



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

STEFFANE RAMIRES DE LIMA CAMPOS

**USO DE INFORMAÇÕES BASEADAS EM *BIG DATA* NO ENFRENTAMENTO DA
PANDEMIA DE COVID-19: UM MAPEAMENTO SISTEMÁTICO DA LITERATURA**

RECIFE, PE

2023

STEFFANE RAMIRES DE LIMA CAMPOS

**USO DE INFORMAÇÕES BASEADAS EM *BIG DATA* NO ENFRENTAMENTO DA
PANDEMIA DE COVID-19: UM MAPEAMENTO SISTEMÁTICO DA LITERATURA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, Centro de Artes e Comunicação, como requisito parcial para a obtenção do título de Mestra em Ciência da Informação.

Área de Concentração: Informação, Memória e Tecnologia.

Orientador: Prof. Dr. Célio Andrade de Santana Júnior

Coorientadora: Prof.^a Dr.^a Edilene Maria da Silva

RECIFE, PE

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Campos, Steffane Ramires de Lima .

Uso de informações baseadas em big data no enfrentamento da pandemia de covid-19: um mapeamento sistemático da literatura / Steffane Ramires de Lima Campos. - Recife, 2023.

90 p., tab.

Orientador(a): Célio Andrade de Santana Júnior

Coorientador(a): Edilene Maria da Silva

Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Artes e Comunicação, Programa de Pós-Graduação em Ciência da Informação, 2023.

1. Big Data. 2. Covid-19. 3. Ferramentas Tecnológicas. 4. Pandemia. 5. Mapeamento Sistemático da Literatura. I. Santana Júnior, Célio Andrade de . (Orientação). II. Silva, Edilene Maria da . (Coorientação). IV. Título.

020 CDD (22.ed.)

UFPE (CAC 2024 - 31)

STEFFANE RAMIRES DE LIMA CAMPOS

**USO DE INFORMAÇÕES BASEADAS EM *BIG DATA* NO ENFRENTAMENTO DA
PANDEMIA DE COVID-19: UM MAPEAMENTO SISTEMÁTICO DA LITERATURA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, Centro de Artes e Comunicação, como requisito parcial para a obtenção do título de Mestra em Ciência da Informação. Área de concentração: Informação, Memória e Tecnologia.

Aprovada em: 07/12/2023

BANCA EXAMINADORA

Prof. Dr. Célio Andrade de Santana Júnior (Orientador)
Universidade Federal de Pernambuco

Prof^ª Dr^a Májory Karoline Fernandes de Oliveira Miranda (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Antônio Souza da Silva Júnior (Examinador Interno)
Universidade Federal de Pernambuco

Prof^ª Dr^a Renata Telles Moreira (Examinador Externo)
Universidade Federal de Lavras

A Deus, minha razão de ser.

Para Raul Levi, meu filho.

À minha família.

*Ela está em todas as coisas, até no vazio que me
dá, quando vejo a tarde cair, e ela não está...*
Jorge Vercillo / Jota Maranhão



In memoriam de

*Tereza Cristina Ferreira de Lima
(mainha)*

*Maria Jocelina Bezerra Cavalcanti
(vó)*

AGRADECIMENTOS

Em primeiro lugar àquele que foi, e é, o meu ponto de equilíbrio, refúgio e fortaleza nos momentos de angústias e dificuldades, que pensei em desistir. Ao Senhor e consumidor da minha fé, Jesus Cristo, responsável por acolher cada um dos meus sonhos e me ajudar a realizar.

Especialmente ao meu filho, Raul Levi de Lima Campos, que foi gerado durante esse processo, e cuja concepção despertou todos os meus instintos para promover este exemplo de perseverança e resiliência, deixando um motivo que desperte seu orgulho em ter a mim como mãe.

Aos meus familiares e amigos, que de alguma forma participaram da construção desta pesquisa, ainda que torcendo pela minha conquista de longe. Principalmente ao meu marido e parceiro de vida, Marcos Augusto Batista Campos, que sempre foi ajuda e incentivo para que eu concluísse esse objetivo.

Ao meu orientador, Célio Andrade de Santana Júnior, que não largou minha mão mesmo essa caminhada tendo apresentado inúmeros obstáculos, e também a minha coorientadora, Edilene Maria da Silva, que se colocou à disposição para ajudar no meu projeto com contribuições sempre que precisei.

Agradeço aos examinadores da minha Banca de Qualificação, que iluminaram minha pesquisa para novos horizontes possíveis.

Às amigas que fiz durante o curso de mestrado, Karen Isabelle Santos D'amorim e Bruna Laís Campos do Nascimento, que foram verdadeiros anjos em forma de gente para me dar apoio na concretização desse sonho.

Muito obrigada a todos!

“O saber se aprende com os mestres. A sabedoria, só com o corriqueiro da vida.”

Cora Coralina

RESUMO

Motivada pela crise sanitária em decorrência da pandemia da Covid-19 e pelos desafios surgidos no cenário pandêmico, pretendendo explorar as medidas de enfrentamento do coronavírus, esta pesquisa tem como objetivo identificar as soluções baseadas em big data que auxiliaram no combate à pandemia da Covid-19. Justifica-se pela importância dada à informação como instrumento de combate à pandemia, bem como pelo caráter interdisciplinar dado às pesquisas relacionadas ao tema Covid-19 e ao elevado número de dados produzidos nesse contexto. Foi utilizado o método de mapeamento sistemático da literatura para identificar as principais ferramentas que foram desenvolvidas e/ou utilizadas naquele contexto a fim de minimizar os impactos negativos causados pela crise sanitária. Realizou-se uma busca por artigos científicos voltados para os temas “*big data* e Covid-19” nas bases de dados BRAPCI (Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação) e *Web of Science (WoS)*, que contou com a recuperação de 60 trabalhos completos, dos quais 29 foram utilizados nesta pesquisa. Os resultados apresentados foram divididos em 17 categorias, 6 temáticas e 11 ferramentas, que foram criadas a partir do fichamento e da análise de conteúdo do material coletado. O presente estudo corrobora para ampliar o conhecimento acerca das possibilidades de soluções que podem assistir à população em momentos de semelhante estado emergencial.

Palavras-chave: *Big Data*; Covid-19; Ferramentas Tecnológicas; Pandemia; Mapeamento Sistemático da Literatura.

ABSTRACT

Motivated by the health crisis resulting from the Covid-19 pandemic and the challenges that have arisen in the pandemic scenario, and with the aim of exploring measures to combat the coronavirus, this research aims to identify solutions based on big data that have helped to combat the Covid-19 pandemic. It is justified by the importance given to information as a tool to combat the pandemic, as well as the interdisciplinary nature given to research related to Covid-19 and the large amount of data produced in this context. The systematic literature mapping method was used to identify the main tools that were developed and/or used in that context in order to minimize the negative impacts caused by the health crisis. A search was carried out for scientific articles on the themes of "big data and Covid-19" in the BRAPCI (Reference Database of Journal Articles in Information Science) and Web of Science (WoS) databases, which retrieved 60 full papers, 29 of which were used in this research. The results presented were divided into 17 categories, 6 themes and 11 tools, which were created based on a summary and content analysis of the material collected. This study contributes to expanding knowledge about the possibilities of solutions that can assist the population in times of such emergency.

Keywords: Big data; Covid-19; Technological Tools; Pandemic; Systematic Mapping of Literature.

LISTA DE QUADROS

Quadro 1 –	Autores e a evolução dos “vs” do <i>big data</i>	25
Quadro 2 –	Delineamento da Pesquisa	35
Quadro 3 –	Engenhos de Busca utilizados na avaliação da necessidade	41
Quadro 4 –	Estudos secundários referentes a <i>Big Data</i> no contexto da Covid	42
Quadro 5 –	Estratégias de busca	44
Quadro 6 –	Critérios de Inclusão e Exclusão	45
Quadro 7 –	Trabalhos Seleccionados	47
Quadro 8 –	Artigos relacionados com as ferramentas	52

LISTA DE FIGURAS

Tabela 1	Medidas de dados de armazenamento digitais e exemplos	23
Figura 1 –	Os v's do <i>Big data</i>	24
Figura 2 –	Fases da Pesquisa	39
Figura 3 –	Sumários da seleção	46

LISTA DE ABREVIATURAS E SIGLAS

ACM – Association Computer Machinery

BDA – Big Data Analytics

SUS – Sistema Único de Saúde

OMS – Organização Mundial da Saúde

ESPII – Emergência de Saúde Pública de Importância Internacional

TIC's – Tecnologias de Informação e Comunicação

GPS – Sistema de Posicionamento Global

MSL – Mapeamento Sistemático da Literatura

RSL – Revisões Sistemáticas da Literatura

EBP – Prática baseada em evidências

IEEE – Electrical and Electronic Engineers

GIS – Sistema de Informação Geográfica

HCP – Computação de Alto Desempenho

SDK – Software Development Kit

SIS – Sistema de Informação Hospitalar

RF – Random Forest

GNN – Graph Neural Networks

CRC – Coronavirus Resource Center

IA – Inteligência Artificial

LGPD – Lei Geral de Proteção de Dados

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

SUMÁRIO

1 INTRODUÇÃO	16
1.1 OBJETIVOS	20
1.2 ESTRUTURA DO TRABALHO	20
2 INFORMAÇÕES BASEADAS EM <i>BIG DATA</i>	21
2.1 CONTEXTUALIZANDO <i>BIG DATA</i>	22
2.1.1 A evolução dos 3vs do <i>big data</i> para os 7vs	25
2.1.2 Concepções de Big Data na Ciência da Informação	27
2.2 BIG DATA E A PANDEMIA DE COVID-19	29
3 PROCEDIMENTOS METODOLÓGICOS	34
3.1 METODOLOGIA DA PESQUISA	34
3.2 CARACTERIZAÇÃO DA PESQUISA	34
3.3 MÉTODO - MAPEAMENTO SISTEMÁTICO DA LITERATURA (MSL)	36
3.3.1 Necessidade do Mapeamento Sistemático da Literatura	39
3.3.2 Análise Preliminar da Literatura (i)	43
3.3.3 Elaboração do Protocolo de Pesquisa (ii)	43
4 RESULTADOS	52
4.1 CATEGORIAS DE FERRAMENTAS	53
4.1.3 Painel de Vigilância Covid-19 (identificador do artigo 15)	57
4.1.4 <i>Pipeline</i> (identificador dos artigos 2, 3 e 21)	57
4.1.5 Repositório MM-Covid (identificador do artigo 25)	58
4.1.6 Dados de Mobilidade Humana (identificador dos artigos 12 e 17)	58
4.1.7 Sistema de Informação Hospitalar (SIS) (identificador do artigo 19)	59
4.1.8 Modelagem de Dados (identificador dos artigos 11 e 22)	60



4.1.9 Modelos de Aprendizado (identificador dos artigos 7 e 16)	61
4.1.10 Modelos de Previsão (identificador dos artigos 6, 13 e 24)	62
4.1.11 Representações Geoespaciais e Temporais (identificador do artigo 28)	63
4.2 TEMA	66
4.2.1 Uso de Inteligência Artificial (IA) (identificador dos artigos 2, 5, 14 e 23)	68
4.2.2 Visualização de Dados (identificador dos artigos 8, 9, 10, 15, 21, 23, 28 e 29)	69
4.2.3 Uso de Redes Sociais (<i>Twitter</i>) (identificador dos artigos 4, 26 e 27)	75
4.2.4 Uso de Modelos Matemáticos no Combate à Covid-19 (identificador dos artigos 3 e 20)	76
4.2.5 Saúde Mental de Idosos na Pandemia (grupo de risco) (identificador dos artigos 26)	77
4.2.6 Impactos Econômicos Provocados pela Pandemia (identificador dos artigos 1)	78
4.3 DISCUSSÃO DOS RESULTADOS	79
5 CONSIDERAÇÕES FINAIS	83
REFERÊNCIAS	86

1 INTRODUÇÃO

O final de 2019, mais precisamente no último dia daquele ano, ficou marcado pelo início do que mais tarde se tornaria uma pandemia que levaria milhões de pessoas a óbito e desafiaria as nações ao apresentar uma nova realidade que não estaria prevista em qualquer planejamento prévio. Naquele dia, foi anunciado pela Organização Mundial de Saúde (OMS) o primeiro caso de Covid-19, doença causada pelo novo Coronavírus.

Cury (2020) afirma que o primeiro infectado se tratava de um homem de cinquenta e cinco anos, residente na cidade de Wuhan, na China, que, segundo o jornal South China Morning, foi diagnosticado com sintomas desta nova doença dias antes da OMS ser notificada, e do caso ganhar destaque na imprensa. O autor aponta que de acordo com Leo Poon, cientista virologista da Escola de Saúde Pública de Hong Kong, o novo coronavírus é primo do vírus responsável por síndromes agudas respiratórias (SARS) e começou a se disseminar entre os seres humanos, após contaminação de origem animal.

Brenda Tesini (2020) sugere que os primeiros casos relacionados à Covid-19 estavam ligados ao comércio de animais vivos, exóticos, que eram vendidos para consumo humano na China, e destaca que a transmissão do vírus acontece entre as pessoas pela exposição a gotículas de espirro e tosse, definindo a doença da seguinte forma: “Covid-19 é uma doença respiratória aguda que pode ser grave e é causada por um coronavírus recentemente identificado, oficialmente chamado SARS-COV-2” (Tesini, 2020, p.1)

Desde então, inúmeros casos de pessoas infectadas pelo novo coronavírus começaram a ser registrados em todo o mundo, levando os órgãos de saúde, governantes e cidadãos a uma condição delicada de sobrevivência. Em 11 de março de 2020, após vários países apresentarem números exorbitantes de infectados e óbitos, atingindo agressivamente a Europa e principalmente na Itália, a Organização Mundial da Saúde declarou que a rápida disseminação da Covid-19 se tratava de uma pandemia.

Naquela data, os números apontavam “118 mil infecções em 114 nações, sendo que 4.291 pessoas já teriam sido vitimadas fatalmente”, segundo Tedros Ghebreyesus (2020), diretor-geral da OMS .

A informação é o principal insumo para o gerenciamento estratégico de processos, pessoas, riscos, custos e da comunicação. Num contexto de crise, torna-se componente

fundamental na tomada de decisões assertivas, como por exemplo, na conjuntura pandêmica de Covid-19, doença causada pelo novo coronavírus (SARS-CoV-2), constituída pela Organização Mundial de Saúde (OMS), em 30 de janeiro de 2020, como a sexta Emergência de Saúde Pública de Importância Internacional (ESPII)¹ da história. O que evidenciou ainda mais a importância de informações confiáveis para enfrentamento de situações adversas (OPAS, 2020).

Iniciava-se uma guerra contra um inimigo invisível e letal rompendo as barreiras do controle da saúde pública e privada, e que inicialmente colocou profissionais das áreas de serviços essenciais, como atuantes na linha de frente da batalha. Tal situação obrigava os indivíduos a reduzir o convívio social, limitando-se às paredes de casa (quarentena) e desafiando as autoridades públicas a elaborarem um plano de enfrentamento, combate e achatamento da curva de contaminação e consequentemente de óbitos.

Ainda em um contexto histórico, ano de 2020, a indústria farmacêutica iniciou uma corrida pelo desenvolvimento de vacinas e medicamentos para o enfrentamento da doença. O empreendimento demandou que centenas de profissionais de saúde, cientistas de dados e profissionais de outras áreas somassem esforços para combater a rápida transmissão do vírus causador da doença. Embora haja o destaque para o trabalho de profissionais da saúde, que atuavam na produção de vacinas e tratamentos eficazes, observa-se que profissionais de diversas áreas também estiveram envolvidos no combate à doença. Por exemplo, Alsunaidi, et al. (2021), aponta que a propagação da pandemia global, Covid-19, gerou uma enorme e variada quantidade de dados, que cresce rapidamente. Esses dados podem ser usados aplicando técnicas de análise de *big data* em várias áreas, incluindo diagnóstico, estimativa ou previsão de pontuação de risco, tomada de decisões em saúde e indústria farmacêutica relativos à Covid-19.

Alsunaidi e colegas (2021) sugerem que questões ligadas à gestão de dados também se formaram como outra fonte em auxílio ao combate a Covid-19. Esta área se aproxima da visão da Ciência da informação, cujo o reconhecimento dos esforços para o gerenciamento das informações e dos fluxos informacionais, visando o uso e a recuperação otimizada, não se traduzir em uma novidade sendo datados desde a década de 1940, período pós Segunda Guerra Mundial (Borko, 1968; Le Coadic, 2004).

¹ Conforme terminologia adotada pelo Ministério da Saúde, na Portaria nº 104, de 25 de janeiro de 2011, uma Emergência de Saúde Pública de Importância Internacional (ESPII) é um “evento extraordinário que constitui risco para a saúde pública de outros países por meio da propagação internacional de doenças e que potencialmente requerem uma resposta internacional coordenada” (Brasil, 2011).

E com os atuais avanços associados às novas Tecnologias de Informação e Comunicação (TIC's), o volume de informações criadas atingiram proporções na escala do que se conhece, hoje por *big data*, o que, no âmbito da Ciência da Informação, já havia sido denominado por “explosão da informação”, ou “crescimento exponencial da informação”, como destaca Barreto (2014). O que tem mudado, no entanto, é a complexidade em lidar com grandes volumes de dados (Santos-D'Amorim, et al., 2020). Nesse sentido, parte-se da premissa de que a informação é objeto de alto valor agregado, capaz de otimizar e modificar cenários, quando utilizada de forma estratégica.

Voltado para o contexto da Covid-19, ainda no início da pandemia foram adotadas medidas restritivas, que promoviam a política de “distanciamento social”, em que as pessoas deveriam ficar em suas casas. Galvão, et al. (2021) apontaram que em algumas cidades houve, por parte do poder público, uso de tecnologias de *big data* para realizar o monitoramento do movimento de pessoas em ambientes abertos. A partir dos dados dos *smartphones* dos cidadãos, era possível determinar, se um certo local estaria apresentando aglomeração de pessoas.

Outro exemplo é apresentado por Alsunaidi et al. (2021), trata-se de um sistema regular de monitoramento e detecção remota para indivíduos, que auxilia no rastreamento rápido de casos suspeitos de Covid-19, através de sistemas vestíveis que monitoram determinados sinais vitais do usuário. Segundo o autor, o uso de tais sistemas gerará uma enorme quantidade de dados, devido ao volume de pessoas que iriam utilizá-lo e que necessitam de técnicas para a realização de análise destas informações, melhorando o nível dos serviços de saúde.

Serviços de informação aplicados ao setor de saúde também não se configuram como uma novidade. O uso de ferramentas de análise de dados, no setor de saúde, estão presentes em áreas como genômica (Wang et. al, 2018), descoberta de medicamentos e pesquisa clínica (Wong et. al, 2015), assistência médica personalizada (Viceconti; Hunter; Rose, 2015), ginecologia (Erekson; Iglesia, 2015), nefrologia (Nadkarni; Coca; Wyatt, 2015), oncologia (Erekson; Iglesia, 2015), (Wang et. al, 2018). No entanto, nesta dissertação, apresentamos os mais proeminentes produtos e serviços de informação encontrados na literatura, que abordam soluções que integram o combate à pandemia.

No campo da ciência da informação, embora os estudos de Lee, et al 2002, Souza (et al., 2013), Pimenta (2013), Milagre e Segundo (2015), Fagundes, Macedo e Dutra (2017) tragam concepções válidas e necessárias acerca das relações entre *big data* e Covid-19, as pesquisas que

mencionam como resultados serviços, produtos ou soluções de informação ainda são escassas, o que amplia as possibilidades de estudos sobre o tema.

Esta pesquisa justifica-se pelo caráter interdisciplinar dado às pesquisas relacionadas ao combate à pandemia. Arencibia-Jorge (2020) aponta que a Covid-19 é uma experiência perturbadora para todos, incluindo cientistas, resultando em aumento maciço de artigos científicos produzidos. Segundo o autor, dezenas de estudos bibliométricos foram publicados após o surgimento e expansão crescente da nova pandemia. A maioria deles inclui indicadores bibliométricos tradicionais, como: principais autores, periódicos, instituições e países. O autor também identificou grupos temáticos, por meio de técnicas de mapeamento de acordo com a ocorrência de palavras-chave, padrões de co-citação ou colaborações internacionais. Alguns deles compararam padrões bibliométricos de doenças epidêmicas/pandêmicas, estatísticas descritivas, análise de redes de co-ocorrência e mineração de texto estão entre os métodos explorados para definir os limites da pesquisa do Covid-19.

Neste contexto, existe um espaço para a discussão do papel das soluções relativas à informação, e mais especificamente, no contexto do *big data*, no combate à pandemia. Nair, et al. (2022) aponta como uma das tendências de pesquisa relativas à Covid-19 a utilização do uso massivo de informação para a atuação em outra frente. Segundo os autores, “[...] as drogas e medicamentos podem tratar da infecção em si, mas o acesso à informação pode evitar que novos casos ocorram, bem como orientar políticas públicas para minimizar o impacto da pandemia” (Nair, et al., 2022, tradução nossa).

Assim, a principal motivação para a realização deste trabalho é trazer à luz, as soluções baseadas em *big data* voltadas ao contexto da Covid-19, como meio de destacar o papel da informação no combate à pandemia e explicitar o campo onde a Ciência da Informação contribui efetivamente para o seu sucesso.

Este trabalho, justifica-se também pela importância e participação de profissionais de outros nichos de atuação, que não sejam da área da saúde, para o enfrentamento ao vírus. Assim, queremos destacar, em especial, o papel dos profissionais da informação, que especificamente se utilizaram de *big data* para contribuir com a sociedade durante a pandemia, diante de uma batalha multidisciplinar em que diversas frentes de atuação se fizeram necessárias. Desta forma, consideramos importante evidenciar quais instrumentos e ferramentas foram utilizadas por estes profissionais e com qual intuito elas foram empreendidas.

1.1 OBJETIVOS

O objeto de estudo desta pesquisa é o uso de ferramentas baseadas em *Big Data* no enfrentamento da pandemia da Covid-19, tendo como **objetivo geral**: identificar as soluções baseadas em big data que auxiliaram no combate à pandemia da Covid-19. Para tanto, foi realizado um mapeamento sistemático da literatura, nas bases de dados BRAPCI e *Web of Science* em busca de fontes primárias que apontem tais soluções de *software*. O tema apresenta certo grau de originalidade, por tratar este assunto sob a ótica dos sistemas de informação voltados ao cenário pandêmico, e esta pesquisa busca enaltecer o papel da informação no enfrentamento à doença e a contribuição de profissionais da informação para as soluções.

A escolha do contexto da Covid-19 dá-se pelo grande número, a princípio, de soluções e processos voltados a preservar vidas neste período tão delicado. Neste sentido, o objetivo geral foi conduzido pelos seguintes **objetivos específicos**:

- a) Definir o termo big data;
- b) Identificar as principais ferramentas e formas de utilização das mesmas durante a pandemia;
- c) Apresentar estas ferramentas por categorias (análise temática).

1.2 ESTRUTURA DO TRABALHO

Está estruturado da seguinte forma, além desta seção introdutória, esta dissertação tem outros quatro capítulos. O capítulo 2 apresenta o referencial teórico contendo o arcabouço de conhecimentos referente ao tema deste trabalho. O capítulo 3 apresenta a metodologia de pesquisa; no capítulo 4 são apresentados os resultados obtidos na discussão, e no capítulo 5 são apresentadas as considerações finais.

2 INFORMAÇÕES BASEADAS EM *BIG DATA*

Na contemporaneidade, cenário em que a informação desempenha um papel fundamental na sociedade, influenciando diversos aspectos da vida cotidiana, desde a tomada de decisões pessoais até o funcionamento de organizações e governo, promovendo produção acelerada de dados e acesso generalizado, surge o termo *Big Data* para ser explorado e definido.

Mauro; Greco; Grimaldi (2015) trazem ao conhecimento do leitor a problemática que envolve o termo *Big Data*, devido ao seu caráter interdisciplinar o que permite sua aplicação em diversos contextos. Na tentativa de delimitar melhor seu significado, a fim de reduzir as ambiguidades e confusões causadas pelos vários sentidos e interpretações que se podem dar ao termo, os autores arriscam compreender o fenômeno *Big Data* a partir da análise de ocorrências do termo em referências da literatura acadêmica e empresarial em quatro temas chaves: Informação, Tecnologias, Métodos e Impacto. O estudo identificou aparição entre dois ou mais desses temas com o *Big Data*, e explicou as possíveis relações de dependência, sendo a “Informação” o combustível, ou seja, a matéria prima que alimenta o *Big Data*, as “Tecnologias” como ferramentas e equipamentos de suporte para armazenamento e distribuição da informação, os “Métodos” estão ligados à maneira que a informação será processada e análise (reconhecimento de padrões, modelagem preditiva, regressão, análise de sentimento, análise espacial, estatística, aprendizagem supervisionada e não supervisionada, simulação, análise de série temporal e visualização, por exemplo) e o “Impacto” diz respeito às mudanças proporcionadas nos ambientes e nas pessoas.

Com base na análise daqueles quatro termos supracitados, a pesquisa de (Mauro; Greco; Grimaldi 2015) sugeriu uma lista de termos consensuais para definir *Big Data* a partir do pressuposto de que esses atributos representam a essência e a centralidade do *Big Data*, sendo alguns deles: “Volume, Velocidade e Variedade” (visando descrever a informação), “Tecnologia e Métodos Analíticos” (para esclarecer a forma de uso da informação), “Valor” (transformação da informação e seu impacto).

Para (Pence, 2015) o próprio termo “Big”, de *Big Data*, sugere conceitos e definições que estejam relacionados com a característica de grandeza no que diz respeito ao tamanho dos dados, observando que o *Big Data* se trata de grande volume de dados, muito maior do que os arquivos de informações que estamos acostumados a acessar e processar através de um simples

computador. Na análise do autor, em outras palavras, pode-se chamar de *Big Data* uma quantidade tão volumosa, grande, de dados que necessite de tecnologias e softwares de informação capazes de armazenar e processar. O autor também enfatiza a importância de compreender o que é o *Big Data* para preparar-se para o futuro, que provavelmente contará com o crescimento exponencial na produção de dados digitais, podendo desfrutar de vantagens e benefícios proporcionados pelo seu uso.

Autores como Davenport (1998), Cunha e Cavalcanti (2008), e Souza e Almeida (2021), partilham do mesmo entendimento quando tratam o dado como sendo uma matéria-prima bruta, que representa uma informação cujo valor e importância informacional precisam de processamento manual ou automático para ser compreendido. Ainda com base na percepção dos autores, a informação seria o resultado do processamento e interpretação de dados, e o conhecimento o produto da junção dos elementos anteriores dotados de significado e representação.

Neste sentido, sugere-se que a atual sociedade está envolvida na produção acelerada de dados, que ao serem interpretados ganham o sentido de valor informacional e geram a noção de apropriação de um novo conhecimento.

2.1 CONTEXTUALIZANDO *BIG DATA*

Alguns autores partilham do mesmo entendimento quando o assunto é conceituar o termo *big data*, ALECRIM (2015); DUNNE (2012); IBM (2017) e KHAN et al.(2014), concordam que "*Big Data*" abrange uma vasta quantidade de dados, tanto estruturais, quanto não estruturados, que exercem influência significativa sobre os processos de tomada de decisão. No final dos anos 1990, o conceito começou a surgir, apontando para a presença de uma grande quantidade de dados sendo gerada cotidianamente por sistemas e outros elementos da tecnologia da informação.

Pence (2015) sugere que a quantidade de bytes necessários para armazenar arquivos de *Big Data* já ultrapassa as unidades mais baixas e convencionais de espaço, sendo o Petabyte a mais apropriada para o uso de ferramentas de *Big Data*.

Neste sentido, "A partir do Petabyte, pode-se dizer que é um tamanho apropriado para o uso do termo *big data*. A título de exemplo, toda a informação acumulada pela humanidade

desde o início da história até o presente, já está na faixa do Zettabyte” (Santos-D’Amorim, *et al.*, 2020, p. 6).

Tabela 1 – Medidas de dados de armazenamento digitais e exemplos

Unidade	Equivalência	Comparação Aproximada
1 Terabyte (TB)	1000 GB	A Biblioteca do Congresso dos EUA mantém 10 Terabytes de informação escrita.
1 Petabyte (PB)	1000 TB	1 Petabyte é equivalente a 250.000 DVDs.
1 Exabyte (EB)	1000 PB	Estima-se que todas as palavras já faladas por seres humanos na história seria cerca de 5 Exabytes.
1 Zettabyte (ZB)	1000 EB	Quinhentos trilhões de cópias de Guerra e Paz, de Tolstói.
1 Yottabyte (YB)	1000 ZB	Mais do que toda informação disponível no mundo.

Fonte: Santos-D’Amorim, *et al.* (2020).

Seguindo a lógica dos autores citados, até agora, faz sentido a busca pela definição mais completa do conceito de *Big Data*, que culminou na seleção de termos, cujos significados representam as características que melhor definem o *Big Data*, observando que este fenômeno abrange o processamento de grandes volumes de dados, que carregam peculiaridades, acerca do que informar, para quem informar e de que forma.

As definições de *big data* na literatura associam-se a um conjunto de características, os chamados, hoje, 7V’s do *big data*: volume, velocidade, variedade, valor, variabilidade, veracidade e visualização (Canbay; Vural; Sagioglu, 2018).

De acordo com os autores, o conjunto de dados que excedem a capacidade de gerenciamento dos sistemas convencionais em termos de volume, velocidade e diversidade são categorizados como dados em grande escala, daí a origem do termo *Big Data*, que na sua primeira versão conceitual, agregou apenas três características básicas, inicialmente: volume, velocidade e variedade. Que ficaram conhecidos como os 3vs do *big data*.

No entanto, ainda na perspectiva dos autores, novos elementos foram acrescentados a estes componentes devido às necessidades de novos desafios, chegando à versão apresentada e defendida de 7vs, conforme a Figura 1 a seguir:

Figura 1 – Os v's do *Big data*



Fonte: Elaborado pela autora (2023).

- a) Volume: esta característica está ligada ao aumento na quantidade de dados disponíveis para serem utilizados;
- b) Velocidade: diz respeito à duração da utilidade dos dados ao longo do tempo;
- c) Variedade: compreende diversas categorias de dados, que podem ser integradas visando aprimorar a análise e o processamento dos resultados de *big data*. Entretanto, quando se mesclam distintos tipos de dados pessoais, emergem novas informações que podem trazer um retrato detalhado de um indivíduo. Esta característica amplifica as preocupações com a privacidade, ao mesmo tempo em que reduz a utilidade dos dados
- d) Valor: quanto mais os dados são utilizados maior é o potencial de valor agregado;

- e) Variabilidade: requer maior cuidado por apresentar dados inconsistentes e discrepantes, comprometendo a utilidade;
- f) Veracidade: avalia a autenticidade das informações;
- g) Visualização: diz respeito à forma clara que os dados são apresentados.

2.1.1 A evolução dos 3vs do *big data* para os 7vs

Apresenta-se no quadro 1 a seguir a relação dos autores e os respectivos elementos do *big data* abordados por eles:

Quadro 1– Autores e a evolução dos “vs” do *big data*

Laney D (2001) Price (2020)	(IBM, 2019)	(Chen; Mao; Liu, 2014)	(Canbay; Vural; Sagiroglu, 2018)
3 Vs	4 Vs	6 Vs	7 Vs
Volume	Volume	Volume	Volume
			Velocidade
Velocidade	Velocidade	Velocidade	Variedade
			Veracidade
	Variedade	Variedade	Variabilidade
Variedade		Veracidade	Valor
	Veracidade	Variabilidade	Visualização
		Valor	

Fonte: Elaborado pela autora (2023)

A origem dos "3Vs" é frequentemente associada a Doug Laney, que os apresentou como Volume, Velocidade e Variedade em Laney D (2001), e pesquisas como a de (Price 2020) também recorreram aos “Vs” iniciais do *big data*. Esses três fatores ressaltaram a natureza

desafiadora dos dados, em termos de escala, taxa de geração e diversidade de tipos. Contudo, à medida que o uso e a análise de dados se tornaram mais complexos, tornou-se evidente que essa tríade não conseguia compreender totalmente a natureza diversa do *Big Data*. A IBM, empresa americana da área de informática, foi uma das protagonistas da revolução do *Big Data*, expandindo a abordagem dos "3Vs" para abranger um quarto V - a Veracidade, (IBM, 2019). Essa característica ressalta a importância da qualidade, precisão e confiabilidade dos dados, reconhecendo que o valor das percepções extraídas está, intrinsecamente, ligado à autenticidade das informações.

Chen; Mao; Liu (2014) acrescentaram mais 3vs com relação aos iniciais, apresentando 6vs que caracterizam o *big data*: veracidade, variabilidade e valor. Essas evoluções, que pararam nos 7vs descritos no subtópico anterior, revelam a complexidade, que tem sido lidar com a extensa produção de dados ao longo do tempo.

Em meio às pesquisas, até agora, consensuais Lugmayr, et al (2017) apresentam uma abordagem diferente para explicar o fenômeno *Big Data*. A pesquisa é direcionada pela proposta de responder uma pergunta que pode intrigar os usuários: “O que realmente é novo em *Big Data*?”. Os autores partem da ideia de estudar o termo, não mais com enfoque tecnológico, mas aproximando-o das investigações epistemológicas. O *Big Data Cognitivo*.

Na concepção de (Lugmayr, et al 2017) a visão que enquadra o *Big Data* em “Vs” é limitante e tendenciosa devido aos apelos tecnológicos e culturais que concentram forças em questões, que envolvem processamento de dados e escala de produtividade. O panorama cognitivo proposto pelos autores aposta na fusão entre tecnologia e ser humano para perceber e compreender os dados e, assim, gerar conhecimento e colaborar com a tomada de decisões.

De acordo com os autores, o *Big Data Cognitivo* abre caminho para ganhos competitivos e impulsiona a inovação em setores de negócios emergentes, introduzindo a cognição como um processo de comunicação interativa entre humanos e máquinas, resultando na formação de modelos mentais de gerenciamento de dados.

Nos últimos anos, o panorama da informação e da tecnologia tem sido palco de uma revolução marcante, impulsionada pela proliferação em larga escala de dados em todos os segmentos da sociedade. Esse crescimento exponencial serviu de alicerce para a concepção do conceito de *Big Data*, o qual se tornou um agente de fomento para a transformação digital, a tomada de decisões embasadas e a promoção da inovação. No entanto, à medida que a

compreensão das complexidades intrínsecas aos conjuntos de dados massivos avançou, os alicerces iniciais dos "3Vs" do *Big Data* (Volume, Velocidade e Variedade) foram enriquecidos para incorporar uma perspectiva mais abrangente - resultando nos "7Vs do *Big Data*", (Canbay; Vural; Sagioglu, 2018).

Afora as discussões, acerca do objetivo de definir e conceituar o termo *Big Data*, a relevância do uso de ferramentas, que favorecem grandes empresas a usufruir de vantagem competitiva e, conseqüentemente de seus impactos positivos, é motivo para incentivar a sociedade a ampliar o conhecimento sobre o tema.

No entanto, as dificuldades de implementação de políticas e ferramentas para o uso de *Big Data*, compreendendo custos com infraestrutura de computação e competências para gerir dados, foram vistas por (Villars; Olofson; Eastwood 2011), por exemplo, que naquele ano já percebiam os efeitos causados pela era da informação somatizando grande produção e acúmulo de dados. Segundo os autores, a taxa de expansão do *Big Data* supera as capacidades das estruturas de TI convencionais, o que implica, em grande parte, em novos desafios, tanto no âmbito computacional, quanto na administração de dados. Os autores falaram sobre a velocidade com que esse processo vem acontecendo e como as organizações são desafiadas a acompanhar o desenvolvimento constante e acelerado do *Big Data*.

Villars; Olofson; Eastwood (2011) indicam que o *Big Data* está, cada vez mais presente no dia a dia das pessoas comuns, permeando diversas áreas e influenciando as interações, decisões e experiências cotidianas, através das mídias sociais e entretenimento, compras online e comércio eletrônico, ensino online, dentre outras. Essa realidade expressada pelos autores sintetiza que o *Big Data* está redefinindo as interações humanas com o mundo, porém, traz consigo desafios relevantes a serem considerados no planejamento de utilização de tecnologias de *Big Data*. Ainda assim, no entendimento deles, os benefícios e implicações positivas, oriundas do uso dessa ferramenta revolucionária sobrepõe os obstáculos que podem ser encontrados na sua implementação.

2.1.2 Concepções de *Big Data* na Ciência da Informação

A conjuntura de produção e uso de grandes volumes de dados nos tempos modernos, mais precisamente, no que diz respeito às tecnologias de *big data*, estabelece um campo de muitas

oportunidades de atuação e estudos para a área da Ciência da Informação. Questões que envolvem, qualquer característica da informação, do ponto de vista qualitativo podem ser instrumento de investigação para a CI.

Alguns aspectos podem ser considerados, para fins de apresentação e conhecimento nesta pesquisa, como o estudo de Souza (et al., 2013) que optaram por analisar os possíveis impactos paradigmáticos, gerados pela tecnologia de *Big Data* na Ciência da Informação. Essa interpretação investiga os efeitos da realidade da ascensão das tecnologias de *Big Data* e seus desdobramentos exercendo influência sobre a área da CI. Os autores explicam como a banalidade, que foi dada ao termo “*Big Data*” para descrever, qualquer tipo de agrupamento de informações, principalmente informações textuais não estruturadas, desafia a Ciência da Informação a adaptar os mecanismos de armazenamento, e de avaliação da qualidade lidando com acervos gigantes e que aumentam o tempo todo.

Neste sentido a pesquisa de Pimenta (2013) explora nuances relativas ao controle de informações digitais na era do *Big Data*, bem como sua cooperação para preservação da memória, e faz reflexão à necessidade da sociedade ter a consciência, de que os dados produzidos e armazenados na *internet* não são esquecidos, ao serem apagados, tudo se torna “patrimônio” dos idealizadores e donos das plataformas digitais. O pesquisador estimula a iniciativa de estudos que analisem melhor questões sensíveis de privacidade de dados, controle e memória informacional no cenário do *Big Data*.

Milagre e Segundo (2015) propõem uma relação de conceitos da Ciência da Informação para ampliar as discussões e soluções, acerca da segurança da informação no ciberespaço, visando combater os crimes digitais praticados no campo de utilização de grandes volumes de dados, o *Big Data*. Os autores apresentam a iniciativa do estudo como uma oportunidade da Ciência da Informação contribuir com a perícia em informática, ou computação forense, para apoiar na formulação de soluções de enfrentamento às violações e fraudes, que ocorrem nas plataformas digitais, agravadas pelo *Big Data*.

Protocolos, metadados, ontologias, arquitetura da informação, métodos de indexação, estudos quantitativos sobre recuperação da informação, e as tecnologias da *web* semântica são exemplos de conceitos da Ciência da Informação, que podem aprimorar os sistemas de análise de *Big Data*, de acordo com Milagre e Segundo (2015). Essa perspectiva que coloca a Ciência da Informação, em destaque na conjuntura do *Big Data* pode ser embasada também pelos atributos

mencionados por (Borko, 1968), que entende a Ciência da Informação como a área de estudo das propriedades, comportamento e fluxos da informação.

Fagundes, Macedo e Dutra (2017) abordam critérios de avaliação da qualidade da informação em *Big Data*. Os autores consideram os 7 Vs do *Big Data* (Volume, Variedade, Velocidade, Veracidade, Valor, Variabilidade e Visualização) para abordar a qualidade da informação, segundo os critérios de (Lee, et al 2002), que são: acessibilidade, suficiência, credibilidade, completeza, representação concisa, representação consistente, facilidade de operação, exatidão, interpretabilidade e objetividade.

A proposta incentiva outras pesquisas, que relacionem critérios para definir se a informação no contexto do *Big Data* possui qualidade para tornar sua utilização eficiente.

Essas pesquisas são resultados de uma busca genérica, que foi realizada na Base de Dados em Ciência da Informação (BRAPCI) por trabalhos que relacionam “*Big Data* e Ciência da Informação”, tendo sido recuperados 83 artigos, no dia 01 de novembro de 2023. Este quantitativo mostra, que mesmo em uma busca genérica os estudos que abordam alguma relação entre os dois temas são escassos, dando margem para pouco embasamento teórico, e apontando para que novas pesquisas sejam realizadas.

2.2 BIG DATA E A PANDEMIA DE COVID-19

Durante a pandemia de Covid-19, a utilização do *Big Data* teve um papel fundamental ao viabilizar a captura, análise e interpretação, em tempo real, de extensos conjuntos de informações. Através dessa abordagem, foram obtidos conhecimentos preciosos que contribuíram para a compreensão da propagação do vírus, a formulação de cenários de previsões, a vigilância da disseminação viral e da eficácia das estratégias de contenção, além de fornecer diretrizes embasadas para a tomada de decisões no âmbito da saúde pública.

A revisão sistemática realizada por Corsi (et al, 2020) aborda o *Big Data Analytics* (BDA) no combate à epidemias e pandemias, uma versão do *Big Data* mais avançada, segundo os autores, que promete tratar os dados não apenas do ponto de vista quantitativo, mas também considerando o viés analítico qualitativo do termo. Essa visão se assemelha àquela apresentada por Lugmayr (et al, 2017) que também explora o *Big Data*, do ponto de vista para além da matéria, incluindo o conteúdo interpretativo e o desenvolvimento de conhecimento.

O objetivo da pesquisa de Corsi (et al, 2020) foi identificar quais tipos de dados e que fontes de dados podem auxiliar em situações de epidemias e pandemias, e como o *Big Data Analytics* (analítico) promove informações relevantes para ajudar no enfrentamento de casos virais de baixa e alta escala. Os autores utilizaram a análise bibliométrica e de conteúdo para quantificar, e em seguida descrever como *Big Data* e o *Big Data Analytics* foram utilizados, mais precisamente, no contexto da pandemia de Covid-19, através de um portfólio com o total de 45 artigos coletados em bases de dados.

Estudos como esse se tornaram muito comuns durante a pandemia de Covid-19, a falta de informação, acerca do novo coronavírus impulsionou pesquisadores na identificação e criação de ferramentas, e métodos para entender, prever e controlar a propagação do coronavírus, além de impeli-los a produzir e gerenciar informações de apoio para tomada de decisões, por parte de governos e a investir em mecanismos de controle e tratamento das pessoas infectadas.

Assim, a pandemia do coronavírus foi tema de tamanha repercussão que a quantidade de dados gerados em tempo real, ao passo que o vírus se alastrava, tornou o tema cada vez mais recorrente nos estudos de *Big Data*.

As medidas adotadas pelas autoridades governamentais em diversos países, bem como por profissionais e pesquisadores da área da saúde, no combate à pandemia de Covid-19 apresentam ações que, pode-se dizer terem sido embasadas pelo tratamento minucioso, analítico e profundo dos dados, que contavam o processo de avanço da pandemia e os números de infectados e de óbitos.

Prova disso é a pesquisa de Mehta e Shukla (2021), que faz um levantamento das iniciativas tomadas pelos países, diante da pandemia de Covid-19. O surgimento do vírus, bem como seu alto índice de contaminação, e a falta de conhecimento prévio a seu respeito, culminaram para que governos e autoridades viessem recorrer às ferramentas tecnológicas existentes, para extrair dados de análise de *big data* na tentativa de conter a proliferação do coronavírus.

Os autores apresentaram exemplos de alguns países, que adotaram aplicativos de rastreamento de contato, como na China, país originário da pandemia, que empresas de tecnologia forneceram dados para rastrear o fluxo populacional de conter o vírus. Em Taiwan, a análise de *big data* contribui para que medidas de distanciamento social fossem tomadas, através do rastreamento de localização de telefones celulares. Com o intuito de disponibilizar dados em

tempo real, para orientar os responsáveis por decisões, o Serviço Nacional de Saúde do Reino Unido fez parceria com a *Microsoft*, *Amazon Web Services*, *Google*, Faculdade e *Palantir* para construir uma plataforma digital baseada em *big data*, IA e nuvem de tecnologia de computação. Outros territórios estudados foram Estados Unidos, Nações da União Europeia, Cingapura, Coreia do Sul, Israel, Austrália, Emirados Árabes Unidos e Índia. Todas essas localidades tiveram em comum o uso de big data e inteligência artificial para providenciar as intervenções contra a alta contaminação pelo coronavírus.

No campo da saúde, a pandemia de Covid-19 impulsionou o uso de *big data* na busca por informações, por exemplo, da sequência genômica do SARS-CoV-2 e assim descobrir novos medicamentos, ou medicamentos conhecidos no tratamento de outras doenças, que pudessem gerar resultados positivos se usados para combater a Covid-19 (Hu, et al, 2020).

Mendonça e Dantas (2020) trazem à luz o conceito de “transformação digital” para falar do uso de *big data*, inteligência artificial e análise de dados na área da saúde. Os autores acreditam que o alto custo para investimento em tecnologia da informação, acessibilidade e as questões de privacidade dos dados são fatores fundamentais que dificultam a expansão do uso de dados eletrônicos do quadro clínico de pacientes, como prontuários eletrônicos, e da telemedicina para apoiar os sistemas de saúde. No entanto, quando há fomentos para promover o acesso a bancos de dados digitais da saúde facilita a condução, principalmente, de eventos emergenciais como a pandemia de Covid-19. Como exemplo disso, foi citada a medida similar adotada no Brasil, por meio da Lei nº. 13.989, que permitiu o uso emergencial da telemedicina durante a crise causada pelo coronavírus (SARS-CoV-2), com financiamento público para os serviços de telemedicina oferecidos pelo SUS (Mendonça; Dantas 2020).

Outra pesquisa, que também evidencia as mudanças experimentadas pela medicina com a ascensão do *big data* e de sistemas de inteligência artificial durante a pandemia, foi conduzida por Neto, Silva e Nogaroli (2020). Os autores versam sobre o uso dessas tecnologias no diagnóstico da Covid-19, através do uso de ferramentas de inteligência artificial, alimentadas por dados clínicos registrados em banco de dados, e como a iniciativa desse método pode se expandir auxiliando no diagnóstico de doenças preexistentes e na escolha de tratamento médico.

Haafza, et al (2021) chamam de tecnologias de *Big Data* as ferramentas que servem para armazenar os grandes volumes de dados produzidos durante a pandemia acerca da propagação do vírus. Os autores explicam que diversas fontes que podem ser acessadas para investigar esses

dados se configuram como *Big Data*, tal como as redes sociais e dispositivos de *internet*. A revisão sistemática da literatura realizada por Haafza (et al, 2021) tem por finalidade investigar qual foi o papel das aplicações de *big data* utilizadas na pandemia, como elas foram utilizadas e quais foram os principais aplicativos, realizando buscas em bases de dados por artigos que correspondem ao tema. A pesquisa contou com a inclusão de 60 artigos, dos quais os autores escolheram alguns para apresentar as principais aplicações de *big data* identificadas como soluções para o enfrentamento da Covid-19 como:

- a) técnicas geoespaciais, que na percepção do pesquisadores pode se configurar como uma tendência;
- b) inteligência de dados em tempo real, que facilitou a vigilância da COVID-19;
- c) Diagnóstico de casos de COVID-19;
- d) tomada de decisão em cuidados de saúde;
- e) estrutura de aplicação de *Big Data* para previsão e monitoramento da pandemia;
- f) Sistema de arquivos distribuídos Hadoop.

Os obstáculos mencionados por Haafza (et al, 2021) durante a pesquisa, dizem respeito às questões de privacidade e da segurança dos dados de saúde, bem como problemas de segurança dos pacientes, e de como se deu a partilha desses dados com organizações empresariais.

O estudo de Lv Y (et al, 2021) se assemelha ao de Haafza (et al, 2021), os autores, de igual modo, conduziram uma revisão sistemática da literatura, desta vez orientada pelas diretrizes PRISMA (revisão sistemática e meta-análise), visando explorar o potencial da abordagem de *Big Data* no gerenciamento da pandemia de Covid-19.

As buscas foram realizadas em bases de dados compreendendo os últimos 10 anos, e foram identificados trinta e dois documentos considerados, mais relevantes para a análise qualitativa.

Os resultados foram apresentados, relacionando às principais fontes de *Big Data* relacionados ao Covid-19 (Mídia social, Bancos de dados de imigração e alfândega, Banco de dados COVID-19/dados de saúde, Dados móveis, Tecnologia móvel, Sistema de transporte público, Transação com cartão bancário, Câmera de circuito fechado/câmera de segurança, Sistema de posicionamento geográfico do carro (GPS), Dispositivo de rastreamento vestível); Técnicas potenciais de *big data* para gestão da pandemia de COVID-19 (Modelagem, Aprendizado de máquina, Mineração de dados, Visualização, Estatísticas, Simulação, Otimização,

Mineração de texto, Previsão, Análise de redes sociais); e Potenciais aplicações de *big data* para gestão da pandemia de COVID-19 (Identificação de infecção, Histórico de viagens, Sintomas de febre, Detecção precoce, Identificação da transmissão, Informações prontas em período de bloqueio, Movimento popular, Desenvolvimento de tratamentos e vacinas). Dentre as dificuldades da pesquisa, os autores também relataram a questão relativa à privacidade dos dados dos envolvidos e ainda às questões ligadas à ética, e acreditam nas contribuições das revisões sistemáticas para ampliar os conhecimentos acerca do assunto estudado.

Da mesma forma Bragazzi (et al, 2020) exploram aplicações de *Big Data* no cenário da pandemia de Covid-19, dessa vez somadas à inteligência artificial. Os autores apresentaram uma revisão sistemática que reúne nomes de pesquisadores, que contribuíram com estudos representados pelas seguintes categorias:

- a) aplicações de curto prazo de inteligência artificial e big data: um alerta pandêmico rápido e eficaz (abordam o monitoramento e a visualização de dados relativos aos casos de Covid-19 em tempo);
- b) aplicações de curto prazo de inteligência artificial e big data: rastreamento e diagnóstico de casos de COVID-19 (tecnologias que ajudam na previsão do vírus e na detecção da doença);
- c) aplicações de Inteligência Artificial e *Big Data* a Médio Prazo: Identificando um Potencial Tratamento Farmacológico (em forma da descoberta de drogas para combater o vírus e vacinas);
- d) aplicações de Inteligência Artificial e *Big Data* a Médio Prazo: Facilitando a Implementação de Intervenções de Saúde Pública (políticas de saúde que visam o controle de infectados);
- e) aplicações de longo prazo de inteligência artificial e *big data*: construindo cidades inteligentes, saudáveis e resilientes (contribuem para a vigilância da população).

3 PROCEDIMENTOS METODOLÓGICOS

3.1 METODOLOGIA DA PESQUISA

Esta seção descreve a metodologia empregada na elaboração deste trabalho. Antes de iniciar a descrição metodológica, é apresentado o quadro 2, que tem como finalidade trazer à memória os elementos que norteiam o delineamento da pesquisa e pode auxiliar na compreensão desta.

Quadro 2 – Delineamento da Pesquisa

Objeto de Estudo	O uso de ferramentas baseadas em <i>Big Data</i> no enfrentamento da pandemia da Covid-19.
Objetivo Geral	Identificar as soluções baseadas em <i>big data</i> que auxiliaram no combate à pandemia da Covid-19
Objetivos Específicos	• Definir o termo <i>big data</i>; • Identificar as principais ferramentas e formas de utilização das mesmas durante a pandemia; • Apresentar estas ferramentas por categorias (análise temática);

Fonte: Elaborado pela autora (2023)

3.2 CARACTERIZAÇÃO DA PESQUISA

A metodologia empregada na pesquisa, quanto aos meios, tem natureza qualitativa, pois o objeto é analisado, por meio da qualidade na atribuição do caráter da importância de outras ciências, em especial a Ciência da Informação, no enfrentamento ao coronavírus, embora tenham ocorrido algumas avaliações quantitativas nas etapas de análise da pesquisa.

Segundo Michel (2009), pesquisas qualitativas auxiliam nas interpretações de fenômenos à luz de um contexto e, são baseadas na interpretação do pesquisador sobre o fenômeno investigado. Denzin e Lincoln (2006) seguem a mesma linha de raciocínio abordada por Michel

(2009), considerando a pesquisa qualitativa como a compreensão dos fenômenos sociais e dos significados atribuídos à informação por meio de investigações interpretativas e contextuais. Laville e Dionne (1999) afirmam que o caráter qualitativo difere-se do quantitativo, por este último focalizar a determinação de quantidade/medição, o que deixa de lado as respostas substanciais que se pode concluir com a análise do objeto de estudo sob o caráter qualitativo.

A caracterização da pesquisa, quanto aos fins a classifica como pesquisa descritiva, pois faz uso das informações coletadas (Marconi; Lakatos, 2003), por exemplo, a representação do conteúdo dos sistemas de informação, em temas distintos localizados em fontes primárias e secundárias. Michel (2009) aponta que a pesquisa descritiva tem como objetivo descrever, explicar e interpretar características de um fenômeno ou de uma população. O principal objetivo desse tipo de pesquisa é obter informações detalhadas e precisas sobre um determinado tema, sem a intenção de estabelecer relações de causa e efeito.

Para tanto, perguntas do tipo "o quê", "quem", "quando", "onde" e "como", são frequentes na pesquisa descritiva com o intuito de fornecer uma visão geral e sistematizada dos fatos ou fenômenos estudados. E assim, este tipo de pesquisa é lastreado na observação, coleta de dados e análise de informações para produzir descrições objetivas e factuais (Michel, 2009).

Ainda segundo a autora, este tipo de pesquisa normalmente é realizada por meio de diferentes técnicas e métodos, como questionários, entrevistas, observações, análise de documentos e estatísticas. Ao conduzir uma pesquisa descritiva, o pesquisador não busca explicar as causas ou relações entre as variáveis, mas sim fornecer uma representação precisa do objeto de estudo.

A pesquisa também é considerada básica, esta é classificada por Gil (2010) como aquela que não tem aplicabilidade imediata e generalista. A pesquisa básica procura os princípios e fundamentos do mundo, das coisas, seu funcionamento e sua intenção é desvendar características, propriedades básicas do fenômeno. Michel (2009) aponta que o foco da pesquisa básica está na geração de conhecimento novo, na exploração de conceitos e na busca de respostas para questões teóricas.

A autora contrapõe este tipo de investigação à pesquisa aplicada, que visa resolver problemas específicos ou desenvolver aplicações práticas, já a pesquisa básica tem um caráter mais teórico e exploratório. Ela busca entender os fenômenos naturais ou sociais em sua essência, sem necessariamente ter uma aplicação imediata (Michel, 2009). A autora também aponta que,

embora a pesquisa básica possa não ter uma aplicação direta e imediata, ela desempenha um papel crucial no avanço científico, fornecendo a base teórica e conceitual para a pesquisa aplicada e contribuindo para a compreensão geral dos fenômenos naturais e sociais.

Quanto aos métodos específicos, Marconi e Lakatos (2003) associam a pesquisa documental ao procedimento de coleta de dados em documentos nas fontes primárias, o que Vechiato (2013) intitula coleta de dados documentais. Outro procedimento empregado na coleta de dados é aquele que Santos (2006) chama de pesquisa bibliográfica ou coleta de dados bibliográficos (Vechiato, 2013). A pesquisa bibliográfica tem por objetivo proporcionar ao investigador os “meios técnicos de garantir a objetividade e a precisão do estudo” (Michel, 2009, p. 52). Ainda segundo a autora, este método fornece a orientação necessária à realização da pesquisa social, sobretudo, no que se refere à obtenção, processamento, análise e validação dos dados pertinentes à problemática que está sendo investigada”.

Ambas tipologias metodológicas estão presentes nas etapas de composição da pesquisa, por exemplo, nas revisões da literatura descritas generalistas, utilizadas para conceber o referencial teórico; e no mapeamento sistemático que descrevem de modo mais sintético os resultados desta pesquisa.

3.3 MÉTODO - MAPEAMENTO SISTEMÁTICO DA LITERATURA (MSL)

O método utilizado nesta pesquisa é o Mapeamento Sistemático da Literatura (MSL), que apresenta um maior grau de formalismo. Segundo Pettigrew e Roberts (2008), o MSL é um método concebido para dar sentido a grandes conjuntos de dados e, ao mesmo tempo, um meio de responder perguntas sobre estas informações.

O MSL é um método de mapeamento de áreas de incerteza, identificando onde pouca, ou nenhuma pesquisa significativa tem sido feita, e onde são necessários novos estudos. De maneira geral, os estudos sistemáticos, que englobam as Revisões Sistemáticas da Literatura (RSL) e o MSL, mostram-se úteis em áreas onde a incerteza predomina e que há pouca evidência para apoiar qualquer teoria.

A diferença entre o MSL e o RSL se dá pelo objetivo da pesquisa. No MSL se deseja fazer um levantamento geral da área de interesse e entender como se construiu o cenário científico, em torno daquele assunto, até aquele momento. Na RSL se deseja responder uma

questão de pesquisa específica, com base em outros trabalhos, que também responderam a mesma questão de pesquisa. Ou seja, o MSL é um apanhado geral das respostas encontradas em diversos trabalhos primários sintetizados em um único trabalho.

Pettigrew e Roberts (2008) elencam alguns motivos, para adotar estudos sistemáticos, ao invés de pesquisas bibliográficas tradicionais. O primeiro ponto é que os resultados de estudos individuais têm uma credibilidade maior do que eles merecem, afinal, poucos estudos são metodologicamente sólidos, e apresentam resultados generalizáveis, a ponto de apresentarem resultados que representem uma boa aproximação da realidade e que possam ser aceitos como suas conclusões efetivas. De forma alguma os autores diminuem a importância dos estudos individuais, mas eles afirmam que a maioria das pesquisas só podem ser entendidas dentro de um determinado contexto.

Além da inconsistência dos trabalhos individuais, um outro problema levantado por Pettigrew e Roberts (2008) é o excesso de informação disponível sobre pesquisas. Com novas revistas sendo lançadas anualmente, e milhares de artigos publicados, é bastante difícil até mesmo para o pesquisador mais enérgico se manter atualizado, com os estudos mais recentes.

Heaton (2008) aponta que os estudos sistemáticos podem ser chamados, também, de estudos secundários por (i) utilizarem fontes de dados primárias, como fonte de evidências se tornando assim uma fonte secundária. Azevedo (2012) afirma que uma fonte secundária utiliza informações sobre documentos primários, que sintetizam os dados, filtrados e organizados, e a relevância dos resultados vem, a partir da diversidade das fontes. E (ii) por ser uma análise secundária. Littel (2006) afirma que análises secundárias envolvem o uso de dados já existentes, recolhidos de estudos prévios, que são utilizados para prosseguir uma pesquisa, que é distinta do trabalho original. Esta nova pesquisa pode ser uma nova pergunta, ou uma perspectiva alternativa sobre a pergunta original.

Tranfield, Denyer e Smart (2003) apontam que os estudos sistemáticos são, particularmente, úteis porque eles sugerem que as conclusões encontradas são baseadas em evidências derivadas de estudos primários. Esse movimento de estudos baseados em evidência, teve início na década de 1980, quando o governo britânico passou a valorizar que as políticas e práticas fossem informadas, através de uma base de evidências mais rigorosa e exigente.

A iniciativa “três E” (economia, eficiência e eficácia) chamou a atenção para a prestação de serviços públicos e têm levado ao desenvolvimento de orientações detalhadas e manuais de

melhores práticas em diversas disciplinas. A eficácia neste contexto está voltada, tanto com a adequação, quanto com a validade dos métodos utilizados pelos profissionais no seu dia-a-dia.

Assim, um movimento de práticas baseadas em evidências foi desenvolvido, e em maio de 1997 Tony Blair aprovou uma abordagem pós-ideológica no serviço público, onde a evidência seria levada ao centro da etapa no processo de tomada de decisão. A partir disso, diversas áreas desenvolveram suas comunidades de práticas baseadas em evidências, tais como a Medicina (Sackett, 2000), Computação (Kitchenham; Dyba; Jorgensen, 2004), Fisioterapia (Herbert, 2005), Administração (Pfeffer; Sutton, 2006) e Biblioteconomia (Madge, 2011).

A Biblioteconomia baseada em evidências é a aplicação da abordagem interdisciplinar conhecida como prática baseada em evidências (EBP) para problemas na área de Biblioteconomia e Ciência da Informação. Isto significa que todas as decisões práticas feitas dentro CI devem: (1) basear-se em estudos de investigação e (2) que estes estudos sejam selecionados e interpretados de acordo com algumas normas específicas para EBP (Roddham, 2004).

Pettigrew e Roberts (2008) argumentam que um protocolo de pesquisa deve incluir detalhes de como diferentes tipos de estudos serão localizados, avaliados e sintetizados. No protocolo de pesquisa, todos os passos do método são definidos e planejados. Por exemplo, engenhos de busca, palavras-chave, restrições, limitações e análise dos resultados. Uma vez realizada a apresentação do MSL e a indicação do uso do mesmo, como método de pesquisa,

acredita-se que, antes de se iniciar esta pesquisa, faz-se necessário aplicar o MSL.

Para esta pesquisa, foram aplicadas as etapas (i) análise preliminar da literatura, (ii) desenvolvimento do protocolo de pesquisa e (iii) execução do MSL. O desenho geral da pesquisa é apresentado na Figura 1.



Fonte: Elaborado pela autora (2023)

A primeira etapa da pesquisa foi baseada em uma pesquisa bibliográfica *ad-hoc* sobre ferramentas de *Big Data* durante a pandemia da Covid-19, abrangendo artigos de diversas áreas de conhecimento. Esta primeira etapa não fez parte do MSL em si, mas, auxiliou para a compreensão do tema.

Em seguida, foi criado um protocolo de pesquisa que guiou a condução do MSL apresentado nesta pesquisa. Após a criação do protocolo, decidiu-se realizar a execução de um piloto para fazer uma validação do mesmo. A principal intenção do piloto foi diminuir o viés da pesquisa, a partir de ajustes no protocolo em caso de problemas (passo iv).

Em seguida, realizou-se a execução do MSL. Para um melhor entendimento de cada atividade a ser realizada nesta pesquisa, é apresentada na Figura 2 a execução detalhada de cada atividade. Estas atividades guiaram toda a concepção desse capítulo de procedimentos metodológicos.

Rememorando que as atividades relativas à análise preliminar da literatura não fazem parte do MSL e, deste modo, adiantou-se para a próxima etapa (desenvolver o protocolo de pesquisa) para apresentar o protocolo do MSL executado nesta pesquisa. Desta forma, as próximas seções apresentam partes do protocolo de fato e englobam as etapas posteriores (execução do piloto até a execução do MSL). Então, é iniciada a apresentação dos elementos do protocolo ao apresentar a necessidade da execução do MSL proposto.

3.3.1 Necessidade do Mapeamento Sistemático da Literatura

Pettigrew e Roberts (2008) afirmam que, antes de se iniciar um estudo secundário, se faz necessário realizar uma pergunta, que é frequentemente esquecida: “é realmente necessária a condução de uma revisão sistemática neste tema?” Os autores sugerem, que atualmente as

revisões sistemáticas estão em alta e, por muitas vezes se assume que uma nova revisão sistemática sobre um determinado conteúdo deve ser conduzida.

Entretanto, em alguns casos, não está claro se conduzir uma nova revisão sistemática é o melhor caminho, a seguir. Mais importante ainda, uma revisão sistemática pode não ser o tipo certo de estudo para responder às necessidades identificadas.

Assim, faz-se necessário ter a certeza de que uma nova revisão sistemática será mais útil do que um novo estudo preliminar. Mas, e se não houver um estudo sistemático sobre o tema? Pettigrew e Roberts (2008) sugerem que uma revisão sistemática pode ser útil para justificar, que novas pesquisas venham a ser conduzidas na área em questão.

Tendo em consideração o detalhamento de informações acerca do método de mapeamento sistemático da literatura, sua abrangência e finalidades, conforme explicitado a partir da seção 3.3, o MSL se adequa a esta pesquisa compreendendo a proposta de um tema pouco estudado, à época pandêmica, a fim de apresentar uma visão panorâmica acerca das soluções de enfrentamento do coronavírus.

Desta forma, o primeiro passo desta pesquisa foi procurar por outras revisões sistemáticas, sejam RSL ou MSL, sobre a utilização de ferramentas *Big Data* aplicadas no contexto da Covid-19, em bases de dados nacionais e internacionais, referentes às diversas áreas de conhecimento. Esta busca em bases de dados diferentes das utilizadas no protocolo de pesquisa se deu devido a ideia inicial da pesquisa de explorar várias bases de dados, concepção que não foi adiante devido ao pouco tempo para explorar os resultados. O Quadro 3 apresenta as bases de dados onde foram buscadas revisões sistemáticas sobre *Big Data* aplicadas à Covid-19 para este momento.

Quadro 3 – Engenhos de Busca utilizados na avaliação da necessidade

Engenho de Busca	Endereço	Termo
Google Acadêmico	https://scholar.google.com/	Covid AND “ <i>Big Data</i> ” AND "Systematic Review"
Scopus	www.scopus.com	TITLE-ABS-KEY("Covid" and “ <i>Big data</i> ” and "Systematic Review")
Web of Science	http://apps.webofknowledge.com/UA_AdvancedSearch_input.do?SID=2CwRnqgKSvtluTFdiLb&product=UA&search_mode=AdvancedSearch	TS=("Covid" and "Big data" and "Systematic Review")
Spring	https://www.springer.com/?SGWID=0-102-13-0-0	"Covid" AND "Big Data" AND "Systematic Review"
E-Lys	http://eprints.rclis.org/	“Covid” AND “ <i>Big Data</i> ”
LISTA	www.libraryresearch.com/	“Covid” AND “ <i>Big Data</i> ”
Scielo	http://www.scielo.br/cgi-SocialMachines/bin/wxis.exe/iah/?IsisScript=iah/iah.xis&base=title&fmt=iso.pft&lang=i	“Covid” AND “ <i>Big Data</i> ”
Brapci	http://www.brapci.ufpr.br/brapci/	“Covid” AND “ <i>Big Data</i> ”

Fonte: Elaborado pela autora (2023)

Esta busca ocorreu em Abril de 2021, e naquele momento não houve nenhum estudo secundário associado. Entretanto, em momentos posteriores, durante esta pesquisa, foram

identificados alguns estudos secundários que tratam deste assunto. Estes estudos são listados no quadro 4 a seguir.

Quadro 4 – Estudos secundários referentes a *Big Data* no contexto da Covid

Ano	Título	Citação
2021	A systematic review and IoMT based big data framework for Covid-19 prevention and detection	HAMID, Soomaiya et al. A systematic review and IoMT based big data framework for Covid-19 prevention and detection. Electronics, v. 11, n. 17, p. 2777, 2022.
2021	Big data analytics as a tool for fighting pandemics: a systematic review of literature.	CORSI, Alana et al. Big data analytics as a tool for fighting pandemics: a systematic review of literature. Journal of ambient intelligence and humanized computing, v. 12, n. 10, p. 9163-9180, 2021.
2021	A systematic review on artificial intelligence/deep learning applications and challenges to battle against Covid-19 pandemic.	MANOJ, Agrawal; SHWETA, Agrawal. A systematic review on artificial intelligence/deep learning applications and challenges to battle against Covid-19 pandemic. Disaster Advances, p. 90-99, 2021.
2021	Big data Covid-19 systematic literature review: Pandemic crisis.	HAAFZA, Laraib Aslam et al. Big data Covid-19 systematic literature review: Pandemic crisis. Electronics, v. 10, n. 24, p. 3125, 2021.
2022	Big Data Analytics in Patient Care During Coronavirus Disease 2019 (Covid-19) Pandemic: A Systematic Review.	ANDY, Andy; EL-HAMZAH, Salsabilla Maula Zalfa; SINATRA, Jadeny. Big Data Analytics in Patient Care During Coronavirus Disease 2019 (Covid-19) Pandemic: A Systematic Review. JMMR (Jurnal Medicoeticolegal dan Manajemen Rumah Sakit), v. 11, n. 3, p. 245-263, 2022.

Fonte: Elaborado pela autora (2023)

Este mapeamento sistemático da literatura difere do trabalho de Hamid, et. al (2021) pois os dispositivos médicos no contexto da *internet* das coisas (IoT) representam, apenas, uma das formas de onde a tecnologia e o *big data* atuaram no combate à pandemia.

Assim, o trabalho do Hamid, et al. (2021) representa um subconjunto das possibilidades aqui investigadas. Situação semelhante é a do trabalho de Manoj e Shweta (2021) que apresentam as potencialidades da inteligência artificial e aprendizado profundo para auxiliar as pessoas no período pandêmico. De igual modo, este também representa um subconjunto das soluções aqui investigadas. O trabalho de Andy, El-Hamzah e Sinatra (2022) também se apresenta como um subconjunto específico de soluções de Covid e *big data*. O trabalho destes autores se concentram nas soluções voltadas aos cuidados aos pacientes já infectados. O trabalho de Corsi et al. (2021) está mais voltado aos dados brutos e conjuntos de dados específicos utilizados pelos pesquisadores, também se configurando como um subconjunto deste trabalho, que está mais próximo de soluções de tecnologias, e não só os dados, para o combate da Covid.

O estudo de Haafza, et al(2021) é o que se assemelha em objetivos a esta pesquisa. Os autores pretendem fazer um levantamento geral de como soluções baseadas em *big data* podem atuar no contexto da pandemia. Ressaltamos que os autores realizaram uma revisão sistemática da literatura em bases de dados tecnológicas: a da *Association Computer Machinery* (ACM) e a da *Institute of Electrical and Electronic Engineers* (IEEE). E por isso, o mapeamento sistemático da literatura compreendendo esta pesquisa, irá em uma linha complementar a do trabalho do Haafza, et al(2021) e iremos olhar para outras bases de dados.

3.3.2 Análise Preliminar da Literatura (i)

Na análise preliminar da literatura (i) pertinente sobre o objeto de pesquisa que, neste caso, engloba a relação entre uso de *Big Data* e a pandemia de Covid-19, foi realizada uma revisão de literatura aberta e sem sistematização (*ad-hoc*), que teve como intuito uma familiarização com o tema, e a identificação dos trabalhos publicados sobre *Big Data* no enfrentamento da Covid-19, na Ciência da Informação. Nesta etapa inicial foi percebida uma característica marcante dos 7 vs do *Big Data*, o grande volume de conteúdo gerado, em torno da temática pandemia e Covid-19. Este fato reforçou a necessidade do uso do MSL como método, por ter o caráter de abrangência,

em relação às revisões sistemáticas. Os objetivos da pesquisa compreendem aqueles citados no Quadro 2 de Delineamento da Pesquisa.

3.3.3 Elaboração do Protocolo de Pesquisa (ii)

Segundo Pettigrew e Roberts (2008), o protocolo de pesquisa é um elemento essencial para a realização de um estudo secundário e deve incluir um ordenamento sistemático de como o mapeamento irá ocorrer. Os autores apontam que um protocolo de pesquisa deve incluir detalhes de como diferentes tipos de estudos serão localizados, avaliados e sintetizados. Neste elemento, que conduz o MSL, todos os passos do método são definidos e planejados. Desta maneira, o protocolo para o MSL contemplou o seguinte planejamento:

A) - Bases de dados consultadas e *Strings* de busca

O Quadro 5 mostra a definição das bases de dados que foram definidas e as *Strings* de busca que recuperaram os documentos, que relacionam *Big Data* e Covid-19, compreendendo o período de 01 de Março de 2020 a 01 de Junho de 2023. A data final da pesquisa se deu pelo momento em que os artigos foram extraídos das bases de dados para passar pela primeira análise.

Quadro 5 – Estratégias de busca

Base	Estratégias de busca	Link
BRAPCI	Covid AND “Big Data”	https://www.brapci.inf.br/
<i>Web of Science</i>	((ALL=(Covid)) AND ALL=("Big Data")) AND ALL=("Information Science")	https://www-webofscience.ez16.periodicos.capes.gov.br/wos/woscc/basic-search

Fonte: Elaborado pela autora (2023)

B) - Execução da Busca

Na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) foram recuperados 2 documentos, utilizando a *string* de busca com delimitação da busca “2020 a 2023” e ordenando por “Mais antigos”. Na *Web of Science (WoS)*, utilizando a *string* de busca no buscador do campo “pesquisa avançada”, um total de 58 artigos foram retornados. As buscas na *WoS* foram possíveis de forma gratuita, através do Portal de Periódicos da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) disponível para pesquisadores, por meio de login e senha, que possuem e-mail institucional na instituição de ensino superior Universidade Federal de Pernambuco (UFPE). A coleta foi realizada no dia 01 de junho de 2023.

C) - Critérios de Inclusão e Exclusão:

Quadro 6 – Critérios de Inclusão e Exclusão

Inclusão	Exclusão
Publicados em Periódicos ou Anais em forma completa;	Publicados em Periódicos ou Anais em forma de resumo ou resenha;
Trabalhos em Língua Portuguesa ou Inglesa;	Publicados em fontes não confiáveis;
Trabalhos publicados a partir de 2020;	Trabalhos em Língua diferente de Português ou Inglês;
Trabalhos que apresentem o uso de alguma ferramenta;	Trabalhos publicados antes de 2020;
Que relacionem <i>Big Data</i> e Covid-19.	Trabalhos que não apresentam o uso de ferramentas;
	Que não relacionem <i>Big Data</i> e Covid-19.

Fonte: Elaborado pela autora (2023)

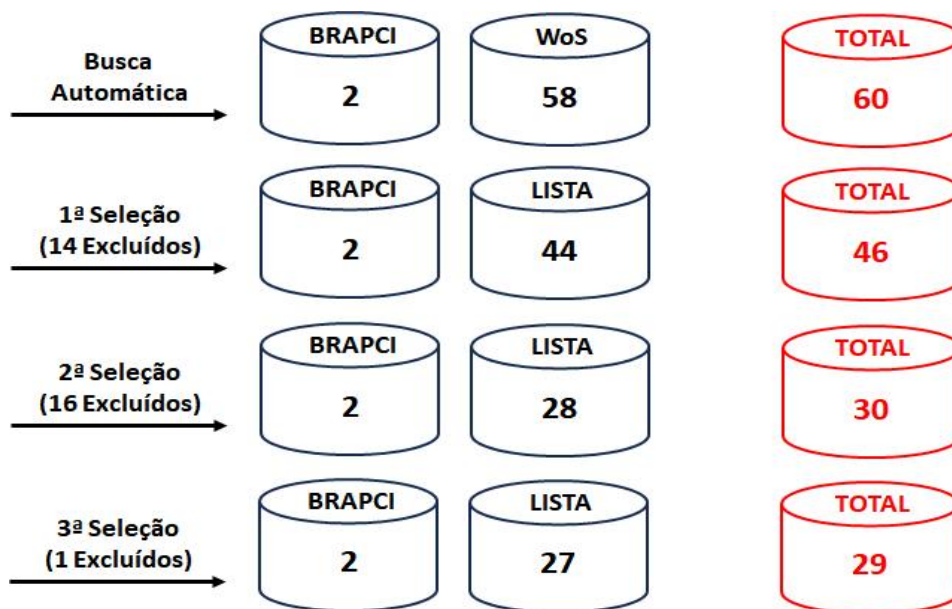
D) - Seleção

A seleção das fontes primárias foi realizada em três passos. Inicialmente, foram observados os quatro primeiros critérios de exclusão, que se referem ao tipo de artigo, idioma da publicação, data da publicação e o veículo, em que ele foi conduzido. O segundo passo se deu com a utilização dos outros dois critérios de exclusão dos trabalhos restantes, entretanto eram observados apenas o título e o resumo dos mesmos. Por fim, nos trabalhos restantes, foram adotados os critérios de inclusão e exclusão considerando todo o texto.

Considerando as duas bases de dados foram recuperados, inicialmente 60 trabalhos, 2 na BRAPCI e 58 na *Web of Science*, a primeira análise, do total de 60 artigos encontrados nas duas bases de dados, 46 artigos foram incluídos para realização de análise mais detalhada dos resumos. Os 14 excluídos nesta primeira etapa eram resumos, ou estavam em idiomas diferentes de português e inglês. Em seguida, foram analisados os títulos e resumos e nesta segunda etapa foram excluídos outros 16 trabalhos, restando 30 trabalhos. Por fim, na última etapa, foram analisados os trabalhos em sua forma completa e apenas 1 foi excluído, totalizando assim 29 trabalhos aceitos para a análise. A Figura 3 apresenta um sumário deste processo de seleção.

As referências foram catalogadas na ferramenta JabRef para o controle e organização das duplicatas e o que foi descartado e aceito.

Figura 3 – Sumário da Seleção



Fonte: Elaborado pela autora (2023)

A relação dos trabalhos utilizados como *corpus* desta pesquisa está presente no Quadro 7 abaixo.

Quadro 7 – Trabalhos Seleccionados

Autores	Título	Ano	Identificador
Sha. <i>et al.</i>	A State-Level Socioeconomic Data Collection of the United States for Covid-19 Research	2020	1
Bhattacharya <i>et. al.</i>	AI-Driven Agent-Based Models to Study the Role of Vaccine Acceptance in Controlling Covid-19 Spread in the US	2021	2
Adiga <i>et. al.</i>	All Models Are Useful: Bayesian Ensembling for Robust High Resolution Covid-19 Forecasting	2021	3
Taeb, Chi e Yan	Applying Machine Learning to Analyze Anti-Vaccination on Tweets	2021	4
Zhou <i>et. el</i>	Artificial intelligence in Covid-19 drug repurposing	2020	5
Wang <i>et al.</i>	CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting	2022	6
Mangalagiri <i>et. al.</i>	Classification of Covid-19 using Deep Learning and Radiomic Texture Features extracted from CT scans of Patients Lungs	2021	7
Zhou <i>et. al.</i>	Covid-19: Challenges to GIS with Big Data	2020	8
Ferlin <i>et. al.</i>	Covid-19, Big Data, Smart City, e Cidade Digital Estratégica na geração de conhecimentos	2021	9
Walker e Sharma	Data Visualization Tool for Covid-19 and Crime Data.	2021	10
Adiga <i>et al.</i>	Data-Driven Modeling for Different Stages of Pandemic Response	2020	11
Mehrab <i>et al.</i>	Evaluating the Utility of High-Resolution Proximity Metrics in Predicting the Spread of Covid-19	2022	12

Wang <i>et. al.</i>	Examining Deep Learning Models with Multiple Data Sources for Covid-19 Forecasting	2020	13
Tan <i>et.al.</i>	Fighting Covid-19 with Fever Screening, Face Recognition and Tracing	2020	14
Peddireddy <i>et. al.</i>	From 5Vs to 6Cs: Operationalizing Epidemic Data Management with Covid-19 Surveillance	2020	15
Menon <i>et. al.</i>	Generating Realistic Covid-19 x-rays with a Mean Teacher + Transfer Learning GAN	2020	16
Hu <i>et al.</i>	Human mobility data in the Covid 19 pandemic characteristics applications and challenges	2021	17
Jiao, Degen, Azimian	Identifying Hospital Deserts in Texas Before and During the Covid-19 Outbreak	2022	18
Abdullah <i>et.al.</i>	Impact of Using Health Information Systems (HIS) among Healthcare Professionals during Pandemic Covid-19	2022	19
Adiga <i>et.al.</i>	Mathematical Models for Covid-19 Pandemic: A Comparative Analysis	2020	20
Machi <i>et. al.</i>	Scalable Epidemiological Workflows to Support Covid-19 Planning and Response	2021	21
Tseng e Chou	Spatiotemporal Data Warehousing for Event Tracking Applications	2022	22
Ezzine e Benhlina	Technology against Covid-19 A Blockchain-based framework for Data Quality	2020	23
Cramer <i>et.al.</i>	The United States Covid-19 Forecast Hub dataset.	2022	24
Li <i>et. al.</i>	Toward A Multilingual and Multimodal Data Repository for Covid-19 Disinformation	2020	25
Cheng <i>et. al.</i>	Towards data-driven tele-medicine intelligence: community-based mental healthcare paradigm shift for smart aging amid Covid-19 pandemic.	2023	26

Leon-Sandrova <i>et. al.</i>	Using Big Data and Serverless Architecture to Follow the Emotional Response to the Covid-19 Pandemic in Mexico	2022	27
Bohnert, Pavão e Silva	Visualização de Dados de Saúde Pública: um estudo de caso sobre a Covid-19	2023	28
Stutz <i>et. al.</i>	Visualizing Internet Measurements of Covid-19 Work-from-Home	2021	29

Fonte: Elaborado pela autora (2023)

E - Extração dos Resultados

A extração dos resultados foi realizada nos vinte e nove trabalhos selecionados, sendo dividida em duas partes: (i) coleta dos dados e (ii) análise dos dados.

A coleta de dados ocorreu a partir de um fichamento onde se extraiu de cada trabalho selecionado todo e qualquer trecho que tratasse de ferramentas *big data* ou finalidades do uso das mesmas no contexto da Covid-19. Então, foi considerado o contexto de cada uma das citações mencionadas no texto e estas eram classificadas de forma livre pela autora. A classificação tinha o intuito de agrupar as ferramentas e finalidades em categorias para posterior agrupamento.

Na análise individual de cada artigo, os trechos selecionados eram marcados na ferramenta *mendeley desktop* (compartilhados com a pesquisadora e com o orientador), esta etapa se deu na primeira parte do fichamento, onde as evidências encontradas eram encontradas. Após o término dessa primeira análise, os trechos selecionados foram catalogados, em uma planilha *excel* e agrupados de acordo com a(s) categorias(s) determinadas pela autora.

A análise dos trechos retirados dos artigos se deu, a partir de uma análise temática, onde cada um desses trechos foi categorizado, primeiramente pelo tipo de ferramenta e ou finalidade. Depois disso, cada trecho foi dado o rótulo de: (i) ferramenta, (ii) finalidade, (iii) tipo, (iv) o que faz.

Muitas vezes, as citações se referem ao uso da ferramenta, em algum contexto específico ou finalidade específica, mas o tema central não se referia, especificamente, ao elemento encontrado. Por exemplo, a implantação de um serviço de contenção da epidemia (tema central) que dentre tantos recursos iria disponibilizar um sistema de georreferenciamento para localizar a identificação de novos casos. As conceituações aconteciam quando o autor, de alguma forma, definia que aquela era a categoria, apresentando-o ao leitor, a partir de algum ponto de vista. As caracterizações aconteciam, assim como nas citações, quando o(s) autor(es) se referiam a alguma característica da ferramenta, seja pelo sentido geral da mesma, seja de alguma característica de uso, mas que não era o tema principal da seção ou artigo.

Por fim, as conclusões/sugestões apresentavam partes, onde as pesquisas tinham como foco central a ferramenta e, ou finalidade identificar e traziam conclusões/resultados relevantes sobre esta ferramenta/finalidade.

Após essa segunda categorização, foi realizada a síntese dos resultados em uma descrição textual presente na seção de resultados desta dissertação, que detalhou como é que cada uma das categorias seja de ferramentas, sejam de finalidades foi citado/conceituado/caracterizado ou concluído nos textos.

Após a etapa de fichamento, o material coletado nas fichas dos 29 artigos passou pela análise de conteúdo, assim como todo o texto foi inspecionado, a fim de atribuir categorias aos assuntos levantados pelos autores. Desta maneira, foram criadas 17 categorias referentes ao conteúdo dos artigos, obedecendo à ideia central de cada um e às ferramentas, que foram elencadas no fichamento. Das 17 categorias 6 foram divididas em temas (finalidades) e 11 em ferramentas, e os artigos foram separados em cada quadro, conforme seu conteúdo se aplicava em uma ou mais de uma categoria.

4 RESULTADOS

Este capítulo apresenta os resultados da análise de conteúdo realizada nos 29 artigos selecionados. Os artigos foram organizados e divididos em categorias, de acordo com o tema abordado e/ou a ferramenta utilizada. Um mesmo artigo pode se enquadrar tanto numa categoria temática quanto em categoria pelo nome da ferramenta.

Foram encontradas 11 categorias temáticas de ferramentas, descritas nas seções 4.1, onde serão apresentadas segundo suas características, vantagens e finalidades, enquanto que as 6 categorias temáticas serão abordadas seção 4.2 onde os artigos foram analisados de acordo com o assunto e ideia central apresentada. A classificação inicial entre temas e ferramentas se deu após a leitura minuciosa dos artigos e baseada na análise temática de Bardin. Àqueles cujo o assunto descreve a utilização, ou criação de uma ferramenta prática, foram colocados em uma categoria que atende pelo nome da própria ferramenta, já os que relacionam o uso de *big data* com soluções mais abrangentes para o combate da Covid-19 receberam uma categoria temática. Os Quadros 7 e 8 a seguir explicitam, respectivamente, os temas abordados pelas ferramentas no contexto da Covid e os tipos de ferramenta encontrados

4.1 CATEGORIAS DE FERRAMENTAS

Quadro 8 – Artigos relacionados com as ferramentas

Ferramentas 1 Novas Tecnologias: BIG Data e Blockchain	Ferramentas 2 Sistemas de Informação Geográfica (GIS)	Ferramentas 3 Painel de vigilância de Covid-19
Identificador 14 - Fighting Covid-19 with Fever Screening, Face Recognition and Tracing. Identificador 23 - Technology against Covid-19 A Blockchain-based framework for Data Quality.	Identificador 8 - Covid-19: Challenges to GIS with Big Data. Identificador 10 - Data Visualization Tool for Covid-19 and Crime Data. Identificador 18 - Identifying Hospital Deserts in Texas Before and During the Covid-19 Outbreak.	Identificador 15 - From 5Vs to 6Cs: Operationalizing Epidemic Data Management with Covid-19 Surveillance.
Ferramentas 4 Pipeline	Ferramentas 5 Repositório MM-Covid	Ferramentas 6 Dados de Mobilidade Humana
Identificador 3 - All Models Are Useful: Bayesian Ensembling for Robust High Resolution Covid-19 Forecasting. Identificador 21 - Scalable Epidemiological Workflows to Support Covid-19 Planning and Response. Identificador 2 - AI-Driven Agent-Based Models to Study the Role of Vaccine Acceptance in Controlling Covid-19 Spread in the US	Identificador 25 - Toward A Multilingual and Multimodal Data Repository for Covid-19 Disinformation.	Identificador 12 - Evaluating the Utility of High-Resolution Proximity Metrics in Predicting the Spread of Covid-19. Identificador 17 - Human mobility data in the Covid 19 pandemic characteristics applications and challenges.
Ferramentas 7 Sistema de Informação Hospitalar (SIS)	Ferramentas 8 Modelagem de Dados	Ferramentas 9 Modelos de Aprendizado

Identificador 19 - Impact of Using Health Information Systems (HIS) among Healthcare Professionals during Pandemic Covid-19.	Identificador 11 - Data-Driven Modeling for Different Stages of Pandemic Response. Identificador 22 - Spatiotemporal Data Warehousing for Event Tracking Applications.	Identificador 7 - Classification of Covid-19 using Deep Learning and Radiomic Texture Features extracted from CT scans of Patients Lungs. Identificador 16 - Generating Realistic Covid-19 x-rays with a Mean Teacher + Transfer Learning GAN.
Ferramentas 10 Modelos de Previsão	Ferramentas 11 Representações Geoespaciais	
Identificador 6 - CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting. Identificador 13 - Examining Deep Learning Models with Multiple Data Sources for Covid-19 Forecasting. Identificador 24 - The United States Covid-19 Forecast Hub dataset.	Identificador 28 - Visualização de Dados de Saúde Pública: um estudo de caso sobre a Covid-19.	

Fonte: Elaborado pela autora (2023).

4.1.1 Novas Tecnologias: *Big Data* e *Blockchain* (identificador dos artigos 14 e 23)

Conforme já observamos nesta pesquisa, as tecnologias da informação desenvolveram um papel importante na luta contra a pandemia da Covid-19. Dentre as principais soluções adotadas, o uso de *Big Data* se mostrou crucial na coleta e análise de grandes volumes de dados referentes ao vírus e a sua disseminação. A visão do Big data, especificamente, foi mais voltada à análise dos dados relacionados à saúde pública, onde se faz necessário uma forte atuação na criação de critérios de qualidade de dados para evitar a disseminação de informações equivocadas.

Aproveitando as vantagens proporcionadas pela tecnologia, Ezzine e Benhlima (2020) apresentaram uma proposta que utilizava uma outra de estrutura de Blockchain para melhorar a qualidade dos dados durante a pandemia. A estrutura foi criada seguindo a forma criteriosa que o uso do *Blockchain* permite, com códigos e registros pré-estabelecidos, imutáveis, dificultando a adulteração ou manipulação das informações. A tecnologia *Blockchain* permite o gerenciamento de dados numa determinada estrutura tecnológica, proporcionando transparência no rastreamento da vigilância de dados de saúde pública, principalmente se tratando de pandemias. Aplicar *Blockchain* em *Big Data* pode gerar relatórios mais precisos e criteriosos acerca de dados referentes ao controle de transmissão de um vírus.

Outra ferramenta tecnológica desenvolvida foi o dispositivo criado por (Weijun Tan *et al* 2020), que consiste em dois módulos de reconhecimento facial no sistema, um no próprio aparelho e outro na nuvem, cujas características do aparelho incluem uma câmera térmica, um corpo preto e uma câmera RGB, sensores com sistema de cores, que demonstram a imagem térmica do indivíduo e o resultado do reconhecimento facial. O aparelho foi elaborado para realizar triagem de febre e reconhecimento facial para auxiliar no rastreamento de possíveis infectados pelo vírus e assim promover o controle de transmissão, conforme mencionado na seção temática 4.2.1.

4.1.2 Sistema de Informação Geográfica – GIS (identificador dos artigos 8,10 e 18)

Durante a pandemia de Covid-19, pesquisadores contaram com dados relativos à localização geográfica para criar sistemas de acompanhamento da dinâmica de transmissão do coronavírus.

Zhou et al (2020) realizaram uma análise da representação espacial da doença Covid-19, considerando aspectos materiais, populacionais e sociais em três escalas: individual, grupo e regional. Com o uso das ferramentas *Internet of Things*, *Tencent* e *Baidu*, os autores criaram um banco de dados baseado em *Big Data* e pandemia, utilizando a tecnologia de Sistema de Informação Geográfica (GIS), cuja visualização de gráficos baseados em análise espaço-temporal do avanço da pandemia gera recursos para apoiar na tomada de decisão. O termo *Internet of Things* se refere à internet das coisas, que agrega elementos físicos e conectados à internet, para coletar e trocar dados entre si, e foi utilizado pelos autores para obtenção de dados de transporte de materiais; A *Tencent* se trata de uma grande empresa de tecnologia, com sede na China, que oferece serviços de jogos *online*, serviços de pagamento, incluindo tecnologias de inteligência artificial, entre outros, e seu banco de dados foi recorrido pelos autores para solicitação de dados de localização; já a *Baidu* representa outra empresa tecnológica que busca integrar a internet das coisas e a inteligência artificial, tendo sido acessada pelos autores para obter dados de migração.

Outro estudo que contou com o uso de GIS foi dirigido por (Walker e Sharma 2021), e contou com a criação de um mapa temático que consiste em apresentar, em realidade virtual, através dos softwares Unity 3D e Maptitude, vários ambientes de visualização de dados relacionados à criminalidade e ao Covid-19. Uma sala dentro do ambiente Unity mostra dados da densidade do crime na cidade de Baltimore (EUA), outra com gráficos de barras apresentando percentual médio de nível de renda e pobreza, e outra permitindo um panorama geral de visualização de todos os mapas envolvidos na pesquisa.

Jiao, Degen e Azimian (2020) abordaram a questão crítica da acessibilidade aos serviços hospitalares no estado do Texas, nos Estados Unidos, tanto antes como durante o surto de Covid-19. O estudo analisa informações de áreas geográficas onde a população tem dificuldade em acessar hospitais e cuidados médicos devido à falta de instalações de saúde próximas, devido a desafios significativos que essas regiões podem enfrentar em termos de acesso a serviços médicos essenciais, o que pode levar a problemas de saúde mais graves e maior vulnerabilidade em situações de emergência, como uma pandemia. Os autores utilizam dados geoespaciais e técnicas de análise espacial para identificar e mapear essas áreas de hospital deserto em todo o estado do Texas, avaliando a disponibilidade de hospitais e serviços de saúde em relação à densidade populacional, distâncias de viagem e outras variáveis relevantes. Esse estudo pode auxiliar nas

tomadas de decisões de implantação de políticas públicas para que haja melhorias na infraestrutura de saúde para a população, principalmente após a realidade caótica vivenciada durante a pandemia. A ferramenta desenvolvida pelos pesquisadores pode ser acessada pelo link <<https://coronavirus-disasterresponse.hub.arcgis.com/signin>>.

4.1.3 Painel de Vigilância Covid-19 (identificador do artigo 15)

O Painel de Vigilância Covid-19 do Biocomplexity Institute & Initiative, desenvolvido por Peddireddy (et al, 2020), oferece uma ampla gama de informações, possibilitando que os usuários visualizem e explorem os casos de Covid-19 em diferentes resoluções temporais e espaciais. Inclui gráficos que mostram as curvas cumulativas e incidentes em todas as regiões, e o uso exclusivo de um controle deslizante de tempo, semelhante a um filme, facilita a análise das mudanças temporais e espaciais da pandemia. Atualmente, o painel suporta a representação de dados a nível de país para todas as nações do mundo, a nível estadual e provincial para 20 países, e estatísticas específicas para os condados dos Estados Unidos (EUA).

Desde o seu lançamento, até 25 de outubro de 2020, o painel foi utilizado por mais de 1,13 milhão de usuários em mais de 220 países, processando mais de 60 milhões de solicitações na camada principal de recursos hospedados em ArcGIS Online.

4.1.4 Pipeline (identificador dos artigos 2, 3 e 21)

Durante uma pandemia, como a Covid-19, a quantidade de dados coletados é vasta e em rápida expansão, tornando essencial o uso de tecnologias e metodologias que possam lidar com essas demandas. Machi et al (2021) desenvolveram um pipeline de clusters de supercomputação geograficamente separados para representar visualização de fluxos de trabalho escalonáveis, baseados em informações em tempo real, para apoiar o gerenciamento da crise sanitária provocada pela pandemia de Covid-19, através da HCP (Computação de Alto Desempenho), processando grandes conjuntos de dados, em resposta aos apelos das autoridades governamentais, que necessitavam utilizar ferramentas de visualização de dados para tomar decisões no período pandêmico. Uma abordagem escalável pode ter a capacidade de adaptar e aplicar métodos epidemiológicos em larga escala, permitindo que os pesquisadores processem grandes volumes

de dados em tempo hábil. A ferramenta apresenta diagramas dos fluxos de trabalho desde a configuração do processo, envolvendo cronogramas de tarefas que envolvem esforços humanos, e a sequência de programação para vários dias.

Já Bhattacharya et al (2021) desenvolveram um pipeline para analisar a relação entre aceitação da vacina contra Covi-19 e aceleração, ou desaceleração, do número de infectados/mortos nos Estados Unidos, através dos fluxos de trabalho de entrada e saída da ferramenta desenvolvida.

A pesquisa de (Adiga et al 2021) também contou com o desenvolvimento de um pipeline de previsão para a Covid-19 baseado no conjunto bayesiano, cujos códigos estão disponíveis no link: <https://github.com/aniruddhadiga/Covid-19_forecast>. A proposta dos autores foi apresentar uma abordagem mais inovadora e precisa, utilizando estatística, aprendizado de máquina e abordagens mecânicas com um conjunto baseado em desempenho para combinar os modelos.

4.1.5 Repositório MM-Covid (identificador do artigo 25)

O estudo de (Li et al 2020) resultou na produção de um repositório multilingual e multimodal para detecção de notícias falsas relacionadas à Covid-19. Os autores desenvolveram esta ferramenta com base em um conjunto de dados que fornece notícias falsas multilíngues e o contexto social relevante, contendo 3.981 notícias falsas e 7.192 informações confiáveis em 6 idiomas diferentes: inglês, espanhol, português, hindi, francês e italiano.

A proposta dos pesquisadores foi ensinar os leitores a identificarem quando uma informação acerca da Covid-19 pode ser falsa, considerando o contexto, o território e o engajamento social de usuários em mídias sociais como facebook, instagram e twitter.

4.1.6 Dados de Mobilidade Humana (identificador dos artigos 12 e 17)

A propagação da Covid-19 se dá principalmente pela transmissão de pessoa para pessoa, geralmente através de gotículas respiratórias e contato próximo com pessoas infectadas. Nesse

contexto, o uso de métricas de proximidade em alta resolução pode ser uma estratégia valiosa para prever a disseminação do vírus e avaliar a eficácia das medidas de controle.

Essas métricas de proximidade podem incluir o monitoramento de interações próximas entre indivíduos, como contato físico, proximidade geográfica e conexões sociais. Ao analisar esses dados em nível individual ou comunitário, os pesquisadores podem obter informações sobre os padrões de contato que influenciam a disseminação do vírus.

Neste sentido, Mehrab et al (2022) desenvolveram uma pesquisa para avaliar a utilidade das métricas de proximidade de alta resolução em termos de sua precisão e capacidade de prever a proliferação da Covid-19, utilizando dados de mobilidade humana para examinar como as interações próximas afetam a transmissão do vírus. O estudo contou com um conjunto de dados de mobilidade que foi adquirido através de um SDK (Software Development Kit) de terceiros operado pela empresa X-Mode, que trabalha em parceria com diversos aplicativos móveis para fornecer um Sistema de Posicionamento Global (GPS) anônimo, rastreando informações contendo identificadores persistentes.

Além do caso supracitado, o estudo de (Hu et al 2021) realizou uma pesquisa de revisão narrativa envolvendo dados de mobilidade e Covid-19 para classificar e descrever os dados de mobilidade utilizados nos estudos, resumir como esses dados foram aplicados nas pesquisas sobre Covid-19 e destacar os desafios, recomendações e direções futuras para a utilização de dados de mobilidade em pesquisas sobre doenças infecciosas. Os autores identificaram três principais fontes de conjuntos de dados de mobilidade humana e resumiram os conjuntos disponíveis publicamente, classificados nessas três fontes principais: (1) sistemas de transporte público; (2) operadoras de redes móveis; e (3) aplicativos de telefone celular. Cada uma dessas fontes oferece uma quantidade valiosa de informações sobre como as pessoas se deslocam e interagem, fornecendo uma visão abrangente da mobilidade humana durante a pandemia.

4.1.7 Sistema de Informação Hospitalar – SIS (identificador do artigo 19)

Abdullah et al (2022) utilizaram um questionário para entrevistar médicos oficiais que trabalham no Ministério da Saúde com mais de dois anos de experiência com uso de Sistema de Informação Hospitalar (SIS) em sua profissão. A finalidade dos autores era investigar a familiaridade desses profissionais de saúde com o SIS, considerando as vantagens que uma

plataforma tecnológica como essa, projetada para gerenciar e organizar informações relacionadas ao atendimento de saúde em um hospital ou instituição de saúde, pode proporcionar principalmente em situações como a pandemia de Covid-19. Os autores defendem a utilidade do SIS em instituições de saúde, uma vez que um sistema como esse desenha um conjunto de recursos que permitem aos profissionais de saúde coletar, armazenar, processar, recuperar e compartilhar informações clínicas e administrativas de pacientes, colaborando para um gerenciamento mais eficaz.

4.1.8 Modelagem de Dados (identificador dos artigos 11 e 22)

Adiga et al (2020) abordam o uso de modelagem baseada em dados em diferentes estágios de resposta a pandemias, enfatizando a importância dessas abordagens para orientar a tomada de decisões efetivas durante a pandemia de Covid-19. No estudo os autores discutiram os desafios e vantagens de utilizar sistemas, bancos de dados, baseados na modelagem lógica de dados para atender as demandas geradas pela pandemia. Os autores enfatizam a modelagem matemática, e o desenvolvimento de modelos computacionais, de dados como ferramentas essenciais para analisar e entender a dinâmica da pandemia desde as fases iniciais até o presente momento, uma vez que esses métodos se utilizam da formulação de problemas e os modelos criados visam responder e encontrar soluções para eles.

Já a pesquisa de Tseng e Chou (2022) apresenta a aplicação de data warehousing para armazenar e analisar dados espaciais e temporais em contextos de rastreamento de eventos, com ênfase em resolver desafios específicos e fornecer percepções valiosas para várias aplicações práticas, como durante a pandemia de Covid-19. Os autores criaram um sistema data warehouse multidimensional, de modelagem espaço-temporal, para rastreamento de eventos dinâmicos em aplicativos contemporâneos, como o rastreamento de contatos em multidões para prevenção do Covid-19. A ferramenta permite uma integração entre dados de fontes diversas, considerando informações de acordo com o espaço e o tempo, gerando um grande sistema de informação que permite analisar e entender os eventos.

4.1.9 Modelos de Aprendizado (identificador dos artigos 7 e 16)

Mangalagiri (et al, 2021) realizaram um estudo baseado em aprendizado profundo e textura radiônica em análise de tomografias computadorizadas (TC) para classificar o diagnóstico dos pacientes com relação ao Covid-19. A técnica consiste em avaliar de maneira automatizada as imagens de TC dos pulmões e prever se o paciente se enquadra nas categorias de positivo, negativo ou outros. Os pesquisadores envolveram 38 pacientes na fase de treinamento dos modelos de aprendizado profundo personalizados. Para isso, utilizaram uma arquitetura de 10 camadas de redes neurais artificiais 3D para extrair recursos de aprendizado profundo a partir dos dados. Além disso, empregaram um modelo de Random Forest (RF) para treinar a combinação das características extraídas. O estudo apresentou resultados eficazes e robustos, no entanto, os autores enfatizam que o diagnóstico de classificação automatizado deve ser usado como uma ferramenta auxiliar e não substituir a avaliação médica profissional.

Outro estudo da área da saúde envolvendo técnica de aprendizado foi conduzido por (Menon et al 2020) que, devido à quantidade limitada de conjunto de dados públicos disponíveis referentes a radiografias de tórax com Covid-19, abordaram a ferramenta MTT-GAN como uma possível solução de transferência de aprendizado de pneumonia Kaggle para Covid-19. Os autores introduziram a MTT-GAN, uma nova arquitetura, com o objetivo de gerar imagens sintéticas de radiografia de tórax de alta qualidade para pacientes com Covid-19. Os resultados qualitativos demonstraram que essa abordagem resulta em maior precisão nos classificadores binários e multiclasse para a triagem automatizada de raios-x de Covid-19. Apesar dos resultados apontarem que a MTT-GAN supera significativamente a linha de base da GAN, e gera imagens que se aproximam das características anatômicas das imagens reais, após as imagens deep fake terem sido apresentadas a dois radiologistas com certificação, com mais de 6 anos de experiência clínica na interpretação de radiografias de tórax, os profissionais avaliaram que a baixa qualidade na resolução das imagens, e a presença de erros sistemáticos na representação da escápula e dos ossos da clavícula, são limitações que precisam de mais pesquisas para que o experimento possa atingir a qualidade diagnóstica comparável ao desempenho humano.

4.1.10 Modelos de Previsão (identificador dos artigos 6, 13 e 24)

Algumas pesquisas foram direcionadas para utilizar ferramentas capazes de entregar modelos precisos de previsão para antever a dinâmica dos dados referentes aos casos de Covid-19.

A estrutura Causal GNN proposta na pesquisa de Wang (et al, 2022), por exemplo, envolve três temas recorrentes nos estudos referentes à pandemia de Covid-19: dados focados em modelos espaço-temporais, modelos de aprendizado de máquina e previsões de pandemias. Com o progresso significativo na disponibilidade de dados espaço-temporais confiáveis, os modelos fundamentados em redes neurais gráficas obtiveram sucesso na representação das inter-relações entre os sinais vindos de diversas regiões, o que resultou em previsões de alta qualidade. Wang (et al, 2022) utilizam os *Graph Neural Networks* (GNNs), que, segundo os autores explicam, são um tipo de algoritmo de aprendizado de máquina criado para trabalhar com dados estruturados em forma de gráficos.

No contexto da pandemia de Covi-19, os GNNs foram utilizados para lidar com dados, que possuem relações complexas entre elementos, para capturar padrões de disseminação da doença ao longo do tempo e espaço. A proposta dos autores foi baseada em identificar relações de causa e efeito para entender a dinâmica entre os dados espaço-temporais dos gráficos e melhorar as previsões acerca da pandemia, construindo uma estrutura Causal GNN que consiste em um módulo causal e um módulo GNN dinâmico baseado em atenção.

Outra pesquisa orientada para previsões acerca da pandemia de Covid-19 foi realizada por Wang (et al, 2022), que examinaram modelos de aprendizado profundo e desenvolveram modelos baseados em redes neurais recorrentes, que foram combinados utilizando a técnica de conjunto de empilhamento. Para capturar os efeitos de vários fatores na propagação da Covid-19, diversas fontes de dados foram consideradas, incluindo contagens de casos confirmados e óbitos, juntamente com informações de testes, com a finalidade de aprimorar as previsões. Os autores realizaram uma análise abrangente comparando diferentes modelos de séries temporais no contexto da Covid-19, o método utilizou diversos modelos mecanicistas, incluindo o SEIR, SIDR, SIDARTHE e outros, com o objetivo de prever os a dimensão que a Covid-19 tomava em várias regiões, e modelos de séries temporais, como estatísticos e de aprendizado, que são reconhecidos por sua simplicidade e precisão na previsão de pandemias. Os resultados mostraram que modelos

de aprendizado profundo simples podem alcançar resultados comparáveis ou até mesmo superiores em relação a modelos mais complexos, contando cada um com suas limitações e fragilidades.

Já Cramer et al (2022) usaram suposições variadas para construir modelos e gerar previsões, acerca de futuras mudanças nas políticas e medidas de distanciamento social e transmissão da Covid-19, bem como taxas de vacinação e a disseminação de novas variantes do vírus nos Estados Unidos. As abordagens de previsão dos autores incluem modelos estatísticos e de aprendizado de máquina, modelos mecanicistas que incorporam doenças, dinâmicas de transmissão e combinações de várias abordagens, contando com o desenvolvimento da ferramenta a seguir que pode ser acessada pelo link [<https://Covid19forecasthub.org/doc/reports/>](https://Covid19forecasthub.org/doc/reports/).

4.1.11 Representações Geoespaciais e Temporais (identificador do artigo 28)

Dentro das soluções para visualização de dados sobre a pandemia de Covid-19, Bohnert; Pavão; Corrêa, (2022) exploraram ferramentas de representações geoespaciais e temporais, conforme descritas a seguir:

a) Johns Hopkins Coronavirus Resource Center (CRC)

Trata-se de uma plataforma criada por especialistas da Universidade Johns Hopkins nos Estados Unidos, cuja finalidade é coletar e analisar dados relativos ao coronavírus, através de um Painel de informações sobre a pandemia Covid-19, servindo de fonte informacional para que o público compreenda a evolução do vírus, bem como acompanhe os números de contaminados, mortes, testes, vacinas, além de disponibilizar esses dados para download.

Esse panorama contribui de forma ativa na comunicação de atualizações acerca do rastreamento do coronavírus.

b) Our World In Data

Pesquisadores da Universidade de Oxford, na Inglaterra, criaram a plataforma pública *online Our World in Data*, que contou com a participação de uma comunidade global de acadêmicos, com o objetivo de tornar público diversos assuntos de problemas globais. O Painel de informações sobre pandemia Covid-19 *Our World in Data* aborda uma estrutura que aborda

especificamente informações relativas à pandemia, cujos dados estão abertos e disponíveis para acesso e download.

c) Painel Coronavírus da Organização Mundial da Saúde (OMS)

Esta ferramenta de acesso aberto informa diariamente, de maneira oficial, as últimas notícias e informações relacionadas ao coronavírus, disseminando dados sobre número de óbitos, contaminados e utilização da vacina, relatados pelos países e territórios.

As três ferramentas apresentadas anteriormente utilizam Representações Geoespaciais, que se tratam de informações representadas em gráficos que exploram uma localização geográfica específica, ou seja, são análises realizadas com base em dados aferidos de um determinado espaço. Por ser um fenômeno global, a pandemia de Covid-19 incentivou governantes e autoridades dos diversos países e organizações a utilizarem mapas temáticos para visualizar e compreender a propagação do coronavírus e orientar os cidadãos com decisões seguras.

Os autores descrevem de forma sucinta os objetivos, bem como a funcionalidade, de algumas ferramentas de Representações Geoespaciais que foram utilizadas, durante a pandemia, a fim de responder, comunicar e sanar dúvidas da sociedade a respeito da crise pandêmica. Sendo elas:

● ***Choropleth Map***

Mapa temático representado através de cores que informam os padrões e variáveis de uma determinada área geográfica. Três mapas são utilizados na composição dessa ferramenta, utilizando cores e sombras para o público visualizar as áreas mais afetadas pelo coronavírus por meio das variáveis número de casos e de óbitos confirmados pelos países.

● ***Bubble Map***

Mapa de bolhas, que permite a visualização de duas variáveis de forma simultânea através da apresentação de bolhas em uma localização. A plataforma Johns Hopkins Coronavirus Resource Center utiliza esse método, adotando a técnica de círculos de vários tamanhos para expandir casos e mortes ocasionados pela Covid-19 em determinado espaço geográfico.

A Johns Hopkins Coronavirus Resource Center, Our World In Data e Painel Coronavírus da Organização Mundial da Saúde utilizam, também, Representações Temporais, que acompanham a evolução do coronavírus ao longo do tempo, a partir de 2019, ano em que

apareceu o primeiro caso de contaminação. As visualizações de gráficos de linha mostram as oscilações da pandemia, as fases de avanço, estabilidade e regressão

4.2 TEMA

Tema 1 Uso de Inteligência Artificial	Tema 2 Visualização de dados para tomada de decisão	Tema 3 Uso de Redes Sociais: Twitter
Identificador 14 - Fighting Covid-19 with Fever Screening, Face Recognition and Tracing. Identificador 23 - Technology against Covid-19 A Blockchain-based framework for Data Quality. Identificador 5 - Artificial intelligence in Covid-19 drug repurposing. Identificador 2 - AI-Driven Agent-Based Models to Study the Role of Vaccine Acceptance in Controlling Covid-19 Spread in the US	Identificador 9 - Covid-19, big data, smart city e cidade digital estratégica na geração de conhecimentos. Identificador 8 - Covid-19: Challenges to GIS with Big Data. Identificador 10 - Data Visualization Tool for Covid-19 and Crime Data. Identificador 15 - From 5Vs to 6Cs: Operationalizing Epidemic Data Management with Covid-19 Surveillance. Identificador 21 - Scalable Epidemiological Workflows to Support Covid-19 Planning and Response. Identificador 23 - Technology against Covid-19 A Blockchain-based framework for Data Quality. Identificador 28 - Visualização de Dados de Saúde Pública: um estudo de caso sobre a Covid-19. Identificador 29 - Visualizing Internet Measurements of Covid-19 Work-from-Home.	Identificador 4 - Applying Machine Learning to Analyze Anti-Vaccination on Tweets. Identificador 27 - Using Big Data and Serverless Architecture to Follow the Emotional Response to the Covid-19 Pandemic in Mexico. Identificador 25 - Toward A Multilingual and Multimodal Data Repository for Covid-19 Disinformation.
Tema 4 Uso de modelos matemáticos no combate à pandemia	Tema 5 Cuidados com a saúde mental de idosos na pandemia (grupo de risco)	Tema 6 Impactos econômicos provocados pela pandemia

Identificador 3 - All Models Are Useful: Bayesian Ensembling for Robust High Resolution Covid-19 Forecasting. Identificador 20 - Mathematical Models for Covid-19 Pandemic: A Comparative Analysis.	Identificador 26 - Towards data-driven tele-medicine intelligence: community-based mental healthcare paradigm shift for smart aging amid Covid-19 pandemic.	Identificador 1- A State-Level Socioeconomic Data Collection of the United States for Covid-19 Research.
---	--	---

Fonte: Elaborado pela autora (2023).

4.2.1 Uso de Inteligência Artificial – IA (identificador dos artigos 2, 5, 14 e 23)

O uso de tecnologias inovadoras no combate à pandemia de Covid-19 como *Big Data*, Inteligência Artificial, pode-se dizer que se tornou tendência na busca por soluções de apoio no controle da taxa de transmissão pelo coronavírus.

Durante a pandemia de Covid-19 organizações da área da saúde, e empreendimentos privados, adotaram a estratégia de utilização de aparelho para aferir a temperatura na entrada dos estabelecimentos, por exemplo, a fim de controlar o acesso de pessoas febris no ambiente, por ter sido a febre incluída como um dos sintomas causados pela infecção humana pelo coronavírus.

Weijun Tan, et al (2020) desenvolveram uma tecnologia de triagem de febre e reconhecimento facial para auxiliar no rastreamento de possíveis infectados pelo vírus e assim promover o controle de transmissão.

Já o estudo de Ezzine e Benhlila (2020) faz uma reflexão sobre a garantia de qualidade no conjunto abundante de dados gerados pelo uso de *Big Data*, considerando que esta tecnologia tem uma característica de grande volume de dados acerca do que se pretende investigar. Os autores apontam que medidas de combate e controle da pandemia de Covid-19 se tornaram temas muito estudados, e que a quantidade de material que aborda esses assuntos é bastante volumosa, podendo levantar questões na utilização desses dados, como acerca da qualidade e confiabilidade, pontos muito pertinentes, já que a extração e seleção desses trabalhos podem resultar em produtos que valham de apoio nas decisões de apoio ao enfrentamento da crise sanitária.

A pesquisa de (Zhou et al 2020) traz um olhar de rentabilidade somado às contribuições de ferramentas tecnológicas da atualidade, apresentando uma proposta de reaproveitamento de medicamentos já existentes para tratar a Covid-19. A ideia dos autores parte do princípio de que pode ser possível tratar doenças atuais utilizando remédios que já foram estudados e desenvolvidos para curar enfermidades preexistentes. Para isso, foi elaborado um sistema de reaproveitamento de medicamentos assistido por inteligência artificial específico para a Covid-19, através da relação entre informações do sistema biológico humano, e uso de aprendizagem de máquina, componente da inteligência artificial.

Bhattacharya (et al, 2021) também apostaram no uso de inteligência artificial para estudar a aceitação da vacina nos Estados Unidos através de fluxos de trabalho baseados em modelos de agentes orientados por IA, por meio de sistemas de computação de alto desempenho (HPC). O estudo mostra que a eficácia da vacina apresenta resultados diferentes em vários estados, e que em estados nos quais a aceitação da vacina foi maior, não necessariamente houve diminuição de infectados e mortos proporcionalmente. Existe uma correlação entre população, aceitação da vacina e achatamento na curva de infectados/óbitos.

Weijun Tan et al (2020) construíram a tecnologia de triagem de febre e reconhecimento facial cujo dispositivo consiste em dois módulos de reconhecimento facial no sistema, um no próprio aparelho e outro na nuvem. Os autores utilizaram algoritmos para realizar testes de aperfeiçoamento nas tecnologias ligadas à detecção da imagem humana e criaram uma plataforma de inteligência artificial e *Big Data, Spring Cloud*, para gerenciar vários tipos de dispositivos na nuvem.

Enquanto a tecnologia desenvolvida por Ezzine e Benhlima (2020) apresenta uma estrutura de *Blockchain* como iniciativa de melhorar o uso de *Big Data* e proporcionar dados para serem utilizados no desenvolvimento de políticas de controle da pandemia, garantindo maior segurança a respeito da qualidade.

A proposta de Zhou (et al, 2020) inclui a construção de gráficos de conhecimento contendo relacionamentos, entre diferentes tipos de entidades médicas (por exemplo, doenças, drogas, e proteínas) e prevê novas ligações entre medicamentos e doenças aprovados, e comprova que a tecnologia de inteligência artificial, é um fator de alta vantagem no reaproveitamento de medicamentos para combater doenças emergentes, como a Covid-19.

4.2.2 Visualização de Dados (identificador dos artigos 8, 9, 10, 15, 21, 23, 28 e 29)

A pandemia de Covid-19 trouxe muitos desafios para o governo nas esferas municipal, estadual e federal, dentre eles utilizar dados relacionados à saúde, mobilidade social e vigilância sanitária para tomar decisões estratégicas de combate à pandemia.

Ferlin et al (2021) apontam *Big Data, Smart City* e Cidade Digital Estratégica como tecnologias facilitadoras para gerar e promover o conhecimento no cenário pandêmico. O grande volume de dados, e a velocidade com que eles aumentam, são características do big data que para

os autores enriquecem a criação de uma smart city e de cidades estratégicas. Quanto mais informações forem armazenadas maior ficará o banco de dados e, conseqüentemente, mais detalhes serão oferecidos para apoiar o gerenciamento de serviços e recursos, como com a experiência da pandemia.

Machi, et al (2021) apontam que outra tecnologia, a HCP (Computação de Alto Desempenho), capaz de processar grandes conjuntos de dados, auxilia nas respostas desejadas pelas autoridades governamentais, que necessitavam utilizar ferramentas de visualização de dados para tomar decisões no período pandêmico.

Neste sentido, Bohnert, Pavão e Corrêa (2022) exploram a visualização de dados como solução que viabiliza a ampla divulgação do conhecimento científico e apropriação social da ciência. Na perspectiva dos autores, a visualização de dados é a forma mais rápida de transformar dados em informação de utilidade pública, uma vez que a linguagem verbal nem sempre alcança o entendimento e compreensão das pessoas no que se pretende comunicar. Os autores apresentaram 3 plataformas de visualização de dados geoespaciais que serviram de fontes de informações para auxiliar os governantes nas tomadas de decisões no intuito de conter o avanço do vírus, assim como também para os cidadãos de forma geral interessados em entender a dinâmica de surgimento, contaminação e mortes. Essas ferramentas são resultados da colaboração interdisciplinar entre cientistas e colaboradores, sendo elas: *Johns Hopkins Coronavirus Resource Center (CRC)*, *Our World In Data* e Painel Coronavírus da Organização Mundial da Saúde (OMS). Além disso, o estudo agregou as ferramentas de visualização baseadas em representações geoespaciais e temporais: *Choropleth Map* e *Bubble Map*. Essas ferramentas foram descritas na seção 4.1.

Stutz, et al (2021) criaram um sistema de visualização baseado na *Web* para medir o momento em que o trabalho em casa (WFH) passou a ser adquirido pela população no período pandêmico. O objetivo dos autores foi fornecer suporte à interação humana com as descobertas acerca da transição do trabalho presencial para o trabalho em casa, bem como acompanhar o momento de volta ao “novo normal” que a pandemia trouxe.

O estudo de Ferlin, et al. (2021) tratou de coletar e analisar grandes volumes de dados (*Big Data*) referentes a casos de Covid-19, como quantitativo de casos de infectados e óbitos em diferentes abrangências territoriais, abrangendo a expansão dos estudos relacionados a projetos de cidade digital estratégica e aplicações de smart city para lidar com a Covid-19, ao mesmo

tempo em que oferece aporte para a definição de serviços e recursos que serão gerenciados nas cidades. Além disso, a pesquisa contribui também para ampliar as possibilidades de como os dados em tempo real podem influenciar as dinâmicas urbanas, facilitando a tomada de decisões por parte dos gestores.

Neste estudo, Zhoua, et al (2020) realizaram uma análise da representação espacial da doença Covid-19, considerando aspectos materiais, populacionais e sociais em três escalas: individual, grupo e regional. Na escala individual, compararam o rastreamento espacial da pandemia com as trajetórias espaço-temporais dos pacientes. Na escala de grupo, foi estimado o fluxo populacional e sua distribuição espacial. Na escala regional, realizaram a segmentação de riscos espaciais, análise do equilíbrio entre oferta e demanda de recursos médicos e análise da diferenciação espacial na capacidade de transporte material e no sentimento social.

Já Walker e Sharma (2021) descrevem o desenvolvimento de uma ferramenta de visualização para análise dos dados criminais e Covid-19, na cidade de Baltimore (EUA), utilizando os softwares Unity 3D e Maptitude.

Peddireddy, et al (2020) criaram um Painel de Vigilância Covid-19 com o objetivo de proporcionar aos usuários acompanhamento da evolução da pandemia. A proposta da ferramenta é preencher lacunas não percebidas nas ferramentas de visualização da pandemia existentes, promovendo maior interatividade através de um painel visual simples e abrangente da Covid-19.

Em outro trabalho, Machi (et al, 2021) desenvolveram um *pipeline* de *clusters* de supercomputação geograficamente separados para representar visualização de fluxos de trabalho por escala, baseados em informações em tempo real, para apoiar o gerenciamento da crise sanitária provocada pela pandemia de Covid-19.

Bohnert, Pavão e Corrêa (2022), discutem a ciência aberta como uma possibilidade de inclusão para a sociedade participar de forma mais ativa em momentos desafiadores como durante a pandemia de Covid-19. Uma vez que informações confiáveis sejam acessíveis para a sociedade, as pessoas podem se sentir seguras para participar das dinâmicas propostas pelas autoridades que trarão benefícios aos cidadãos, além de torná-las capazes de disseminar conhecimento científico. Em situações desafiadoras como a pandemia provocada pela transmissão do coronavírus, munir a sociedade de dados que podem ajudar no combate a propagação do vírus pode ser uma boa estratégia para minimizar os impactos devastadores para a sociedade.

Stutz (et al, 2021) desenvolveram um mapa-múndi online, exibindo uma grade geográfica de observações sobre o WFH. Além das interações típicas de zoom e panorâmica do mapa, foram adicionadas duas novas formas de informações pop-up que permitem aos usuários investigar detalhadamente os dados subjacentes. Utilizaram minigráficos para mostrar as alterações nos primeiros seis meses de 2020 em um local específico, facilitando a identificação e navegação em pontos críticos. Alternativamente, os usuários podem visualizar redes específicas de Provedores de Serviços de Internet que mostram o WFH em um dia específico.

Em seu estudo, Ferlin (et al, 2021) afirma que ainda é necessário democratizar o acesso a esses dados, a fim de amplificar os benefícios que o *big data* pode oferecer à sociedade. A grande questão do objetivo dos autores não trata apenas de ter um grande volume de dados disponíveis, mas sim de ter dados confiáveis e acessíveis em tempo real que sejam realmente úteis para a sociedade, e que possam gerar conhecimento.

No cenário descrito por Peddireddy (et al, 2020), foi percebido que não existia um padrão definitivo para o gerenciamento e a visualização de dados relativos a pandemias. Apesar do reconhecimento da importância de um dos painéis de visualização de dados da pandemia mais amplamente utilizado, que foi desenvolvido e mantido pela Universidade Johns Hopkins (*Johns Hopkins Coronavirus Resource Center*), os autores apontaram que essa plataforma exibia algumas limitações, incluindo mudanças frequentes no formato e na plataforma de compartilhamento de dados, o que acarretava em dificuldades para a adaptação por parte dos usuários. Essa situação ressalta a importância da consistência dos dados como um dos objetivos centrais a serem almejados.

Além disso, o painel carecia de informações temporais detalhadas e visualizações em níveis regionais, assim como da habilidade de realizar consultas ou pesquisas específicas em uma determinada área. As distintas categorias de dados (tais como casos, óbitos, recuperações, entre outros) estavam divididas em abas e painéis separados, o que tornava mais difícil a obtenção de uma compreensão abrangente da situação em uma região específica.

Portanto, Peddireddy (et al, 2020) julgaram necessária a criação de um sistema centrado no usuário, com uma solução completa apoiada por fontes de dados confiáveis, juntamente com visualizações ricas e uma interface de fácil utilização, acessível tanto ao público quanto aos pesquisadores.

Outro problema é elencado pelos pesquisadores Bohnert, Pavão e Corrêa (2022) que defendem a ciência aberta como instrumento de colaboração da sociedade, de forma democrática, com o cumprimento de políticas governamentais de enfrentamento ao caos. A ciência aberta propõe uma estrutura colaborativa de informações científicas, minimizando as dificuldades na disseminação e compartilhamento dessas informações tanto pelos cientistas quanto pelos cidadãos comuns.

Zhoua (et al, 2020) se utilizam das ferramentas *Internet of Things* para obtenção de dados de transporte de materiais, *Tencent* para solicitação de dados de localização e *Baidu* para dados de migração, os autores criaram um banco de dados baseado em *Big Data* e pandemia, utilizando a tecnologia de Sistema de Informação Geográfica (GIS), cuja visualização de gráficos baseados em análise espaço-temporal do avanço da pandemia gera recursos para apoiar na tomada de decisão. A proposta dos autores foi monitorar os sentimentos da sociedade nas mídias sociais através da construção de um banco de dados de opiniões, fornecendo informações cruciais para orientar as ações governamentais no período pandêmico.

A iniciativa se mostrou satisfatória, uma vez que, inovações da tecnologia de *big data*, como agilidade na coleta e integração de dados, enriqueceram o âmbito dos sistemas de informação geográfica (GIS).

Para outra ferramenta proposta por Walker e Sharma (2021) foram utilizados dados do Painel de Coronavírus da cidade, e gráficos de visualização em barra foram elaborados com dados do *site Open Baltimore*, pretendendo mostrar como ferramentas de visualização de dados agregam informação relacionando aumento, ou diminuição, do crime durante a pandemia. Os autores exploram como questões étnicas-raciais, e econômicas, influenciaram nos dados de criminalidade e de Covid-19.

Por meio dos gráficos de barras e do mapa temático criado no Maptitude, percebeu-se que os cuidados nas tomadas de decisões pelos governantes, relacionadas a medidas de segurança, como maior apoio de profissionais da área de saúde, foram concentrados nas regiões centrais da cidade, enquanto que nas áreas periféricas o número desses profissionais tende a diminuir.

Peddireddy (et al, 2020) propôs a criação do Painel de Vigilância Covid-19 do *Biocomplexity Institute & Initiative*, que tem por base seis métricas identificadas pelos autores que podem estabelecer um padrão para o gerenciamento de dados de vigilância em pandemias, que chamaram de padrão 6Cs.

Os elementos que compõem esse padrão determinam critérios de qualidade na criação de painéis de vigilância sanitária, e são eles: que os dados sejam Consistentes, Corretos, Atualizados, Abrangentes, devidamente Curados e Legíveis por computador. Segundo os autores a consistência pode estar ligada tanto ao formato dos dados quanto ao seu conteúdo, garantindo que sua firmeza e originalidade permaneçam sem alterações ao longo do tempo; os dados corretos significa garantir que sejam precisos, para evitar situações desastrosas em caso de dados equivocados; os dados precisam ser, frequentemente atualizados para que expressem a realidade presente em um ambiente de mudanças rápidas como no contexto da pandemia de Covid-19; o elemento de dados abrangentes é o que mais difere o painel de acompanhamento Covid-19 criado por Peddireddy (et al, 2020) daqueles que focam em fragmentos específicos, como o *Johns Hopkins Coronavirus Resource Center*, já que uma visão panorâmica dos casos de Covid-19 traz mais detalhes da amplitude da pandemia; a curadoria complementa a abrangência, viabilizando que os dados sejam coletados de uma variedade de fontes, de modo a atender às diversas e variadas necessidades da população; e, por fim, a legibilidade dos dados por computador, ou seja, que sejam facilmente modelados em sistemas computacionais. Para os autores, esse conjunto de elementos podem dar maior confiabilidade aos painéis de acompanhamento de pandemias.

A pandemia de Covid-19 despertou a necessidade de preencher as lacunas de informação acerca do coronavírus (Bohnert; Pavão; Corrêa, 2022) apresentam as ferramentas de visualização de dados geoespaciais, *Johns Hopkins Coronavirus Resource Center (CRC)*, *Our World In Data* e Painel Coronavírus da Organização Mundial da Saúde (OMS), que serviram de fontes de informações para auxiliar os governantes nas tomadas de decisões no intuito de conter o avanço do vírus, assim como também para os cidadãos de forma geral interessados em entender a dinâmica de surgimento, contaminação, mortes. Essas ferramentas são resultados da colaboração interdisciplinar entre cientistas e colaboradores. Além disso, as ferramentas de visualização baseadas em representações geoespaciais, *Choropleth Map* e *Bubble Map* foram abordadas neste estudo. Essas ferramentas foram descritas na seção 4.1.

O *web site* implantado por (Stutz et al 2021) possibilita o monitoramento de dados que detectam o trabalho remoto relacionado à Covid-19, baseado em alterações no uso da internet, podendo proporcionar grande utilidade para visualizações de dados semelhantes no futuro, segundo os autores.

No estudo de Walker e Sharma (2021), foi observado que na pandemia do coronavírus houve uma queda do número de crimes na cidade de Baltimore devido às ações legislativas tomadas pelas autoridades locais e federais do governo de Maryland (EUA), que afetaram as práticas de crimes. Outra percepção dos autores é que a adoção de ferramentas de visualização de dados, como os softwares Unity 3D e Maptitude, aproximam as informações obtidas para a perspectiva humana promovendo maior interatividade e compreensão nos cenários onde tais ferramentas atuam a respeito dos problemas que se pretende analisar.

4.2.3 Uso de Redes Sociais: Twitter (identificador dos artigos 4, 26 e 27)

A iniciativa de Taeb, Chi e Yan (2021) serve de exemplo para que os ambientes virtuais sociais sejam levados, em consideração na disseminação e compartilhamento massivo de informações, e que em situações de interesse público, como a pandemia da Covid-19, as autoridades precisam dar atenção a políticas de conscientização, sensibilização e esclarecimentos para combater a circulação de informações que possam prejudicar o bem estar comum da sociedade. Os autores apresentam um framework de análise de tweets para medir a aceitação, ou negação, da vacinação contra a Covid-19. A proposta concluiu que as especulações baseadas em teorias da conspiração foram fatores que reforçaram a recusa por parte dos usuários com relação à vacina, e perceber os sentimentos e manifestações anti-vacina na rede social twitter contribui para a elaboração de políticas pró vacinas em situações de pandemia.

Já a pesquisa de León-Sandova (et al, 2022) utiliza a rede social twitter para analisar aspectos mais profundos da sociedade durante a pandemia. Considerando que os impactos negativos foram muitos e em diversos segmentos relacionados à sociedade, apesar dos números ligados à taxa de mortalidade terem sido o fator mais alarmante nesse contexto emergencial de saúde, há ainda causas complexas e íntimas dos seres humanos que foram expostas, voluntariamente ou não, durante a pandemia, como questões de cunho psicológico que pesaram muito na sociedade nesse período, pois muitas pessoas que já tinham problemas de saúde psicológica receberam gatilhos devido à necessidade de isolamento social exigida pela crise em longos períodos de quarentena e *block down*.

Além do que, curiosamente, indivíduos que nunca tiveram casos de fragilidade psíquica desencadearam quadros de ansiedade e depressão, por exemplo, devido às pressões exercidas pelo alarmante avanço do coronavírus.

Nesta perspectiva, León-Sandova, et al (2022) desenvolveram um extenso sistema de dados utilizando servidores modernos baseados em arquiteturas de nuvem, para analisar 760.064.879 *tweets* de domínio público coletados de um repositório de acesso público. Esses dados foram utilizados para examinar as mudanças coletivas no humor geral da população mexicana, no período de fevereiro a dezembro de 2020, em relação à evolução da pandemia, aos ciclos de notícias e às políticas, por meio do uso de ferramentas de análise de sentimentos abertas. O estudo apresentou métricas, vantagens e desafios do desenvolvimento de arquiteturas baseadas em nuvem sem servidor para um projeto de processamento de linguagem natural de grande magnitude. Os autores demonstraram que esse sistema pode processar eficientemente uma quantidade vasta de dados, tanto em lote quanto em modo de fluxo, possibilitando a utilização de diversos modelos de linguagem e facilitando a exploração e visualização dos dados.

O estudo de Li et al (2020) considerou a propagação de notícias falsas relativas à Covid-19, e perceberam a necessidade de criar e divulgar uma ferramenta de identificação dessas informações, em vários idiomas, para guiar a sociedade durante a pandemia.

Li (et al, 2020) criaram um repositório coletando dados através da captura de tela de imagens inverídicas, e analisaram pistas de como identificar quando uma informação é falsa, avaliando o uso de termos relacionados aos países onde os casos de Covid-19 foram mais críticos, em três tipos de contexto social: perfis de usuários, postagens de tweets e estrutura de rede social.

4.2.4 Uso de Modelos Matemáticos no Combate à Covid-19 (identificador dos artigos 3 e 20)

Adiga (et al, 2020) realizaram um estudo para destacar alguns dos principais modelos computacionais tradicionais atualmente utilizados para apoiar a resposta à pandemia de Covid-19, e apresentar ao leitor uma variedade das opções que podem ser utilizados em tais situações. Para realizar uma comparação cuidadosa entre os modelos, os autores sugerem considerar: (a) o propósito do modelo, (b) o usuário final a quem o modelo se destina, (c) a resolução espacial e temporal do modelo e (d) as suposições e limitações subjacentes. Deste modo, a pesquisa foi

organizada apresentando definições preliminares na Seção 2; discussão dos modelos centrados nos EUA e no Reino Unido desenvolvidos por pesquisadores do Imperial College na Seção 3;

Na Seção 4, apresentaram abordagens dos modelos de metapopulação focados nos EUA, desenvolvidos pelo próprio grupo, e os modelos desenvolvidos por pesquisadores da Northeastern University. A Seção 5 descreve os modelos desenvolvidos por pesquisadores suecos para estudar o surto na Suécia; na Seção 6, discutiram os métodos desenvolvidos para previsão; e na Seção 8 realizaram a discussão, limitações dos modelos e considerações finais.

Já em 2021, Adiga et al (2021) observaram a alta complexidade na dinâmica de expansão da pandemia, e a necessidade de utilizar dados confiáveis para uma previsão mais precisa, os autores utilizaram o conjunto matemático bayesiano para desenvolver um pipeline de previsão de doenças infecciosas, especificamente com o foco na previsão dos casos confirmados de Covid-19 em nível municipal nos Estados Unidos.

O estudo apresenta uma abordagem inovadora baseada em estatística, aprendizado de máquina e abordagens mecánísticas com um conjunto baseado em desempenho para combinar os modelos, ressaltando que a combinação de diferentes métodos pode levar a previsões mais precisas e robustas para enfrentar a pandemia. A iniciativa surge como um instrumento de apoio ao Centro de Previsão Covid-19, desenvolvido em abril de 2020 de uma parceria entre os Centros de Controle e Prevenção de Doenças (CDC) com instituições acadêmicas.

4.2.5 Saúde Mental de Idosos na Pandemia - grupo de risco (identificador dos artigos 26)

A pandemia da Covid-19 promoveu impactos negativos de grandes proporções na sociedade. Profissionais da saúde atuaram na linha de frente de combate, lutando para minimizar as consequências catastróficas de pacientes infectados pelo coronavírus.

No decorrer desse processo, o fator saúde mental foi um elemento que trouxe desafios ainda maiores para os profissionais da saúde, uma vez que os períodos de quarentena e confinamento obrigaram os cidadãos a limitarem o convívio social às pessoas que moravam na mesma casa. Os idosos foram incluídos nos grupos de riscos de pessoas que possuem características que os classificam como indivíduos de saúde frágil, que são mais propensos a desenvolver um quadro mais grave caso sejam infectados pelo coronavírus. Essa condição colaborou para que este grupo fosse submetido a uma situação mais rígida de isolamento social.

Pensando nas necessidades relativas à promoção da saúde mental de pessoas idosas em situações de estresse como o da pandemia, (Cheng, et al, 223) desenvolveram uma pesquisa com grupos de idosos e cuidadores de Hong Kong, na China, utilizando o protocolo SMART com o intuito de melhorar a qualidade da saúde mental durante o envelhecimento.

A pesquisa foi conduzida obedecendo os critérios do protocolo SMART, que é um conceito que se refere a uma abordagem específica utilizada em pesquisas e projetos para definir metas e estratégias de intervenção. Incluindo o público alvo, idosos e cuidadores, e contando com os feedbacks para ajudar na implantação do projeto.

A sigla SMART representa cinco características fundamentais que as metas devem ter para serem eficazes: *Specific* (Específico); *Measurable* (Mensurável); *Achievable* (Alcançável); *Relevant* (Relevante); *Time-bound* (Com prazo definido). Os autores utilizaram essa abordagem pela proposta de garantir que as metas sejam bem formuladas, atingíveis e que haja clareza sobre o que se espera alcançar. Ao adotar o protocolo SMART, os projetos têm maior probabilidade de sucesso, pois as metas são mais bem planejadas, mensuráveis e orientadas para resultados concretos.

4.2.6 Impactos Econômicos Provocados pela Pandemia (identificador dos artigos 1)

Sha (et al, 2020) desenvolveram uma pesquisa concentrada na coleta de dados socioeconômicos em nível estadual nos Estados Unidos, com o intuito de fornecer informações importantes para a pesquisa sobre a pandemia de Covid-19. Esta iniciativa importante reúne diversas perspectivas socioeconômicas para auxiliar na compreensão dos impactos da pandemia de Covid-19

Os autores fazem uso da abordagem espaço-temporal, agregando conjuntos de dados de diferentes fontes, como política, economia e determinantes socioeconômicos, abrangendo a coleta de uma ampla gama de informações relevantes, como índice de rigor de políticas, atributos econômicos com múltiplas frequências de tempo e fatores socioeconômicos. A coleção de dados é atualizada diariamente e envolve todos os estados dos EUA.

O trabalho é conduzido em um contexto colaborativo, com especialistas em controle de doenças, tomadores de decisão, funcionários do governo, sociólogos, economistas e humanistas

envolvidos no projeto. Esse método multidisciplinar garante que os dados coletados sejam abrangentes e de alta qualidade, fornecendo percepções significativas para análises e pesquisas relacionadas à pandemia.

O armazenamento e compartilhamento dos dados são feitos em um modo operacional, permitindo que os usuários acessem e utilizem as informações de forma eficiente. O artigo destaca o uso de técnicas de computação em nuvem para gerenciar a infraestrutura computacional e garantir a disponibilidade dos dados. Além disso, o estudo prevê o enriquecimento da coleção de dados por meio de metadados adicionais oriundos de crowdsourcing, modelo aberto e compartilhado, o que permite a integração de fatores e atributos socioeconômicos inclusivos relacionados ao tema Covid-19.

4.3 DISCUSSÃO DOS RESULTADOS

A proposta de criar categorias de ferramentas e categorias de temas foi prevista com o intuito de agrupar os artigos extraídos de acordo com a relação que eles possuem entre si de acerca do conteúdo.

Abordando os trabalhos pelas ferramentas, os artigos com identificador 14 e 23 abordam, na categoria “Novas Tecnologias: *Big Data* e *Blockchain*”, estruturas baseadas em códigos de *blockchain* para possibilitar informações de big data com maior grau de segurança e transparência, levando em consideração os dados sensíveis relativos aos números da covid-19.

As ferramentas de “Sistemas de Informação Geográfica (GIS)”, “Dados de Mobilidade Humana” e “Representações Geoespaciais”, representadas em três grupos de artigos (8, 10 e 18), (12 e 17) e (28), respectivamente, são complementares, pois nos seis trabalhos envolvidos nestas ferramentas são utilizados dados de localização humana para rastrear os indivíduos e os casos de covid-19, visando acompanhar e conter a evolução do vírus.

O artigo de identificador com número 15 foi incluído na categoria que representa a ferramenta que foi desenvolvida especialmente para que os usuários pudessem acompanhar a dinâmica da pandemia “Painel de vigilância de Covid-19”.

As ferramentas de “*Pipeline*” consistem em técnicas estruturadas para apresentar instruções de fluxos e etapas, podendo servir de apoio para orientar os usuários acerca dos

processos seguidos desde a contaminação pelo coronavírus. Essas estruturas foram utilizadas nos artigos 2, 3 e 21.

O “Repositório MM-Covid” foi uma ferramenta criada para instruir usuários de mídias sociais a identificarem notícias falsas relativas à Covi-19, trazida no artigo identificado pelo número 25.

Outra ferramenta apresentada nesta pesquisa foi o “Sistema de Informação Hospitalar (SIS)”, artigo 19, que pode auxiliar o gerenciamento de hospitais, compreendendo os insumos, pacientes e profissionais da saúde, principalmente em momentos de emergência sanitária como durante a pandemia.

As ferramentas de “Modelagem de Dados”, “Modelos de Aprendizado” e “Modelos de Previsão”, representadas em três grupos de artigos (11 e 22), (7 e 16) e (6, 13 e 24), respectivamente, são modelos mais complexos que carecem de maior conhecimento específico na área da computação para serem facilmente assimilados, no entanto, para os fins desta pesquisa, pode-se dizer que foram estruturas propostas pelos autores com a intenção de apresentar contribuições da área para criar bancos de dados, prever padrões de contaminação pelo coronavírus e prever diagnóstico da doença, covid-19.

Do ponto de vista das categorias temáticas, os temas foram apresentados considerando a ideia central dos artigos, de forma mais teórica e analítica, trazendo o sentido mais genérico de cada tema. Como por exemplo, os artigos de identificador 2, 5, 14 e 23, que discutem o tema “Uso de Inteligência Artificial” para abordar as contribuições que podem ser aproveitadas pelo uso da IA durante momentos como a pandemia.

O tema “Visualização de dados” pode ser observado nos artigos 8, 9, 10, 15, 21, 23, 28 e 29, e foi a temática mais explorada dentro dos estudos apresentados nesta pesquisa, inferindo a importância das ferramentas que utilizaram representações gráficas como soluções no combate à pandemia.

“O uso de Redes Sociais: *Twitter*” parte das propostas de três artigos, sob os identificadores 4, 26 e 27, em utilizar as ferramentas de mineração de dados oferecidas pela plataforma *twitter* para mapear as informações compartilhadas entre os usuários, a fim de investigar o comportamento, os sentimentos e as decisões adotadas pelas pessoas durante as políticas públicas desenvolvidas no contexto pandêmico.

O tema “Uso de modelos matemáticos no combate à pandemia” foi criado a partir da leitura e identificação da ideia central dos artigos 3 e 20, que pretendem convencer acerca das vantagens da utilização desses modelos lógicos para lidar com momentos de incerteza como aquele da pandemia.

Os artigos com identificadores 26 e 1 e abordam temas mais específicos que foram considerados importantes serem incluídos nesta pesquisa: “Cuidados com a saúde mental de idosos na pandemia” e “Impactos econômicos provocados pela pandemia, respectivamente”. O primeiro considera os impactos psicológicos negativos que foram experimentados por muitos idosos durante a pandemia, devido ao fato de terem sido incluídos no grupo de risco nos casos de covid-19. O outro discorre acerca dos impactos que a crise sanitária causou no setor econômico dos Estados Unidos, mas que podem ser considerados a nível mundial, uma vez que as medidas de distanciamento social adotadas em diversos países durante a pandemia culminou no fechamento do comércio nas etapas mais críticas de transmissão do coronavírus.

Vale lembrar que as contribuições elencadas pelas soluções identificadas nesta pesquisa não se configuram como regra, ou um padrão, que se aplicada a qualquer realidade, localidade e contexto surtirão efeitos eficazes, pois cada país vivenciou suas próprias dificuldades, mas são estudos que naquela conjuntura somatizam esforços válidos na tentativa de minimizar os malefícios causados pelo coronavírus, e que podem ser revistos de forma aprofundada e experimentados.

É fundamental reconhecer que, como em qualquer pesquisa científica, este estudo apresenta algumas limitações que devem ser consideradas ao interpretar seus resultados. Embora os resultados encontrados nesta pesquisa tenham se limitado ao uso de artigos científicos de apenas 2 bases de dados, Brapci e *Web Of Science*, pode-se considerar satisfatórios uma vez que foi utilizada quase a metade desses produtos, a saber: 29 dos 60 artigos.

Uma dessas limitações está relacionada ao fator tempo, que não foi o suficiente para explorar outras bases de dados ou para verificar se os resultados da implementação das ferramentas apresentadas naqueles cenários foram satisfatórios, no entanto, acredita-se que não haveria quantidade significativa de trabalhos apresentando ferramentas diferentes das que foram apresentadas nessa pesquisa. Outra questão importante seria abordar as questões de confidencialidade de dados pessoais e da manipulação desses dados por máquinas. Estas

limitações oferecem oportunidades para reflexão e sugerem áreas que podem ser exploradas em futuras investigações a fim de garantir a validade desta pesquisa.

5 CONSIDERAÇÕES FINAIS

A adoção de ferramentas de *big data* no combate à pandemia representa um marco significativo na resposta a crises de saúde pública. Essas ferramentas desempenharam um papel crucial ao fornecer às autoridades de saúde e pesquisadores informações valiosas em tempo real, permitindo uma abordagem mais eficaz para lidar com a Covid-19 e suas implicações. Como visto na etapa anterior, os aspectos vantajosos dessa atuação contemplam, predominantemente, ferramentas de análise e interpretação de dados em tempo real, modelagem preditiva, rastreamento de contatos eficiente e uma aceleração considerável da pesquisa científica.

O uso de ferramentas de big data na pandemia permitiu às autoridades de saúde monitorar a propagação do vírus, identificar e prever tendências da disseminação e das necessidades de recursos médicos com muito mais precisão, o que viabilizou a tomada de decisões rapidamente para que os sistemas de saúde se preparassem adequadamente para a demanda crescente, evitando a sobrecarga dos hospitais. Sem as ferramentas de big data para promover o processamento e a extração dos grandes volumes de dados produzidos durante a pandemia, as respostas acerca de pesquisas que contribuíram para a criação de vacinas, bem como a respeito da disseminação do coronavírus, seriam mais lentas, a falta de modelos de previsão poderia levar a uma sobrecarga inesperada dos sistemas de saúde, com consequências potencialmente devastadoras.

É compreensível, que as questões relativas ao uso excessivo de dados pessoais extraídos através das tecnologias de big data e inteligência artificial no período pandêmico não tenham sido o foco de muitos pesquisadores, como os citados nos resultados desta pesquisa. Naquela altura, a

busca incansável por soluções para minimizar os impactos negativos provocados pelo coronavírus era o mais importante. Objetivou-se conter o vírus, prever as dinâmicas de contágio e morte, correr contra o tempo nas investigações científicas para dar origem às vacinas que poderiam salvar vidas. Passados os momentos mais caóticos da crise sanitária, e tendo os ciclos de vacinação colaborado efetivamente para frear o avanço da pandemia, ou pelo menos para colocar a sociedade em uma situação mais cômoda de redução significativa do número de mortes, vale rever o que foi negligenciado acerca de dados pessoais sensíveis para que em situações semelhantes seja levado em consideração.

Durante a pandemia, tornou-se evidente a importância fundamental da informação na orientação das ações governamentais e privadas para enfrentar a Covid-19, a desordem causada pelo significativo aumento de casos de contaminação pelo coronavírus revelou que o uso apropriado de dados tinha o potencial de mitigar os impactos trágicos que o vírus poderia infligir à população. Também expôs fragilidades decorrentes do compartilhamento indevido de informações falsas, uma das faces negativas da manipulação desenfreada de dados. Este cenário transparece a necessidade de profissionais da informação para destacar que a reputação da informação é tão essencial quanto o seu papel de informar, e que os desafios diante da propagação de informações duvidosas, tendenciosas, ou mesmo inverídicas, aumentam concomitantemente à produção e ao avanço do compartilhamento dessas informações.

A tentativa de minimizar o número das contaminações e de disseminar informações seguras e confiáveis promoveu buscas intensivas de dados relacionados à Covid-19, colocando profissionais acadêmicos e técnicos, não só da área da saúde, numa corrida contra o tempo para ampliar as pesquisas orientadas a entender e tratar o Coronavírus, produzindo centenas de artigos científicos e soluções tecnológicas que revelaram as sequelas e impactos sociais causados pelo coronavírus.

Diante dos expostos, em contextos de significativa necessidade informacional, rodeados de autores que manipulam de maneira duvidosa a informação, como durante a pandemia da Covid-19, o profissional da informação ganha notória visibilidade e pode contribuir de maneira relevante no combate a informações falsas e na distribuição daquelas providas de elementos que garantam seu caráter verdadeiro, com poder decisório e transformador da realidade em vigor. Este traço de mediador que o profissional da informação exerce pode suscitar desconfiança, ou mesmo estranheza, com relação ao uso da tecnologia de algoritmos para determinar planos de

ação durante a pandemia baseados em análises automatizadas de dados pessoais desprovidos de supervisão humana, como será possível acompanhar mais a frente neste estudo.

Os objetivos desta pesquisa foram alcançados mediante a descrição dos resultados em toda a seção 4, uma vez que foram identificadas soluções baseadas em big data que auxiliaram no combate à pandemia da Covid-19, através do mapeamento do uso de tecnologia de big data e da identificação das principais ferramentas utilizadas durante a pandemia, tendo sido apresentadas 11 categorias de ferramentas e 6 categorias temáticas.

Pensando em expandir e enriquecer esta pesquisa futuramente, propõe-se considerar estudos como o de Neto; Silva e Nogaroli (2020), para abordar as vantagens e desvantagens, bem como os riscos, na utilização de inteligência artificial e big data no diagnóstico de saúde, conforme foi mostrado na seção 2 e 4 desta pesquisa. Outro ponto importante de discutir é a Lei Geral de Proteção de Dados (LGPD) através do estudo de Nascimento e Silva (2023), que traz muitas contribuições a respeito desse tema sensível.

Esta pesquisa tem profunda relevância em termos práticos e teóricos para a ciência da informação, pois o uso de grandes volumes de dados envolve muitas características da CI acerca do trato com a informação como: gestão e organização, acesso, avaliação da confiabilidade e qualidade, privacidade e ética, representação. Contribui para a valorização do profissional da informação como ator qualificado e consciente para os desdobramentos que a manipulação de informações pode acarretar, principalmente, tratando-se de proteger a individualidade humana. Apoiar pesquisadores interessados em desenvolver ou utilizar ferramentas tecnológicas para solucionar problemas semelhantes aos causados pela pandemia.

REFERÊNCIAS

- ALECRIM, E. **O que é big data?** Infowester. 2015 [acesso 19 ago 2023]. Disponível em: <http://www.infowester.com/big-data.php>.
- ALSUNAI, Shikah J. et al. *Applications of big data analytics to control Covid-19 pandemic*. Sensors, v. 21, n. 7, p. 2282, 2021.
- ANDY, Andy; EL-HAMZAH, Salsabilla Maula Zalfa; SINATRA, Jadeny. *Big Data Analytics in Patient Care During Coronavirus Disease 2019 (Covid-19) Pandemic: A Systematic Review*. JMMR (Jurnal Medicoeticolegal dan Manajemen Rumah Sakit), v. 11, n. 3, p. 245-263, 2022.
- ARENCIBIA-JORGE, Ricardo et al. *The multidisciplinary nature of Covid-19 research*. Iberoamerican Journal of Science Measurement and Communication, p. v.1, 2020. 11 p.
- AZEVEDO, A. W. Metodologia de identificação de fonte e coleta de informação: Uma proposta de modelo para cadeia produtiva do couro, calçados e artefatos. **Perspectivas em Gestão & Conhecimento**, v. 2, p. 149-158, 2012.
- BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2016.
- BARRETO, Aldo. A aventura de perceber significados. **DataGramZero**, v. 15, 2014.
- BORKO, H. *Information science: what is it?* American Documentation, v.19, n.1, p.3-5, 1968.
- BRAGAZZI, NL; DAI, H.; DAMIANI, G.; BEHZADIFAR, M.; MARTINI, M.; WU, J. Como Big Data e Inteligência Artificial podem ajudar a gerenciar melhor a pandemia de COVID-19. *Internacional J. Meio Ambiente*. **Res. Saúde Pública** 2020 , 17 , 3176. <https://doi.org/10.3390/ijerph17093176>
- BRASIL. **Ministério da Saúde**. Gabinete do Ministro. Portaria nº 104, de 25 de janeiro de 2011. Define as terminologias adotadas em legislação nacional, a relação de doenças, agravos e eventos

em saúde pública de notificação compulsória em todo o território nacional e estabelece fluxo, critérios, responsabilidades e atribuições aos profissionais e serviços de saúde. Brasília, 2011.

CANBAY, Y.; VURAL, Y.; SAGIROGLU, S. *Privacy Preserving Big Data Publishing*. In: *18th International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism* (IBIGDELFT), Ankara, Turkey, 2018, pp. 24-29, doi: 10.1109/IBIGDELFT.2018.8625358. (ver como citar corretamente esse artigo publicado em um proceeding de conferência segundo a ABNT).

CHEN, M., MAO, S., LIU, Y. *Big Data: A Survey*. *Mobile Netw Appl*, v. 19, n. 1, p. 171–209, 2014.
<https://doi.org/10.1007/s11036-013-0489-0>.

CORSI, Alana et al. *Big data analytics as a tool for fighting pandemics: a systematic review of literature*. *Journal of ambient intelligence and humanized computing*, v. 12, n. 10, p. 9163-9180, 2021.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008.

CURY, Maria Eduarda. **Novo coronavírus pode ter surgido antes de dezembro**. 2020. Disponível em <<https://exame.com/ciencia/novo-coronavirus-pode-ter-surgido-antes-de-dezembro/>> . Último acesso em 19/07/2023.

DAVENPORT, T. E. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação**. São Paulo: Futura, 1998.

DE MAURO, ANDREA ; GRECO, MARCO ; GRIMALDI, MICHELE. What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings*, Volume 1644, Issue 1, p.97-104, February 2015.

DENZIN, N. K; LINCOLN, I. **O planejamento da pesquisa qualitativa: teorias e abordagens**. Porto Alegre: Artmed, 2006.

DUNNE, T. *Big data, analytics, and energy consumption*. *Lavastorm Agle Analytics, Boston, Mass, USA, 2012*.

EREKSON, Elisabeth A.; IGLESIA, Cheryl B. *Improving patient outcomes in gynecology: the role of large data registries and big data analytics*. *Journal of minimally invasive gynecology*, v. 22, n. 7, p. 1124-1129, 2015.

FAGUNDES, P. B.; MACEDO, D. D. J.; DUTRA, M. L. **Uma análise das relações entre a qualidade da informação e big data**. , p. 206-220, . Disponível em: <http://hdl.handle.net/20.500.11959/brapci/101607>. Acesso em: 31 out. 2023.

GALVÃO, Maria Cristiane Barbosa et al. Usos de tecnologias da informação e comunicação no ensino superior em Enfermagem durante a pandemia da Covid-19. *Brazilian Journal of Information Science*, v. 15, n. 1, p. 8, 2021.

GHEBREYESUS, T. A. **OMS decreta pandemia do novo coronavírus: saiba o que isso significa.** Veja, Saúde, v. 11, 2020. Disponível em <<https://saude.abril.com.br/medicina/oms-decreta-pandemia-do-novo-coronavirus-saiba-o-que-isso-significa/>>. Último acesso 12/07/2023.

GIL, A. **Métodos e técnicas de pesquisa social.** Atlas, 2010.

HAAFZA, Laraib Aslam et al. Big data Covid-19 systematic literature review: Pandemic crisis. *Electronics*, v. 10, n. 24, p. 3125, 2021.

HAMID, Soomaiya et al. *A systematic review and IoMT based big data framework for Covid-19 prevention and detection.* *Electronics*, v. 11, n. 17, p. 2777, 2022.

HEATON, J. *Secondary analysis of qualitative data: An overview.* *Historical Social Research/Historische Sozialforschung*, p. 33-45, 2008.

HERBERT, R. *Practical evidence-based physiotherapy.* Elsevier Health Sciences, 2005.

HU F, JIANG J E YIN P. (2020). *Prediction of potential commercial inhibitors against SARS-CoV-2 by multi-task deep model.* arXiv preprint arXiv:2003.00728.

IBM. *Big Data.* 2017 [acesso em 19 ago 2023]. Disponível em: <http://www.ibm.com/big-data/us/en/>.

IBM. *The Four V's of Big Data and Its Value.* 2019 [acesso em 19 ago 2023]. Disponível em: https://www.researchgate.net/figure/The-Four-Vs-of-Big-Data-and-Its-Value-IBM-2019_fig1_344303432.

KFOURI NETO, M.; SILVA, R. D. G.; NOGAROLI, R. **Inteligência artificial e big data no diagnóstico e tratamento da covid-19 na América Latina:** novos desafios à proteção de dados pessoais. *Revista Brasileira de Direitos Fundamentais & Justiça*, [S. l.], v. 14, n. 1, p. 149–178, 2020. DOI: 10.30899/dfj.v0i0.974. Disponível em: <https://dfj.emnuvens.com.br/dfj/article/view/974>. Acesso em: 25 ago. 2023.

KHAN, N. *Big data: survey, technologies, opportunities, and challenges.* *The Scientific World Journal*, v. 18, n. 1, 2014.

KITCHENHAM, B. A.; DYBA, T.; JORGENSEN, M. *Evidence-based Software engineering.* In: *Proceedings Of The 26th International Conference on Software Engineering.* IEEE Computer Society, 2004. p. 273-281.

LANEY D (2001) 3D data management: Controlling data volume, velocity and variety. In: Meta Group. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 16 January 2013).

LAVILLE, Christian; DIONNE, Jean. **A construção do saber: manual de metodologia da pesquisa em ciências humanas**. Porto Alegre: Artmed; Belo Horizonte: Editora UFMG, 1999.

LE COADIC, Y-F. **A ciência da informação**. 2 ed. rev. amp. Brasília: Brinquet de Lemos, 2004.

LEE, Yang W. et al. AIMQ: *A methodology for information quality assessment*. *Information and Management*, v. 40, n. 2, 2002, p. 133-146. Disponível em: <<http://mitiq.mit.edu/Documents/Publications/TDQMpub/2002/AIMQ.pdf>>. Acesso em: 20 maio 2017.

LITTELL, J. H. *Systematic Reviews in the Social Sciences: A Review*. 2006.

LV Y, MA C, LI X, WU M. *Big data driven COVID-19 pandemic crisis management: potential approach for global health*. *Arch Med Sci*. 2021 Mar 20;17(3):829-837. doi: 10.5114/aoms/133522. PMID: 34025856; PMCID: PMC8130465.

LUGMAYR, A., STOCKLEBEN, B., SCHEIB, C. AND MAILAPARAMPIL, M.A. (2017), "Cognitive big data: survey and review on big data research and its implications. What is really "new" in big data?", *Journal of Knowledge Management*, Vol. 21 No. 1, pp. 197-212. <https://doi.org/10.1108/JKM-07-2016-0307>

MADGE, O. L. *Evidence Based Library and Information Practice*. *Library & Information Science Research*, v. 15, 2011.

MANOJ, A.; SHWETA, A. *A systematic review on artificial intelligence/deep learning applications and challenges to battle against Covid-19 pandemic*. *Disaster Advances*, p. 90-99, 2021.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2003.

MEHTA N, SHUKLA S. *Pandemic Analytics: How Countries are Leveraging Big Data Analytics and Artificial Intelligence to Fight Covid-19?* *SN Comput Sci*. 2022;3(1):54. doi: 10.1007/s42979-021-00923-y. Epub 2021 Nov 9. PMID: 34778841; PMCID: PMC8577168.

MENDONÇA, F. M.; DANTAS, M. A. R. Covid-19: Where is the Digital Transformation, Big Data, Artificial Intelligence and Data Analytics?. **Revista do Serviço Público**, [S. l.], v. 71, p. 212-234, 2020. DOI: 10.21874/rsp.v71i0.4770. Disponível em: <https://revista.enap.gov.br/index.php/RSP/article/view/4770>. Acesso em: 23 ago. 2023.

MICHEL, M. **Metodologia e Pesquisa Científica em Ciências Sociais**. Atlas, 2009.

MILAGRE, J. A.; SEGUNDO, J. E. S. **As contribuições da ciência da informação na perícia em informática no desafio envolvendo a análise de grandes volumes de dados - big data**. , p.

35-48, . Disponível em: <http://hdl.handle.net/20.500.11959/brapci/41287>. Acesso em: 01 nov. 2023.

MOURA, Luciano Raizer. **Gestão estratégica da informação: a proposição de um modelo de gestão. 150 Maiores Empresas**: Espírito Santo. Vitória, ano VIII, n. 8, 04/11/2004.

NADKARNI, Girish N.; COCA, Steven G.; WYATT, Christina M. *Big data in nephrology: promises and pitfalls*. *Kidney International*, v. 90, n. 2, p. 240-241, 2016.

NAIR, Chithra S. et al. RESEARCH TRENDS OF OPEN ACCESS PUBLICATIONS IN LIBRARY AND INFORMATION SCIENCE DURING Covid-19 PANDEMIC: A BIBLIOMETRIC ANALYSIS. *Journal of Positive School Psychology*, v. 6, n. 7, p. 1517-1524, 2022.

NASCIMENTO, B. L. C.; SILVA, E. M. Lei geral de proteção de dados (lgpd) e repositórios institucionais: reflexões e adequações. **Em Questão**, v. 29, p. 127314, 2023. DOI: 10.1590/1808-5245.29.127314 Acesso em: 01 set. 2023.

OPAS, 2020. **Histórico da Pandemia Covid-19**. Disponível em: <<https://www.paho.org/pt/Covid19/historico-da-pandemia-Covid-19>>. Último acesso em 12/07/2023

PENCE, HARRY E. *What Is Big Data And Why Is It Important?* J. Educational Technology Systems, Vol. 43(2) 159-171, 2014-2015

PETTIGREW, MARK; ROBERTS, Helen. *Systematic reviews in the social sciences: a practical guide*. John Wiley & Sons, 2008.

PIMENTA, R. M. *Big data e controle da informação na era digital*. , . Disponível em: <http://hdl.handle.net/20.500.11959/brapci/151577>. Acesso em: 31 out. 2023.

PFEFFER, J.; SUTTON, R. I. *Evidence-based management*. *Harvard Business Review*, v. 84, n. 1, p. 62, 2006.

PRICE, William Nicholson. *Artificial Intelligence in Health Care: Applications and Legal Issues*. University of Michigan Public Law Research Paper, n. 599, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3078704. Acesso em: 25 agos. 2023.

RODDHAM, M. Evidence- based practice for information professionals: a handbook. *Health Information & Libraries Journal*, v. 21, n. 4, p. 276-277, 2004.

SACKETT, D. L. *Evidence- based medicine*. John Wiley & Sons, Ltd, 2000.

SANTOS, Antonio Raimundo dos. **Metodologia científica: a construção do conhecimento**. Rio de Janeiro: DP&A, 2006.

SANTOS-D'AMORIM, K.; CRUZ, R. W.; SILVA, M. L.; CORREIA, A. E. G. C. Dos dados ao conhecimento: tendências da produção científica sobre Big Data na Ciência da Informação no Brasil. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 25, p. 01–23, 2020. <https://doi.org/10.5007/1518-2924.2020.e70527>

SOUZA, M.; ALMEIDA, F. G. **O comportamento do termo dado na ciência da informação**. *Ciência da Informação em Revista*, v. 8, n. 2, p. 39-54, 2021. DOI: 10.28998/cirev.2021v8n2c Acesso em: 20 mar. 2023.

SOUZA, R. R.; ALMEIDA, M. B.; PORTO, R. M. A. B. Ciência da informação em transformação: **big data, nuvens, redes sociais e web semântica**. , . DOI: 10.18225/ci.inf.v42i2.1379 Acesso em: 01 nov. 2023.

TESINI, Brenda L. **Coronavírus e síndromes respiratórias agudas (Covid-19, Mers e Sars). Manual MSD para profissionais da saúde**, 2020. Disponível em <<https://www.msmanuals.com/pt-br/casa/infec%C3%A7%C3%B5es/v%C3%ADrus-respirat%C3%B3rios/coronav%C3%ADrus-e-s%C3%ADndromes-respirat%C3%B3rias-agudas-Covid-19-mers-e-sars>>. Último acesso em 12/07/2023

TRANFIELD, D.; DENYER, D.; SMART, P. *Towards a methodology for developing evidence-informed management knowledge by means of systematic review*. *British Journal of Management*, v. 14, n. 3, p. 207-222, 2003.

VECHIATO, Fernando Luiz. **Encontrabilidade da informação: contributo para uma conceituação no campo da Ciência da Informação**. 2013. Tese (Programa de PósGraduação em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Marília

VICECONTI, Marco; HUNTER, Peter; HOSE, Rod. *Big data, big knowledge: big data for personalized healthcare*. *IEEE journal of biomedical and health informatics*, v. 19, n. 4, p. 1209-1215, 2015.

VILLARS; Olofson; Eastwood. *WHITE PAPER Big Data: What It Is and Why You Should Care* - AMD. Publicação: June 2011.

WANG, Yichuan et al. *An integrated big data analytics-enabled transformation model: Application to health care*. *Information & Management*, v. 55, n. 1, p. 64-79, 2018

WONG, Ho Ting et al. Big data as a new approach in emergency medicine research. *Journal of Acute Disease*, v. 4, n. 3, p. 178-179, 2015.