference (Hussain et al., 2016), big data usage in the finance and insurance sector has a great impact on the level of performance of the company itself. An examination of cluster four in the cluster analysis supports these findings by addressing the application of big data technologies to improve the current insurance industry practices. Reference (Ellili, 2022) recommends a new customer profitability prediction method, while Reference [57]recommends a new classification model for insurance based on big data technologies. A bibliographic coupling analysis rendered the top documents on the subject. Several papers with a high number of citations were recognized, indicating a growth in the topic. Additionally, the results of the identified authors in the co-authorship analysis are the same as those of the most influential authors identified in the citation analysis. Furthermore, the top 20 organizations were identified based on the number of citations, with the United States dominating all the analyses. This indicates that the majority of research papers on big data and insurance are conducted in the United States. In addition, based on the number of documents, the United States came in first place with 286 different documents (30.36% of total publications), followed by China with 149 documents (around 15.81% of total publications), and finally South Korea with 148 documents (around 15.71% of total publications). This leads to the conclusion that the United States is producing more relevant research findings. In keyword analysis, keyword co-occurrence analysis was used to identify research trends within the subject area of big data and insurance by exploring published documents. This finding indicates that most studies on the relationship between big data and insurance examine the impact of this relationship on human-related aspects. This is further supported by cluster analysis, which dedicates two clusters to the medical field, linking big data to medical insurance as a major topic. The gap in the existing literature can be reduced with future

Table 17
Cluster analysis

Stream/Cluster	Author	Purpose	Findings	Suggestions for future research
Cluster 1: Big data in the medical field	Kim et al. (2018a)	Research statistics on breast reconstruction were obtained from the Health Insurance Review and Assessment Services (HIRA) big data hub in Korea.	•The use of data obtained from the Health Insurance Review and Assessment Services (HIRA) big data hub resulted in more accurate breast reconstruction statistics and therefore improved research results compared to the traditional studies that relied on survey-based statistics. •This method proved to be more objective and will prove to be a good basis for future studies.	•How does big data impact the accuracy of statistical data in different fields? •How will the application of big data affect medical insurance policies? •What is the impact of big data on healthcare insurance?
	Jeong et al. (2020)	This paper investigates the use of big data from health insurance data to detect childhood development delays before registration.	•With the use of big data, disabilities were detected as early as 3years old based on medical records alone. This surpasses the average diagnostic age of 5 years through traditional methods. •Enabling early treatment options which can reduce the degree of disability in children.	Will feature engineered algorithms with a more detailed classification of a diverse range of disabilities provide for more accurate data analysis?
	Kim et al. (2018b)	Linkage of acute stroke registry and the national health claim databases to establish high-quality big data on stroke patients in Korea	By using claims data without personal identifiers, the feasibility of linking administrative data and stroke registry data was determined. This big data will allow for the analysis and development of prognostic predictions.	How can the accuracy of data linkage be improved using big data and machine learning?
Cluster 2: The use of big data to combat COVID 19 mortality and spread	Kim et al. (2020a)	Analyze big data to link preexisting comorbidities affecting covid19 deaths to determine high-risk groups in Korea.	The findings suggest that there is a high positive correlation between comorbidities and patient mortality. The results can be used to determine high-risk groups and recommend vulnerable target groups for vaccination.	How effective is early clinical intervention on decreasing high-risk group mortality?
	Byeon et al. (2021)	Analyze data to determine the survival rate as well as the factors affecting COVID-19 patients in South Korea.	The study identified several factors that contribute to a high mortality rate such as gender, where men were found to have a lower survival rate than women. In addition, malignant neoplasms of the respiratory system, diseases of the urinary system as well as diabetes were leading causes of low survival rates related to COVID-19.	How do the main challenges in using big data analytics affect the management of the COVID-19 pandemic?
	Kim et al. (2020b)	To assess the risk of infection of COVID-19 in patients being treated with antihypertensive medications by analyzing big data.	The study was conducted based on information obtained from December 31, 2019, to April 2, 2020, on adults over 40 years. The study found no correlation between patients taking antihypertensive medications and the risk of contracting the COVID-19 virus in Korea.	**Limitations of the study**: •Indirect data was used which may not reflect real compliance within the population. •The analysis does not reflect the risk of developing fatal COVID-19 as hospital stay data was not obtained. •Asymptomatic COVID -19 cases were not included in the study as only RT-PCR test data were analyzed.
Cluster 3: Big data and the impact on individualized insurance	Barry and Charpentier (2020)	To assess the impact of Big Data technologies for classifying risk to produce personalized insurance	The findings suggest that personalized insurance would not be realistic and jeopardize the very existence of insurance as it is based on the pooling of risk and demands uncertainty.	•Can individualization be obtained through the classification of risk classes? •How does the personalization of risk impact the business model of insurers?
	Arumugam and Bhargavi (2019)	To analyze Manage-How-You -Drive (MHYD) data for Usage-Based-Insurance (UBI)	The study found that by monitoring a driver's behavior more accurately a personalized premium can be determined. Therefore, the	•Can the predictions obtained from machine learning technologies and big data reduce the risk for insurers?

(*continued on next page*)

Table 17 (*continued*)

Stream/Cluster	Author	Purpose	Findings	Suggestions for future research
			adaption of big data could bridge the gap between the insurer and customer.	•Is the adaption of big data technologies essential for modern-day insurers?
	Barry (2020)	Aims to indicate the shift of insurance mechanisms from fairness based on solidarity towards an idealistic individual risk.	The research found that due to new big data technologies, a finer personalization of product can be established which increase fairness. Therefore, based on scores and predictions from big data analysis there has been a shift from collective apprehension of risk to a more individual one which allows for the development of new models in which fairness will be reduced to an algorithmic calculation.	Can the risk be managed through big data applications by predicting individual behavior rather than predicting average cost?
Cluster 4: The use of big data technologies to improve the insurance industry's current practices	Fang et al. (2016)	By adding liability reserve, this paper proposes a new customer profitability prediction method for the insurance industry.	The empirical research study found that, compared to other traditional analytics methods such as linear regression, decision tree and generalized boosted method, the random forecast regression method of big data analytics was able to predict insurance customer profitability more efficiently by measuring the real insurance customer contribution.	•**Limitations of this study**: The accuracy of the model could be influenced by the absence of customer income data. The reserve fund calculation only adopts the twenty-fourth method of calculation, therefore might lack accuracy. •How might the accuracy of the random forecast regression method of big data analytics be affected when customer income data is considered?
	Lin et al. (2017)	Provide a new classification model for traditional insurance business databases by combining it with big data technologies.	The experiment result shows that within imbalanced data the random forest algorithm rendered better results than the support vector machine in both performance and accuracy. This can be used for improving the accuracy of product marketing.	•How will the proposed algorithm respond to different types of big data analytics? •Can the accuracy of big data analysis predictions be improved by applying deep learning to the algorithm?

research expanding on big data and its application to improve the insurance industry's current practices, as this was also the topic of the most cited papers. This research offers deeper insights and highlights possible future research topics on big data and insurance.

7. Suggestions for future research

Big data and insurance have become important subject matters that researchers have focused on because not only does it include the business and specifically the financial world, but it also incorporates new technological phenomena to better bind the best of both worlds and advance majorly in research and others. Although the growing interest in research on Big Data and insurance has been increasing, there is always more potential for future research ideas that build more on the current findings and might lead to unanticipated knowledge, thereby changing the overall field. Future research is vital as it targets unanswered aspects of every research problem. Despite the growing number of studies and research, when it comes to big data and insurance, there is still a big hole that has the potential to be filled. The following suggestions for future research were made because of the massive gap found in the analysis, especially in the cluster analysis.

Broadening the scope of future research contributions is essential to strengthen the research field, build a solid practical background, and gather up-to-date data and evidence. Researchers interested in insurance and technology are encouraged to consider how to guide the private and public insurance sectors on the usage of Big Data. Researchers can direct their efforts to develop Big Data approaches and frameworks that could help insurance companies and public organizations optimally utilize their unstructured and massive data. Another research trend is to enhance the practicality of research papers on this topic. As discussed in the cluster analysis, most relevant publications discuss the relationship between Big Data and the insurance industry. Therefore, researchers should start examining the application of Big Data technologies in specific insurance markets, either at the local, sub-regional, or regional levels. For instance, a possible future topic is to investigate the performance of the insurance market in a developing country, as a case study, and how Big Data technology can be utilized to improve it.

As seen in the cluster analysis, most previous research articles focus mainly on discussing the role of Big Data in health insurance. Hence, future research projects are expected to focus on this subject at a wider level. This means exploring the significance of Big Data for insurance types other than health and medical insurance, such as property insurance. Finally, technological and economic research activities have been known as pioneers in the research field, especially during the last decade. Ensuring a strong scientific collaboration between practitioners and scholars in both fields would foster the innovation of the studied research topic "Big Data and insurance industry."

8. Limitations

As for any research, this research paper comes with several limitations. First, all documents used in the bibliometric analysis were generated only from Scopus, as it is the most popular research database. Therefore, all documents published in other databases and websites were omitted. Another limitation is that the dataset used for the analysis was limited only to the title, abstract, keywords, and English language. Therefore, this research cannot be considered as an in-depth review, and for that reason, future research projects should consider building on it. Finally, the clustering method is limited in terms of the generalized structure and the slightly tight picture of the research area. However, this initial structure was adequate to meet the purpose of this research.

9. Conclusion

Big data and insurance have gained the interest of researchers in the past five years from all around the globe and are still growing steadily. Big data has been demonstrated as a key element in facilitating the work for the insurance sector through a better understanding of customer needs, making the process of fraud detection easier, and many more benefits are yet to be revealed in the future. Thus, this study presents a comprehensive bibliometric analysis to pave the way for future researchers to conduct their studies more efficiently. The Scopus database was used to collect all relevant papers published in this regard. VOSviewer was also used to visualize the link between the datasets retrieved from Scopus.

Results demonstrated that "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis" is the top document based on the number of citations, thus making the authors and their affiliations the top authors and organizations based on the number of citations, contributing all in placing IEEE journal the first place based on the number of citations. Additionally, China was placed first based on the number of citations, and South Korea was placed first with regard to the number of documents. Although China was identified as the dominator of co-authorship publications based on the number of citations, it is not strongly connected to higher levels of collaboration. Conversely, authors and organizations from South Korea have the greatest total link strength at the level of authorship collaboration. At a wider level, the United States is the strongest connected country concerning international collaborations of authorship related to publications on big data and insurance. An additional finding is that human-related aspects are considered an integral part of the relationship between big data and insurance.

Big data has revolutionized the insurance sector, improving customer experiences, risk assessment, operations, and fraud detection. Insurers analyze structured and unstructured data to enable accurate risk pricing and personalized offerings. Understanding customer behavior patterns enhances tailored policies and marketing, improving relationships and retention. Big data aids fraud detection, quicker investigations, and reduced losses. Predictive analytics improves customer service and streamlines claims processing, enhancing satisfaction. Additionally, it predicts future trends, optimizing processes and reducing operational costs. In health and life insurance, big data monitors customer health data, leading to personalized plans. Reinsurers benefit from accurate risk assessment. The extensive impact of big data in insurance has led to a comprehensive bibliometric and systematic review. It explores various applications, providing valuable insights for researchers, practitioners, and policymakers. However, a gap exists regarding the need for a bibliometric review study to understand the structural development of the existing literature and to suggest future research. Addressing this gap is vital for informed decisions and ensuring big data's responsible use in the insurance industry's future success.

The implications of big data applications in the insurance industry, derived from a comprehensive bibliometric and systematic review, are profound and diverse. By understanding customer behavior, insurers can enhance experiences and retention rates. Improved risk assessment through data analysis leads to more accurate pricing and reduced financial exposure. Big data streamlines operations, automating tasks, and enhancing efficiency. Fraud detection benefits from early identification of suspicious patterns, minimizing losses. Personalized health and life insurance plans are possible through big data analysis of customer health and behavior. Reinsurance companies benefit from more precise risk assessment, ensuring sustainability. Predictive analytics aids in making informed strategic decisions by anticipating future trends and market conditions. However, ethical and privacy considerations demand responsible data use to maintain trust and data security. Embracing these implications can transform the insurance industry, fostering customer satisfaction, optimizing operations, and empowering decision-making while addressing ethical challenges responsibly.

Furthermore, a cluster analysis was conducted to reveal the main streams in the topic of big data and insurance that were "Big data in the medical field," "The use of big data to combat COVID 19 mortality and spread," "Big data and the impact on individualized insurance," and "The use of big data technologies to improve insurance industry's current practices." The results are discussed in further detail to provide an effective research recommendation to ensure that the gap will be resolved in future research. Encouraging further research in this field will aid both the insurance sector and consumers, as the former will have the information and data necessary to best meet the expectations and needs of the market.

Conflicts of Interest Statement

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgements

We would like to thank the Editor and reviewers for taking the time and effort necessary to review this manuscript. We would like also to thank Editage for their extensive support in editing this article.

References

Arumugam, S., Bhargavi, R., 2019. A survey on driving behavior analysis in usage based insurance using big data. J Big Data 6.
Baker, H.K., Kumar, S., Pandey, N., 2020. A bibliometric analysis of managerial finance: a retrospective. Manag. Finance 46, 1495–1517.
Banu, A., 2022. Big data analytics-tools and techniques-application in the insurance sector. Big Data: A Game Changer for Insurance Industry. https://doi.org/10.1108/978-1-80262-605-620221013.
Barry, L. Insurance, 2020. Big data and changing conceptions of fairness. Arch. Eur. Sociol. 61, 159–184.
Barry, L., Charpentier, A., 2020. Personalization as a promise: can Big Data change the practice of insurance? Big Data Soc 7.
Bauder, R.A., Khoshgoftaar, T.M., 2016. A novel method for fraudulent medicare claims detection from expected payment deviations. In: Proceedings - 2016 IEEE 17th International Conference on Information Reuse and Integration. IRI, pp. 11–19. https://doi.org/10.1109/IRI.2016.11.
Boyack, K.W., Klavans, R., 2010. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? J. Am. Soc. Inf. Sci. Technol. 61, 2389–2404.
Branting, L.K., Reeder, F., Gold, J., Champney, T., 2016. Graph analytics for healthcare fraud risk estimation. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, 2016, pp. 845–851. https://doi.org/10.1109/ASONAM.2016.7752336.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Buzydlowski, J.W., 2015. Co-occurrence analysis as a framework for data mining. Journal of Technology Research 6. http://www.aabri.com/copyright.html.
Byeon, K.H., et al., 2021. Factors affecting the survival of early COVID-19 patients in South Korea: an observational study based on the Korean National Health Insurance big data. Int. J. Infect. Dis. 105, 588–594.
Chandola, V., Sukumar, S.R., Schryver, J., 2013. Knowledge discovery from massive healthcare claims data. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Part F1288, pp. 1312–1320.
Citron, D.K., Pasquale, F., 2014. The scored society: due process for automated predictions. Wash. Law Rev. 89, 1–33.
Corbett, P., Schroeck, M., Shockley, R., 2018. Analytics: the real-world use of big data in insurance. Executive Report. IBM Institute for Business Value.
Dutta, A., Ang, S.S., 2016. A 3-D stacked wire bondless silicon carbide power module. In: WiPDA 2016 - 4th IEEE Workshop on Wide Bandgap Power Devices and Applications, pp. 11–16. https://doi.org/10.1109/WiPDA.2016.7799902.
E Fonseca, B.P.F., Sampaio, R.B., Fonseca, M.V.A., Zicker, F., 2016. Co-authorship network analysis in health research: method and potential use. Health Res. Pol. Syst. 14.
Ellili, N.O.D., 2022. Is there any association between FinTech and sustainability? Evidence from bibliometric review and content analysis. J. Financ. Serv. Market. https://doi.org/10.1057/s41264-022-00200-w.
Fang, K., Jiang, Y., Song, M., 2016. Customer profitability forecasting using Big Data analytics: a case study of the insurance industry. Comput. Ind. Eng. 101, 554–564.
Frees, E.W., Huang, F., 2023. The discriminating (pricing) actuary. North Am. Actuar. J. 27, 2–24.
Fröhlich, H., et al., 2018. From hype to reality: data science enabling personalized medicine. BMC Med. 16.
Gai, K.A., 2014. Review of leveraging private cloud computing in financial service institutions: value propositions and current performances. Int. J. Comput. Appl. 95.
Gai, K., Qiu, M., Tao, L., Zhu, Y., 2016. Intrusion detection techniques for mobile cloud computing in heterogeneous 5G. Secur. Commun. Network. 9, 3049–3058.
Glänzel, W., Schubert, A., 2005. Domesticity and internationality in co-authorship, references and citations. Scientometrics 65, 323–342.
Hanafy, M., Ming, R., 2021. Machine learning approaches for auto insurance big data. Risks 9, 1–23.
Hassani, H., Unger, S., Beneki, C., 2020. Big data and actuarial science. Big Data and Cognitive Computing 4, 1–29.
Herland, M., Khoshgoftaar, T.M., Bauder, R.A., 2018. Big Data fraud detection using multiple medicare data sources. J Big Data 5.
Ho, C.W.L., Ali, J., Caals, K., 2020. Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance | Garantir un usage fiable de l'intelligence artificielle et de l'analyse des big data dans le domaine de l'assurance maladie | Garantizar el uso fiable de la i. Bull. World Health Organ. 98, 263–269.
Hsieh, C.-Y., et al., 2019. Taiwan's national health insurance research database: past and future. Clin. Epidemiol. 11, 349–358.
Hussain, K., Prieto, E., 2016. In: Cavanillas, J.M., Curry, E., Wahlster, W. (Eds.), Big Data in the Finance and Insurance Sectors. New Horizons for a Data-Driven Economy [Online]. Springer International Publishing, Cham, pp. 209–223. https://doi.org/10.1007/978-3-319-21569-3_12.
Jeong, S.-H., Lee, T.R., Kang, J.B., Choi, M.-T., 2020. Analysis of health insurance big data for early detection of disabilities: algorithm development and validation. JMIR Med Inform 8.
Keller, B., Eling, M., Schmeiser, H., Christen, M., Loi, M., 2018 Big Data and Insurance: Implications for Innovation, Competition and Privacy. www.genevaassociation.org.
Khandelwal, C., Kumar, S., Sureka, R., 2022. Mapping the intellectual structure of corporate risk reporting research: a bibliometric analysis. Int. J. Discl. Gov. 19, 129–143.
Khatib, S.F.A., Abdullah, D.F., Hendrawaty, E., Elamer, A.A., 2022. A bibliometric analysis of cash holdings literature: current status, development, and agenda for future research. Management Review Quarterly 72, 707–744.
Kim, J.-W., Lee, J.-H., Kim, T.-G., Kim, Y.-H., Chung, K.J., 2018a. Breast reconstruction statistics in Korea from the big data hub of the health insurance review and assessment service. Arch Plast Surg 45, 441–448.
Kim, T.J., et al., 2018b. Building linked big data for stroke in Korea: linkage of stroke registry and national health insurance claims data. J. Kor. Med. Sci. 33.
Kim, D.W., Byeon, K.H., Kim, J., Cho, K.D., Lee, N., 2020a. The correlation of comorbidities on the mortality in patients with COVID-19: an observational study based on the Korean national health insurance big data. J. Kor. Med. Sci. 35.
Kim, J., et al., 2020b. Compliance of antihypertensive medication and risk of coronavirus disease 2019: a cohort study using big data from the Korean National Health Insurance Service. J. Kor. Med. Sci. 35.
Landsman, Z., Sherris, M., 2001. Risk measures and insurance premium principles. Insur. Math. Econ. 29, 103–115.
Lehrer, C., Wieneke, A., vom Brocke, J., Jung, R., Seidel, S., 2018. How big data analytics enables service innovation: materiality, affordance, and the individualization of service. J. Manag. Inf. Syst. 35, 424–460.
Lim, K.W., Buntine, W., 2016. Bibliographic analysis on research publications using authors, categorical labels and the citation network. Mach. Learn. 103, 185–213.
Lin, W., Wu, Z., Lin, L., Wen, A., Li, J., 2017. An ensemble random forest algorithm for insurance big data analysis. IEEE Access 5, 16568–16575.
Manoj Kumar, M., Tejasree, S., Swarnalatha, S., 2016. Effective implementation of data segregation and extraction using big data in E - Health insurance as a service. In: ICACCS 2016 - 3rd International Conference on Advanced Computing and Communication Systems: Bringing to the Table. Futuristic Technologies from Around the Globe. https://doi.org/10.1109/ICACCS.2016.7586323.
Massie, A.B., Kuricka, L.M., Segev, D.L., 2014. Big data in organ transplantation: registries and administrative claims. Am. J. Transplant. 14, 1723–1730.
Muppidi, S., Reddy, K.T., 2020. Co-occurrence analysis of scientific documents in citation networks. Int. J. Knowl. Base. Intell. Eng. Syst. 24, 19–25.
Nobanee, H., 2021. A bibliometric review of big data in finance. Big Data 9, 73–78.
Nobanee, H., Ellili, N.O.D., 2023. What do we know about meme stocks? A bibliometric and systematic review, current streams, developments, and directions for future research. Int. Rev. Econ. Finance 85, 589–602.
Nobanee, H., Alhajjar, M., Abushairah, G., Al Harbi, S., 2021. Review reputational risk and sustainability: a bibliometric analysis of relevant literature. Risks 9.
Patil, H.K., Seshadri, R., 2014. Big data security and privacy issues in healthcare. In: Proceedings - 2014 IEEE International Congress on Big Data. https://doi.org/10.1109/BigData.Congress.2014.112. *BigData Congress 2014* 762–765.

Ponomariov, B., Boardman, C., 2016. What is co-authorship? Scientometrics 109, 1939–1963.
Price, W.N., Cohen, I.G., 2019. Privacy in the age of medical big data. Nat. Med. 25, 37–43.
Qiu, M., Ming, Z., Li, J., Gai, K., Zong, Z., 2015. Phase-change memory optimization for green cloud with genetic algorithm. IEEE Trans. Comput. 64, 3528–3540.
Raghupathi, W., Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. 2.
Rana, A., Bansal, R., Gupta, M., 2022. Big data: a disruptive innovation in the insurance sector. Big Data Analytics in the Insurance Market. https://doi.org/10.1108/978-1-80262-637-720221009.
Sagiroglu, S., Sinanc, D., 2013. Big data: a review. In: Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS, pp. 42–47. https://doi.org/10.1109/CTS.2013.6567202.
Scopus. How do Author/Indexed keywords work? https://service.elsevier.com/app/answers/detail/a_id/21730/supporthub/scopus/. (Accessed 20 October 2021).
Senousy, Y., Shehab, A., Riad, A.M., Elkhamisy, N., 2020. A smart social insurance big data analytics framework based on machine learning algorithms. J. Theor. Appl. Inf. Technol. 98, 232–244.
Small, H., 1973. Co-citation in the scientific literature: a new measure of the relationship between two documents. J. Am. Soc. Inf. Sci. 24, 265–269.
Telecom SNC & IT, 2018. Big Data in Insurance.
Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5, 99–114.
van Eck, N.J., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84, 523–538.
Vanhala, M., et al., 2020. The usage of large data sets in online consumer behaviour: a bibliometric and computational text-mining–driven analysis of previous research. J. Bus. Res. 106, 46–59.
Xu, J., et al., 2021. Federated learning for healthcare informatics. J Healthc Inform Res 5.