



BIG DATA ANALYTICS APPLIED TO HEALTH SERVICES: A LITERATURE REVIEW

ANÁLISE DE BIG DATA APLICADA A SERVIÇOS DE SAÚDE: UMA REVISÃO DE LITERATURA

Recebido em: 26 maio 2020


Aprovado em: 14 set. 2020

Versão do autor aceita publicada online: 14 set. 2020

Publicado online: 04 jun. 2021

Como citar esse artigo - American Psychological Association (APA):

Gomes, M. A. S., Silva, V. L., Kovalski, J. L., & Pagani, R. N. (2022, jul./set.). Big data analytics applied to health services: a literature review. *Exacta*, 20(3), 647-665.
<https://doi.org/10.5585/exactaep.2021.17297>.

Submeta seu artigo para este periódico 



Dados Crossmark



BIG DATA ANALYTICS APPLIED TO HEALTH SERVICES: A LITERATURE REVIEW

ANÁLISE DE BIG DATA APLICADA A SERVIÇOS DE SAÚDE: UMA REVISÃO DE LITERATURA

 Myller Augusto Santos Gomes¹

 Vander Luiz da Silva²

 João Luiz Kovalski³

 Regina Negri Pagani⁴

¹ PhD candidate. in Production Engineering
Universidade Tecnológica Federal do Paraná – UTFPR
Ponta Grossa, Paraná - Brasil
myller@unicentro.br

² Msc. in Production Engineering
Universidade Tecnológica Federal do Paraná - UTFPR
Ponta Grossa, Paraná - Brasil
luizvnder@gmail.com

³ PhD in Industrial Instrumentation
Universidade Tecnológica Federal do Paraná - UTFPR
Ponta Grossa, Paraná - Brasil
kovaleski@utfpr.edu.br

⁴ Ph.D in Production Engineering
Universidade Tecnológica Federal do Paraná - UTFPR
Ponta Grossa, Paraná - Brasil
reginapagani@utfpr.edu.br

Recebido em: 26 maio 2020

Aprovado em: 14 set. 2020

Abstract: The purpose of this study is to understand the concepts and evolution of big data analytics applied to health services, considering activities that involve the diagnosis, treatment, and management of the patient. The literature review, consulting the databases Science Direct, Scopus and Web of Science and employing the keywords health analytics and big data analytics with no time restrictions, found papers that approach, specifically, the use of big data analytics in the context of healthcare, represented by examples and related analyses. Time and decision-making appear as actions developed by both the information technology team and the clinical team, may consider variables such as cost, time, decision, and functional structure performance as the main determining factors aligned with corporate strategy. This work expects to foster research on aspects of public health, in addition to considering the concern with the survival of affected people.

Keywords: Big data analytics. Health services. Health analytics. Technology transfer.

Resumo: O objetivo deste estudo é compreender os conceitos e a evolução da análise de big data aplicada aos serviços de saúde, considerando as atividades que envolvem o diagnóstico, tratamento e manejo do paciente. A revisão da literatura, consultando as bases de dados Science Direct, Scopus e Web of Science e empregando as palavras-chave health analytics e big data analytics sem restrições de tempo, encontrou trabalhos que abordam, especificamente, o uso de big data analytics no contexto da saúde, representados por exemplos e análises relacionadas. O tempo e a tomada de decisão aparecem como ações desenvolvidas tanto pela equipe de tecnologia da informação quanto pela equipe clínica, podendo considerar variáveis como custo, tempo, decisão e desempenho da estrutura funcional como os principais determinantes alinhados à estratégia corporativa. Este trabalho espera fomentar pesquisas sobre aspectos da saúde pública, além de considerar a preocupação com a sobrevivência das pessoas afetadas.

Palavras-chave: Análise de big data. Serviços de saúde. Análise de saúde. Transferência de tecnologia.

1 Introduction

Organizational contexts, especially environments containing a large and often-overlooked volume of data and information, require techniques and tools capable of analyzing and interpreting this volume of data so as to improve the performance of certain activities. Known as big data, this computer science problem has been widely discussed and employed, giving a real meaning to information storage (Shafqat et al, 2020).

Volume, variety, velocity and veracity are the defining characteristics of big data applications (Moutselos; Kyriazis; Maglogiannis, 2018; Oussous et al, 2018). The public health context, the reach and coverage of the Single Health System, and the private clinics, laboratories, and specialized centers produce a large volume of data and information that may be employed in predictive analyses that aim for strategic decision-making.

The research problem is to understand the concepts and evolution that surround health analytics based on big data analytics, whose data management and interpretation may generate several challenges related to size, format, heterogeneity, and sample incongruity Zhou et al, (2019), allowing for interpretations represented by knowledge and expressed as decision-making actions in patient treatment and management.

2 Big data

Collecting and handling big data is a challenge but also an opportunity for an in-depth analysis, with an economic perspective, of the emerging problems in healthcare (Moutselos; Kyriazis; Maglogiannis, 2018).

Several authors define and elucidate the concept of big data. In the view of Manjika (2011), it is a data set whose size is beyond the capacity of the traditional tools of data storage with initiatives of collecting, storing, managing, and analyzing possibilities of information extraction. The literature often defines big data in terms of data size, generally big and multidimensional as its central characteristics (Emani, Cullot, Nicolle, 2015).

Big data requires significant resources and robust methods and technologies that are capable of cleaning, processing, analyzing, protecting, and providing granular access to big, evolving data sets from different providers and institutions. It is directly influenced by emerging technologies like cloud computing and internet of things (Oussous et al, 2018).

Moreover, the aforementioned authors point out that the nature of the data treated by big data solutions is structured, that is, data from which information or knowledge can be obtained directly, yet unstructured and semi-structured data can be susceptible to analysis for extracting information,



highlighting its heterogeneity. Nevertheless, due to their complex nature, traditional tools of business intelligence cannot be superior to big data applications.

Big data is characterized by the volume, seeing that digital data are continuously generated by different information technology devices and their growth is extremely accelerated, illustrating the velocity. These data need to be processed to extract relevant information and construct insights. Finally, it is characterized by the variety, due to the heterogeneous nature of the data (Oussous et al, 2018).

Authors like Gandomir and Haider (2015) present other characteristics that concisely define big data:

Vision: Definition of a goal;

Verification: Data processed according to specifications;

Validation: Fulfillment of the goal;

Value: Pertinence of the information;

Complexity: Evolution and relationship of the data;

Immutability: It may be permanent, as long as it is managed.


Due to their characteristics and processing methods, big data applications present high hardware and software requirements. Seeing that the actions of collecting, integrating, and storing represent challenges in data management, it is essential to have a reliable process of information extraction and significantly lower defenses in order to ensure efficiency (Emani, Cullot, Nicolle, 2015).

Given this complexity, the challenge of dealing with big data lies in making it manageable, that is, facilitating the extraction of reliable information for an acceptable cost, moment in which the aggregation of data from different encoded sources entail efficiency gains in terms of storage, access, management, and security (Oussous et al, 2018).

With that understanding, the next steps will explore how big data analytics based on analytical techniques allows achieving significant findings that may reveal noteworthy strategic conditions in management, operations, businesses, and especially sustainable competitive advantages.

3 Health analytics through big data analytics

Linking health analytics to big data applications may contribute to the collective knowledge, track the results of prevention-based strategies, and improve the efficiency in patient management. This context, overlooked by managers, engineers, and policy-makers, may represent an avenue to be explored in healthcare (Moutselos; Kyriazis; Maglogiannis, 2018).



The big data generated in the context of health analytics may represent a way to personalize services, constantly seeking efficiency (Nambiar et al, 2013). Starting from different, heterogeneous data sources, the analyses show that the beneficial conditions are remarkable, such as the adjustment of online medical prescriptions, adaptation of health plans to the population's symptoms, evolution of diseases, enhancements in hospital treatments and operations, besides the reduction of healthcare expenses (Oussous et al, 2018).

Analytical techniques lead to big data analytics, whose range of tools and storage resources may contribute to its development. The authors Emani, Cullot and Nicolle (2015) point to a few techniques that stand out due to data variety:

- Association rules: seeking relationships between entities
- Machine learning: learning complex patterns and making intelligent, machine-based decisions
- Data mining: developing statistical combinations, machine learning, and database management
- Cluster analysis: employed in unsupervised machine learning, its purpose is to divide the data into smaller groups with the same set of still-unknown characteristics
- Crowdsourcing: employed in the collection of data, metadata, and/or resources seeking to enhance the current data semantics
- Text analytics: analyzing large collections of texts to extract information, with the possibility of use for topic modeling, with the question-answer relation.

Besides the analytical techniques, the term analytics gathers data from management information systems, operational research, and statistics (Galetsji; Katsaliaki, 2019).

Notably, interested parties (including doctors, patients, healthcare organizations, governments, and universities) are in some way invested in maximizing the potential of big data-based health analytics. However, due to the lack of qualified professionals, they often resort to data management automation from other organizations, which hinders quick responses to existing demands (Galetsji; Katsaliaki, 2019).

In addition to the hardware and software requirements, health analytics through big data analytics requires qualified professionals capable of integrating data from several sources and applying the analytical techniques that allow information extraction.

4 Methodology

The systematic review of the literature uses the adequacy of the Prisma and Methodi Ordinatio methodologies, with materials already published and found through a search in the databases. A search



using keywords without time restrictions due to the limited phenomenon of the study. The results included published articles, which were identified to verify whether they met the research question and the phenomenon studied. The intention of the systematic literature review is to find variables that can be analyzed in a given context, taking into account the research question. Developed by Pagani, Kovaleski, Resende, (2015), the Methodi Ordinatio it was applied to classify articles by scientific research, taking into account the impact variables, year of publication and number of citations.

Aiming to compose a portfolio of papers promoting new reflections, this study restricted itself to finding publications that specifically approach health analytics and big data analytics. Its strategy was based on stages that concisely enabled the fulfillment of the research problem:

1st stage: establishing the research purpose – **understanding the concepts and evolution of big data analytics applied to health analytics.**

2nd stage: defining the keyword combination – **health analytics, big data analytics.**

3rd stage: selecting the databases for the investigation – **Science Direct, Scopus and Web of Science.**

4th stage: selecting the time restrictions for the search – **2015 to 2020.**

5th stage: selecting the publication types – **papers and conference papers.**

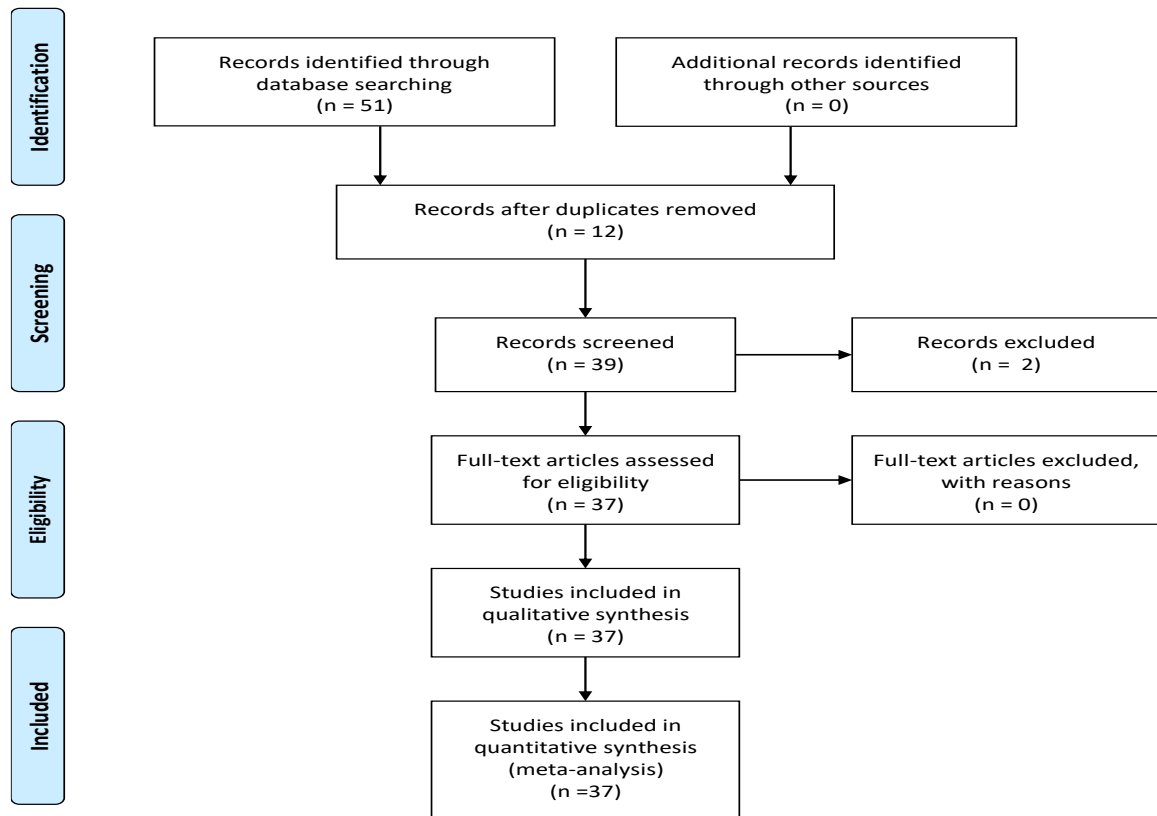
The following section exhibits the papers found, discussing the techniques employed and the contexts to which they were applied, in addition to how the big data analytics process was established with a link to health analytics.

5 Results

Systematizing the relevant findings, the searches on the Science Direct, Scopus and Web of Science databases returned, respectively, 21, 14 and 16 publications that approach the theme at hand. Checking for duplicate results led to the exclusion of 12 works, reaching a portfolio of 37 publications, one of which removed because it was not presented in the event to which it was submitted. Final search in the databases and gross results are described in Figure 1. Afterward, each study was analyzed to ascertain specific:

Figure 1

Flow chart of studies included in the review



Source: research data (2020).

Afterward, each study was analyzed to ascertain specific information, namely:

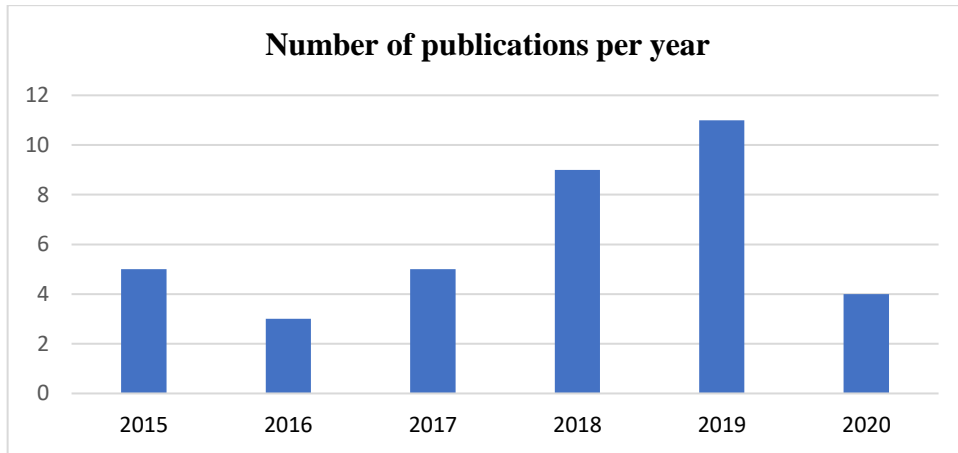
- Number of publications per year;
- Main authors and their publications;
- Countries that publish about the theme;
- Main journals;
- Field of the study object.

Graph 1 presents the number of publications per year on the Scopus database:



Graph 1

Number of publications per year

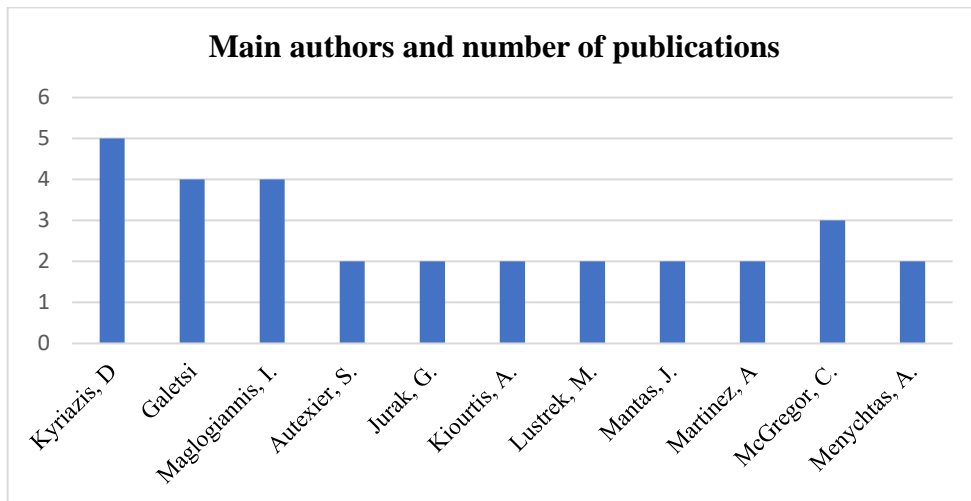


Source: Research data (2020).

Graph 2 displays the main authors and their number of publications.

Graph 2

Main authors and their number of publications

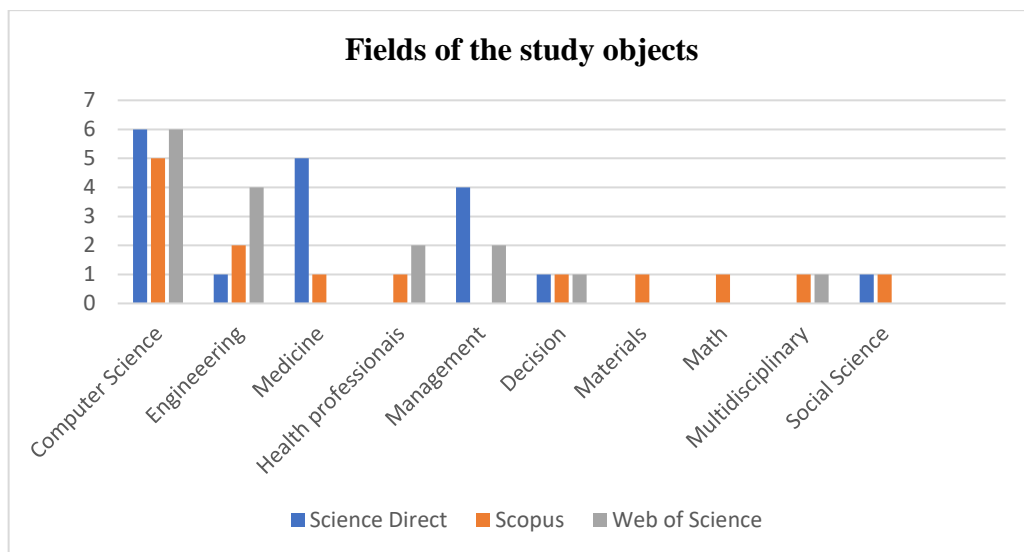


Source: Research data (2020).

Graph 3 presents the main fields of research development that had studies on the theme published by international journals, pointing to the percentage corresponding to each field.

Graph 3

Fields of the study objects



Source: Research data (2020).

Table 1 presents the main journals that have published about the research theme. Due to the number of publications found, all of the journals and conferences are listed.

Table 1

Main journals and conferences

Main Journals and Conferences
Big Data Research
Scientific Reports
International Journal of Human-Computer Studies
Information Processing & Management
Proceedings – IEEE Symposium on Computer-Based Medical Systems
9th International Conference on Information, Intelligence, Systems and Applications, IISA 2018
Proceedings – 2018 IEEE International Conference on Big Data, Big Data 2018
Journal of the Operational Research Society
Journal of Ambient Intelligence and smart environments
Journal of Biomedical Informatics
Health Information and libraries journal
Studies in Health Technology and Informatics
Procedia Computer Science
IEEE Access
International Journal of Information Management
Social Science & Medicine
Journal of the American College of Radiology
International Journal of Human-Computer Studies
Journal of Infection and Public Health
Clinical Therapeutics
Pediatric Clinics of North America
Future Generation Computer Systems



Main Journals and Conferences
Studies in Health Technology and Informatics
Proceedings – IEEE International Congress on Big Data, BigData Congress 2015
Advances in Intelligent Systems and Computing
Lecture Notes in Computer Science
Journal of the American College of Radiology
Technological Forecasting and Social Change

Source: research data (2020).

5.1 Discussion


In an initial analysis, the publications found illustrate that the theme itself has been significantly approached over the last five years, highlighting its current relevance due to its relation to computer science. The technological evolution of the use of analytical techniques and big data accompanies fields like engineering, business, and medicine. However, the number of studies decreases considerably when it comes to big data analytics in the health field.

Galetsis, Katsaliaki, Kumar, (2020), in a study published in the International Journal of Information Management, called Big data analysis in the health sector, theoretical framework, techniques and perspectives presented a systematic overview of the literature in order to determine the how big data analytics have been able to improve the health domain, supported by resource-based theory to identify the sources of big data and the analytical techniques that allow big data capabilities to create values, using content analysis, present the most popular analytical techniques that scientists use to make meaningful interpretations of data are: modeling, machine learning, data mining, visualization and statistical analysis.

Observing phenomena on social media can contribute to the identification of health behavior in certain localities. Given this, Nhuyen et al., (2020) presents and proposes the application of statistical resources built on primary resources to predict health indexes at the health care level. county, entitled Using spatiotemporal distribution of geocoded Twitter data to predict US county-level health indexes indexed in the journal Future Generation Computer Systems.

Prepared by Stephaine, Sharma (2020) entitled Digital health eco-systems: An epochal review of practice-oriented research, this paper presents a historical review of the research literature on practice-oriented digital health ecosystems, in the period 1998-2018 along with monitoring the evolution of digital health and developments evident in the field - crucial steps to gain an understanding of the key issues surrounding healthcare delivery, published in the International Journal of Information Management.

Galetsis, Katsaliaki, Kumar, (2019) developed a study based on a literature review entitled Values, challenges and future directions of big data analysis in health: a systematic review, published in the journal Social Science & Medicine whose objective was to systematically review 804 publications



academic studies related to big data analysis in health, in order to identify the organizational and social values together with the associated challenges. Organized methodologically by (PRISMA), the relevant findings, mainly values created, are the development of analytical techniques that provide personalized health services to users and support human decision making through automated algorithms, challenging the issues of power in the medical relationship -patient and creating new working conditions.

One of the initial challenges is the availability of the variety of data that circulates in this environment. Focusing on that matter, the investigation of Zhou et al, (2019), published by the journal Scientific Reports, selected the Biobank in the United Kingdom as its object of analysis. It is a database of the National Health Service that provides valuable information about human health divided into data related to demographics, health, and diseases.

The authors sought to develop the analytical technique of machine learning geared toward the main determining factors of mental illnesses, with a specific focus on depression. Other analytical challenges included resource selection, data harmonization, and data analysis.

The analysis of the information from the Biobank employed demographic, clinical, biological, and genomic data, as well as data from questionnaires and images, specifically magnetic resonance, comprising 3,297 morphometric measurements. Afterward, it developed a decision support system to detect and prevent common diseases.

Publishing at the Proceedings – IEEE Symposium on Computer-Based Medical Systems, Mcgregor and Majola, (2019) approached traditional male circumcision and how health analytics with a cloud-based service may help save lives and preserve customs. The authors proposed a platform that aids in the collection of physiological data, supporting real-time decision-making and providing information about conditions that may develop during the process. Every initiative is related to big data analytics, given that a repository is constructed.

The expository assessment of health data supports the analysis of the relations between negative health effects and the levels of environmental factors. Alkobaisi et al, (2019) applied predictive and expository analysis to conduct the simultaneous monitoring of behaviors and mobility beyond the measurement of environmental conditions to determine their impact on human health. Their purpose was to promote management guided by health surveillance. The authors presented a framework that combines expository analysis and health analytics, displaying their managerial processes and underlying concepts. The study focused on asthma development influenced by environmental conditions.

Blanford (2019) in his work entitled HCI for health and wellbeing: Challenges and opportunities where it presents environments of interactive technologies in health, users and uses, and interdependencies between them, published in the International Journal of Human-Computer Studies, a literature review was developed, highlighted opportunities and challenges for the application of HCI methods in the design and implementation of digital health technologies.




Holistic health records and big data analytics were discussions presented by Gallos et al, (2019) in their paper published in the journal *Studies in Health Technology and Informatics*. The authors demonstrated the use of big data techniques in the stored data, that is, in holistic health records. The correlations and extractions from different data sources may be employed by different analyses, such as risk analysis, path mining, prediction, and causal analysis.

Galetsis and Katsaliaki, (2019) published a literature review about big data analytics and health analytics in the *Journal of the Operational Research Society*, a journal with an impact factor. In summary, the authors created profiles to provide a general view of the publishing dynamics of big data analytics in the field of health and discuss related examples based on the literature review.

The same authors developed a bibliometric study about big data applied to health analytics, seeking to report the new systems and tools of information technology attained by the literature. They highlighted the context in which those studies are developed and published, pointing to the United States as the leading country in the research development. The authors also drew attention to other nations and research institutes, especially in the fields of medicine and computer science. Finally, they pointed out that several interested parties appear, emphasizing the patient as the beneficiary of this relation between big data analytics and health analytics.

Technological Forecasting and Social Change a relevant journal in the field of technological research published a study entitled *An empirical study of patients' privacy concerns for health informatics as a service*, developed by Xu (2019) empirically tested a developed model that examines predictors and consequences of patients' concerns about the privacy of information in HlaaS (health informatics and cloud computing). Considering that each aspect of the HlaaS application has the possibility of raising concerns about the privacy of information for users, we formulate the hypothesis of the antecedents of concerns with six dimensions privacy, including privacy awareness (individual level), perceived informativeness (service level), information sensitivity (information contingency), social norm privacy (socio-relational level), regulatory expectations (macro-environmental level) and importance transparency of information (level organizational).

Research presented in the journal *Information Processing & Management*, by authors Saggi, Jain, (2018) presents a comprehensive set of realistic analysis of the implementation of successful big data analysis in companies, yet still provide an overview of the architecture of the BDA, including six components, namely:

- 
1. Data generation,
 2. Data acquisition,
 3. Data storage,
 4. Advanced data analysis,
 5. Data visualization and
 6. Decision making

In this work, seven V 's characteristics of big data are presented, being Volume, Speed, Variety, Valency, Veracity, Variability and Value.

Developed by Alharthi (2018), this review presents the current state of knowledge about “big data” and predictive analysis approaches in the health sector. Application-specific case studies currently in effect in the for-profit business sector are presented to help consider what may be possible in specific health applications, finally highlighting the unique challenges inherent in the development and application of predictive health analytic applications effective in Saudi Arabia, where progress towards a unified electronic medical record system has been slow and limiting the scalability of decision making, called Healthcare predictive analytics: An overview with a focus on Saudi Arabia released in the Journal of Infection and Public Health.

Bathazar et al., (2018), Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics, published in the Journal of the American College of Radiology addresses the main principles of the structures applicable in the imaging environment and radiology, explored known problems in dealing with big data and machine learning in healthcare, examined the perspectives of key stakeholders, especially for healthcare professionals, and needed to embark on their own big data and artificial intelligence ventures.

Health management and policy-making in public health based on big data analytics were approached by Moutselos, Kyriazis, Maglogiannis, (2018) in their work published at the 9th International Conference on Information, Intelligence, Systems and Applications IISA 2018. The authors developed a conceptual model of a modular web environment that aims to explore health data and preexisting knowledge of public health policy models. They considered mechanisms of prediction, causal analysis, and risk.

Health analytics and its tools were concerns listed by Moutselos, Kyriazis, Diamantopoulou and Maglogiannis, (2018). In their work, presented at the Proceedings – IEEE International Conference on Big Data, Big Data 2018, the authors pointed out that the tools are dissociated from the database and user interface. The initiative offers a process of modification and construction of new health policy models based on the association of health analytics tools.



Electronic health records and the application of analytical tools focused on enhancing the work of healthcare professionals comprised the objective of the paper authored by Khennou, Khamlichi and Chaoui, (2018) published in the journal *Procedia Computer Science*. The authors explored, with prediction and adaptation, the analytical tools applied to electronic health records seeking to achieve new perspectives of the patient based on the healthcare professionals. However, the exploration of this resource is modest.

Harerimana, Jang, Kim and Park, (2018) in a paper published by the journal *IEEE Access*, highlighted the concept of internet of medical things. The authors focused on technological aspects, considering challenges, data sources, technologies, and orientation of big data analytics. They encourage the production of an integrated analytical health app geared toward analyzing the vast availability of existing data.

The health activities and their operationalization depend on solutions that process a large amount of data; therefore, the authors Pashazadeh, Navimipour (2018) developed a study entitled *Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review* indexed in the *Journal of Biomedical Informatics* its goal is to provide the comprehensive, detailed and systematic study of cutting edge big data mechanisms related to healthcare applications in five categories, including machine learning, cloud based, heuristic based, agent based and hybrid mechanisms.

Supported by the development of a semantic extraction framework and the assessment of risk factor feelings based on machine learning algorithms aimed at a cardiovascular disorder, the vector support machine structure stood out, a study published in the journal *Computers in Biology and Medicine* by the authors Sabra, Malik, Alobaidi (2018), entitled *Prediction of venous thromboembolism using semantic and sentiment analyzes of clinical narratives*.

Concerned with the service quality and the efficiency and sustainability of the healthcare and social assistance systems, big data analytics and its technologies can potentially process and analyze these data in order to produce relevant ideas and aid the decision-making process.

Report the main dimensions of hemorrhagic dengue provide unified access to health records with all information about the patient's health determinants, yet present the current and future potential applications to gain a new understanding of the patient's health through big data and modeling tools are the object of study by authors Cano et al., (2017) entitled *Perspectives on Big Data applications of health information* published in *Current Opinion in Systems Biology*, in terms of findings, latent sources of information on health determinants of the patient are highlighted and potential strategies are proposed within a digital health framework that aims to support the emerging requirements of medicine. The development of risk assessment and risk stratification models for patient health, integrating health information from informal care, formal health and biomedical research, while

creating new knowledge about the mechanisms of the disease at a time when strategies are deliberate and intended.

Gonzalez, Vilar and Lupianez-Villanueva, (2017) presented a paper about computer-based medical systems at the Proceedings – IEEE Symposium on Computer-Based Medical Systems. Their purpose was to present a methodology to approach big data analytics geared toward integrated care-taking.

At the same IEEE Symposium mentioned above, Balaji, Patil and Macgregor, (2017) presented the Artemis project applied to a hospital in Toronto, Canada, involving big data analytics employed to feed the system of clinical decision support. The authors revealed opportunities and challenges related to intensive pediatric care, seeking to improve medical care in low-resource locations based on remote, real-time patient monitoring.

Personalized healthcare based on health policies and big data analytics applied to holistic records were the focus of the study conducted by Kbioassist, (2017), published by the journal Studies in Health Technology and Informatics. Based on the Crowd health approach, it pointed to health determinants and holistic health records, grouped into clusters considering the clinical, social, and human context of the population segments, producing collective knowledge about different factors.

Assessing the Quality of Service Using Big Data Analytics: With Application to Healthcare prepared by researchers Batarseh, Latif, (2016) and published in the magazine Big Data Research, the focus of the study is on the analysis of big data is applied to health data highlighting the different data sources from which they can contribute to new quality insights and apprehend best field practices using new health-specific big data tools, in particular by focusing on the data life cycle in the organizational environment.

Concern with patients with systemic lupus erythematosus (SLE), Kan et al., (2016) presented a study in the journal Clinical Therapeutics, entitled Longitudinal Treatment Patterns and Associated Outcomes in Patients With Newly Diagnosed Systemic Lupus Erythematosus, its objective is focused on complexity treatment process, with a wide range of commonly prescribed drugs. The objectives of this study were to identify longitudinal treatment patterns in patients with incident SLE and to estimate the associations of treatment patterns with clinical and economic results.

Prescriptive analysis applied to data processing is the objective of the study developed by Poornima, Pushpalatha, (2016) published in the International Journal of Intelligent Networks, to optimize decision making based on prescriptive analysis are explored from comparative studies, demonstrating the efficiency in a context of big data analysis, however, has no specific focus on health data or organizations.

Titled Big Data and Predictive Analytics: Applications in the Care of Children, published by Pediatric Clinics of North America, Suresh (2016), presents research on the combination of big data and



predictive analysis for treatments that work specifically for individual children, yet provide the ability to prescribe drugs that work for each individual and are not generalized data just because that drug works for most people, highlight the role of the patient and family and collaborative terms.

At the Proceedings – IEEE International Congress on Big Data, Big Data Congress 2015, the authors Abusharekh, Stewart, Hashemian and Abidi, (2015) published their investigation, in which they presented a big data-based health analysis represented by the H-drive platform, an environment of inference and visualization for data integrity reconciliation. This platform is a work environment that provides analytical experiments and complex analyses of health data, aiming for evidence-based decision-making. The authors reported the case of a pathology laboratory, whose performance is described in operational terms.

Published in the journal *Advances in Intelligent Systems and Computing*, the paper of Kakhki, Singh and Loyd, (2015) focused on patient activation, that is, patients who perform an active role in their health management and care with the perception of confidence in their capability to manage their health. The authors discussed this aspect considering big data analytics to identify, monitor, and enhance the health organization in patient activation, promoting health analytics.

Considering the mobile technologies utilized by individuals, Wlodarczak, Soar and Ally, (2015) associated the analysis of behavioral data as a method through which personalized healthcare and monitoring become tools to improve the quality of healthcare. Employing mobile phones, the behavioral analysis developed a social sensor fed automatically with daily data, which were analyzed through techniques of machine learning, data mining, and predictive analysis. These techniques identify atypical behaviors and the initiatives that health centers may develop.

The same authors also highlighted, in another publication with the same research object, that reality mining enables monitoring, a method that may follow the aging of the population and reduce impacts in medical care cases. Reality mining contributes to the predictability of a population in the event of an epidemic.

Comprising diverse techniques in different contexts, big data analytics applied to health analytics may significantly contribute to stages like diagnosis, treatment, and monitoring of several diseases. Information technology improves medical decisions in terms of time, cost, and options, considering the relations between data produced and inferred from different sources, enabling a better way to conduct and manage healthcare.

6 Final considerations

Fulfilling its purpose, this paper explored the concepts and evolution of health analytics based on big data analytics. Nations that have a wide-coverage healthcare system may employ big data

analytical techniques to guide health analytics focusing on the operational and strategic aspects of health policies.

The examined publications revealed significant aspects represented by technology, applied context, and analytical techniques, as well as their relation to patient treatment and management, providing healthcare professionals with analyses of cause, risk, prediction, electronic records, and decision-making support.

The contexts to which big data analysis was applied included hospitals, pathology laboratories, patient activation management, integrated care, government health databases, holistic health records and electronic health records.

The analysis of all publications indexed in the Science direct, Scopus and Web of Science database with the keywords health analytics and big data analytics showed that the number of publications is small due to the complexity of the problem, however an evolution in the investigations is being perceived from 2018, a limitation presented is the behavior of keywords in the databases containing themselves in the analytical context, in addition, few articles were published in journals with a high impact factor, which is an opportunity for researchers develop new investigations on the subject.

The challenge of health analytics based on big data analytics lies in the complexity of data from different sources. However, it is possible to establish health analytics in systems that complement the central one, enabling the application of analytical techniques capable of influencing the time of treatment and the decisions involved in patient management.

References

- Alharthi, H. (2018). Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *Journal of infection and public health*, 11(6), 749-756. <https://doi.org/10.1016/j.jiph.2018.02.005>.
- Alkobaisi, S., Bae, W. D., Horak, M., Narayanappa, S., Lee, J., AbuKhoua, E., ... & Bae, D. J. (2019). Predictive and exposome analytics: A case study of asthma exacerbation management. *Journal of Ambient Intelligence and Smart Environments*, (Preprint), 1-26. <https://doi.org/10.3233/AIS-190540>
- Abusharekh, A., Stewart, S. A., Hashemian, N., & Abidi, S. S. R. (2015, June). H-DRIVE: A Big Health Data Analytics Platform for Evidence-Informed Decision Making. In 2015 IEEE International Congress on Big Data (pp. 416-423). IEEE. <https://doi.org/10.1109/BigDataCongress.2015.68>
- Balaji, S., Patil, M., & McGregor, C. (2017, June). A cloud based big data based online health analytics for rural nicus and picus in india: Opportunities and challenges. In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 385-390). IEEE. <https://doi.org/10.1109/CBMS.2017.112>
- Batarseh, F. A., & Latif, E. A. (2016). Assessing the quality of service using big data analytics: with application to healthcare. *Big Data Research*, 4, 13-24. <https://doi.org/10.1016/j.bdr.2015.10.001>



- Balthazar, P., Harri, P., Prater, A., & Safdar, N. M. (2018). Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. *Journal of the American College of Radiology*, 15(3), 580-586. <https://doi.org/10.1016/j.jacr.2017.11.035>
- Blandford, A. (2019). HCI for health and wellbeing: Challenges and opportunities. *International Journal of Human-Computer Studies*, 131, 41-51. <https://doi.org/10.1016/j.ijhcs.2019.06.007>.
- Cano, I., Tenyi, A., Vela, E., Miralles, F., & Roca, J. (2017). Perspectives on big data applications of health information. *Current Opinion in Systems Biology*, 3, 36-42. <https://doi.org/10.1016/j.coisb.2017.04.012>
- Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable big data: a survey. *Computer science review*, 17, 70-81. <https://doi.org/10.1016/j.cosrev.2015.05.002>
- Galetsis, P., Katsaliaki, K., & Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management*, 50, 206-216. <https://doi.org/10.1016/j.ijinfomgt.2019.05.003>
- Galetsis, P., Katsaliaki, K., & Kumar, S. (2019). Values, challenges and future directions of big data analytics in healthcare: A systematic review. *Social Science & Medicine*, 241, 112533. <https://doi.org/10.1016/j.socscimed.2019.112533>
- Galetsis, P., & Katsaliaki, K., & Kumar, S. (2019). Big Data Analytics in Health: an overview and bibliometric study of research activity. *Health Information & Libraries Journal*. <https://doi.org/10.1111/hir.12286>
- Galetsis, P., & Katsaliaki, K., & Kumar, S. (2019). A review of the literature on big data analytics in healthcare. *Journal of the Operational Research Society*, 1-19. <https://doi.org/10.1080/01605682.2019.1630328>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gonzalez-Alonso, P., Vilar, R., & Lupiáñez-Villanueva, F. (2017, June). Meeting Technology and Methodology into Health Big Data Analytics Scenarios. In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 284-285). IEEE. <https://doi.org/10.1109/CBMS.2017.71>
- Harerimana, G., Jang, B., Kim, J. W., & Park, H. K. (2018). Health big data analytics: A technology survey. *IEEE Access*, 6, 65661-65678. <https://doi.org/10.1109/ACCESS.2018.2878254>
- Kakhki, M. D., Singh, R., & Loyd, K. W. (2015). Developing Health Analytics Design Artifact for Improved Patient Activation: An On-going Case Study. In *New Contributions in Information Systems and Technologies* (pp. 733-739). Springer, Cham. https://doi.org/10.1007/978-3-319-16486-1_72
- Kan, H., Nagar, S., Patel, J., Wallace, D. J., Molta, C., & Chang, D. J. (2016). Longitudinal treatment patterns and associated outcomes in patients with newly diagnosed systemic lupus erythematosus. *Clinical Therapeutics*, 38(3), 610-624. <https://doi.org/10.1016/j.clinthera.2016.01.016>

- 
- kBioAssist, S. (2017). CrowdHEALTH: Holistic Health Records and Big Data Analytics for Health Policy Making and Personalized Health. *Informatics Empowers Healthcare Transformation*, 238, 19. <http://dx.doi.org/10.3233/978-1-61499-781-8-19>
- Khennou, F., Khamlichi, Y. I., & Chaoui, N. E. H. (2018). Improving the Use of Big Data Analytics within Electronic Health Records: A Case Study based OpenEHR. *Procedia Computer Science*, 127, 60-68. <https://doi.org/10.1016/j.procs.2018.01.098>
- Manyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation
- McGregor, C., & Majola, P. X. (2019, June). Opportunities for a Cloud Based Health Analytics as a Service for Eastern Cape Initiation Schools in South Africa. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 531-534). IEEE. <https://doi.org/10.1109/CBMS.2019.00108>
- Moutselos, K., Kyriazis, D., & Maglogiannis, I. (2018, July). A Web Based Modular Environment for Assisting Health Policy Making Utilizing Big Data Analytics. In 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-5). IEEE. <https://doi.org/10.1109/IISA.2018.8633625>
- Moutselos, K., Kyriazis, D., Diamantopoulou, V., & Maglogiannis, I. (2018, December). Trustworthy data processing for health analytics tasks. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 3774-3779). IEEE. <https://doi.org/10.1109/BigData.2018.8622449>
- Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013, October). A look at challenges and opportunities of big data analytics in healthcare. In 2013 IEEE international conference on Big Data (pp. 17-22). IEEE. <https://doi.org/10.1109/BigData.2013.6691753>
- Nguyen, T., Larsen, M., O'Dea, B., Nguyen, H., Nguyen, D. T., Yearwood, J., ... & Christensen, H. (2018). Using spatiotemporal distribution of geocoded Twitter data to predict US county-level health indices. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2018.01.014>
- Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Pagani, R. N., Kovaleski, J. L., & Resende, L. M. (2015). Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. *Scientometrics*, 105(3), 2109-2135. <https://doi.org/10.1007/s11192-015-1744-x>
- Pashazadeh, A., & Navimipour, N. J. (2018). Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review. *Journal of biomedical informatics*, 82, 47-62. <https://doi.org/10.1016/j.jbi.2018.03.014>
- Poornima, S., & Pushpalatha, M. (2020). A survey on various applications of prescriptive analytics. *International Journal of Intelligent Networks*, 1, 76-84. <https://doi.org/10.1016/j.ijin.2020.07.001>



- Sabra, S., Malik, K. M., & Alobaidi, M. (2018). Prediction of venous thromboembolism using semantic and sentiment analyses of clinical narratives. *Computers in biology and medicine*, 94, 1-10. <https://doi.org/10.1016/j.combiomed.2017.12.026>
- Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, 54(5), 758-790. <https://doi.org/10.1016/j.ipm.2018.01.010>
- Stephanie, L., & Sharma, R. S. (2020). Digital health eco-systems: An epochal review of practice-oriented research. *International Journal of Information Management*, 53, 102032. <https://doi.org/10.1016/j.ijinfomgt.2019.10.017>
- Shafqat, S., Kishwer, S., Rasool, R. U., Qadir, J., Amjad, T., & Ahmad, H. F. (2020). Big data analytics enhanced healthcare systems: a review. *The Journal of Supercomputing*, 76(3), 1754-1799. <https://doi.org/10.1007/s11227-017-2222-4>
- Suresh, S. (2016). Big data and predictive analytics: applications in the care of children. *Pediatric Clinics*, 63(2), 357-366. <https://doi.org/10.1016/j.pcl.2015.12.007>
- Xu, Z. (2019). An empirical study of patients' privacy concerns for health informatics as a service. *Technological Forecasting and Social Change*, 143, 297-306. <https://doi.org/10.1016/j.techfore.2019.01.018>
- Zhou, Y., Zhao, L., Zhou, N., Zhao, Y., Marino, S., Wang, T., ... & Dinov, I. D. (2019). Predictive Big Data Analytics using the UK Biobank Data. *Scientific reports*, 9(1), 6012. <https://doi.org/10.1038/s41598-019-41634-y>
- Wlodarczak, P., Soar, P., & Ally, M. (2015). Behavioural health analytics using mobile phones. *EAI Endorsed Trans. Scalable Information Systems*, 2(5), e6. <https://doi.org/10.4108/sis.2.5.e6>
- Wlodarczak, P., Soar, J., & Ally, M. (2015, May). Reality mining in eHealth. In *International Conference on Health Information Science* (pp. 1-6). Springer, Cham. https://doi.org/10.1007/978-3-319-19156-0_1