ISSN: 2710-2599

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Data Science and Big Data Analytics

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper | Open Access

The Bibliometric Overview of Research on Healthcare Information Systems Using Big Data Analytics

Christine Nya-Ling Tan[1*] and Muhammad Ashraf Fauzi[2]

[1]School of Management, Massey University, Albany, Auckland, New Zealand. E-mail: n.l.tan@massey.ac.nz
[2]Faculty of Industrial Management, Universiti Malaysia Pahang, Kuantan, Malaysia. E-mail: ashrafauzi@ump.edu.my

Article Info

Volume 3, Issue 1, May 2023
Received : 18 February 2023
Accepted : 22 April 2023
Published : 05 May 2023
doi: 10.51483/IJDSBDA.3.1.2023.45-57

Abstract

The significance of Big Data Analytics (BDA) has long been acknowledged by academics exploring Healthcare Information Systems (HIS). This paper aims to analyze and categorize the literature linking BDA to HIS research. From the Web of Science (WoS) database, the bibliometric data of BDA on HIS scholarly papers published between 1999 and 2022 were obtained, processed and analysed using VOS viewer. Based on the analysis of 266 publications, the study provides a summary of the impact of BDA on HIS research from the perspectives of (1) the number of publications and citations; (2) the top 10 papers with the highest co-citation indicating their essential impact on past and present; and (3) the top 15 keywords in the co-occurrence of keywords analysis with the highest co-occurrence keywords being big data (115 occurrences), artificial intelligence (111 occurrences), and the internet (20 occurrences). This paper also depicts the co-word analysis network topology, with four clusters of literature relating to BDA and HIS identified. The publications on BDA in HIS research were examined using a bibliometric analytic method, as this study is the first attempt at doing so—to the best of the author's knowledge.

Keywords: *Big Data Analytics (BDA), Healthcare Information Systems (HIS), Bibliometric, VOSviewer, Overview*

1. Introduction

The healthcare system has rapidly changed over the past 20 years to become more valuable and practical due to the new developments in Big Data Analytics (BDA). With the expansion of Health Information Systems (HIS), numerous researchers and healthcare providers have investigated the most recent developments in big data management and the use of BDA in the modern healthcare sector (Pramanik *et al.*, 2020). In the current COVID-19 ecosphere, the increasing amounts of structured and unstructured complex data, such as patient diagnoses and treatments, depend heavily on BDA's evidence-based decision-making (Sheng *et al.*, 2021). In the modern information system environment, medical practitioners rely on 'big data' in healthcare to compile,

* *Corresponding author: Christine Nya-Ling Tan, School of Management, Massey University, Albany, Auckland, New Zealand. E-mail: n.l.tan@massey.ac.nz*

decipher, and visually portray information that will help them comprehend the demands of their patient's state of health (Renault, 2021).

In HIS, 'big data' is described as high volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals/subjects to large cohorts concerning their state of wellness at a single point in time (Pastorino *et al.*, 2019). As a result, medical claims, patient summaries, local population health data, electronic health records, and other types of healthcare data can all be gathered using BDA in the healthcare industry, despite its size and complexity (Casey *et al.*, 2016). Pastorino *et al.* (2019) have discovered that BDA has impacted HIS in four distinctive ways: (1) by increasing the effectiveness and caliber of treatments, as well as earlier diagnosis; (2) by increasing opportunities for illness prevention by identifying disease risk factors; (3) by making better medical judgements based on information that is given directly to patients that would improve pharmacovigilance and patient safety; and (4) by improving the overall prognosis of results.

Since researchers must examine or analyze exhaustive sources and databases of papers released in the field to find the gaps in BDA research trends in HIS, this study narrows the gap by presenting an overview of the research related to HIS that deals with BDA. It is vital to better understand BDA in HIS studies and inform BDA academics about this field's implications and future potential in HIS's research planning. Moreover, this research's implications are to support academics' knowledge creation by discovering and examining publication trends in BDA and HIS research. Therefore, this study uses a popular strategy, i.e., bibliometric analysis, that offers a preliminary understanding of the enormous volume of published papers linking BDA-related research in the field of HIS.

The bibliometric analysis would be beneficial in this study to profile the current and general research on BDA on HIS as it allows for the objectivity of a field of study compared to a literature review (Farhan and Iqbal, 2021). This paper also provides more detailed descriptive data and co-citation analysis, deemed essential to locate, analyze, and track the connections between the ideas in a particular academic area and direct future growth. By using a bibliometric technique to assess works of literature, this study aims to analyze the current state of BDA in HIS and provide guidance to aspiring researchers interested in the subject for all research works published between 1999 and 2022. Two primary goals form the foundation of this paper. Firstly, the study determines past critical research on BDA in HIS through co-citation analysis. Secondly, this research assesses the directions and trends of BDA in HIS through co-word analysis that can aid students, researchers, and policymakers in determining the ideal ecosystem for BDA research in HIS.

This paper is structured as follows: the bibliometric analysis methodology will be presented in Section 2, while in Section 3, the findings and discussions will be provided, with the implications in Section 4, the limitations and future research in Section 5 and the conclusion in Section 6.

2. Methodology

2.1. Bibliometric Approach

The bibliometric approach is a quantitative method for examining how disciplines change through time based on their social, conceptual and intellectual structures (Zupic and Cater, 2015). It analyzes and categorizes bibliographic material using representative summaries of the available literature (Suban *et al.*, 2021). Interested researchers may use bibliometric analysis to guide their future work (Mavric *et al.*, 2021) and utilise it to identify changing journal trends, intellectual structure, and research aspects and to delve deeper into collaboration patterns in a given topic in the existing literature. Evidently, bibliometric analysis uses mathematical and statistical techniques to analyse scientific publications using articles, books, and other forms of communication (Repanovici, 2011). The bibliometric analysis uses mathematical and statistical techniques to analyse scientific publications using articles, books, and other forms of communication (Repanovici, 2011). By offering a clear, organised, and repeatable review procedure, bibliometric analysis can significantly raise the caliber of literature reviews. Without relying on the researcher's intuition, it offers ways to map the corpus of publications and find significant studies (Ellegaard and Wallin, 2015). Therefore, bibliometric analysis is advantageous to researchers by identifying highly cited articles, authors, research trends, and upcoming themes (Mao *et al.*, 2015) as it can help researchers choose the ideal subject area or journal to publish in Ebrahim *et al.* (2014). Moreover, it assists scholars in getting a general understanding of the papers and studies that impact a particular area of interest (Gomez-Jauregui *et al.*, 2014).

This study provides a science mapping of the subject based on co-citation and co-occurrence of keywords analysis using VOSviewer software (Perianes-Rodriguez *et al.*, 2016). It includes network visualisation and is frequently used in conjunction with bibliometric analysis (van Eck and Waltman, 2017).

2.2. Research Design and Data Collection Procedure

The current study employed the following search string (see Table 1) to extract relevant documents in the Web of Science (WoS) database. With its Web of Science Core Collection, WoS, which Clarivate Analytics owns, covers more than 74.8 million academic data sets, 1.5 billion cited references and 254 different subject areas. It is chosen in this study as it provides extensive coverage going back to 1990.

Table 1: Search String in WoS Database

No	Keywords	Justification
1	"data analytic*" OR "big data*" OR "data science*" OR "massive data"	To identify literature-related data analytics and related terminologies
2	"healthcare IS*" OR "healthcare information system*" OR "medic* IS*" OR "medic* information system*" OR "health care IS*" OR "health care information system*"	To identify literature related to healthcare information systems and related terminologies.

3. Findings and Discussion

The search was performed on November 24, 2022. The initial search returned 542 documents. After filtering for only journal publications, the final publications were 266. Total citations were 5,726 and 5,680 (without self-citations). The H-index was 33, with an average of 21.53 per item. From Figure 1, the first identified publication was in 1999, but since then, no publication has been produced until only 2008. The publications increased significantly since 2013 (4 publications) and reached more than 35 publications since 2018. Due to the topic's relevancy and the sophisticated diseases and illness, studies in health BDA will be increased. Hence, the number of publications is expected to increase in the coming years.

3.1. Co-citation Analysis

Of the 12,084 cited references, 56 met the four thresholds in the co-citation analysis. The threshold was determined based on several trials to ensure the network map produces a robust and appropriate number of

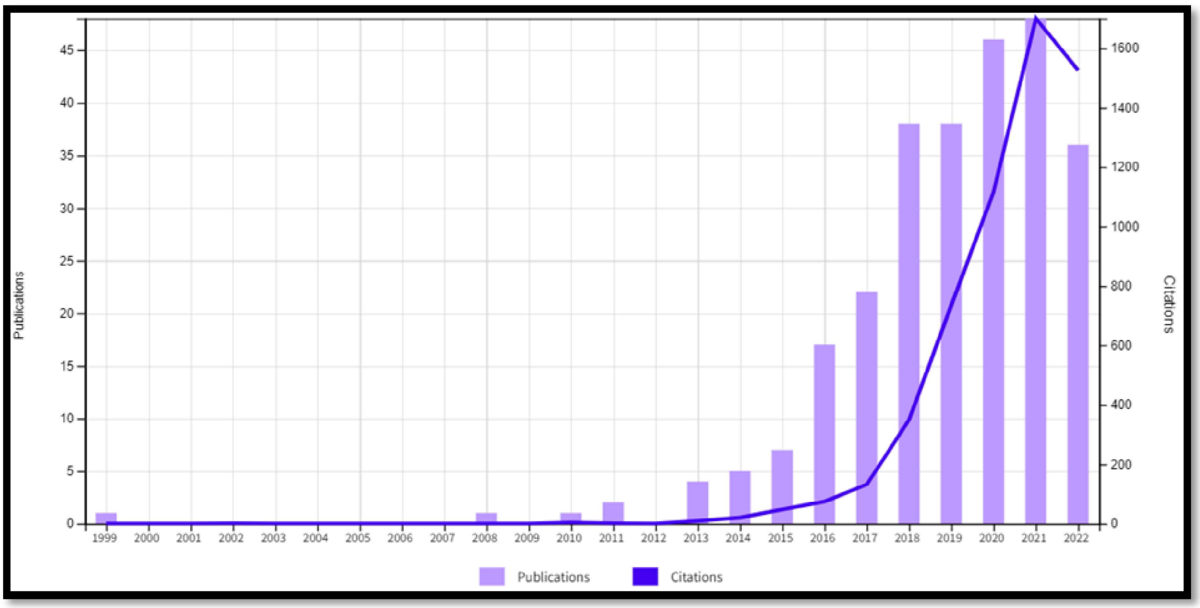

Figure 1: Number of Publications and Citations on BDA in HIS

clusters. The highest number of co-cited documents in the analysis are (Raghupathi and Raghupathi, 2014) (24 citations), (Bates *et al.*, 2014) (11 citations) and (Andreu-Perez *et al.*, 2015) (8 citations). An excessive threshold might result in over-filtering, while too low would result in under-filtering, leading to excessive clusters (Geng *et al.*, 2020). The top 10 documents with the highest co-citation are presented in Table 2. These documents are ranked based on the number of citations received, indicating their influential impact on past and present research streams of BDA-related research in HIS.

The network map of the co-citation analysis is presented in Figure 2. The figure suggests that there are six identified clusters. Cluster 1 (red) and Cluster 2 (green) are estranged from other clusters on the map. In contrast, other clusters are closely interconnected (blue, yellow, purple and light blue), suggesting the clusters produced are highly correlated. The clusters are discussed and labeled according to the author's inductive interpretation.

Table 2: Top 10 Documents with the Highest Co-citation and Total Link Strength

Documents	Citation	Total Link Strength
Raghupathi *et al.* (2014), *Health Inf. Sci. Syst.*, Vol. 2	24	72
Bates *et al.* (2014), *Health Affair*, Vol. 33, p.1123	11	42
Andreu-Perez *et al.* (2015), *IEEE J. Biomed Health*, Vol. 19, p.1193	8	21
Islam *et al.* (2015), *IEEE Access*, Vol. 3, p.678	8	20
Collins and Varmus (2015), *New Engl. J. Med.*, Vol. 372, p.793	7	22
Murdoch and Detsky (2013), *JAMA-J AM. Med. Assoc.*, Vol. 309, p.1351	7	21
Obermeyer and Emanuel (2016), *New Engl. J. Med.*, Vol. 375, p.1216	7	13
Breiman (2001), *Mach Learn*, Vol. 0045	7	12
Archenaa and Anita (2015), *Procedia. Comput. Sci.*, Vol. 50, p.408	6	18
Atzori *et al.* (2010), *Comput. Netw.*, Vol. 54, p.2787	6	17

Source: Author interpretation based on VOSviewer analysis

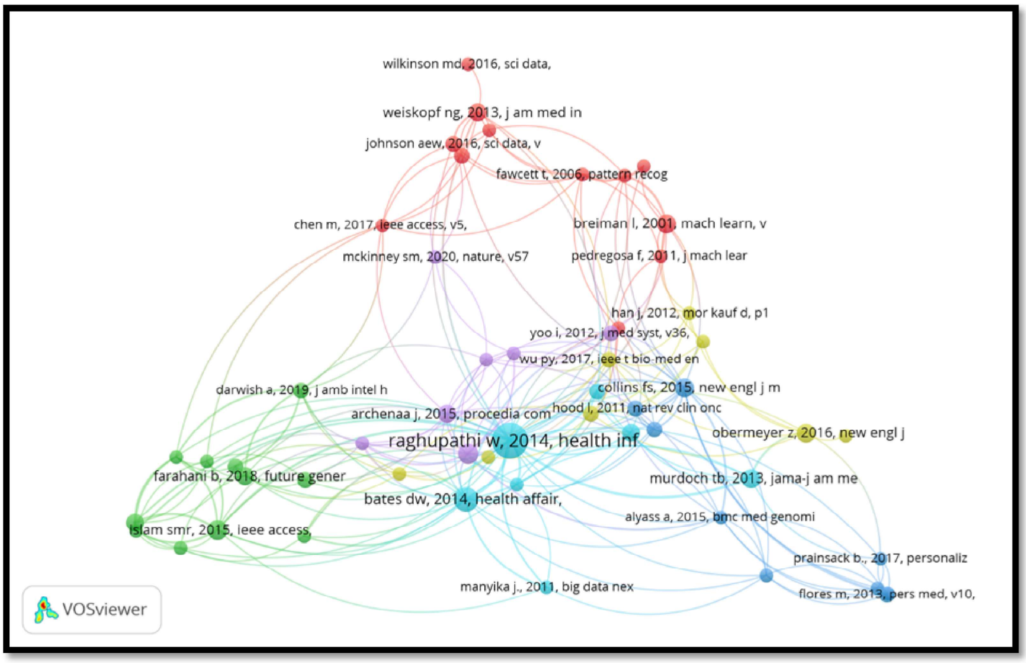

Figure 2: Co-citation Analysis on the BDA in HIS

- **Cluster 1 (Red):** With 12 articles, cluster 1 is labeled "Fundamental of big data analytics in healthcare". The fundamental of BDA in healthcare is central to the discussion of the topic, and it primarily discusses the fundamental function and role of BDA in healthcare. Weiskopf and Weng (2013) identified dimensions for data quality assessment for electronic health records: correctness, completeness, plausibility, concordance and currency. Chen *et al.* (2017) evaluate the effectiveness of disease prediction using machine learning over big data in patient care and community services. Johnson *et al.* (2016) present an extensive single-center database comprising patient information admitted to tertiary care hospitals. The database supports various applications, including research, quality management and higher education coursework.

- **Cluster 2 (Green):** Cluster 2, with 11 documents, is centralised towards the theme of "IoT in healthcare". The Internet of Things (IoT) comprises a network of physical devices and software applications embedded within electronics, network connectivity and sensor, allowing faster data collection and exchange. Dimitrov (2016) reviewed medical IoT and big data in healthcare. Such a tool is crucial in healthcare transformation, allowing the new business model to emerge and stimulating changes in work processes, cost reduction, enhanced customer satisfaction and productivity improvements. Farahani *et al.* (2018) present the applicability of IoT in medicine and healthcare through a holistic architecture of the IoT eHealth ecosystem. It was suggested that healthcare requires the transition from clinic-centric to patient-centric treatment, requiring multi-layer architecture consisting of the device, fog computing and cloud to empower complex data handling. Rahmani *et al.* (2018) proposed a concept of Fog computing in Healthcare IoT systems through a Geo-distributed intermediary intelligence layer between Cloud and sensor nodes. This fog-assisted system can cope with many challenges in challenging healthcare systems regarding energy efficiency, reliability, mobility and scalability issues.

- **Cluster 3 (Blue):** Cluster 3, with nine articles, is labelled as "Big data in precision medicine". Precision medicine is a concept that takes individual variability into account in taking prevention and treatment strategies (Collins and Varmus, 2015). Precision medicine will provide a basis for the patient to improve their health through the impact of lifestyle decisions (Flores *et al.*, 2013). Precision, personalised and individualised medicine are used interchangeably (Jameson and Longo, 2015). It is a treatment targeted towards individual needs underlie by the genetic, phenotypic, psychosocial and biomarkers characterizing and distinguishing a patient from others with similar clinical presentations. Alyass *et al.* (2015) present the challenges of personalized medicine: data storage and processing; data integration and interpretation; cost-effective generating high-throughput data; hybrid education, and multidisciplinary data.

- **Cluster 4 (Yellow):** Cluster 4, with eight articles, is labeled as "Challenges of BDA in healthcare". Belle *et al.* (2015) stated that the fundamental problems within the big data paradigm hinder the development of BDA in healthcare. As patients' conditions and medical technologies become more complex, BDA and machine learning in healthcare will grow (Obermeyer and Emanuel, 2016). Despite the challenge posed by machine learning in medicine, patients emerge as the biggest winners in clinical transformation. Sivarajah *et al.* (2017) present a systematic literature review on big data challenges and analytical methods employed by organisations to facilitate making business decisions.

- **Cluster 5 (Purple):** Cluster 5, with seven articles, is labelled as "Fundamental of BDA in healthcare". Andreu-Perez *et al.* (2015) outline BDA characteristics and how they can benefit medical and health information, sensor informatics, translational bioinformatics and imaging informatics by integrating personalized information from many data sources. BDA reveals hidden patterns and links unknown correlations and functions based on large-scale data sets (He *et al.*, 2017). Despite the challenges in applying BDA, it can provide an effective and efficient approach to identifying clinically actionable genetic variants to perform therapy and diagnosis on patients. McKinney *et al.* (2020) present an Artificial Intelligence (AI) system that can detect breast cancer by reducing false positives and negatives.

- **Cluster 6 (Light Blue):** Cluster 6, with seven articles, is labelled as "Potential of BDA in healthcare". This cluster also presents the fundamental BDA in healthcare by presenting its potential in advanced technologies and applications. Bates *et al.* (2014) pointed out six opportunities for reducing cost using BDA in healthcare: triage, readmissions, decomposition, high-cost patient, treatment optimization and adverse events for diseases affecting multiple organ systems. BDA can transform healthcare providers and applies sophisticated technologies to make informed decisions based on clinical and repository data (Raghupathi and Raghupathi, 2014). Despite that, healthcare practitioners must address security and privacy challenges and establish standards and governance to improve the technologies and tools that come with them. Krumholz (2017) explores how BDA can be utilised to advance the performance, prediction, comparative research and discovery to solve complex public health concerning the larger population. Incorporating BDA within research and practice requires relevant data sources and new thing paradigms, tools and training to fuel knowledge generation in the healthcare learning system.

Table 3 summarizes the co-citation analysis cluster on BDA in the healthcare information system. It includes the cluster number and colour, labels, number of publications, and representative publications.

Table 3: Co-citation Clusters on BDA in HIS

Cluster	Cluster Label	Number of Publications	Representative Publications
1 (red)	Fundamental of big data analytics in healthcare	12	Weiskopf and Weng (2013), Chen *et al.* (2017) and Johnson *et al.* (2016)
2 (Green)	IoT in healthcare	11	Farahani *et al.* (2018), Rahmani *et al.* (2018) and Dimitrov (2016).
3 (Blue)	Big data in precision medicine	9	Alyass *et al.* (2015), Collins and Varmus (2015) and Jameson and Longo (2015)
4 (yellow)	Challenges of BDA in healthcare	8	Belle *et al.* (2015), Obermeyer and Emanuel (2016) and Sivarajah *et al.* (2017)
5 (purple)	Fundamental of BDA in healthcare	7	He *et al.* (2017), McKinney *et al.* (2020) and Andreu-Perez *et al.* (2015)
6 (light blue)	Potential of BDA in healthcare	7	Raghupathi and Raghupathi (2014), Krumholz (2014) and Bates *et al.* (2014)

3.2. Co-word Analysis

Applying the same database, the co-word analysis presents 56 out of 1,513 keywords that met the six thresholds, resulting in four clusters. The highest co-occurrence keywords are big data (115 occurrences), artificial

Table 4: Top 15 Keywords in the Co-occurrence of Keywords Analysis

Rank	Keyword	Occurrences	Total Link Strength
1	Big data	115	333
2	Artificial intelligence	21	56
3	Internet	20	110
4	Precision medicine	20	48
5	Machine learning	19	71
6	Healthcare	18	78
7	Care	18	52
8	Health	17	60
9	Things	16	91
10	Data science	15	43
11	IoT	14	60
12	Framework	14	58
13	Big data analytics	13	41
14	Cloud computing	13	41
15	Data mining	13	27

intelligence (21 occurrences) and the internet (20 occurrences). The keywords are centralised within big data as the mediator keywords in the network map. Table 4 presents the top 15 keywords in the co-occurrence of keywords analysis.

The network structure of the co-word analysis (see Figure 3) shows compact and closely connected clusters. Three clusters are identified (red, green and blue), while the fourth cluster (yellow) is slightly interconnected with all the other clusters within the network. Similarly, all the clusters are discussed and labeled according to the author's inductive interpretation.

- **Cluster 1 (Red):** Cluster 1, with 20 keywords, Cluster 1 is labeled as "Personalised medicine based on BDA". The core of the medical field is individualised patient treatment. Diagnosis, monitoring and disorders treatment requires advances in biomarker discovery (Ho *et al.*, 2020). Effective application of BDA in personalised medicine requires scientific and technical developments comprising engineering, infrastructure and financial management (Cirillo and Valencia, 2019). These non-medical support systems serve as the fundamental personalised medicine management to sustain in the long run. An essential component of personalized medicine is mobile health (mHealth). Myriad sensory devices are associated with mHealth, such as Fitbit, that monitor real-time data for evaluating users' health and personal data (Saxena and Saxena, 2020). mHealth can be applied to clinical and non-clinical contexts for patient health and well-being.
- **Cluster 2 (Green):** Cluster 2 discusses the "Advanced BDA technology in healthcare" theme. This cluster addressed the application of technology in support of BDA. Advanced technology includes disease surveillance (Amirian *et al.*, 2017) and alert system (Manogaran *et al.*, 2018; Kumar *et al.*, 2022). Through predictive analytics via machine learning, BDA enables users to make smarter decisions reliably and swiftly. The technology includes meeting healthcare needs through e-health and m-health (Khanra *et al.*, 2020). Fog computing implementation in e-healthcare to perform pre-processing of raw healthcare data and manage healthcare data analytics (Wang and Alexander, 2020). Fog computing is privacy protected and allows efficient information sharing.

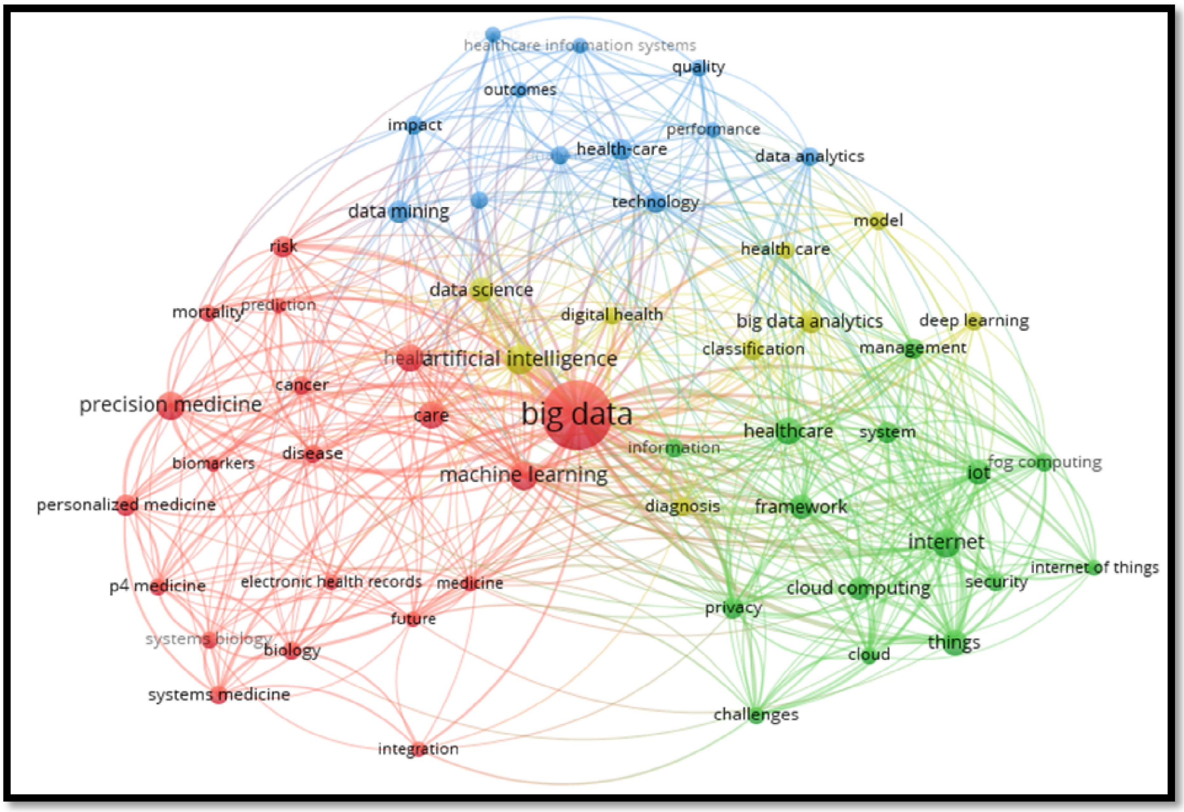

Figure 3: Co-word Analysis on BDA in HIS

- **Cluster 3 (Blue):** With 12 keywords, Cluster 3 is labelled "Impact of BDA on healthcare". Wang *et al.* (2019) studied the impact of BDA in healthcare through a fuzzy-set qualitative comparative analysis of multi-source data. It was found that BDA alone is not sufficient but requires synergy with human analytical skills, organisational resources and capabilities as support for the healthcare system. BDA has proven to improve healthcare organisational performance (Kamble *et al.*, 2018). Gravili *et al.* (2021) present a data-driven model by presenting intellectual capital performance in healthcare BDA. The human, relational and structural capital positively impact care management efficiency performance indicators in terms of cost. Wu *et al.* (2017) investigated the impact of BDA on health IT market competition and its influence on providers' BDA adoption decisions. Within the scope of marketing, pricing strategies are a crucial predictor of BDA's efficiency and privacy risk.
- **Cluster 4 (Yellow):** With ten keywords, Cluster 4 is labeled as "Diagnosis and classification through BDA". BDA in healthcare facilitate the development of analytical techniques for personalised health services and supports decision-making through automated algorithms (Galetsi *et al.*, 2019). BDA and other smart healthcare systems, i.e., IoT, cloud computing and wireless sensor networks, play crucial roles in diagnosing and detecting diseases, disorders, viruses and illnesses (Minopoulos *et al.*, 2022). The diagnosis is facilitated with the integration of visualisation techniques such as augmented reality, virtual reality and mixed reality to provide accurate detection towards patient treatment. Shafqat *et al.* (2021) apply a hybrid deep learning technique to diagnose the real-time dataset of validated patients with dengue fever. The method was effective, having more than 70% accuracy and suggested for future algorithms in healthcare big data diagnostics.

A summary of the co-word analysis is presented in Table 5, comprising cluster number and colour, cluster labels, number of keywords, and representative keywords.

Table 5: Summary of Co-word Analysis on BDA in HIS

Cluster No and Color	Cluster Label	Number of Keywords	Representative Keywords
1 (Red)	Personalized medicine based on BDA	20	Big data, machine learning, personalized medicine, precision medicine, predictions,
2 (Green)	Advanced BDA technology in healthcare	15	Cloud computing, internet of things, information, healthcare, fog computing
3 (Blue)	Impact of BDA on healthcare	12	Healthcare information system, impact, performance, outcomes
4 (Yellow)	Diagnosis and classification through BDA	9	Data science, diagnosis, classification, artificial intelligence, digital health, deep learning

4. Implications

4.1. Theoretical Implications

This study sheds light on the scope, prominence, and significance of BDA-related research in HIS that has been included in the WoS database. The present work adds to the body of literature on BDA-related research in the area of HIS in several ways. First, to the author's knowledge, this is the first study that uses the bibliometric analysis method to provide an overview of the current research stage around BDA in the HIS environment. Second, it suggests and uses bibliometric and network analysis tools to find and compare the most influential works (based on citations, co-citations, and co-word analysis), which goes beyond a simple systematic review of the literature in the area of research. Third, the findings of this study indicate that BDA-related studies in the HIS research topic (based on the number of citations and publications) have been rapidly expanding over the last twelve years (between the years 2010 to 2022); which indicates that it becoming a well-liked research subject with plenty of research opportunities for both new and existing scholars. Fourth, this research reveals the top 10 papers with the highest co-citation indicating their essential impact on past and present BDA research conducted in HIS. Fifth, big data (115 occurrences), artificial intelligence (111 occurrences), and the internet (20 occurrences) are the three most popular terms out of the 15 keywords in the co-occurrence of keywords analysis of BDA articles in the HIS sector. Next, based on the co-word analysis conducted, this research finds and suggests four clusters (i.e., "Personalised medicine based on BDA", "Advanced BDA technology in healthcare", "Impact of BDA on healthcare", and "Diagnosis and classification through BDA") that concentrate on specific areas of BDA in HIS. Lastly, this study provides an overview of the BDA in HIS research area, which has not been thoroughly investigated in the literature and offers a sizable area for future research.

4.2. Managerial Implications

The practitioners utilising BDA in HIS can take advantage of several prospects provided by this study. First, it gives excellent knowledge of how healthcare organisations may use BDA to change HIS and create business value by delivering crucial BDA capabilities such as meaningful clinical reports. These healthcare organisations can provide their employees, who will play essential support roles in the future information-rich HIS workplace, with analytical training in subjects like fundamental statistics, data mining, and business intelligence. Moreover, it gives managers access to many schools of thought that help them take advantage of BDA in HIS daily work. Second, with efficient skills and information retrieval methodologies, this research could guide healthcare professionals to locate pertinent health information from BDA. Traceability, which allows for tracking output data from all the HIS components across the organization's service units, could aid in maintaining real-time changes. It responds to the call for further study to determine the knowledge, skills, and abilities required for health information professionals' occupations. Third, the results of this study will help healthcare workers and information professionals fulfil the information-intensive needs of the healthcare industry because of the rapid growth of information systems and the evolving decision-making environment. Lastly, the academics who have spent years focusing on BDA can carry out several pertinent research projects and amass real-world experience in HIS. These scholars, who have solid theoretical and technological backgrounds, will be able to provide insightful advice on how to overcome big data challenges.

5. Limitations and Suggestions for Future Works

In this research, it is vital to highlight its limitations. First, even though this study concentrates on a single database, i.e., WoS Core Collection, which is sizable enough to provide a wide diversity of publications necessary for the analyses. However, future research work will need to employ other databases (e.g., Scopus, ProQuest, ERIC, PubMed, and PsycInfo) to examine a more comprehensive outlook of the literature of BDA in HIS. Second, the WoS typically covers English-language journals that were found to be overrepresented at the expense of other languages. As a result, it is recommended that future research works should be carried out in various languages from non-English speaking countries. Third, grey literature, such as conference proceedings, theses, and books, were other types of publications that were not examined. As a result, future work can be expanded to incorporate additional relevant research to enhance the body of literature by integrating books and conferences. Fourth, thisstudy's coverage period, centred on the WoS database, is from 1999-2022. Future researchers may consider expanding the timespan of coverage. Lastly, the Emerging Sources Citation Index (ESCI) of WoS was not included since its impact factors for the journals it contains are not available for publication; and it is not a part of the core collection of WoS. A remedy would be to include it as part of future research work.

6. Conclusion

Using data from the WoS database, this study presents an overview of the development and state of BDA-related research in HIS. This analysis offers a thorough summary of the BDA-related research done in the HIS discipline. This research is confident that it has found inclusive BDA-related HIS studies thanks to numerous searching and screening rounds, ten finalised search phrases, and a 23-year timespan from 1999 to 2022. BDA-related research will be the forefront topic in health technologies contributed by the emergence of complicated diseases and illnesses as modern lifestyle develops.

Acknowledgment

The authors wish to thank the editor and reviewers for their insights.

References

Alyass, A., Turcotte, M. and Meyre, D. (2015). From Big Data Analysis to Personalised Medicine for All: Challenges and Opportunities. *BMC Medical Genomics*, *8*(1). https://doi.org/10.1186/s12920-015-0108-y

Andreu-Perez, J., Poon, C.C.Y., Merrifield, R.D., Wong, S.T.C. and Yang, G.-Z. (2015). Big Data for Health. *IEEE Journal of Biomedical and Health Informatics*, *19*(4), 1193-1208. https://doi.org/10.1109/JBHI.2015.2450362

Amirian, P., van Loggerenberg, F., Lang, T., Thomas, A., Peeling, R., Basiri, A. and Goodman, S.N. (2017). Using Big Data Analytics to Extract Disease Surveillance Information from Point of Care Diagnostic Machines. *Pervasive and Mobile Computing*, 42, 470-486. https://doi.org/10.1016/j.pmcj.2017.06.013

Archenaa, J. and Anita, E.M. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50, 408-413. https://doi.org/10.1016/j.comnet.2010.05.010

Atzori, L., Iera, A. and Morabito, G. (2010). The Internet of Things: A Survey. *Computer Networks*, 54(15), 2787-2805.

Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A. and Escobar, G. (2014). Big Data in Health Care: Using Analytics to Identify and Manage High-risk And High-cost Patients. *Health Affairs*, 33(7), 1123-1131. https://doi.org/10.1377/hlthaff.2014.0041

Belle, A., Thiagarajan, R., Soroushmehr, S.M., Navidi, F., Beard, D.A. and Najarian, K. (2015). Big Data Analytics in Healthcare. *BioMed Research International*. https://doi.org/10.1155/2015/370194

Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Casey, J.A., Schwartz, B.S., Stewart, W.F. and Adler, N.E. (2016). Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annual Review of Public Health*, 37, 61-81. https://doi.org/10.1146/annurev-publhealth-032315-021353

Chen, M., Hao, Y., Hwang, K., Wang, L. and Wang, L. (2017). Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access*, 5, 8869-8879. https://doi.org/10.1109/ACCESS.2017.2694446

Cirillo, D. and Valencia, A. (2019). Big Data Analytics for Personalised Medicine. *Current Opinion in Biotechnology*, 58(2019), 161-167.

Collins, F. S. and Varmus, H. (2015). A New Initiative on Precision Medicine. *New England Journal of Medicine*, 372(9), 793-795.

Dimitrov, D.V. (2016). Medical Internet of Things and Big Data in Healthcare. *Healthcare Informatics Research*, 22(3), 156-163. https://doi.org/10.4258/hir.2016.22.3.156

Ebrahim, N.A., Salehi, H., Embi, M.A., Danaee, M., Mohammadjafari, M., Zavvari, A., Shakiba, M. and Shahbazi-Moghadam, M. (2014). Equality of Google Scholar with Web of Science Citations: Case of Malaysian Engineering Highly Cited Papers. *Modern Applied Science*, 8(5), 63-69. https://doi.org/10.5539/mas.v8n5p63

Ellegaard, O. and Wallin, J. A. (2015). The Bibliometric Analysis of Scholarly Production: How Great is the Impact? *Scientometrics*, 105(3), 1809-1831. https://doi.org/10.1007/s11192-015-1645-z

Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N. and Mankodiya, K. (2018). Towards Fog-driven Iot Ehealth: Promises and Challenges of IoT in Medicine and Healthcare. *Future Generation Computer Systems*, 78(Part 2), 659-676. https://doi.org/10.1016/j.future.2017.04.036

Farhan, M. and Iqbal, M. K. (2021). Twenty Years of Sustainable Supply Chain: Past Trends and Future Research Suggestions. *International Journal of Business and Psychology*, 3(1), 1-16.

Flores, M., Glusman, G., Brogaard, K., Price, N. D. and Hood, L. (2013). P4 Medicine: How Systems Medicine will Transform the Healthcare Sector and Society. *Personalised Medicine*, 10(6), 565-576. https://doi.org/10.2217/PME.13.57

Galetsi, P., Katsaliaki, K. and Kumar, S. (2019). Values, Challenges and Future Directions of Big Data Analytics in Healthcare: A Systematic Review. *Social Science and Medicine*, 241, 112533. https://doi.org/10.1007/s11356-020-09283-1

Geng, D., Feng, Y. and Zhu, Q. (2020). Sustainable Design for Users: A Literature Review and Bibliometric Analysis. *Environmental Science and Pollution Research*, 27(24), 29824-29836.

Gomez-Jauregui, V., Gomez-Jauregui, C., Manchado, C. and Otero, C. (2014). Information Management and Improvement of Citation Indices. *International Journal of Information Management*, *34*(2), 257-271. https://doi.org/10.1016/j.ijinfomgt.2014.01.002

Gravili, G., Manta, F., Cristofaro, C.L., Reina, R. and Toma, P. (2021). Value that Matters: Intellectual Capital and Big Data to Assess Performance in Healthcare. An Empirical Analysis on the European Context. *Journal of Intellectual Capital*, 22(2), 260-289. https://doi.org/10.1108/JIC-02-2020-0067

He, KY., Ge, D. and He, M.M. (2017). Big Data Analytics for Genomic Medicine. *International Journal of Molecular Sciences*, 18(2), 412. https://doi.org/10.3390/ijms18020412

Ho, D., Quake, S.R., McCabe, E.R.B., Chng, W.J., Chow, E.K., Ding, X., Gelb, B.D., Ginsburg, G.S., Hassenstab, J., Ho, C.M., Mobley, W.C., Nolan, G.P., Rosen, S.T., Tan, P., Yen, Y. and Zarrinpar, A. (2020). Enabling Technologies for Personalised and Precision Medicine. *Trends in Biotechnology*, 38(5), 497-518. https://doi.org/10.1016/j.tibtech.2019.12.021

Islam, S.R., Kwak, D., Kabir, M.H., Hossain, M. and Kwak, K.S. (2015). The Internet of Things gor Health Care: A Comprehensive Survey. *IEEE Access*, 3, 678-708.

Jameson, J.L. and Longo, D.L. (2015). Precision Medicine—Personalised, Problematic, and Promising. *The New England Journal of Medicine*, 372(23), 2229-2234. https://doi.org/10.1056/NEJMsb1503104

Johnson, A., Pollard, T. and Shen, L. (2016). MIMIC-III, A Freely Accessible Critical Care Database. *Sci Data,* 3, 160035. https://doi.org/10.1038/sdata.2016.35

Kamble, S. S., Gunasekaran, A., Goswami, M. and Manda, J. (2018). A Systematic Perspective on the Applications of Big Data Analytics in Healthcare Management. International *Journal of Healthcare Management*, 12(3), 226-240. https://doi.org/10.1080/17517575.2020.1812005

Khanra, S., Dhir, A., Islam, A.N. and Mäntymäki, M. (2020). Big Data Analytics in Healthcare: A Systematic Literature Review. *Enterprise Information Systems*, 14(7), 878-912.

Krumholz, H.M. (2014). Big Data and New Knowledge in Medicine: The Thinking, Training, and Tools Needed for a Learning Health System. *Health Affairs*, 33(7), 1163-1170.

Kumar, A., Alghamdi, S.A., Mehbodniya, A., Webber, J.L. and Shavkatovich, S. N. (2022). Smart Power Consumption Management and Alert System Using IoT on Big Data. *Sustainable Energy Technologies and Assessments*, 53(Part C). https://doi.org/10.1016/j.seta.2022.102555

Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P.M., Sundarasekar, R. and Thota, C. (2018). A New Architecture of Internet of Things and Big Data Ecosystem for Secured Smart Healthcare Monitoring and Alerting System. *Future Generation Computer Systems*, 82, 375-387. https://doi.org/10.1016/j.future.2017.10.045

Mao, G., Zou, H., Chen, G., Du, H. and Zuo, J. (2015). Past, Current and Future of Biomass Energy Research: A Bibliometric Analysis. *Renewable and Sustainable Energy Reviews*, 52(2015), 1823-1833.

Mavric, B., Öretmenolu, M. and Akova, O. (2021). Bibliometric Analysis of Slow Tourism. *Advances in Hospitality and Tourism Research (AHTR)*, 9(1), 157-178.

McKinney, S.M., Sieniek, M. and Godbole, V. (2020). International Evaluation of an AI System for Breast Cancer Screening. *Nature,* 577, 89-94. https://doi.org/10.1038/s41586-019-1799-6

Minopoulos, G.M., Memos, V.A., Stergiou, C.L., Stergiou, K.D., Plageras, A.P., Koidou, M.P. and Psannis, K.E. (2022). Exploitation of Emerging Technologies and Advanced Networks for a Smart Healthcare System. *Applied Sciences*, 12(12), 5859.

Murdoch, T.B. and Detsky, A.S. (2013). The inevitable application of big data to health care. *JAMA*, *309*(13), 1351-1352. https://doi.org/10.1001/jama.2013.393

Obermeyer, Z. and Emanuel, E.J. (2016). Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *The New England journal of medicine*, 375(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W. and Boccia, S. (2019). Benefits and Challenges of Big Data in Healthcare: An Overview of the European Initiatives. *European Journal of Public Health*, 29, 23-27. https://doi.org/10.1093/eurpub/ckz168

Perianes-Rodriguez, A., Waltman, L. and van Eck, N. J. (2016). Constructing Bibliometric Networks: A Comparison Between Full and Fractional Counting. *Journal of Informetrics*, 10(4), 1178-1195. https://doi.org/10.1016/j.joi.2016.10.006

Pramanik, M.I., Lau, R.Y.K., Azad, M.A.K., Hossain, M.S., Chowdhury, M.K.H. and Karmaker, B.K. (2020). Healthcare Informatics and Analytics in Big Data. *Expert Systems with Applications*, 152. https://doi.org/10.1016/j.eswa.2020.113388

Raghupathi, W. and Raghupathi, V. (2014). Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems*, 2, 3. https://doi.org/10.1186/2047-2501-2-3

Rahmani, A.M., Gia, T.N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M. and Liljeberg, P. (2018). Exploiting Smart e-Health Gateways at the Edge of Healthcare Internet-of-Things: A Fog Computing Approach. *Future Generation Computer Systems*, 78, 641-658. https://doi.org/10.1016/j.future.2017.02.014 %R 10.1016/j.future.2017.02.014

Renault, M. (2021). How Patient Data Underpin COVID-19 Research. *Nature Medicine*, 27(9), 1486-1488. https://doi.org/10.1038/s41591-021-01493-5

Repanovici, A. (2011). Measuring the Visibility of the University's Scientific Production Through Scientometric Methods. *Performance Measurement and Metrics*, 12(2), 106-117. https://doi.org/10.1108/14678041111149345

Saxena, M. and Saxena, A. (2020). Evolution of mHealth Ecosystem: A Step Towards Personalised Medicine. In: Khanna, A., Gupta, D., Bhattacharyya, S., Snasel, V., Platos, J., Hassanien, A. (Eds.), International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, 1087. Springer, Singapore. https://doi.org/10.1007/978-981-15-1286-5_30

Shafqat, S., Fayyaz, M., Khattak, H.A., Bilal, M., Khan, S., Ishtiaq, O., Abbasi, A., Shafqat, F., Alnumay, W.S. and Chatterjee, P. (2021). Leveraging Deep Learning for Designing Healthcare Analytics Heuristic for Diagnostics. *Neural Processing Letters*, 1-27. https://doi.org/10.1007/s11063-021-10425-w

Sheng, J., Amankwah-Amoah, J., Khan, Z. and Wang, X. (2021). COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions. *British Journal of Management*, 32(4), 1164-1183. https://doi.org/10.1111/1467-8551.12441

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. (2017). Critical Analysis of Big Data Challenges and Analytical Methods. *Journal of Business Research*, 70, 263-286. https://doi.org/10.1016/j.jbusres.2016.08.001

Suban, S.A., Madhan, K. and Shagirbasha, S. (2021). A Bibliometric Analysis of Halal and Islamic Tourism. *International Hospitality Review*. https://doi.org/10.1108/ihr-05-2021-0038.

van Eck, N.J. and Waltman, L. (2017). Citation-based Clustering of Publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2), 1053-1070. https://doi.org/10.1007/s11192-017-2300-7

Wang, Y., Kung, L., Gupta, S. and Ozdemir, S. (2019). Leveraging Big Data Analytics to Improve Quality of Care in Healthcare Organisations: A configurational perspective. *British Journal of Management*, 30(2), 362-388. https://doi.org/10.1111/1467-8551.12332

Wang, L. and Alexander, C.A. (2020). Big Data Analytics in Medical Engineering and Healthcare: Methods, Advances and Challenges. *Journal of Medical Engineering and Technology*, 44(6), 267-283. https://doi.org/10.1080/03091902.2020.1769758

Weiskopf, N.G. and Weng, C. (2013). Methods and Dimensions of Electronic Health Record Data Quality Assessment: Enabling Reuse for Clinical Research. *Journal of the American Medical Informatics Association*, 20(1), 144-151. https://doi.org/10.1136/amiajnl-2011-000681

Wu, J., Li, H., Liu, L. and Zheng, H. (2017). Adoption of Big Data and Analytics In Mobile Healthcare Market: An Economic Perspective. *Electronic Commerce Research and Applications*, 22, 24-41. https://doi.org/10.1016/j.elerap.2017.02.002

Zupic, I. and Èater, T. (2015). Bibliometric Methods in Management and Organisation. *Organisational Research Methods*, 18(3), 429-472. https://doi.org/10.1177/1094428114562629

Cite this article as: Christine Nya-Ling Tan and Muhammad Ashraf Fauzi (2023). The Bibliometric Overview of Research on Healthcare Information Systems Using Big Data Analytics. *International Journal of Data Science and Big Data Analytics,* 3(1), 45-57. doi: 10.51483/IJDSBDA.3.1.2023.45-57.