

APLICABILIDADES E VANTAGENS DO USO DO BIG DATA NO SETOR DE SAÚDE: UM ESTUDO BIBLIOMÉTRICO

Luana Da Silva Alves - UFRRJ

Saulo Barbara De Oliveira - Universidade Federal Rural do Rio de Janeiro

Sandro Luís Freire De Castro Silva - Instituto Nacional de Câncer

Resumo

O crescente uso dos recursos do Big Data nos diversos setores da sociedade, impõe a necessidade de se compreender as vantagens e benefícios desta tecnologia, especialmente no setor de saúde, onde a vida humana é o bem mais valioso da sociedade contemporânea. A preocupação no presente artigo foi fazer a revisão bibliométrica da literatura científica sobre o Big Data na área da saúde, publicada no período de 2018 a 2022, com o fim de melhor compreender a sua aplicabilidade e principais benefícios nesta área. Para isso, foi feito o levantamento da literatura usando o motor de busca Scopus. Os artigos selecionados foram analisados descritivamente quanto à frequência de publicações, as áreas de pesquisa e os principais autores sobre o tema. Os resultados apontaram que na área da saúde, o Big Data se apresenta como um mecanismo estratégico de gerenciamento dos dados. Isso porque, a análise do Big Data permite a interpretação dos dados coletados e, a partir disso, é possível obter insights importantes sobre questões relacionadas a saúde. Identificou-se que os países norte-americanos, europeus e asiáticos são predominantes quanto as publicações sobre o tema, aumentando a importância de pesquisas voltadas para a divulgação do conhecimento em países sul americanos.

Palavras-chave: Gestão estratégica; Big Data na saúde; Tecnologia em saúde.**Abstract**

The growing use of Big Data resources in different sectors of society imposes the need to understand the advantages and benefits of this technology, especially in the health sector, where human life is the most valuable asset of contemporary society. The concern in this article was to carry out a bibliometric review of the scientific literature on Big Data in the health area, published from 2018 to 2022, in order to better understand its applicability and main benefits in this area. For this, a literature survey was carried out using the Scopus search engine. The selected articles were descriptively analyzed regarding the frequency of publications, the research areas and the main authors on the topic. The results showed that in the health area, Big Data presents itself as a strategic data management mechanism. This is because the analysis of Big Data allows the interpretation of the data collected and, from this, it is possible to obtain important insights on health-related issues. It was identified that North American, European and Asian countries are predominant in terms of publications on the subject, increasing the importance of research aimed at disseminating knowledge in South American countries.

Keywords: Strategic management; Big Data in health; Health technology.

APLICABILIDADES E VANTAGENS DO USO DO BIG DATA NO SETOR DE SAÚDE: UM ESTUDO BIBLIOMÉTRICO

1 INTRODUÇÃO

Nos últimos anos, observou-se um aumento da busca de diferentes setores por ferramentas analíticas eficazes para o gerenciamento estratégico de grandes volumes de dados, utilizando os recursos do *Big Data*. Em parte, isso se deve ao fato de as organizações estarem procurando maneiras de usar essa tecnologia para melhorar a tomada de decisão, aumentar a vantagem competitiva e potencializar o desempenho do negócio (FAVI et al, 2020; HAJJAJI et al, 2021).

Essa tendência, também é perceptível nas instituições de saúde, onde houve uma mudança no modelo de gestão, que deixou de ser centrado na doença e passou a ser centrado no paciente (SENTHILKUMAR et al, 2018). A complexidade desse setor aumentou a necessidade das instituições gerenciarem e analisarem suas bases de dados para ter uma visão sistêmica do cenário e, conseqüentemente, tomar decisões orientadas pelas informações extraídas do *Big Data*, para melhorar seus processos e serviços (WANG et al, 2020).

Uma das questões frequentemente levantadas quando se trata do uso de dados no setor de saúde refere-se ao uso adequado do *Big Data*. A introdução de prontuários eletrônicos, o compartilhamento de informações por meio de sensores e equipamentos médicos e a disseminação de informações do paciente em mídias sociais, fizeram com que o fluxo de dados médicos aumentasse exponencialmente (LUPTON; MICHAEL, 2016; CARDONA, 2022).

Além disso, a indústria médica gera quantidades significativas de dados por meio da inclusão de imagens, registros clínicos, dados genômicos e comportamentos de saúde (GUPTA; RANI, 2019). Buscando extrair valor dessas massas de dados, aumenta a tendência de as organizações adotarem as tecnologias desenvolvidas para a análise do *Big Data*, também chamadas *analytics*, isso se deve, principalmente, ao seu potencial para gerar conhecimentos úteis para a área da saúde (RESNYANSKY, 2019; GHASSEMI; MOHAMED, 2022).

Os desafios para implementação do *Big Data* na saúde vão além da dificuldade de análise dos dados estruturados e não estruturados, e incluem a necessidade de um investimento inicial maciço para aquisição dos recursos (tecnológicos e humanos) necessários para viabilizar e potencializar os benefícios do *Big Data* no setor. Além disso, é preciso que as instituições se preocupem com as boas práticas para manipulação dos dados dos pacientes, a fim de garantir a segurança e privacidade na coleta e compartilhamento de dados sensíveis (KHANRA, 2020; GRIVA, 2022).

Do ponto de vista da Gestão de Saúde Pública, os dados sobre saúde são utilizados como subsídio para a gestão do planejamento estratégico e a tomada de decisões (WONG et al, 2022). Um exemplo disso, está na análise de dados geográficos, que permite ao gestor público identificar as regiões com menores taxas de encaminhamento populacional aos serviços básicos de assistência (D'AGOSTINO et al, 2021).

Esses conhecimentos, são utilizados para a implementação de Políticas Públicas e para a promoção de ações de intervenção de saúde populacional, congruentes com as particularidades de cada região, garantindo assim, o direito de acesso a saúde do cidadão de forma igualitária (SAUNDERS et al, 2020). Tendo em vista que o *Big Data* é uma ferramenta

estratégica para a gestão das instituições de saúde, este estudo se propõe a responder a seguinte questão norteadora: **“Quais as aplicabilidades e vantagens do *Big Data* no setor de saúde?”**.

Para tal, um estudo bibliométrico foi planejado e executado com base na conceituação de bibliométrica proposta por Zupic e Carter (2015). O artigo está organizado da seguinte forma: A seção 2 apresenta as cinco etapas do estudo bibliométrico; a seção 3 analisa os resultados bibliométrico; a seção 4 apresenta as reflexões teóricas a respeito do uso do *Big Data* no setor de saúde e a percepção dos autores a respeito dos principais desafios e as projeções futuras para a implementação do *Big Data* na área saúde; A seção 5 mostra como este trabalho analisou o *Big Data* aplicado na área da saúde e seus respectivos benefícios.

2 DEFINIÇÃO DAS FONTES DE INFORMAÇÃO – ESTUDO BIBLIOMÉTRICO

A partir do cenário descrito na introdução, buscou-se na literatura artigos publicados entre 2018 e 2022, que discorressem sobre o uso do *Big Data* na área da saúde pública e privada. A delimitação do período de publicação dos estudos, se deve a expectativa de recuperar o estado da arte sobre o tema nos últimos cinco anos.

De modo a alcançar o objetivo proposto, buscou-se amparar a realização do estudo na metodologia proposta por Zupic e Cater (2015). A Figura 1 ilustra as etapas adotadas neste artigo para elaboração do estudo bibliométrico.

Figura 1 — Metodologia do estudo bibliométrico.



Fonte: Adaptado de Zupic e Cater (2015).

2.1 Questão de pesquisa

O Quadro 2 apresenta a questão de pesquisa (QP) que será respondida, através do estudo bibliométrico sobre o tema de pesquisa.

Quadro 2 — Questão de pesquisa do estudo.

Questão de Pesquisa
(QP). Quais as aplicabilidades e vantagens do <i>Big Data</i> no setor de saúde?

Fonte: Elaboração própria (2022).

A questão de pesquisa aqui posta, não é diretamente respondida por métricas de produção. Ela foi desmembrada em métricas subjacentes, para ser respondida sob diferentes perspectivas, a partir dos resultados encontrados no estudo bibliográfico. Desta forma, espera-se uma maior aproximação com os interesses de possíveis leitores, acerca do assunto abordado.

2.2 Escolha das bases de fontes informacionais

A base de dados *Scopus* foi escolhida por ser considerada uma base multidisciplinar, com um extenso banco de dados internacional, além do fato de oferecer ferramentas inteligentes (*ranking* de periódicos, perfis de autores e frequência de uso de termos científicos), capazes de auxiliar na análise do portfólio final dos estudos levantados. Esses recursos, são um diferencial no momento da pesquisa, uma vez que se busca identificar as tendências de publicações sobre o tema.

2.3 Identificação dos Documentos

Para selecionar estudos capazes de responder a QP, foram considerados artigos publicados em periódicos, na língua inglesa e portuguesa, eliminando-se possíveis itens duplicados. A estratégia de busca abrangeu estudos indexados entre 2018-2022, na base selecionada. O estudo bibliométrico foi realizado em julho de 2022.

2.4 Elaboração da *String* de Busca

Inicialmente, para a elaboração da *string* de busca (SS), optou-se por executar a calibragem inicial na base *Scopus* em inglês, onde foi utilizada a SS (“*big data AND health*”). Porém, esta iteração retornou um elevado número de resultados (8.088 artigos), em sua maioria abordando o uso do *Big Data* no setor de saúde privada. Assim, optou-se por escolher uma cadeia que observasse o *Big Data* e a saúde como algo específico. Desta forma, optou-se por escolher uma cadeia que buscasse recuperar as tendências do uso do *Big Data*, sem excluir o setor de saúde pública, com ênfase nos estudos em tecnologia, organizações e saúde.

Para isso, foi feito o refinamento com base num termo que melhor representasse o propósito da pesquisa, resultando nos seguintes operadores lógicos: (“*big data*”) AND (“*public*”) AND (“*health*”). Em seguida, a SS foi executada nas bibliotecas digitais. O Quadro 3 apresenta a *string* de busca final.

Quadro 3 – Definição da *string* de busca.

<i>String</i> de Busca
(“ <i>big data</i> ”) AND (“ <i>public</i> ”) AND (“ <i>health</i> ”)

2.5 Busca de informações nas bases de dados

Por meio da utilização das palavras-chave (“*big data*”), (“*public*”) e (“*health*”), foram feitas as buscas de artigos, na base de dados *Scopus*, aplicando como critério de refinamento o período de publicação (2018-2022), o tipo do documento (artigo) e a área de conhecimento (tecnologia, organizações e saúde). A base retornou o total de 4.219 documentos. Após o processo de refinamento, obteve-se um resultado de 538 artigos, conforme exposto no Quadro 4.

Quadro 4 — Resultado das buscas na base *Scopus*.

Data da Busca	Filtro 1	Result	Filtro 2	Result.	Filtro 3	Result. Final
Jul/2022	Tipo de acesso: Aberto Período 2018-2022	2.425	Tipo de doc: Artigo	1.717	Categorias (Scopus): Computer Science; Medicine; Business, Management and Accounting.	538

2.6 Critérios de seleção, inclusão e exclusão

Seguindo as recomendações de Kitchenham e Charters (2007) e Petersen et al. (2015), para o processo de seleção de artigos foram definidos os critérios de inclusão (CI) e exclusão (CE) descritos no Quadro 5, com o objetivo de responder à questão de pesquisa com base em estudos relevantes e dar confiabilidade ao estudo.

Quadro 5 - Critérios de inclusão e exclusão

ID	Critério de Inclusão
CI	O estudo aborda o uso do <i>Big Data</i> no setor de saúde pública e privada.
ID	Critérios de Exclusão
CE1	O artigo é uma revisão
CE2	O artigo trata o problema de forma lateral.
CE3	O artigo não trata o <i>Big Data</i> na área da saúde pública.

Após definir os critérios de seleção dos estudos, iniciou-se o processo de seleção dos artigos. Esta etapa foi dividida em quatro fases: (1) leitura de títulos, resumos e palavras-chave; (2) leitura da introdução e conclusão; (3) leitura completa; e (4) aplicação do CI e CE.

A etapa 1, teve como objetivo identificar os artigos duplicados. Para isso, foi feita a leitura dos títulos de todos os 685 artigos levantados no processo de busca, o número foi reduzido para 537 artigos diferentes entre si. A partir da remoção dos arquivos duplicados, foi aplicada uma nova filtragem dos trabalhos por meio da leitura do resumo e das palavras-chave, resultando em 389 artigos. Na segunda etapa, foi feita a leitura da introdução e conclusão dos artigos, resultando em 93 estudos. Na terceira etapa, foi realizada a leitura completa dos 93 artigos resultantes da etapa 2, que após a aplicação do CI e CE obteve-se um portfólio final de 34 estudos, que serviram de base para a discussão sobre o tema proposto.

3 ANÁLISES DOS DADOS

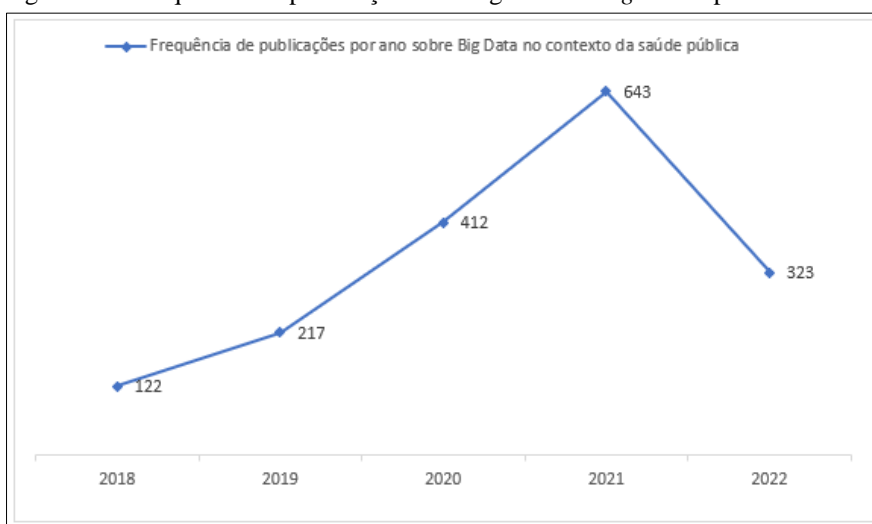
As unidades de análise adotadas para a o cálculo das métricas bibliométricas foram os documentos resultantes do processo de busca, na base de dados *Scopus* (1.717 artigos) e seus conjuntos de palavras-chave. Foram considerados apenas artigos publicados em periódicos, entre 2018 e 2022, sem aplicar filtros de categorias. Os metadados dos registros foram coletados a partir da biblioteca digital e baixados em lotes de 500 registros por vez.

Estes metadados foram importados para a ferramenta *VOSviewer*, um *software* que permite a criação de uma rede com palavras-chave e autores, apontando tendências e conexões em uma determinada área de pesquisa (VAN ECK; LUDO, 2010). Para identificar a frequência de publicações anuais, os países com maior número de publicações sobre o tema e os principais autores de acordo com fonte de dados pesquisada, foi utilizada a ferramenta de análise de dados da biblioteca digital *Scopus*.

3.1 Análises dos resultados do estudo bibliométrico

Para compreender como se deu a evolução do *Big Data* na área da saúde, foi realizado o levantamento da frequência de publicações de artigos sobre o tema, entre os anos de 2018 e 2022. Os registros de publicações foram agregados por ano e exportados para o Excel para gerar a Figura 2.

Figura 2 — Frequência de publicações de artigos sobre *Big Data* aplicado a saúde (2018-2022)



Fonte: Elaboração própria, com base nos dados da biblioteca digital *Scopus* (2022)

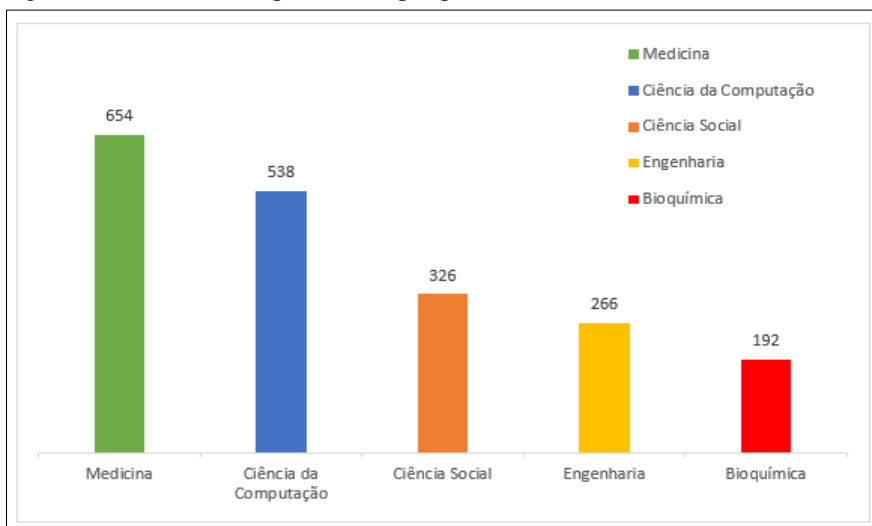
Os resultados mostram que, entre os anos de 2018 e 2022, houve um crescimento de 427,04% do número de publicações sobre o tema. O ano com maior número de publicações foi 2021 (643 artigos), e o ano com menor número de publicações foi 2018 (122 artigos). Proporcionalmente, considerando os anos de 2020 e 2021, houve um crescimento absoluto de 56,06% do número de publicações e, conseqüentemente, um crescimento da área de pesquisa como um todo.

É interessante observar que, os artigos publicados entre janeiro e julho de 2022, ou seja, num período de sete meses, quando comparado aos registros de publicações do ano de 2018 (num período de 12 meses), apresentaram um crescimento relativo de 164,75% do número de publicações, evidenciando que a área de estudo se encontra em ascensão.

Destaca-se que, a quantidade dos artigos referentes a 2022 não representa a totalidade, pois a busca foi realizada no mês de julho, logo, a redução do número de artigos publicados no ano 2022 comparado com 2021, não indica uma redução no interesse de estudo no tema. As cinco principais áreas de pesquisa sobre o tema são a área Médica, de Ciência da Computação, Ciências Sociais, Engenharia e Bioquímica, e estão representadas na Figura 3.

Segundo Van Eck & Waltman (2010), por meio da análise de redes de coautoria é possível identificar como pesquisadores, instituições de pesquisa ou países se relacionam de acordo com a quantidade de estudos que realizam e publicam em cooperação. Com isso, a partir do *software VOSviewer*, foi realizada a análise de cooperação entre 945 autores, dos artigos selecionados para análise bibliométrica, onde cada nó representa um autor, e os tamanhos dos nós indicam o número de artigos publicados, conforme mostra a Figura 3.

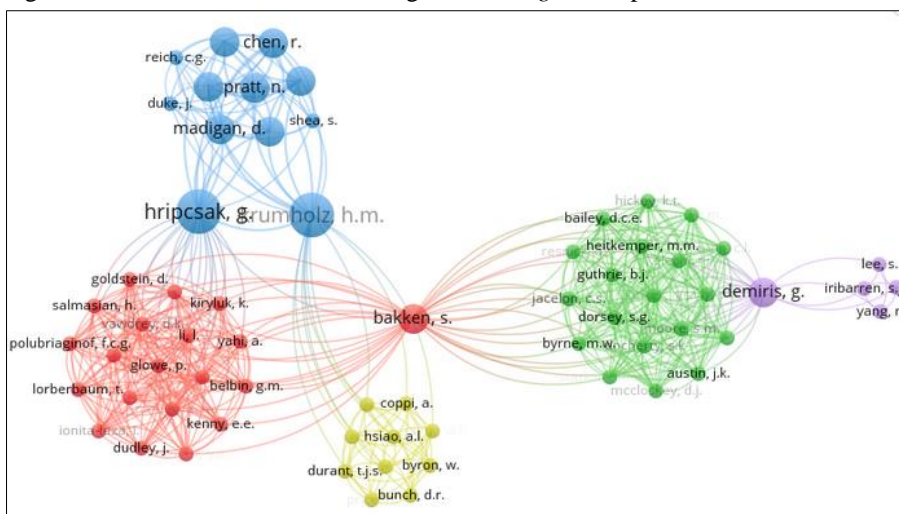
Figura 3 — Documentos por área de pesquisa



Fonte: Elaboração própria, com base nos dados da biblioteca digital Scopus (2022).

O link que conecta dois nós, representa a relação de cooperação entre dois ou mais autores, e a espessura do *link* indica a intensidade da cooperação. Assim, para mapear as principais redes de coautoria, foram considerados apenas os autores com pelo menos duas citações na amostra. Tal restrição, produziu uma rede com sessenta e seis nós (autores), distribuídos em cinco *clusters* de colaboração em pesquisa (Figura 4).

Figura 4 — Rede de coautoria em artigos sobre *Big Data* aplicado a saúde, entre os autores.



Fonte: Elaboração própria, utilizando o *software VOSviewer* versão 1.6.18 (2022).

O Quadro 5, apresenta os autores apontados como mais relevantes pela rede de coautoria, gerada no *VOSviewer*, e o total de artigos publicados e de citações de cada autor, na base de dados *Scopus*.

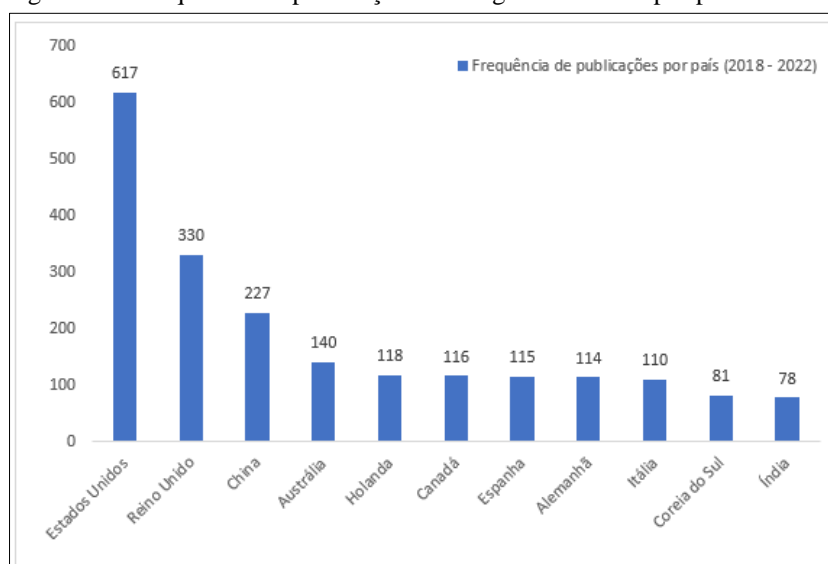
Quadro 5 — Principais autores da rede de coautoria.

Autor (a)	Total de artigos	Total de citações (Scopus)
CHEUNG, Sarah	342	128.236
HOSSAIN, Sakir	149	17.021
HRIPCSAK, George	355	14.098
CHRISTENSEN, H. Jens	278	8.825
GUPTA, Deepak	271	3.923

Fonte: Base de dados Scopus e *software* VOSviewer versão 1.6.18 (2022).

De acordo com as métricas bibliométricas usadas neste estudo, considerando os 10 países com maior número de artigos científicos publicados sobre o tema (Figura 5), na base de dados *Scopus*, percebe-se que aproximadamente 46% foram publicados no continente europeu (787 artigos), seguido do continente norte americano com 733 artigos (43%), o continente asiático com 386 artigos (23%), e o continente oceânico, com 140 artigos (7%). No *ranking*, o Brasil ocupa a décima sétima posição, com 50 artigos publicados.

Figura 5 — Frequência de publicações de artigos científicos por país.



Fonte: Dados da *Scopus* (2022).

Vale destacar que, os países sul americanos não estão entre os dez que mais publicam artigos científicos sobre o tema, e que autores brasileiros não foram citados na rede de colaboração entre os pesquisadores da área, o que é preocupante para a produção científica sobre o tema, tendo em vista que, pesquisas sobre o uso de *Big Data* na área da saúde são cada vez mais relevantes, devido ao aumento da demanda global pela utilização de técnicas e

ferramentas de análise de dados, a fim de gerar informação úteis e identificar *insights* que possam ser utilizados para auxiliar a tomada de decisão dos profissionais da área (SANG, 2021).

4 DISCUSSÕES

Esta seção apresenta reflexões teóricas a respeito das definições e do uso do *Big Data* na saúde e a percepção dos autores a respeito dos principais desafios para a sua implementação e as projeções futuras para o uso do *Big Data* na área da saúde.

4.1 Conceitos de *Big Data*

Na literatura, pode-se encontrar diversas definições de *Big Data*. Este conceito evoluiu nos últimos anos, no entanto, ainda não é claramente compreendido e não há uma unanimidade quanto a sua definição. Dada a abrangência do termo, o *Big Data* pode ser tratado como: uma grande quantidade de dados digitais, um grande repositório de dados, uma ferramenta, tecnologia ou um fenômeno (cultural ou tecnológico) (GUPTA et al, 2019; ZHIHAN, 2020; ZHOU et al, 2020; HRIPCSAK, 2021; CARDONA et al, 2022; YUNITTA et al, 2022).

Na visão de Lupton e Mike (2016), o *Big Data* pode ser considerado como conjuntos massivos de dados digitais, gerados continuamente e que são produzidos por meio das interações de tecnologias online. De acordo com esta ideia, Gupta et al (2019), definem o *Big Data* como uma volumosa base de dados, que representa um desafio para as técnicas tradicionais de armazenamento e análise de dados.

Uma opinião semelhante foi apresentada por Ohlhorst (2012), autor seminal sobre o tema pesquisado e que percebeu o *Big Data* como um conjunto de dados extremamente grande, impossível de gerenciar ou analisar com ferramentas tradicionais de processamento de dados. Na sua opinião, quanto maior o conjunto de dados, mais difícil é obter algum valor a partir dele, daí a importância do *Big Data* para o tratamento e análise dos dados.

Por sua vez, Knapp (2018) percebe o *Big Data* como ferramentas, processos e procedimentos que permitem a uma organização criar, manipular e gerenciar volumosos conjuntos de dados e bases de armazenamento. Sob este ponto de vista, o *Big Data* é identificado como uma ferramenta capaz de reunir diferentes tipos e fontes de dados, que permitem aos usuários gerenciar grandes quantidades de informações. Percepção semelhante é mostrada por Daniel Power (2014), que se referiu as tecnologias de *Big Data* como parte de uma geração de tecnologias e arquiteturas, projetadas para extrair valor de um conjunto de dados, em alta velocidade.

Combinando essas duas abordagens, Cheung (2020) define o *Big Data* como um sistema complexo, onde o gerenciamento destes dados é dependente de *hardwares* e *softwares* robustos, sendo fundamental que a organização conte com grupo de profissionais capacitados e com expertise para extrair informações relevantes e de fácil compreensão.

Embora existam diversas definições, o conceito estabelecido por Beyer e Laney (2012) que se concentra em listar as características do *Big Data* com base no modelo "3 V's": Volume, Velocidade e Variedade, é provavelmente a mais popular dentre elas, podendo ser observada em parte significativa dos estudos voltados para esta área de conhecimento (ZHIHAN, 2020; LIU et al, 2021). Posteriormente, outros autores estenderam o modelo dos "3 V's" e, como resultado, uma gama de recursos foram adicionados à lista, sendo: variabilidade, veracidade,

disponibilidade e valor, sendo estas dimensões do *Big Data* amplamente aceitas na comunidade acadêmica (YUNITTA et al, 2022).

É importante compreender que, o *Big Data* é composto por dados estruturados e não-estruturados como textos, vídeos, imagens, registros das redes sociais e informações geradas em tempo real, por meio de sensores e dispositivos interconectados (LEE, 2020). Portanto, ao descrever o *Big Data*, não se pode esquecer que o termo se refere mais a um fenômeno social do que a uma tecnologia específica (CARDONA et al, 2022).

Autores como Agrawal et al (2020) Bhardwaj (2020), Zhang (2018) e Gandomi (2018) descrevem o fenômeno do *Big Data* a partir de características relacionadas a sua natureza, incluindo uma coleção de V's. O Quadro 6, descreve os recursos apontados pelos autores como fatores capazes de transformar um ativo de informação de alto volume em valor.

Quadro 6 - Características do *Big Data* com base na coleção de V's.

Identificação	Definição	Desafio
Volume	Refere-se à quantidade de dados armazenados em um repositório de dados.	Analisar e extrair informações relevantes de um conjunto de dados complexos.
Velocidade	Rapidez com que novos dados são gerados.	Gerenciar dados não-estruturados de forma eficaz e em tempo real.
Variedade	Heterogeneidade de dados. Diferentes tipos de dados de saúde.	Obter <i>insights</i> a partir da análise de grandes volumes de dados heterogêneos, disponíveis de maneira holística.
Variabilidade	Inconsistência de dados.	Corrigir a interpretação de dados que podem variar de forma significativa, dependendo do contexto.
Veracidade	Refere-se a confiabilidade e qualidade dos dados.	Adotar estratégias e práticas seguras para coletar, armazenar, recuperar e analisar os dados.
Visualização	Capacidade de interpretar os dados e criar conhecimentos.	O desafio está em gerar insights a partir das informações extraídas do <i>Big Data</i> .
Valor	Descoberta de <i>insights</i> e conhecimentos que aumentam as chances de uma organização alcançar um determinado objetivo.	Descobrir os conhecimentos ocultos em grandes quantidades de dados complexos.

Fonte: Adaptado de, Zhang (2018), Gandomi (2018), Agrawal (2020) e Bhardwaj (2020).

O valor do *Big Data* na área saúde está diretamente relacionado com a análise dos dados e extração de informações úteis do *Big Data* para as partes interessadas. Estes *insights* são capazes de auxiliar no processo de tomada de decisões estratégicas das instituições e facilitar a resolução de problemas estruturais (CHRISTENSEN et al, 2019).

Jindal et al (2018) complementam esta ideia, ao abordar o conceito de valor sobre a perspectiva do uso do *Big Data* no setor de saúde por meio das tecnologias. O autor cita como exemplo os modelos assistenciais de saúde, onde os serviços prestados ao paciente foram ampliados por meio da computação em nuvem, que forneceu escalabilidade e flexibilidade para

as plataformas digitais de serviços médicos, aumentando a segurança e a velocidade do compartilhamento das informações sensíveis (KHANRA et al, 2020).

Possivelmente, o mais importante a observar é que o conceito de *Big Data* evoluiu ao longo do tempo, e atualmente não se concentra em uma ferramenta para o armazenamento de grandes quantidades de dados, mas sim no processo de criação de valor por meio do gerenciamento, processamento e análise dos dados. Desta forma, autores como Hossain et al (2019) definem o *Big Data* como fluxos e não estoques de informações, reforçando a importância da mudança de visão das organizações sobre a amplitude e aplicabilidade do *Big Data* em diferentes setores da área da saúde (CHEUNG, 2020).

4.2 Aplicabilidades do *Big Data* na área da saúde

A pesquisa conduzida por Wang et al (2018), abordou os impactos da adoção do *Big Data* na área da saúde em quatro frentes: (1) qualidade dos serviços em saúde; (2) práticas médicas; (3) atividades científicas e de pesquisa; e (4) gestão das instituições de saúde.

Neste estudo, foram citados como fatores de contribuição para a melhoria da qualidade dos serviços em saúde a difusão da prática médica orientada a dados e os recentes avanços no diagnóstico e detecção precoce de doenças, potencializados pela medicina de precisão e o uso do *Big Data* na composição de conhecimentos médicos (WANG et al, 2018).

A medicina de precisão une os dados comumente utilizados para diagnosticar e tratar doenças conhecidas e relacionadas ao perfil genético do indivíduo, com o objetivo de determinar os tratamentos mais eficientes para os pacientes, respeitando as diferenças biológicas que tornam cada indivíduo único (KARPATHAKIS et al, 2021).

Para alcançar este objetivo, as instituições de saúde precisam ter acesso a um vasto acervo de amostras de dados populacionais distintos, obtidos por meio dos registros médicos e de pesquisas geograficamente abrangentes (BAINBRIDGE, 2019). Neste cenário, o *Big Data* se apresenta como uma ferramenta robusta para o armazenamento dos dados coletados, e seus benefícios podem ser potencializados pela fusão com outras tecnologias, como a Inteligência Artificial (IA), utilizada para maximizar a capacidade do cruzamento dos dados (CHEUNG, 2020).

Durante o período mais crítico da pandemia do COVID-19, esse conjunto de tecnologias se apresentou como ativos fundamentais para compreensão do comportamento e dos modos de transmissão do vírus SARS-CoV-2, até então relativamente desconhecidos pela área médica (LONG et al, 2021). Como resultado da análise do *Big Data*, novos conhecimentos sobre o vírus emergiram, apoiando as decisões e orientando as práticas dos profissionais de saúde que atuaram na linha de frente do combate à pandemia global (CARDONA et al, 2022).

A área de pesquisa em saúde também é apoiada pelo *Big Data*. Com a integração dos dados médicos, é possível reunir no *Big Data* os prontuários dos pacientes das redes públicas e privadas, viabilizando a realização de estudos mais assertivos e com maior impacto para a população como um todo (BATKO; SLEZAK, 2022). Em paralelo, o cruzamento de dados clínicos, biométricos e dos registros de informações fornecidas pelos próprios pacientes, podem gerar conhecimentos relevantes para fomentar pesquisas destinadas ao tratamento de doenças complexas e multifatoriais (ZHENG et al, 2022).

A área de negócios e gestão de saúde foi fortemente impactada pela adoção do *Big Data*, porém é preciso ter em mente que a sua natureza se transforma de acordo com o contexto em que os dados estão inseridos (PAIGE et al, 2021). No setor público, por meio dos resultados extraídos do *Big Data*, é possível identificar as atividades e os procedimentos médicos duplicados ou desnecessários, diminuindo custos e mitigando o risco de desperdício de materiais, melhorando, assim, a gestão da verba pública destinada para as instituições de saúde (D'AGOSTINO et al, 2021).

Na saúde privada, sob a perspectiva do aumento da lucratividade, a partir da análise dos dados médicos, é possível detectar pacientes cujos tratamentos geram altos custos para a instituição, além dos métodos de trabalho que requerem um maior aporte financeiro (CHEUNG, 2020). Neste sentido, as instituições de saúde, podem lançar mão destas informações para oferecer aos profissionais e paciente soluções personalizadas, que beneficiem ambas as partes (WESSON et al, 2022).

4.3 Desafios para implementação do *Big Data* na saúde

O setor de saúde sempre gerou grandes quantidades de dados e isso está relacionado, entre outros fatores, com a necessidade de armazenamento de prontuários do paciente. No entanto, seu armazenamento não é unificado em uma única base e nem segue um padrão de registros das informações (SHUBHAM et al, 2021). Assim, os estabelecimentos de saúde que vislumbram utilizar o potencial do *Big Data* em seus processos, precisam enfrentar novos desafios tecnológicos e organizacionais, que vão além da dificuldade de análise dos dados estruturados e não estruturados (CHEUNG, 2020).

Tais desafios, incluem um investimento inicial maciço para aquisição dos recursos (tecnológicos e humanos) necessários para viabilizar e potencializar os benefícios do *Big Data* na área da saúde (BATKO et al, 2022). Estudiosos como Yunitta (2022), Liang (2021) e Khanra (2020) alertam sobre os cuidados com a segurança e privacidade dos dados dos pacientes, especificamente com relação a exposição, extração, manipulação e compartilhamento não autorizado de dados sensíveis.

Devido a sua faceta disruptiva, o entendimento sobre o potencial do *Big Data* pode não ficar claro para os profissionais da saúde, e esta falta de clareza, pode gerar resistências quanto a sua aceitação (TALWAR et al, 2020); (KHANRA et al, 2020). Desta forma, é fundamental que os setores de tecnologia, gestão hospitalar e demais alas administrativas, se esforcem mutuamente para conscientizar os profissionais sobre o valor que do uso do *Big Data* pode agregar para a área médica, sendo esta, uma ação tão necessária quanto preparar a infraestrutura para dar suporte operacional para sua implementação (WU et al, 2022).

4.4 Projeções futuras para o uso do *Big Data* na saúde

Com base nos estudos selecionados, foram identificadas duas oportunidades auspiciosas para potencializar os benefícios da adoção do *Big Data* na área da saúde: *Machine Learning* (ML) e Inteligência Artificial (IA), que compreendem métodos que permitem que os computadores executem tarefas geralmente associadas à inteligência humana, de forma automatizada (ARORA; SHARMA, 2022).

Os conceitos de ML e IA são facilmente confundidos, devido à similaridade dos termos. No entanto, para perceber os impactos desses eventos no setor de saúde, é importante delimitar seus conceitos (WERUTSKY, 2021). Tecnicamente, o *Machine Learning* é considerado uma subcategoria da Inteligência Artificial, mas apesar disso, conceitualmente é compreendido como um conjunto de técnicas desenvolvidas para aprender por meio do processamento de uma grande massa de dados, de maneira automatizada e em constante evolução (CHEUNG, 2020). Em contrapartida, Inteligência Artificial pode ser definida como o ramo da ciência da computação, que desenvolve sistemas que simulam a capacidade de aprendizagem humana para a resolução de um problema (SUPRIYA; SASIDARAN, 2021).

Quando aplicados no setor de saúde, a Inteligência Artificial e o *Machine Learning* utilizam as máquinas para definir os algoritmos adequados para a análise do *Big Data*, com o objetivo de identificar os componentes do problema, suas possíveis soluções e ao final, gerar *insights* capazes de auxiliarem a tomada de decisões com precisão, sem a necessidade de intervenção humana (CHRISTENSEN et al, 2019). Estes recursos tecnológicos, tem como principais características a flexibilidade e interoperabilidade, que no setor de saúde significa a troca de dados simultânea e em alta velocidade entre os diferentes sistemas de saúde, gerando mais informações e relações significativas para o cuidado do paciente (RANA et al, 2022).

Partindo deste pressuposto, autores como Silvennoinen (2022) e Singh (2021) acreditam que a adoção do *Machine Learning* e da Inteligência Artificial somadas aos fluxos de informações geradas por meio do *Big Data*, contribuem para a criação de conhecimentos com valor e identificação de tendências, diante de um grande volume de dados médicos. Além disso, por meio deste conjunto de tecnologias é possível prever a propagação de doenças, promover tratamentos personalizados e aumentar a precisão dos diagnósticos (CARDONA et al, 2022).

5 CONCLUSÃO

Este trabalho analisou como o *Big Data* é aplicado na área da saúde e seus respectivos benefícios. Para isso, foi realizado um estudo bibliométrico, cujo os resultados mostraram que houve um crescimento exponencial do número de publicações sobre o tema, entre os anos de 2018 e 2022 (+ 427,04%).

Destaca-se que os artigos publicados entre janeiro e julho de 2022 (323 estudos num intervalo de sete meses) quando comparados as publicações do ano de 2018 (122 estudos num período de 12 meses), apresentaram um crescimento relativo de 164,75% do número total de publicações, evidenciando que a área de estudo se encontra em ascensão.

Adicionalmente, foi possível identificar que os estudos que abordaram o *Big Data* na saúde apresentam características multidisciplinares, englobando diferentes áreas de pesquisa. Um dos reflexos da construção multidisciplinar do conhecimento sobre o tema, pode ser percebido ao buscar definições mais precisas para o termo *Big Data* na literatura, onde a natureza do conceito não é claramente compreendida e não há uma unanimidade quanto a sua definição.

De acordo com os dados extraídos da biblioteca digital *Scopus*, no período analisado, nenhum autor brasileiro foi citado na rede de coautoria dos principais autores que pesquisam sobre o tema, evidenciando uma lacuna e a oportunidade de desenvolvimento de estudos inéditos sobre o tema na literatura nacional, e ao mesmo tempo reforçando a importância do

presente artigo. Por outro lado, chama-se a atenção para o fato de que os países norte-americanos, europeus e asiáticos são predominantes quanto ao desenvolvimento de estudos relevantes nesta área. Ao analisar o *ranking* de publicações por países, até o mês de julho do ano de 2022, o Brasil ocupava a décima sexta posição. Estes fatos evidenciam também a importância de os pesquisadores desenvolverem estudos destinados para a divulgação do conhecimento nos países sul americanos.

Dada a sua importância, o conceito de *Big Data* evoluiu ao longo do tempo, passando a considerar outras vertentes na sua definição, sendo atualmente visto não apenas como uma ferramenta para o armazenamento de grandes volumes de dados, mas sim como um processo de criação de valor por meio do armazenamento, gerenciamento, processamento e análise dos dados coletados.

Neste contexto, os benefícios do *Big Data* são potencializados por meio da fusão com outras tecnologias, a exemplo do *Machine Learning* e da Inteligência Artificial, impactando diretamente na qualidade dos serviços prestados pelas instituições de saúde. Percebe-se a tendência do uso dos dados para orientar a prática médica, potencializando o diagnóstico e a detecção precoce de doenças por meio da medicina preventiva de precisão.

De modo geral, os desafios que as instituições de saúde enfrentam para implementar o *Big Data* vão além da dificuldade de análise dos dados estruturados e não estruturados que compõem estas bases. Passam pela necessidade de aporte financeiro para aquisição de *softwares* e *hardwares* desenvolvidos para suportar a tecnologia, além dos custos para a contratação e treinamento de profissionais responsáveis por gerenciar o *Big Data*, com *expertise* para integrar estas tecnologias e extrair informações de qualidade da base de dados.

6 REFERÊNCIAS

AGRAWAL, R., PRABAKARAN, S. Big data in digital healthcare: lessons learnt and recommendations for general practice. **Heredity**, v. 124, n. 4, p. 525-534, 2020.

ARORA, Manpreet e SHARMA, Roshan Lal. Artificial intelligence and big data: ontological and communicative perspectives in multi-sectoral scenarios of modern businesses. **foresight**, n. ahead-of-print, 2022.

BAINBRIDGE, Michael. Big data challenges for clinical and precision medicine. In: **Big data, big challenges: a healthcare perspective**. Springer, Cham, 2019. p. 17-31.

BATKO, K., ŚLĘZAK, A. The use of Big Data Analytics in healthcare. **Journal of big Data**, v. 9, n. 1, p. 1-24, 2022.

BHARDWAJ, P., BALIYAN, N. Big data analytics in healthcare. In: **Smart Healthcare Systems**. Chapman and Hall/CRC, 2019. p. 1-15.

CARDONA, L. A., FRANCO OCAMPO, D. F., ESTRADA ESPONDA, R. D. Aplicaciones de la Datificación y Big Data en América Latina entre el 2015 y 2019. *Revista Logos Ciencia & Tecnología*, v. 14, n. 2, p. 125–143, 30 jun. 2022.

CHEUNG, Sarah. Disambiguating the benefits and risks from public health data in the digital economy. **Big Data & Society**, v. 7, n. 1, p. 2053951720933924, 2020.

- CHRISTENSEN, J. H, ANDREASEN, M. M., RUDEBUSCH, G. D. Term structure analysis with big data: one-step estimation using bond prices. **Journal of Econometrics**, v. 212, n. 1, p. 26-46, 2019.
- D'AGOSTINO, M., MARTI, M., MEJIA, F. M., MALEK, V., e SAISO, S. G. Salud pública y la interdependencia digital: evolución tecnológica, sostenibilidad tecnológica y la revolución del usuario. **Revista Panamericana de Salud Pública**, v. 45, 2021.
- FAVI, C., MARCONI, M., MANDOLINI, M. e GERMANI. Big data analysis for the estimation of disassembly time and de-manufacturing activity. **Procedia CIRP**, v. 90, p. 617-622, 2020.
- GANDOMI, A.H. Deep learning in medical imaging: fmri big data analysis via convolutional neural networks. In: **Proceedings of the Practice and Experience on Advanced Research Computing**. 2018. p. 1-4.
- GHASSEMI, Marzyeh e MOHAMED, Shakir. Machine learning and health need better values. **npj Digital Medicine**, v. 5, n. 1, p. 1-4, 2022.
- GRIVA, A., DENNEHY, D., PAPPAS, I., MÄNTYMÄKI, M., POULOUDI, N., DWIVEDI, Y. K., e SCHMARZO, B. Artificial intelligence and analytics in practice. **Journal of Decision Systems**, p. 1-7, 2022.
- GUPTA, D., RANI, R. A study of big data evolution and research challenges. **Journal of information science**, v. 45, n. 3, p. 322-340, 2019.
- HAJJAJI, Y, BOULILA, W., FARAH, I. R., ROMDHANI, I., e HUSSAIN, A. Big data and IoT-based applications in smart environments: A systematic review. **Computer Science Review**, v. 39, p. 100318, 2021.
- HOSSAIN, M. M., TASNIM, S., SHARMA, R., SULTANA, A., SHAIK, A. F., FAIZAH, F. e BHATTACHARYA, S. Digital interventions for people living with non-communicable diseases in India: A systematic review of intervention studies and recommendations for future research and development. **Digital Health**, v. 5, p. 205520761989615, jan. 2019.
- JANETT, R. S., YERACARIS, P.P. Electronic Medical Records in the American Health System: challenges and lessons learned. **Ciencia & saude coletiva**, v. 25, p. 1293-1304, 2020.
- JINDAL, A., DUA, A., KUMAR, N., DAS, A. K., VASILAKOS, A. V., e RODRIGUES, J. J. Providing healthcare-as-a-service using fuzzy rule based big data analytics in cloud computing. **IEEE Journal of Biomedical and Health informatics**, v. 22, n. 5, p. 1605-1618, 2018.
- KANNAP, S. Understanding forest health with remote sensing, part III: requirements for a scalable multi-source forest health monitoring network based on data science approaches. **Remote sensing**, v. 10, n. 7, p. 1120, 2018.
- KARPATHAKIS, K, LIBOW, G., POTTS, H. W., DIXON, S., GREAVES, F., & MURRAY, E. An evaluation service for digital public health interventions: user-centered design approach. **Journal of medical Internet research**, v. 23, n. 9, p. e28356, 2021.

KHANRA, S., DHIR, A., ISLAM, A. N., e MÄNTYMÄKI, M . Big data analytics in healthcare: a systematic literature review. **Enterprise Information Systems**, v. 14, n. 7, p. 878-912, 2020.

KITCHENHAM, B. A., CHARTERS, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. [s.l.]: Keele University and Durham University Joint Report, 2007. Disponível em:
https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf

LANEY, D (2001) 3D data management: controlling data volume, velocity and variety. META Group, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (acesso em 18 de maio de 2022).

LIU, W., LONG, S., XIE, D., LIANG, Y., e WANG, J. How to govern the big data discriminatory pricing behavior in the platform service supply chain? An examination with a three-party evolutionary game model. **International Journal of Production Economics**, v. 231, p. 107910, 2021.

LUPTON, Deborah e MICHAEL, Mike. Toward a manifesto for the ‘public understanding of big data’. **Public Understanding of Science**, v. 25, n. 1, p. 104-116, 2016.

OHLHORST, Frank J. **Big data analytics: turning big data into big money**. John Wiley & Sons, 2012.

PAIGE, E., DOYLE, K., JORM, L., BANKS, E., HSU, M. P., NEDKOFF, L. e FIGTREE, G. A . A versatile big data health system for Australia: driving improvements in cardiovascular health. **Heart, Lung and Circulation**, v. 30, n. 10, p. 1467-1476, 2021.

PETERSEN, Kai; VAKKALANKA, Sairam; KUZNIARZ, Ludwik. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and software technology**, v. 64, p. 1-18, 2015.

POWER, Daniel J. Using ‘Big Data’ for analytics and decision support. **Journal of Decision Systems**, v. 23, n. 2, p. 222-228, 2014.

RANA, A., CHAKRABORTY, C., SHARMA, S., DHAWAN, S., PANI, S. K., e ASHRAF, I. Internet of Medical Things-Based Secure and Energy-Efficient Framework for Health Care. **Big Data**, v. 10, n. 1, p. 18-33, 2022.

RESNYANSKY, Lucy. Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. **Big Data & Society**, v. 6, n. 1, p. 2053951718823815, 2019.

SANG, Nguyen Minh. Publications on the Big data research. Mendeley, 11 de junho de 2021. DOI.org (Datacite) <https://doi.org/10.17632/H8BBD67DJP.1>

SAUNDERS, G. H., CHRISTENSEN, J. H., GUTENBERG, J., PONTOPPIDAN, N. H., SMITH, A., SPANOUDAKIS, G., e BAMIOU, D. E. Application of Big Data to Support Evidence-Based Public Health Policy Decision-Making for Hearing. **Ear & Hearing**, v. 41, n. 5, p. 1057–1063, 23 jan. 2020.

SENTHILKUMAR, S. A., RAI, B. K., MESHRAM, A. A., GUNASEKARAN, A. e CHANDRAKUMARMANGALAM, S. Big data in healthcare management: a review of literature. **American Journal of Theoretical and Applied Business**, v. 4, n. 2, p. 57-69, 2018.

SINGH, R.P., HALEEM, A., JAVAID, M., KATARIA, R., e SINGHAL, S. Cloud computing in solving problems of COVID-19 pandemic. **Journal of Industrial Integration and Management**, v. 6, n. 02, p. 209-219, 2021

SUPRIYA, M. S. e SASIDARAN, K. Machine Learning for Big Data. In: **Applications of Big Data in Large-and Small-Scale Systems**. IGI Global, 2021. p. 56-76.

VAN ECK, Nees Jan, e LUDO Waltman. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics*, vol. 84, no 2, agosto de 2010, p. 523–38. Springer Link, <https://doi.org/10.1007/s11192-009-0146-3>

WANG, S., PERSHING, LEE A., ARON Y. Big data requirements for artificial intelligence. **Current opinion in ophthalmology**, v. 31, n. 5, p. 318, 2020.

WANG, Y., KUNG, L., BYRD, T. A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. **Technological forecasting and social change**, v. 126, p. 3-13, 2018.

WERUTSKY, G., BARRIOS, C. H., CARDONA, A. F., ALBERGARIA, A., VALENCIA, A., FERREIRA, C. G., ... e CAZAP, E. Perspectives on emerging technologies, personalised medicine, and clinical research for cancer control in Latin America and the Caribbean. **The Lancet Oncology**, v. 22, n. 11, p. e488-e500, 2021.

WESSON, P., HSWEN, Y., VALDES, G., STOJANOVSKI, K., e HANDLEY, M. A. Risks and Opportunities to Ensure Equity in the Application of Big Data Research in Public Health. **Annu Rev Public Health [Internet]**, v. 43, n. 1, p. 59-78, 2022.

WU, Xiaobo, LIANG, Liping, CHEN Siyuan. How big data alters value creation: through the lens of big data competency. **Management Decision**, 2022.

YUNITA, A.; SANTOSO, H. B.; HASIBUAN, Z. A. “Everything is data”: towards one big data ecosystem using multiple sources of data on higher education in Indonesia. *Journal of Big Data*, v. 9, n. 1, 14 jul. 2022.

ZHANG, Y., HUANG, T., BOMPARD, E. F. Big data analytics in smart grids: a review. **Energy informatics**, v. 1, n. 1, p. 1-24, 2018.

ZHOU, Xiaokang. Variational LSTM enhanced anomaly detection for industrial big data. **IEEE Transactions on Industrial Informatics**, v. 17, n. 5, p. 3469-3477, 2020.

ZUPIC, Ivan, ČATER, Tomaž. Bibliometric methods in management and organization. **Organizational research methods**, v. 18, n. 3, p. 429-472, 2015.