

# PRODUÇÃO CIENTÍFICA SOBRE HOSPITAIS NO CONTEXTO DA CIÊNCIA DE DADOS: UM ESTUDO A PARTIR DA WEB OF SCIENCE

Scientific production about hospitals in the context of Data Science: a study from the Web of Science

**Natanael Vitor Sobral**

Doutor em Ciência da Informação  
Universidade Federal da Bahia, Instituto de Ciência da  
Informação, Salvador, Brasil  
natanvsobral@gmail.com  
<https://orcid.org/0000-0003-2410-494X> 

**Gillian Leandro de Queiroga Lima**

Doutor em Difusão do Conhecimento  
Universidade Federal da Bahia, Instituto de Ciência da  
Informação, Salvador, Brasil  
gillianqueiroga@gmail.com  
<https://orcid.org/0000-0003-4736-893X> 

**Ana Sara Pereira de Melo Sobral**

Mestre em Ciência da Informação  
Instituto Federal Baiano, Biblioteca, Salvador, Brasil  
anasarap@hotmail.com  
<https://orcid.org/0000-0002-8794-0905> 

A lista completa com informações dos autores está no final do artigo 

## RESUMO

**Objetivo:** realizar análise bibliométrica sobre as aplicações da ciência de dados no âmbito das organizações hospitalares.

**Método:** por meio de pesquisa na base de dados Web of Science, verificou-se a existência de termos relacionados à ciência de dados, tais como “big data”, “data analytics”, “business intelligence”, “data mining”, “data warehouse”, “text mining” e “data science”, relacionando-os a hospitais. A análise de dados pautou-se na técnica de análise de redes sociais. O período considerado foi de 2015 a 2019.

**Resultado:** “machine learning” e “electronic health records” despontam como assuntos relevantes. As interações mais expressivas refletem a inclinação da informática médica em assuntos relacionados à tomada de decisão, sistemas de informação para hospitais e unidade de cuidados intensivos. Sobre os campos científicos, nota-se a predominância esperada da área de saúde e dos domínios pertencentes ou fronteiros à tecnologia. No mais, vê-se que a grande variedade de áreas encontradas acusa a natureza multidisciplinar do assunto, inclusive com importante participação da Ciência da Informação (CI). Em relação à geografia do conhecimento, observa-se um razoável grau de descentralização, havendo produções representativas na América do Norte, Europa e Ásia. Quanto aos veículos de publicação, destaque para os Studies in Health Technology and Informatics, que compreendem uma série de publicações. Os dois periódicos mais representativos da lista, integram, respectivamente, os grupos Springer Nature e Elsevier, grandes players do mercado editorial científico.

**Conclusões:** por fim, evidencia-se a multidisciplinaridade existente em torno do assunto estudado e a relevância da tecnologia para o progresso das organizações hospitalares.

**PALAVRAS-CHAVE:** Ciência de Dados. Bibliometria. Web of Science. Big Data. Machine Learning.

## ABSTRACT

**Objective:** to carry out the bibliometric analysis on the applications of Data Science in the context of hospital associations.

**Methods:** Through research in the Web of Science database, it was verified the existence of terms related to Data Science, such as, big data, data analysis, business intelligence, data mining, data warehouse, text mining and data science, relating them to hospitals. Data analysis was based on the social network analysis technique. The period considered was from 2015 to 2019.

**Results:** Machine learning and electronic health records emerge as relevant issues. The most expressive interactions reflect the inclination of Medical Informatics in matters related to decision making, information systems for hospitals and intensive care units. Regarding the fields, it is noted the expected predominance of the Health area and of the domains belonging or bordering on Technology. In addition, it can be seen that the wide variety of areas found accuses the interdisciplinary nature of the subject, including, with an important participation of Information Science. Regarding the geography of knowledge, there is a reasonable degree of decentralization, with representative productions in North America, Europe and Asia. As for the publication vehicles, emphasis is given to Studies in Informatics and Health

Technology, which comprise a series of publications. The two most representative journals on the list are, respectively, members of the Springer Nature and Elsevier groups, major players in the scientific publishing market. Conclusions: Finally, there is evidence of the multidisciplinary around the subject studied and the technology company for the progress of hospital associations.

**KEYWORDS:** Data Science. Bibliometrics. Web of Science. Big Data. Machine Learning.

## 1 INTRODUÇÃO

Ao longo dos anos, vem se notando um aumento na quantidade de dados gerados pelos processos de assistência em saúde. Esses conjuntos de dados, armazenados nos sistemas de registro eletrônico de saúde, são recursos subutilizados, considerando-se as potencialidades dos atuais métodos desenvolvidos no âmbito da ciência de dados, sobretudo com o propósito de melhorar a prestação de serviços em saúde (DALIANIS et al., 2015). Sobre isso, entende-se que a utilização efetiva desses registros é uma demanda crescente, por proporcionar condições de personalização da assistência provida aos cidadãos e pelas possibilidades de formulação de políticas públicas em saúde para grupos específicos, permitindo a adoção de uma perspectiva gerencial no âmbito dos serviços e sistemas de saúde, pautando-se em dados e indicadores que possibilitem incremento à qualidade do funcionamento desses organismos.

No que concerne ao estudo da aplicação da ciência de dados no contexto hospitalar, em uma perspectiva praxeológica, nota-se que as ações da interlocução entre os domínios da Saúde, Tecnologia e Ciência da Informação (CI), centradas em sistemas e serviços de saúde, podem potencializar a concepção, planejamento e desenvolvimento de ferramentas e métodos capazes de aperfeiçoar os produtos e serviços em saúde entregues à população, evitando e solucionando problemas (GALVÃO; VALENTIM, 2019), com base em *data analytics*, *big data* e outras disciplinas adjacentes (CAO, 2016).

Motivado pelas possibilidades oferecidas pela ciência de dados à gestão dos equipamentos hospitalares, assunto de grande ressonância no tempo presente, uma vez que essas técnicas são propulsoras da qualidade desses organismos, questiona-se, aqui, qual é o perfil e quais são as características da produção científica indexada em veículos internacionalizados sobre o tema selecionado. Para isso, buscou-se examinar o estado da arte relacionado ao assunto, com o intuito de se obter uma visão geral sobre o que vem sendo produzido, proporcionando uma ordenação que permita conhecer a evolução das pesquisas, bem como suas características (ROMANOWSKI; ENS, 2006).

Assim, este artigo tem por objetivo realizar uma análise bibliométrica sobre as aplicações da ciência de dados no âmbito das organizações hospitalares. Tendo em vista a contemporaneidade desse tópico na literatura científica, o recorte temporal selecionado

considera o período entre 2015 e 2019. A base de dados utilizada foi a Web of Science (WoS), dada sua característica pluridisciplinar, sendo referência na indexação de periódicos científicos e demais bibliografias de circulação internacional.

## 2 CIÊNCIA DE DADOS EM SAÚDE

A ciência de dados aplicada à área de saúde tem obtido progresso rápido a partir de três frentes principais: a primeira, baseada na agregação de grandes e complexos volumes de dados, incluindo registros advindos de prontuários médicos eletrônicos, mídias sociais, bancos de dados genômicos e dados fisiológicos digitalizados de dispositivos móveis de saúde sem fio; a segunda, relacionada às iniciativas de acesso aberto que buscam disponibilizar fontes de dados sobre estudos clínicos, pesquisas e ciência cidadã<sup>1</sup> para compartilhamento de dados; e a terceira, que enfatiza técnicas analíticas, como *big data*, aprendizado de máquina e inteligência artificial, que permitem obter conhecimento a partir de dados estruturados e não estruturados (BHAVNANI; MUÑOZ; BAGAI, 2016).

De um modo geral, “a Ciência de Dados é um conceito que emerge das aplicações ligadas ao comércio, saúde, meio ambiente, governo e outros domínios” (MOSTAFA, ROCHA; 2020, p. 128) e que pode ser compreendida como uma disciplina emergente

[...] que analisa e gerencia dados, combinando aplicações estatísticas, *machine learning*, mineração de dados e tecnologias de base de dados, para responder aos desafios apresentados pelo Big Data, tais como, grandes volumes de dados, a velocidade de atualização e a variedade dos suportes informacionais em diversas área e setores, mas, principalmente, em empresas e na comunidade científica (MOSTAFA; ROCHA, 2020, p. 129).

*Big data* “requer formas inovadoras de processamento de grandes volumes de dados heterogêneos, amparando o processo de tomada de decisão guiado por dados” (RAUTENBERG; CARMO, 2019, p. 57). Além disso, também é fundamental que os profissionais atuantes nesse segmento possuam competências associadas à organização, representação, recuperação e visualização das informações provenientes das bases de dados, de modo a potencializar o processo decisório (RAUTENBERG; CARMO, 2019). Assim, compreende-se que o uso do *big data* permite

---

<sup>1</sup> Segundo Elliott e Rosenberg (2019), a ciência cidadã é reconhecida como uma abordagem importante para a coleta de dados, baseada em necessidades comunitárias e potencializada pelo engajamento entre cientistas profissionais e cidadãos. Bonney, Cooper e Ballard (2016) vão além, afirmando que a ciência cidadã abrange projetos nos quais os voluntários participam de funções que vão além da coleta e análise de dados, em equipe ou isoladamente, com ou sem a colaboração de cientistas. Esses projetos podem enfatizar questões levantadas pela própria comunidade, e não apenas por cientistas.

[...] tomadas de decisões melhores e mais rápidas para uma ampla gama de indústrias e aplicações. A tomada de decisão automatizada pode reduzir as complexidades para os cidadãos e permitir que empresas e governos prestem serviços em tempo real e ofereçam suporte para tudo, desde interações com o cliente até o preenchimento de documentos tributários e o pagamento de impostos (SCHWAB, 2016, p. 137).

Rautenberg e Carmo (2019, p. 57) compreendem que “salvaguardar volumosas coleções de dados (*Big Data*) distingue-se da produção de informação a partir dessas coleções”. Nessa perspectiva, a ciência de dados é “um conceito conexo à camada dos métodos, na qual os *softwares* são empregados para transformar dados em informação, resultando no apoio à tomada de decisão” (RAUTENBERG, CARMO, 2019, p. 57). Baseados em Bugnion, Manivannan e Nicolas (2017), Rautenberg e Carmo (2019) apresentam sete passos que podem ser executados interativamente em soluções de ciência de dados (Figura 1).

Figura 1 – Ciclo de vida da ciência de dados



Fonte: Rautenberg e Carmo (2019, p. 60)  
Baseado em Bugnion, Manivannan e Nicolas (2017).

Nesse contexto, também orientado a propósitos gerenciais de tomada de decisão, o *Business Intelligence* pode ser entendido como uma ferramenta usada pelas organizações para coletar, gerenciar e analisar dados estruturados e não estruturados, utilizando-se das mais modernas tecnologias da informação (LIN et al., 2009), sendo requisito essencial para o processo de incorporação da ciência de dados nos ambientes organizacionais. Segundo Arnott, Lizama e Song (2017), *Business Intelligence* é o ramo de maior investimento em tecnologia da informação nas organizações, e, durante muitos anos, ele foi classificado como a principal prioridade dos executivos de tecnologia da informação em todo o mundo.

Na área da saúde, e principalmente no contexto das organizações hospitalares, no qual existe uma grande necessidade de interconectar os conhecimentos produzidos, destaca-se a importância e a necessidade de iniciativas que possibilitem o uso, o tratamento e a visualização dos dados produzidos e acumulados, de modo a potencializar o processo decisório e, conseqüentemente, proporcionar melhorias nos sistemas de saúde por meio da adoção de políticas públicas. Dessa forma, é contínua a necessidade de avanços tecnológicos que permitam a integração, a interoperabilidade, a comunicação e a segurança dos dados existentes nos Sistemas de Informação em Saúde (SIS), de modo a garantir o processamento, a exploração e a análise dos dados produzidos nesses ambientes (GALVÃO; VALENTIN, 2019).

Por outro lado, apesar da variedade de ferramentas associadas ao uso do *big data* e de seus avanços tecnológicos, a área da saúde ainda possui uso pouco significativo dessas ferramentas, principalmente em razão da complexidade dos sistemas analíticos. Além disso, também existe a necessidade de profissionais qualificados para atuar com *big data* nesses locais (GALVÃO; VALENTIM, 2019; JOKONYA, 2014). Em paralelo, há um crescente processo de digitalização da área da saúde, no qual os dados até então produzidos convencionalmente passaram a integrar os sistemas informacionais e plataformas de bancos de dados. Ziviani (2019, p. 17-18), elenca alguns dos desafios técnicos da ciência de dados aplicada à saúde digital, a saber:

(i) a heterogeneidade dos dados, havendo diferentes formatos de dados e níveis de acurácia; (ii) a fragmentação dos dados, havendo múltiplas bases de dados, além de diferentes controladores e tomadores de decisão dessas bases de dados, o que dificulta seu uso integrado e transparente; (iii) a disponibilidade de dados, uma vez que devido à sensibilidade de dados de saúde há diversas políticas de proteção dos dados por razões comerciais ou culturais, ou ainda relacionadas à privacidade pessoal, o que leva à necessidade de regulamentação e clareza dos usos éticos e autorizados desses dados; (iv) o processamento de dados, que envolve desafios específicos na capacitação e viabilização de formas eficientes de realizar o gerenciamento dos dados, dar acesso aos dados, lidar com diferentes níveis de qualidade de dados, a consulta aos dados e o compartilhamento de dados; (v) privacidade e integridade, bem como uso ético dos dados, gerando desafios para a prevenção de corrupção dos dados, o impedimento de vazamento de dados dada a sensibilidade dos mesmos e também a inibição de mal uso dos dados de forma antiética; (vi) conceituação dos dados, integração e interoperabilidade, incitando a necessidade de criação de ontologias e outras formas de construção de bases de conhecimento, além da viabilização de aspectos ligados à interoperabilidade semântica entre as diferentes bases de dados envolvidas.

Existem diversos obstáculos a serem superados na utilização da ciência de dados na área de saúde, mas é possível apontar que eles vêm sendo solucionados nos mais diversos níveis. Entretanto, o maior desafio é fazer com que toda essa transformação digital gere resultados positivos, não somente profissionalizando a gestão e operação do sistema, mas melhorando as condições de atuação dos profissionais de saúde, refletindo positivamente nos serviços e no atendimento (ZIVIANI, 2019).

### 3 METODOLOGIA

Com o intuito de cumprir o objetivo proposto – de realizar uma análise bibliométrica sobre as aplicações da ciência de dados no âmbito das organizações hospitalares – buscou-se evidenciar, por meio de pesquisa bibliográfica, a existência de termos relacionados à ciência de dados, tais como, *big data*, *data analytics*, *business intelligence*, *data mining*, *data warehouse*, *text mining* e *data science*, associados a hospitais, com base em informações presentes nos títulos, resumos e palavras-chave das publicações científicas. Dessa forma, para o desenvolvimento desta pesquisa, cumpriram-se, respectivamente, as seguintes etapas metodológicas:

- a) Recuperação de informações científicas na base de dados WoS: para tanto, utilizou-se a expressão: TÓPICO: (((*"hospital"* OR *"hospitals"*) AND (*"big data"* OR *"data analytics"* OR *"business intelligence"* OR *"data mining"* OR *"data warehouse"* OR *"text mining"* OR *"data science"*))). Os termos foram definidos conforme as principais tendências presentes na literatura científica no que tange ao escopo da ciência de dados, relacionando-os com os hospitais, objetos de interesse deste estudo. Sendo esta uma pesquisa exploratória, pretende-se, mais adiante, avançar na identificação de novos termos pertencentes ao domínio analisado com o objetivo de aperfeiçoar a expressão de busca para trabalhos futuros. O período determinado para a busca foi de 2015 a 2019, visando obter as publicações mais recentes frente à data da consulta. Com isso, encontraram-se 1.291 registros, que compuseram o *corpus* definitivo do estudo. Ressalta-se que o ano 2019 foi considerado até o início de setembro, haja vista que este foi o mês da coleta de dados e realização da pesquisa. Todo o acesso às plataformas foi realizado via Comunidade Acadêmica Federada (CAFe) do Portal de Periódicos Capes.

- b) **Processamento dos dados:** os metadados dos trabalhos e artigos encontrados foram compilados e processados na ferramenta de mineração de dados e textos The Vantage Point®, aplicando-se instrumentos de limpeza com o propósito de lidar com as possibilidades de dispersão causadas por questões linguísticas comuns aos estudos bibliométricos. Posteriormente, foram elaboradas matrizes, tendo como cerne as palavras-chave, porém, associando-as às áreas de pesquisa definidas pela WoS e às afiliações institucionais dos autores. As fontes também foram levantadas para fins de elaboração de um *ranking* sobre os veículos de publicação. Durante a padronização dos registros, agregaram-se os conteúdos conforme as similaridades sintáticas encontradas, desta feita, uniformizaram-se as palavras-chave, áreas de pesquisa, instituições e fontes.
- c) **Análise dos dados:** para a análise, optou-se pela técnica de Análise de Redes Sociais (ARS), com amparo da teoria dos grafos, fazendo-se uso da ferramenta VOSviewer (VANECK; WALTMAN, 2010), baseando-se nas matrizes elaboradas na etapa anterior. Nos grafos, foram consideradas as relações, sua intensidade, o quantitativo de produções e clusterização, variáveis expressadas nos nós, vínculos e cores. Com o propósito de proporcionar uma visualização legível, alguns filtros foram aplicados com base no grau de cada nó.

## 4 RESULTADOS E DISCUSSÃO

As palavras-chave de maior representatividade nos resultados, por ocorrência, foram: *data mining* (230), *big data* (166), *machine learning* (75), *electronic health records* (59), *healthcare* (44), *text mining* (40), *classification* (39), *decision tree* (30), *electronic medical records* (28), *data analytics* (27), *hospital* (26), *natural language processing* (25), *predictive modeling* (24), *business intelligence* (22), *cloud computing* (22), *data warehouse* (22), *mortality* (22), *artificial intelligence* (20) e *heart failure* (20). Dentre as dez primeiras colocadas, não surpreende o posicionamento de *data mining*, *big data*, *text mining* e *data analytics*, haja vista a presença delas na expressão de busca. Por outro lado, *machine learning*, *electronic health records*, *classification*, *decision tree* e *electronic medical records* despontaram como assuntos relevantes no âmbito da aplicação da ciência de dados em hospitais, podendo ser acrescentadas à busca em estudos futuros. Desautels e demais autores (2017) afirmam que “aprendizado de máquina”, tradução livre de *machine learning*, refere-se a uma coleção de técnicas algorítmicas para representação e análise de dados com

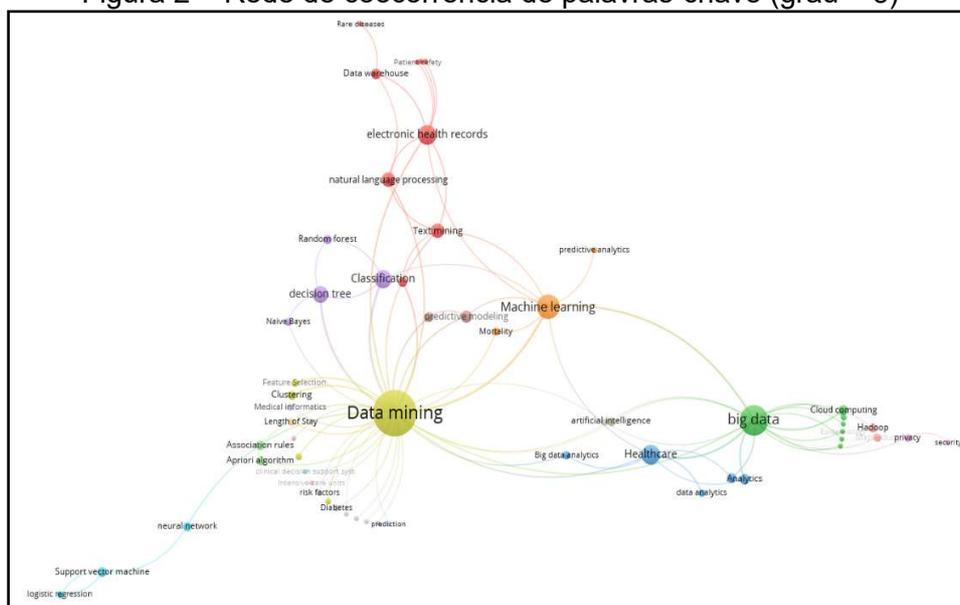
aplicações em muitos domínios, sendo capaz de lidar com a não linearidade, alta dimensionalidade e heterogeneidade de dados complexos. No estudo que realizaram, identificaram o uso dessa técnica no contexto hospitalar para a detecção de sepse, mortalidade e previsão de estabilidade de pacientes (DESAUTELS et al., 2017).

Em relação ao desempenho de *electronic health records* – registros eletrônicos em saúde – e *electronic medical records* – registros médicos eletrônicos –, percebe-se um grande destaque, tendo em vista que, em caso de agregação semântica, poderiam, a depender do rigor empregado, constarem como expressão única. Na literatura, tanto a primeira como a segunda podem aparecer no sentido de “prontuário eletrônico”, tendo aplicações na construção de grandes bancos de dados de pacientes (NISSEN et al., 2017) e no apoio a julgamentos mais precisos na tomada de decisão médica (HUI et al., 2020).

No que concerne à presença relevante da palavra “*healthcare*”, entende-se que, em buscas futuras, ela pode ser agregada a “hospitais”, tendo em vista que seu sentido é pertinente aos cuidados em saúde prestados à população, seja em sistemas ou serviços de saúde.

A Figura 2 apresenta a rede de coocorrência de palavras-chave (grau > 5). Os nós foram dispostos no grafo pela centralidade de grau, considerando-se a quantidade de vértices adjacentes (BARABÁSI, 2016), sendo os tamanhos dos nós proporcionais à ocorrência, igualmente às outras figuras que serão apresentadas nesta seção.

Figura 2 – Rede de coocorrência de palavras-chave (grau > 5)



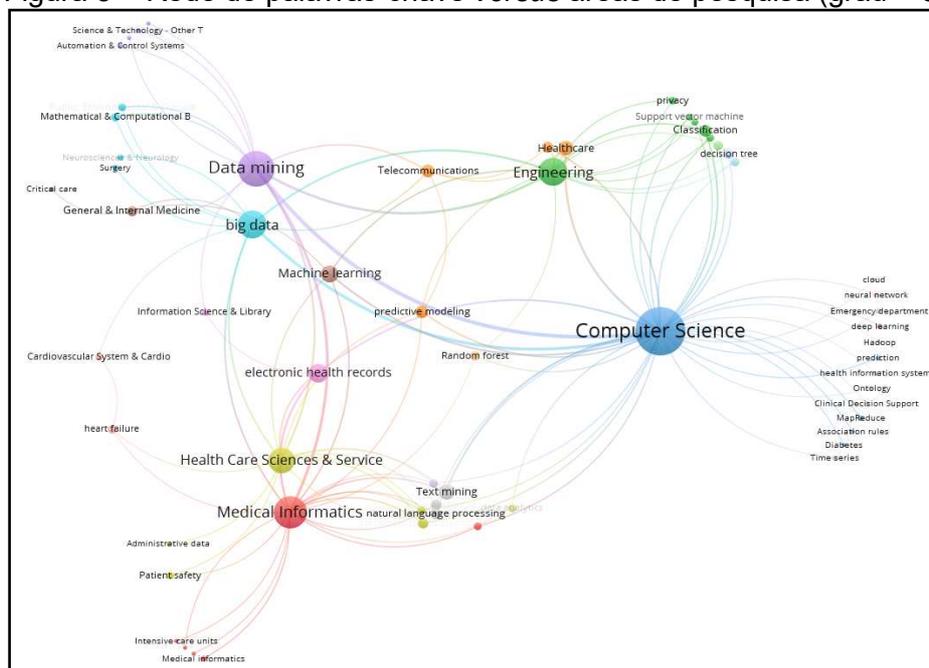
Fonte: dados da pesquisa (2020).

Os pares de vértices (Figura 2) que apresentam maior grau de coocorrência, respectivamente, são “*data mining & machine learning*” (24), “*data mining & classification*”

(23), “data mining & decision tree” (19), “big data & healthcare” (15), “big data & machine learning” (14), “big data & cloud computing” (12), “electronic health records & data mining” (11), “data mining & clustering” (10), “big data & analytics” (10), “data mining & predictive modeling” (9), “data mining & natural language processing” (9), “data mining & electronic medical records” (9), “data mining & healthcare” (8), “data mining & hospital” (8). As associações de termos mais evidentes relacionam-se à discussão trazida por Suchting e demais autores (2018), quando afirmaram que “ciência de dados” é um termo amplo para metodologias pautadas em algoritmos de aprendizado de máquina com o propósito de proporcionar previsões amparadas em conjuntos de dados grandes, complexos e com muitas variáveis. Essas ferramentas auxiliam na solução de problemas de *big data* em uma vasta gama de fenômenos, com aplicações promissoras e multiformes no campo hospitalar, haja vista que, no estudo citado, os autores exploraram a ciência de dados para prever eventos agressivos em um hospital psiquiátrico (SUCHTING et al., 2018).

Na Figura 3, encontram-se as áreas de pesquisa associadas às palavras-chave. O esquema da WoS compreende, aproximadamente, 250 áreas de conhecimento<sup>2</sup> em ciências, ciências sociais, artes e humanidades.

Figura 3 – Rede de palavras-chave versus áreas de pesquisa (grau > 3)



Fonte: dados da pesquisa (2020).

As áreas de pesquisa mais proeminentes na Figura 3 são as seguintes: *computer science* (464), *engineering* (227), *medical informatics* (220), *health care sciences and*

<sup>2</sup> Disponível em: <https://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm>. Acesso em: 20 nov. 2020.

*services* (195), *general and internal medicine* (84), *telecommunications* (69), *public, environmental and occupational health* (59), *mathematical and computational biology* (48), *pharmacology and pharmacy* (42), *science and technology - other topics* (37), *surgery* (37), *information science and library science* (31), *operations research and management science* (29), *cardiovascular system and cardiology* (27), *neurosciences and neurology* (27), *nursing* (25), *business and economics* (24), *radiology, nuclear medicine and medical imaging* (24). Considerando que a classificação da WoS se baseia no perfil temático dos veículos indexados em que os artigos são publicados, nota-se a predominância esperada da área de saúde e dos campos pertencentes ou fronteiriços à tecnologia. No mais, vê-se que a grande variedade de áreas acusa a natureza multidisciplinar do assunto em questão, constatando, inclusive, uma importante participação da CI.

Conforme apresentado na Figura 3, foram identificados 28 *clusters* de associações entre palavras-chave e áreas de pesquisa, sendo os mais representativos: *medical Informatics* (área) e “*decision support systems*”, “*health informatics*”, “*hospital information systems*”, “*intensive care units*”, “*medical informatics*” (termos); *engineering* (área) e “*big data analytics*”, “*classification*”, “*feature selection*”, “*privacy*”, “*support vector machine*” (termos); *computer science* (área) & “*association rules*”, “*diabetes*”, “*mapreduce*”, “*prediction*”; *health care sciences and service* (área) & “*administrative data*”, “*clinical decision support systems*”, “*electronic medical records*”, “*patient safety*”; “*data mining*” (termo) & *materials science, mathematics, science and technology – others, automation and control systems* (áreas); “*big data*” (termo) & *mathematical and computational biology, neurosciences and neurology, public, environmental and occupational health, surgery* (áreas); *telecommunications* (área) & “*healthcare*”, “*internet of things*”, “*predictive modeling*” (termos).

Assim, observa-se que as interações mais expressivas refletem a inclinação da informática médica a assuntos relacionados à tomada de decisão, sistemas de informação para hospitais e unidade de cuidados intensivos. Gómez-Vallejo e demais autores (2016), ao discutirem esse espectro temático, afirmam que a alta disponibilidade de registros eletrônicos em saúde tem estimulado o desenvolvimento de sistemas de tomada de decisão no campo hospitalar, tendo como argumento para o seu desenvolvimento a possibilidade de reduzir custos e a implementação de ações de vigilância.

Sobre a CI, destaca-se sua associação aos *electronic health records* (Figura 3), o que revela a importância de processos organizacionais relacionados à gestão de documentos e à gestão da informação no âmbito dos sistemas e serviços de saúde. Ao



*Univ Paris V, Univ Wurzburg* (instituições); “*data mining*” (termo) e *Florida Polytech Univ, Mahidol Univ, Natl Changhua Univ, Natl Chi Nan Univ, Natl Chung Cheng Univ, Natl Taiwan Univ, Providence Univ* (instituições); “*electronic health records*” (termo) e *1265 Welch Rd, Hop Europeen Georges Pompidou, Kings Coll London, Univ Florida* (instituições); “*big data*” e “*cloud computing*” (termos) & *Cent S Univ, Univ Maryland, Univ Messina* (instituições); “*machine learning*” (termo) & *MIT, Univ Pittsburgh, Vanderblit Univ* (instituições); “*clinical data warehouse*” (termo) e *CHU Rennes, Inserm, Univ Rennes* (instituições).

Enquanto mérito dessas instituições, destaca-se a criação de um contexto interdisciplinar para o estudo da saúde relacionada à tecnologia. Essa conexão disciplinar tem um alto potencial de crescimento nos próximos anos, devido à interdependência dos campos. Apesar de Estados Unidos (384), China (181) e Índia (82) serem os países mais representativos no cômputo geral, nota-se o destaque das academias francesas de saúde no tocante à especialização temática. Em suma, percebe-se que o conhecimento produzido no domínio possui um razoável grau de descentralização, sendo representativo na América do Norte, na Europa e na Ásia.

Quanto às fontes, que refletem os veículos de publicação, destacaram-se: *Studies in Health Technology and Informatics* (59), *BMC Medical Informatics and Decision Making* (21), *International Journal of Medical Informatics* (21), *Book Series: Lecture Notes in Computer Science (LNCS)* (21), *Plos One* (19), *Book Series: Procedia Computer Science* (17), *IEEE Access* (16), *Journal of Medical Systems* (15), *Journal of Biomedical Informatics* (13), *Journal of the American Medical Informatics Association* (11).

Segundo dados da IOS Press (2020), a série *Studies in Health Technology and Informatics* foi iniciada em 1990, em colaboração com programas da União Europeia para promover a pesquisa em informática em saúde e biomédica. Ela se tornou uma plataforma global altamente visível para a disseminação de pesquisas originais nesse campo, contendo mais de 250 volumes de trabalhos de alta qualidade de todo o mundo (IOS PRESS, 2020). Explica-se sua alta representatividade em número de publicações pela sua natureza, enquanto uma compilação de textos oriundos de vários veículos.

Já os dois periódicos mais representativos – *BMC Medical Informatics and Decision Making* e *International Journal of Medical Informatics* –, integram, respectivamente, os grupos Springer Nature e Elsevier. Adiante, nota-se a série de livros *LNCS*, publicada desde 1973 e pertencente à Springer Nature. No mais, há uma diversificada lista de editoras, tais como, PLOS, IEEE – organização profissional técnica, sem fins lucrativos, dedicada ao avanço da tecnologia –, Oxford University Press, além da repetição de veículos da Springer

Nature e Elsevier. Tal comportamento evidencia que o escoamento da produção científica, nesse campo, possui uma significativa diversidade, porém, sendo dominada pelos grandes *players* do mercado editorial científico, detentores dos direitos sobre a produção científica no assunto.

### 3 CONSIDERAÇÕES FINAIS

O presente artigo objetivou realizar uma análise bibliométrica sobre as aplicações da ciência de dados, no âmbito dos hospitais, na literatura recente. Com isso, verificou-se a proeminência do “*data mining*” e “*big data*” como os principais assuntos, porém, chamando bastante a atenção o posicionamento de “*machine learning*” e “*electronic health records*”, em especial por não serem termos adotados na expressão de busca, o que diminuía suas possibilidades de recuperação e representação nos resultados.

Em relação ao termo “*electronic health records*” (documentos eletrônicos em saúde), destaca-se a sua associação com a CI, demonstrando a importância e o potencial desse domínio do conhecimento nos estudos sobre as organizações produtoras de serviços de saúde, principalmente por meio de processos organizacionais associados à gestão de documentos e à gestão da informação, já inseridos no paradigma contemporâneo da ciência de dados.

A partir desta pesquisa exploratória, foi possível identificar palavras-chave que podem possibilitar, em trabalhos futuros, a realização de um estudo mais aprofundado acerca das tecnologias e métodos informacionais com potencial de utilização nos sistemas e serviços de saúde. Ademais, o artigo evidenciou a multidisciplinaridade existente em torno do assunto analisado, passando por áreas como a ciência da computação, engenharia, telecomunicações, saúde pública, matemática, CI, dentre outras. Por fim, ressalta-se a necessidade de realização de estudos mais aprofundados acerca da ciência de dados em organizações hospitalares, expandido o escopo temático para *healthcare*.

### REFERÊNCIAS

AMARAL, F. **Introdução à Ciência de Dados**: mineração de dados e big data. Alta books: Rio de Janeiro, 2016.

ARNOTT, D.; LIZAMA, F.; SONG, Y. Patterns of business intelligence systems use in organizations. **Decision Support Systems**, [s. l.], v. 97, p. 58-68, maio 2017.

BARABÁSI, A. L. **Network science**. Cambridge: University Press, 2016.



BHAVNANI, S. P.; MUÑOZ, D.; BAGAI, A. Data Science in Healthcare. **Circulation: Cardiovascular Quality and Outcomes**, Dallas, TX, v. 9, n. 6, p. 683-687, nov. 2016. Ovid Technologies (Wolters Kluwer Health). Disponível em: <https://www.ahajournals.org/doi/epub/10.1161/CIRCOUTCOMES.116.003081>. Acesso em: 01 nov. 2020.

BONNEY, R.; COOPER, C.; BALLARD, H. The theory and practice of citizen science: Launching a new journal. **Citizen Science: Theory and Practice**, [s. l.], v. 1, n. 1, 2016. Disponível em: <https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.65/>. Acesso em: 20 nov. 2020.

BUGNION, P.; MANIVANNAN, A.; NICOLAS, P. R. **Scala: Guide for Data Science Professionals**. Birmingham: Packt Publishing, 2017.

CAO, L. Data science and analytics: a new era. **International Journal of Data Science and Analytics**, [s. l.], v. 1, p. 1-2, 2016. Disponível em: <https://link.springer.com/content/pdf/10.1007/s41060-016-0006-1.pdf>. Acesso em: 20 nov. 2020.

DALIANIS, H. *et al.* Health bank-A - Workbench for data science applications in healthcare. *In: CAiSE 2015 INDUSTRY TRACK*, 27., 2015, Estocolmo, Suécia. **Proceedings** [...]. Estocolmo, Suécia: CAiSE, 2015. p. 1-18.

DESAUTELS, T. *et al.* Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. **BMJ open**, [s. l.], v. 7, n. 9, 2017. Disponível em: <https://bmjopen.bmj.com/content/bmjopen/7/9/e017199.full.pdf>. Acesso em: 20 nov. 2020.

ELLIOTT, K. C.; ROSENBERG, J. Philosophical foundations for citizen science. **Citizen Science: Theory and Practice**, [s. l.], v. 4, n. 1, 2019. Disponível em: <https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.155/print/>. Acesso em: 20 nov. 2020.

FORD, E. *et al.* Extracting information from the text of electronic medical records to improve case detection: a systematic review. **Journal of the American Medical Informatics Association**, [Oxford], v. 23, n. 5, p. 1007-1015, 2016.

GALVÃO, A. B.; VALENTIM, R. A. M. Desafios para os avanços da análise de big data na Saúde. *In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE*, 19., 2019, Niterói. **Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde**. Niterói: SBC, 2019. p. 155-160.

GÓMEZ-VALLEJO, H. J. *et al.* A case-based reasoning system for aiding detection and classification of nosocomial infections. **Decision Support Systems**, [s. l.], v. 84, p. 104-116, 2016.

HUI, C. *et al.* Research on diabetes prediction method based on electronic medical record data analysis. *In: INTERNATIONAL CONFERENCE ON ENERGY, ENVIRONMENT AND*

BIOENGINEERING, 1., 2020, XI'An. **Proceedings** [...]: XI'An, 2020. p. 1-4. Disponível em: [https://www.e3sconferences.org/articles/e3sconf/pdf/2020/45/e3sconf\\_iceeb2020\\_03001.pdf](https://www.e3sconferences.org/articles/e3sconf/pdf/2020/45/e3sconf_iceeb2020_03001.pdf). Acesso em: 01 nov. 2020.

HWANG, U. *et al.* Disease prediction from electronic health records using generative adversarial networks. **arXiv preprint**, 2017. Disponível em: <https://arxiv.org/pdf/1711.04126.pdf>. Acesso em: 01 fev. 2019.

IOS PRESS. **Studies in Health Technology and Informatics**. 2020. Disponível em: <https://www.iospress.nl/bookserie/studies-in-health-technology-and-informatics/#scope>. Acesso em: 10 dez. 2020.

JOKONYA, O. Towards a Big Data Framework for the Prevention and Control of HIV/AIDS, TB and Silicosis in the Mining Industry. **Procedia Technology**, [s. l.], v. 16, p. 1533-1541, 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2212017314004022>. Acesso em: 20 nov. 2020.

KOHLI, R.; TAN, S. S. L. Electronic health records: how can IS researchers contribute to transforming healthcare?. **Mis Quarterly**, [s. l.], v. 40, n. 3, p. 553-573, 2016.

LIN, Y. *et al.* Research on using ANP to establish a performance assessment model for business intelligence systems. **Expert Systems with Applications**, [s. l.], v. 36, p. 4135-4146. 2009.

MOSTAFA, S. P.; ROCHA, E. S. S. O rizoma e as voltas que o mundo dá. **Texto Digital**, Florianópolis, v. 16, n. 1, p. 124-140, 2020. Disponível em: <https://periodicos.ufsc.br/index.php/textodigital/article/view/1807-9288.2020v16n1p124>. Acesso em: 20 nov. 2020.

NISSEN, F. *et al.* Validation of asthma recording in electronic health records: a systematic review. **Clinical epidemiology**, [s. l.], v. 9, p. 643-656, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5716672/>. Acesso em: 20 nov. 2020.

RAUTENBERG, S.; CARMO, P. R. V. do. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Science: research trends**, Marília, v. 13, n. 1, p. 56-67, 2019. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/8315>. Acesso em: 20 nov. 2020.

ROMANOWSKI, J. P.; ENS, R. T. As pesquisas denominadas do tipo “estado da arte” em educação. **Revista Diálogo Educacional**, v. 6, n. 19, p. 37-50, set./dez. 2006.

SCHWAB, K. **A Quarta Revolução Industrial**. São Paulo: Edipro, 2016.

SUCHTING, R. *et al.* A data science approach to predicting patient aggressive events in a psychiatric hospital. **Psychiatry research**, [Limerick], v. 268, p. 217-222, 2018.

VAN ECK, N.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, [Amsterdam], v. 84, n. 2, p. 523-538, 2010.

ZIVIANI, A. Desafios de ciência de dados aplicada à saúde digital. In: BRASIL. Ministério da Saúde. **Reflexões sobre ensino e pesquisa no SUS: experiência no contexto hospitalar de alta complexidade**. Brasília: Ministério da Saúde, 2019. p. 14-19. Disponível em: [http://bvsms.saude.gov.br/bvs/publicacoes/reflexoes\\_ensino\\_pesquisa\\_sus.pdf](http://bvsms.saude.gov.br/bvs/publicacoes/reflexoes_ensino_pesquisa_sus.pdf). Acesso em: 20 nov. 2020.

## NOTAS

### CONTRIBUIÇÃO DE AUTORIA

**Concepção e elaboração do manuscrito:** N. V. Sobral, G. L. Q. Lima, A. S. P. M. Sobral

**Coleta de dados:** N. V. Sobral, G. L. Q. Lima, A. S. P. M. Sobral

**Análise de dados:** N. V. Sobral, G. L. Q. Lima, A. S. P. M. Sobral

**Discussão dos resultados:** N. V. Sobral, G. L. Q. Lima, A. S. P. M. Sobral

**Revisão e aprovação:** N. V. Sobral, G. L. Q. Lima, A. S. P. M. Sobral

### CONJUNTO DE DADOS DE PESQUISA

Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no próprio artigo.

### LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

### PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

### EDITORES

Enrique Muriel-Torrado, Raymundo N. Machado, Valdineia Barreto Ferreira, Kátia de Oliveira Rodrigues, Susane Barros e Genilson Geraldo.

### HISTÓRICO

Recebido em: 16-12-2020 – Aprovado em: 16-07-2021 – Publicado em: 08-10-2021.