



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE CENTRO DE
TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO**

**TRILHAS PARA AUDITORIA APLICADA AO SISTEMA ÚNICO DE SAÚDE NO
CONTEXTO DAS ÓRTESES, PRÓTESES E MATERIAIS ESPECIAIS (OPMES):
ESTRUTURAÇÃO DE UMA BASE DE DADOS INTEGRADA A PARTIR DO
PROJETO FISCALIZASUS**

LUCA PAREJA CREDIDIO FREIRE ALVES

NATAL/RN

2025

LUCA PAREJA CREDIDIO FREIRE ALVES

**TRILHAS PARA AUDITORIA APLICADA AO SISTEMA ÚNICO DE SAÚDE NO
CONTEXTO DAS ÓRTESES, PRÓTESES E MATERIAIS ESPECIAIS (OPMES):
ESTRUTURAÇÃO DE UMA BASE DE DADOS INTEGRADA A PARTIR DO
PROJETO FISCALIZASUS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da UFRN (área de concentração: Engenharia de Computação) como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica e de Computação.

Orientador: Prof. Dr. Ricardo Alexsandro de Medeiros
Valentim

NATAL/RN

2025

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Alves, Luca Pareja Credidio Freire.

Trilhas para auditoria aplicada ao sistema único de saúde no contexto das órteses, próteses e materiais especiais (OPMES): estruturação de uma base de dados integrada a partir do projeto fiscalizatus / Luca Pareja Credidio Freire Alves. - 2025.

114 f.: il.

Dissertação (mestrado) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Natal, 2025.

Orientação: Prof. Dr. Ricardo Alexsandro de Medeiros Valentim.

1. Auditoria em saúde - Dissertação. 2. SUS - Dissertação. 3. Órteses, próteses, e materiais especiais - Dissertação. 4. Big Data - Dissertação. 5. irregularidade - Dissertação. I. Valentim, Ricardo Alexsandro de Medeiros. II. Título.

RN/UF/BCZM

CDU 657.6:614

RESUMO

Contexto: O Sistema Único de Saúde (SUS) foi um marco histórico firmado pela Constituição Federal de 1988. Ao longo dos anos, o SUS passou por diversas transformações, sobretudo no que diz respeito à quantidade de dados gerados a respeito de seus processos internos. Para comandar os processos de controle e auditoria interna no SUS, foi criado o Departamento Nacional de Auditoria do SUS (DenaSUS). Mediante isso, novas demandas referentes à auditoria no SUS foram surgindo. Uma parte dessas atividades foi disposta em parceria entre o DenaSUS, a Universidade Federal do Rio Grande do Norte (UFRN), o Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), o Ministério Público Federal do Rio Grande do Norte (MPF/RN), o Tribunal de Contas da União (TCU) e demais autoridades competentes para a construção do projeto FiscalizaSUS, que visa a proporcionar métodos inteligentes para a análise de grandes volumes de dados no campo da saúde, tendo como objetivo final proporcionar trilhas de auditoria focadas em Regulação e Órteses, Próteses e Materiais Especiais (OPME) no domínio do SUS, pois são insumos que já foram identificados como objetos de interesse em escândalos de corrupção envolvendo profissionais da saúde, como foi o caso da “máfia das próteses”. **Objetivo:** este trabalho busca demonstrar como o uso de métodos estatísticos e aplicações de Big Data podem otimizar e viabilizar a construção de trilhas de auditoria em OPME por meio da integração de bases de dados públicos e privados. **Metodologia:** inicialmente, foi conduzido um mapeamento sistemático da literatura com o intuito de analisar de forma qualitativa quais métodos estão sendo empregados para pesquisas de auditoria em saúde baseadas em Big Data. A partir disso, foi possível refletir a respeito de quais tipos de abordagem seriam ideais para o desenvolvimento de um modelo relacional para trilhas de auditoria em OPME, e, conseqüentemente, para a elaboração de métricas pertinentes de análise adequada entre as diversas bases de dados adotadas. **Resultados:** O Mapeamento Sistemático da Literatura (MSL), que contava inicialmente com um total de 344 artigos, obtidos a partir da string de busca utilizada. Com isso, sobraram 40 artigos para avaliação e coleta de dados. Foram identificadas soluções heterogêneas para abordar a temática de auditoria em saúde, apresentando técnicas também relacionadas a métodos computacionais e tratamentos de grandes volumes de dados. Uma das necessidades do ponto de vista operacional para o funcionamento das trilhas de auditoria foi a organização das tabelas de interesse orientadas a um esquema-estrela, contendo tabelas de fato e dimensão. Desse modo, é factível a agregação das diferentes informações e bases de dados utilizadas no projeto, respeitando os objetivos de cada trilha de auditoria e avaliando a presença

de valores comuns entre as bases. Não obstante, a integração entre as bases de dados possibilitou a visualização dos indícios encontrados na interface visual do Apache Superset, permitindo uma visualização abrangente no que se refere à localização geográfica, aos vínculos entre os dados e à identificação por meio das métricas propostas na metodologia. **Conclusão:** a integração de diferentes fontes de dados relacionadas à temática de OPME podem colaborar para a identificação de irregularidades em processos transacionais envolvendo esse tipo de insumo médico. Além disso, o intermédio de uma arquitetura de Big Data fornecida pelo projeto FiscalizaSUS viabilizou a agregação desses diferentes conjuntos de dados e trouxe novas perspectivas para as trilhas de auditoria em saúde no escopo do projeto.

Palavras-chave: auditoria em saúde; SUS; órteses, próteses e materiais especiais; Big Data; irregularidade.

ABSTRACT

Background: The Unified Health System (SUS) was a historic milestone established by the 1988 federal constitution. However, over the years the SUS has undergone several transformations, especially concerning the amount of data generated about its internal processes. The National Audit Department of SUS (DenaSUS) was created to manage the SUS's internal control and audit processes. As a result, new demands regarding auditing in the SUS have emerged. One of these activities was organized in partnership between DenaSUS, the Federal University of Rio Grande do Norte (UFRN), the Federal Institute of Rio Grande do Norte (IFRN), the Federal Prosecution Office of Rio Grande do Norte (MPF/RN), the Federal Court of Auditors (TCU) and other competent authorities to build the FiscalizaSUS project, which aims to provide intelligent methods for analyzing large volumes of data in the field of health, with the ultimate goal of delivering audit trails focused on Regulation and Special Orthotics, Prosthetics and Special Materials (OPSM) in the SUS domain, as these are inputs that have already been identified as objects of interest in corruption scandals involving health professionals, as was the case with the “máfia das próteses”. **Goal :** Thus, the objective of this work is to demonstrate how the application of statistical methods and Big Data can enhance the feasibility for developing OPSM audit trails through the integration of public and private databases. **Methods:** On the first hand, a systematic mapping of the literature was conducted in order to evaluate which methods are being applied for Big Data-based health audit research. From this point of view, it was possible to analyse and adopt approaches to develop a relational model for OPSM audit trials, and also develop metrics based on statistical analysis of the collected data samples. **Results :** The Systematic Literature Mapping (SLM) initially included a total of 344 articles, obtained from the search string used. This left 40 articles for evaluation and data collection, after the inclusion and exclusion steps. Heterogeneous solutions were identified to address the issue of health auditing, presenting techniques also related to computational methods and Big Data approaches. From an operational perspective, the tables were organized oriented towards a star schema, containing fact and dimension tables, which enable the integration of different data from various sources, also concerning the goals of each audit trail and assessing the presence of common values between the databases. Nevertheless, the integration between the databases also provided a visualization layer by using Apache Superset's virtual interface, allowing a comprehensive view of the anomalies through geographical location, links between the data sources, and also through the proposed metrics. **Conclusion:** The results allow us to conclude that the integration of different data sources

related to OPSM can contribute to the identification of irregularities in transactional processes involving this type of medical supply. In addition, the Big Data architecture provided by the FiscalizaSUS project also made it possible to aggregate these different data sets and brought new perspectives to the health audit trails within the scope of the project.

Keywords: Health auditing, SUS, orthoses, prostheses and special materials, Big Data.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer a Deus e minha família pelo apoio e encorajamento em minha jornada acadêmica durante o mestrado.

À minha namorada e aos meus amigos, que também acompanharam minha trajetória.

Ao LAIS/UFRN, ao NAVI/IFRN, ao Ministério da Saúde e às demais instituições parceiras do projeto FiscalizaSUS.

Aos professores Ricardo Valentim e Methanias Colaço, por me acompanharem e me guiarem durante este período, mas acima de tudo pela paciência e encorajamento.

Por fim, aos meus amigos e colegas do projeto FiscalizaSUS, Caldeira, Raphael, Helder, Wesckley e Gabriel, que sempre estenderam a mão para tirar dúvidas e me auxiliar.

LISTA DE FIGURAS

Figura 1 – Arquitetura de Big Data do projeto FiscalizaSUS	22
Figura 2 – Exemplo de procedimento SIGTAP registrado	32
Figura 3 – Quantitativo de Artigos publicados em Journals coletados por base	49
Figura 4 – Quantitativo dos artigos restantes para análise	50
Figura 5 – Fluxograma para exclusão de materiais não pertinentes.....	51
Figura 6 – Métodos mais utilizados.....	52
Figura 7– Ranqueamento por publicação	53
Figura 8 – Áreas ou tipo de dado em saúde identificados	54
Figura 9 – Quantitativo de publicações por ano	55
Figura 10 – Frequência de países na amostra selecionada	56
Figura 11 – Fluxograma de trabalho para estruturação de dados	72
Figura 12 – Distribuição do modelo relacional de dados mediante a arquitetura do FiscalizaSUS	77
Figura 13 – Teste do modelo relacional para Bolsa Família e OPME	87
Figura 14 – Índícios em procedimentos SIGTAP (Nível-Estadual).....	88
Figura 15 – Índícios em procedimentos SIGTAP (Nível-Municipal)	88
Figura 16 – Índícios em procedimentos com Stents recorrentes (nível Estadual).....	89
Figura 17 – Índícios em procedimentos com Stents recorrentes (nível Municipal).....	90
Figura 18 – Índícios em procedimentos com Stents recorrentes (Nível Estabelecimento de Saúde) -	90

LISTA DE QUADROS

Quadro 1 – Trilhas de auditoria analisadas como objeto de estudo	20
Quadro 2 – Aplicação dos conceitos de Big Data - Fonte: Elaborada pelo autor, 2025	34
Quadro 3 – Trabalho Correlatos	41
Quadro 4 – Protocolo PICO	43
Quadro 5 – Formulário de Extração de dados	47
Quadro 6 – Amostra dos artigos selecionados e seus desafios/limitações	58
Quadro 7 – Amostra de contribuições dos artigos selecionados	59
Quadro 8 – Dados referentes a Bolsa Família e Cadastro da Pessoa Jurídica.....	64
Quadro 9 – Origem das trilhas de Stents recorrentes e procedimentos SIGTAP	66
Quadro 10 – Dimensão sócio	79
Quadro 11 – Dimensão Auxílio.....	79
Quadro 12 – Dimensão Funcionário Público	80
Quadro 13 – Fato Sociedade -	80
Quadro 14 – Fato Bolsa Família.....	81
Quadro 15 – Dimensão Fornecedor.....	82
Quadro 16 – Dimensão Cidade.....	83
Quadro 17 – Dimensão Estado	83
Quadro 18 – Dimensão tempo	84
Quadro 19 – Fato Notas.....	85

LISTA DE ABREVIATURAS E SÍMBOLOS

ANVISA	Agência Nacional de Vigilância Sanitária
AIH	Autorização de Internação Hospitalar
API	Interface de Programação de Aplicação
CATMAT	Cadastro Oficial de Materiais do Governo Federal
CNES	Cadastro Nacional de Estabelecimentos de Saúde
DENASUS	Departamento Nacional de Auditoria do Sistema único de Saúde
DATASUS	Departamento de Informação e Informática do SUS
ETL	Extração, Transformação e Carregamento
LAIS	Laboratório de Inovação Tecnológica em Saúde
LDO	Lei de Diretrizes de Orçamentarias
LGPD	Lei Geral de Proteção de Dados
IFRN	Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte
NAVI	Núcleo Avançado de Inovação Tecnológica
OPME	Órteses, Próteses e Materiais Especiais
UFRN	Universidade Federal do Rio Grande do Norte
MPF/RN	Ministério Público Federal do Rio Grande do Norte
ODS	Objetivos de Desenvolvimento Sustentável
ONU	Organização das Nações Unidas
PAMDAS	Plataforma de Análise e Monitoramento de Dados Saúde
RNA	Redes Neurais Artificiais
SIGTAP do SUS	Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM
SNA	Sistema Nacional de Auditoria

SUMÁRIO

1 INTRODUÇÃO	14
1.1 CONTEXTUALIZAÇÃO	14
1.1.1 Sistema Único de Saúde do Brasil	14
1.1.2 Departamento Nacional de Auditoria do SUS (DenaSUS)	16
1.1.3 Projeto FiscalizaSUS	17
1.1.4 Trilhas de auditoria do Projeto FiscalizaSUS em Órteses, Próteses e Materiais Especiais (OPME)	18
1.1.5 Arquitetura FiscalizaSUS	20
1.2 PROBLEMÁTICA	22
1.2.1 Problemática em corrupção em saúde	23
1.2.2 Problemática em informações fragmentadas	23
1.3 OBJETIVOS	24
1.3.1 Objetivo Geral	24
1.3.2 Objetivos específicos	24
1.3.3 Hipótese da Pesquisa	25
1.3.4 Questões de Pesquisa	25
2 REFERENCIAL TEÓRICO	26
2.1 DIMENSÕES DE PESQUISA	26
2.1.1 Auditoria	26
2.1.2 Auditoria no SUS	28
2.1.3 Órteses, Próteses e Materiais Especiais (OPME)	30
2.1.4 Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)	31
2.1.5 Big Data	33
2.1.6 Arquitetura de referência para Big Data	34
2.1.7 Detecção de valores anômalos	36
3 TRABALHOS CORRELATOS	38
4 MAPEAMENTO SISTEMÁTICO DA LITERATURA	42
4.1 OBJETIVOS	42
4.2 ESTRATÉGIAS DE BUSCA E SELEÇÃO	44
4.3 CRITÉRIOS DE SELEÇÃO DE FONTES	45

4.4 ESTRATÉGIA DE EXTRAÇÃO DE DADOS	46
4.5 CONDUÇÃO DO MAPEAMENTO SISTEMÁTICO	48
4.6 RESULTADOS E DISCUSSÃO DO MAPEAMENTO	50
4.7.1 Quais os principais algoritmos e métodos estatísticos utilizados?	51
4.7.2 Quais os tipos de dados ou qual a área da saúde dos estudos analisados?	54
4.7.3 Em quais anos foram publicados mais artigos?	55
4.7.4 Quais os meios de publicação mais populares?	56
4.7.5 Quais países têm mais publicações nessa área?	56
4.7.6 Quais os principais desafios e limitações identificados?	57
4.7.7 Quais as contribuições desenvolvidas pelas pesquisas?	58
4.8 SÍNTESE NARRATIVA.....	60
4.9 AMEAÇAS À VALIDADE	61
4.9.1 Validação da string de busca	62
4.9.2 Variedade de Algoritmos, técnicas e tipos de dado em saúde	62
4.9.3 Viés de Inclusão e Exclusão	62
4.9.4 Viés de Seleção Temporal	63
4.10 CONSIDERAÇÕES FINAIS	63
5 METODOLOGIA.....	64
5.1 COLETA DA BASE DE DADOS	64
5.1.1 Trilha de <i>Stents</i> recorrentes	67
5.1.2 Trilha de procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)	68
5.2 TRATAMENTO DOS DADOS DA TRILHA BOLSA FAMÍLIA E MÉDICO FORNECEDOR	69
5.2.1 Tratamento dos dados para trilha de <i>Stents</i> recorrentes	70
5.2.2 Tratamento dos dados para trilha de Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)	71
5.3 ESTRATIFICAÇÃO DOS MODELOS DE DADOS.....	73
5.4 DICIONARIZAÇÃO DE DADOS	73
5.5 ESTUDO DE CASO	74
5.5.1 Elementos a ser criticados.....	75
5.5.2 Resultados do Estudo de Caso Qualitativo	75
5.5.3 Discussão do Estudo de Caso.....	76
6 RESULTADOS	77

6.1 AGREGAÇÃO DE DIVERSAS BASES DE DADOS EM ESQUEMA ESTRELA.....	77
6.2 DICIONARIZAÇÃO E VALIDAÇÃO DO MODELO ESTRATIFICADO PARA AS TRILHAS QUE ENVOLVEM AS BASES DE BOLSA FAMÍLIA E PESSOA JURÍDICA	78
6.3 IDENTIFICAÇÃO DE INDÍCIOS NA TRILHA DE PROCEDIMENTOS SIGTAP	87
6.4 IDENTIFICAÇÃO DE INDÍCIOS NA TRILHA DE <i>STENTS</i> RECORRENTES.....	89
6.5 RELACIONAMENTO ENTRE AS TRILHAS DE AUDITORIA DE ÓRTESES, PRÓTESES E MATERIAIS ESPECIAIS.....	91
7 DISCUSSÃO	92
7.1 CONSIDERAÇÕES DO MAPEAMENTO SISTEMÁTICO DA LITERATURA.....	92
7.2 ANÁLISE DE BIG DATA.....	92
7.3 AGREGAÇÃO DE BASES DE DADOS	93
7.4 ANÁLISE DAS TRILHAS DE AUDITORIA DO PROJETO.....	94
7.5 LIMITAÇÕES E PERSPECTIVAS FUTURAS.....	95
8 CONCLUSÃO.....	97
8.1 IMPACTOS SOCIAIS	98
REFERÊNCIAS	101
APÊNDICE A – MODELO DE DADOS ESTRATIFICADO COMPLETO, UTILIZANDO AS PRINCIPAIS TABELAS PARA AS TRILHAS DE AUDITORIA DE OPME	112

1 INTRODUÇÃO

Esta seção visa a introduzir a linha de raciocínio e o cenário presentes nesta pesquisa. Inicialmente, é apresentada a contextualização do trabalho em volta do Sistema Único de Saúde (SUS), a participação do Departamento Nacional de Auditoria do SUS (DenaSUS) e o escopo em que esta dissertação está inserida, mediante o projeto FiscalizaSUS.

1.1 CONTEXTUALIZAÇÃO

A seguir, procederemos à contextualização deste estudo, descrevendo o SUS, o FiscalizaSUS e em que consiste uma auditoria.

1.1.1 Sistema Único de Saúde do Brasil

O primeiro passo para o entendimento deste trabalho de dissertação é discutir a respeito do Sistema Único de Saúde (SUS) do Brasil. Isso pode ser explicado por meio do contexto de sua atuação ao longo da história e porque constitui um elemento fundamental para a sociedade brasileira. Também será importante apresentar a relevância de órgãos de controle e vigilância e o papel dos dados produzidos nesse ecossistema para processos que envolvem auditorias e análise de dados em saúde.

A promulgação da Constituição Federal de 1988 marcou um período de modernização institucional no Brasil, catalisando mudanças profundas em múltiplos setores da sociedade. No âmbito da saúde pública, a criação do Sistema Único de Saúde (SUS) foi um marco singular para a população brasileira, estabelecendo um sistema abrangente a fim de promover e regular o acesso equitativo aos serviços de saúde para toda a população brasileira (BRASIL, 1988; Lima *et al.*, 2019).

O Sistema Único de Saúde do Brasil é conhecido internacionalmente por seus princípios, especialmente pelo acesso universal e integral à saúde. Ele se destaca principalmente por atuar em um país como o Brasil, de tamanho continental (com cerca de 8.510.000 km²), com uma das maiores populações do mundo (cerca de 211 milhões de habitantes), e que, apesar de estar, atualmente, entre as 10 maiores economias do mundo (Nakamura, 2025), ainda apresenta grandes assimetrias regionais e profundas desigualdades sociais. Por isso, desperta o interesse de vários grupos de pesquisas de outras nações, especialmente depois do início da pandemia de HIV no mundo, quando o programa brasileiro ganhou destaque internacional,

conforme afirmam Marques (2002), Mendes (2013) e Teixeira *et al.* (2018). Não obstante, o SUS não se restringe ao campo epidemiológico e atinge outros setores da saúde que envolvem diferentes níveis de profundidade.

Além disso, é importante mencionar a atuação do SUS mediante procedimentos considerados de alta complexidade, sobretudo no que tange ao uso de insumos médicos como Órteses, Próteses e Materiais Especiais (OPME). Para isso, pode-se encontrar na literatura trabalhos que exponham a relevância desse tipo de tecnologia para a população brasileira. A publicação de Lima *et al.* (2024) é um estudo que traz dados estatísticos do Instituto Brasileiro de Geografia e Estatística (IBGE) referentes ao quantitativo de indivíduos que necessitam de algum tipo de auxílio que pode ser mediado pelo uso de OPME, sendo essa uma marca de aproximadamente 13 milhões de brasileiros. Além disso, o estudo faz um recorte baseado no estado do Maranhão a respeito de como profissionais da área de fisioterapia interagem mediante o uso desses equipamentos para o atendimento à população.

Outras publicações também evidenciam o papel e o impacto que as OPME possuem na saúde brasileira, como o trabalho de Siqueira, Siqueira-Filho e Land (2017), no qual os autores desenvolvem uma análise a respeito dos impactos econômicos causados pela incidência de doenças cardiovasculares no Brasil. O artigo também menciona o impacto nos custos referentes a procedimentos que envolvem OPME, como uma angioplastia. Nesse sentido, os autores apontam um aumento significativo nos gastos envolvendo esse tipo de insumo, no intervalo de 5 anos, de 2010 a 2015, estimado em mais de 150 milhões de reais. Dessa forma, pode-se identificar a significância que a aplicação desses materiais possui para a saúde brasileira.

De nos últimos anos, o Brasil enfrentou a pandemia de covid-19 e, com ela, diversos desafios surgiram. Entre eles, a governança e a gestão de recursos médicos, como OPME, sendo essa temática de interesse do trabalho apresentado por Pinheiro, Busato e Caveião (2021). Os autores fizeram um estudo comparativo entre o comportamento dos preços desses insumos médicos no intervalo do segundo semestre de 2019 ao segundo semestre de 2021 no estado do Amazonas. Além disso, o artigo enfatiza a importância de boas práticas de governança, controle e auditoria em saúde. Como resultado, identificou-se que os valores mais inflacionados para OPME mostraram-se nos meses de maio e junho de 2021, devido a assuntos de ordem política e financeira no país. Os autores também constataram a importância que a auditoria em saúde pode desempenhar para o gerenciamento adequado desse tipo de cenário, não só pandêmico, mas no campo de insumos.

Nos processos de cuidado e atenção à saúde, o SUS destaca-se na vida cotidiana dos brasileiros por desempenhar um papel-chave. Porém, ele não se resume à interação entre médicos, enfermeiros e pacientes. Com todo o seu alcance, vale questionar quais recursos o sistema utiliza para exercer controle e vigilância sobre suas despesas e atividades. Desse modo, o trabalho irá abordar o papel do Departamento Nacional de Auditoria do SUS (DenaSUS) como um órgão de controle e vigilância, além de destacar o papel do projeto FiscalizaSUS nesse setor. Esse projeto é oriundo de uma necessidade de auxiliar o trabalho de auditores nessa esfera, por meio de mediação tecnológica em saúde. Nesta seção, abordaremos ainda as trilhas de auditoria contempladas neste trabalho, a arquitetura de Big Data utilizada e seus componentes, e, por fim, a problemática e objetivos da pesquisa.

1.1.2 Departamento Nacional de Auditoria do SUS (DenaSUS)

A expansão do Sistema Único de Saúde do Brasil possibilitou a criação do Departamento Nacional de Auditoria do SUS (DenaSUS) (Silva *et al.*, 2022). O DenaSUS é responsável pela auditoria interna do SUS, mediante avaliação independente e objetiva das políticas públicas de saúde e a aplicação dos recursos federais executados no âmbito do SUS, tendo suas competências estabelecidas pelo Decreto nº 11.798, de 28 de novembro de 2023 (Fontes *et al.*, 2023).

Nessa perspectiva da auditoria, do controle e do monitoramento do Sistema Único de Saúde (SUS), o trabalho desenvolvido por Gomes *et al.* (2023) apresenta, em sua discussão, um relato a respeito de um evento sobre fraudes em processos transacionais relacionados a Órteses, Próteses e Materiais Especiais (OPME). Essa situação, discutida no ano de 2015, envolveu médicos, hospitais e outros atores que atuam na área da saúde. Situações como essa levam à reflexão acerca da importância das auditorias no SUS, dos impactos que elas podem promover em benefício do sistema de saúde no Brasil e, consequentemente, dos seus usuários. Desse modo, faz-se necessário discutir conceitos, métodos, técnicas, ferramentas e estratégias que aprimorem e favorecem os mecanismos de controle no SUS.

O SUS tem um dos maiores orçamentos da União sendo, portanto, responsável por um grande volume de recursos públicos (mais de R\$ 240 bilhões para o ano de 2025) (Agência Senado, 2025) vinculados a políticas públicas, programas e projetos. Segundo apresentação disponibilizada no Portal da Câmara dos Deputados (Unasus, 2025), o DenaSUS conta atualmente com 400 auditores em todo o país. Logo, é difícil que esse contingente de trabalhadores consiga desenvolver, de forma eficiente e oportuna, as atividades de controle

interno, tendo em vista as dimensões continentais do Brasil. Portanto, é fundamental para o desenvolvimento das atividades no âmbito do DenaSUS a apropriação de novas metodologias e ferramentas que auxiliem na execução das tarefas de controle interno.

1.1.3 Projeto FiscalizaSUS

Considerando o contexto em que o DenaSUS se insere, mediante seu efetivo de auditores e a quantidade de dados gerados no âmbito da saúde regularmente, soluções tecnológicas que possam auxiliar e qualificar os processos de auditoria no sistema de saúde se apresentam como uma alternativa para abordar essa problemática. Com base nessa realidade, foi firmado o projeto FiscalizaSUS, oriundo de uma parceria entre diversas instituições, como o Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), a Universidade Federal do Rio Grande (UFRN), o Ministério Público Federal do Rio Grande do Norte (MPF/RN), o Ministério da Saúde do Brasil (MS), o DenaSUS, a Universidade de Brasília (UNB).

No âmbito do IFRN, o motor responsável por conduzir o projeto é o Núcleo Avançado de Inovação Tecnológica (Navi), que opera no âmbito de processos de inovação tecnológica em saúde desde 2014, quando foi fundado por pesquisadores do IFRN. Desde então, o laboratório conta com 10 projetos de pesquisa, 6 patentes registradas e 53 registros de software. Portanto, é um dos parceiros com especialidade no desenvolvimento de soluções tecnológicas em saúde, sendo membro do projeto FiscalizaSUS.

Outro parceiro para a construção do Projeto FiscalizaSUS é o Laboratório de Inovação Tecnológica em Saúde (Lais) da UFRN. Situado no Hospital Universitário Onofre Lopes, o Lais vem desenvolvendo soluções de inovação em saúde mediadas por tecnologia aplicada desde sua criação, em 2010, nos mais variados escopos, como Informática em Saúde; Audição, Linguagem e Cognição na área de Fonoaudiologia; Bioengenharia e Tecnologias assistivas. Desse modo, são contabilizados cerca de 60 projetos em andamento ou concluídos.

Além disso, a importância do projeto FiscalizaSUS e suas atribuições mediante auditorias pode ser explicada por meio de um acontecimento recente em nossa história, a pandemia de covid-19. Desse modo, o Projeto tem como princípio a implementação de modelos analíticos escaláveis para a plataforma de gestão de dados do Ministério da Saúde, viabilizando análises relativas à comercialização de Órteses, Próteses e Materiais Especiais (OPME) e Regulação, por meio de trilhas de auditoria, métricas e parâmetros definidos atribuídos pelo DenaSUS.

1.1.4 Trilhas de auditoria do Projeto FiscalizaSUS em Órteses, Próteses e Materiais Especiais (OPME)

No que se refere ao planejamento e aos objetivos do projeto FiscalizaSUS, foi necessário estabelecer uma metodologia concisa e objetiva para execução de suas atividades em relação aos auditores do Sistema Único de Saúde. Para isso, foram construídas trilhas de auditoria com o intuito de agregar diversas fontes de dados para diferentes casos de uso. O conceito de trilha de auditoria pode ser definido como um processo metodológico direcionado para analisar e avaliar o desempenho de atividades (Carcary, 2020). O intuito é o de contribuir para um fluxo de trabalho resiliente e descrever detalhadamente as etapas e decisões tomadas ao longo da auditoria para a criação de indicadores (Johnson; Adkins; Chauvin, 2020). Desse modo, foi importante refletir a partir desse raciocínio para a elaboração tanto das trilhas de auditoria do projeto FiscalizaSUS como de métricas que serão discutidas na metodologia do trabalho.

As trilhas de auditoria agregadas ao projeto FiscalizaSUS também podem ser sustentadas por meio de elementos do cotidiano. Um deles diz respeito à divulgação de notícias na mídia tradicional a respeito de indícios de fraude que envolvem a OPME. O *Fantástico*, programa jornalístico da rede *Globo*, apresentou diversos problemas em uma reportagem que apontava fraudes relacionadas a procedimentos e superfaturamento de valores por uma empresa prestadora de serviços da ala ortopédica em hospitais públicos de três estados brasileiros (Fantástico, 2024¹). A reportagem de 2024 também apresenta uma série de prejuízos em relação aos pacientes afetados por essas relações anômalas. São reportadas situações em que pacientes sequer são notificados a respeito do tipo de procedimento cirúrgico ao qual estão sendo submetidos. Notícias como essa podem despertar diversos questionamentos e fomentar a busca por esclarecimentos. Um dos meios que será discutido para auxiliar na avaliação de fenômenos como esse é o desenvolvimento de trilhas de auditoria fundamentadas por questões associadas a vínculos entre profissionais de diferentes perfis e procedimentos médicos irregulares.

Além do ponto destacado anteriormente, faz-se necessário avaliar questões que podem estar relacionadas a benefícios pagos pelo governo. Um dos programas com mais notoriedade no Brasil é o Bolsa Família, que auxilia inúmeras famílias pelo território nacional. Apesar do papel social dessa iniciativa, já ocorreram situações adversas mediante vínculos anômalos associados a pagamentos do programa. O portal G1, no estado da Bahia, divulgou em 2016 um

¹ Disponível em: <https://globoplay.globo.com/v/12366776/>

escândalo envolvendo irregularidades na saúde associadas ao programa Bolsa Família (Mendes, 2016). A situação consistia em uma investigação associada a desvios de verba em municípios do estado e ao uso de “laranjas” para pagamentos de parcelas do Bolsa Família. Casos como esse também despertam um alerta e podem provocar análises mediante esses vínculos associados a programas sociais.

O primeiro passo para a construção de uma trilha do projeto é a identificação do problema, a qual escopo e realidade aquela situação pertence, e quais são as metas desejadas para serem atingidas. Nesta dissertação de mestrado, a esfera de Órteses, Próteses e Materiais Especiais (OPME) construída pela equipe do projeto FiscalizaSUS foi adotada como objeto de estudo. Assim, as seguintes trilhas serão discutidas e analisadas:

- procedimentos realizados com stents recorrentes;
- aquisição de material de empresa cujo sócio é beneficiário de programa de transferência de renda (Bolsa Família);
- aquisição de material de empresa de profissionais da saúde;
- OPMEs vendidas com valores anormalmente altos por procedimentos que envolvem o Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP);

Cada atividade detém suas devidas atribuições, tendo seus escopos distribuídos e definidos. Nesse processo, é preciso ressaltar que essas trilhas de auditoria apresentam relação direta entre si, devido a fatores como o uso de fontes de dados conjuntas, por exemplo: fornecedores de OPME, base de médicos que atendem no Sistema Único de Saúde, base de Beneficiários do Programa de Transferência de Renda do Governo Federal, Malha de cidades e estados, base de pessoas jurídicas e notas fiscais de OPMEs dos estados do Rio Grande do Norte e Ceará.

Nesse caso, as trilhas de auditoria relativas ao projeto FiscalizaSUS que se enquadram no campo de OPME, e que serão debatidas neste trabalho, podem ser identificadas no Quadro 1, a seguir. Nele, estão descritos indícios elaborados para a execução das atividades do projeto. Na seção 5, esses elementos serão apresentados como base para o desenvolvimento de métricas.

Quadro 1 – Trilhas de auditoria analisadas como objeto de estudo

Trilha de Auditoria	Descrição de indícios
Procedimentos realizados com stents recorrentes	Verificar o número de pacientes que passaram por procedimentos com stents e, em um curto período de tempo, foram submetidos a um novo procedimento que também requereu o uso de stents em menos de 30 dias.
Aquisição de material de empresa cujo sócio é beneficiário de programa de transferência de renda	Para identificação de um possível indício, o primeiro passo é verificar o vínculo entre uma pessoa jurídica e um beneficiário do programa de transferência de renda do governo federal. Além disso, os valores recebidos em transações e os valores oriundos do programa serão comparados entre si e filtrados para visualização de casos específicos.
Identificação de possíveis relações entre fornecedores de OPME e médicos funcionários públicos que atendem no SUS	Identificar a relação de vínculo entre um médico que atende no Sistema Único de Saúde e um fornecedor de OPMEs, por meio de variáveis-chave, como CNPJ.
OPMEs vendidas com valores anormalmente altos por Procedimento SIGTAP	Verificar a ocorrência de insumos vendidos com valores acima da média aplicados pelas unidades de saúde responsáveis.

Fonte: Diretrizes do projeto FiscalizaSUS (2025)

As trilhas dependem da integração entre diferentes conjuntos de dados, que apresentam perfis distintos. Por isso, também se faz necessária a existência de uma arquitetura de Big Data capaz de agregar todas essas informações de forma eficiente e otimizada para a consulta. Além disso, é preciso disponibilizar ferramentas capazes de suportar o trabalho de pesquisadores associados ao projeto FiscalizaSUS no tratamento e na interação com os dados.

1.1.5 Arquitetura FiscalizaSUS

Para lidar com as diversas fontes de dados adquiridas por meio de parcerias firmadas ao longo do planejamento e da implementação do projeto FiscalizaSUS, assim como pelo armazenamento de fontes de dados de origem pública, foi projetada uma arquitetura de Big

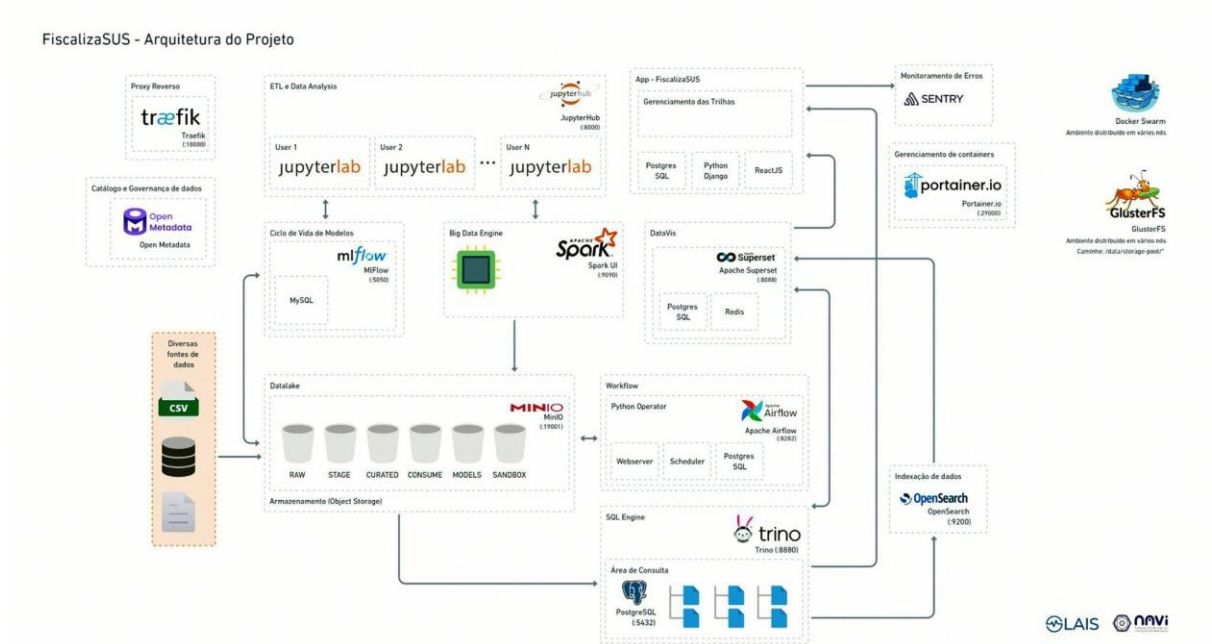
Data para dar suporte ao desenvolvimento das trilhas de auditoria solicitadas pelos auditores do DenaSUS.

As principais ferramentas contidas na arquitetura de Big Data do FiscalizaSUS e utilizadas para a parte prática desta dissertação podem ser listadas a seguir, com suas respectivas funções no âmbito do processamento de grandes volumes de dados:

1. MinIO: consiste em uma ferramenta de armazenamento para grandes quantidades de dados e pode ser agregada a diversos serviços externos (MinIO, 2021);
2. Jupyter Lab: ambiente de programação web voltado para os mais diversos contextos de desenvolvimento, como ciência de dados e aprendizado de máquina (Jupyterlab, 2025);
3. Apache Airflow: ferramenta destinada a organizar fluxos de rotina por meio de Gráficos Acíclicos Direcionados, conhecidos como DAGs, que atuam como mecanismos de execução de tarefas. Devido aos moldes de arquitetura do Airflow, essas tarefas podem ser conduzidas de forma distribuída (Benítez-Hidalgo *et al.*, 2021);
4. Apache Superset: ferramenta de código aberto utilizada para visualização de dados, integração com bancos de dados e criação de gráficos (Apache, 2024);
5. Opensearch: ferramenta de código aberto capaz de interagir com grandes volumes de dados e ser utilizada para consumo via API (Amazon, 2024);
6. Apache Spark: ferramenta que atua como um motor para execução de múltiplas tarefas relacionadas à manipulação de dados, podendo ser executada em um único nó ou em clusters (Apache, 2024);
7. pgAdmin 4: interface utilizada para trabalhar com banco de dados PostgreSQL e operações com dados;
8. Gitlab: sistema para armazenamento dos códigos relativos às DAGs do projeto FiscalizaSUS.

A Figura 1, a seguir, demonstra, de forma resumida o fluxo de comunicações entre todos os elementos que compõem a arquitetura de Big Data do projeto.

Figura 1 – Arquitetura de Big Data do projeto FiscalizaSUS



Fonte: produzida pelos pesquisadores do projeto FiscalizaSUS (2025)

Considerando a disponibilidade de uma arquitetura de referência capaz de integrar todas essas ferramentas, além de conter conjuntos de dados necessários para condução e elaboração das trilhas de auditoria, pode-se afirmar que essa funcionalidade é de fundamental importância para a garantia de eficiência no cumprimento dos objetivos do projeto FiscalizaSUS.

1.2 PROBLEMÁTICA

A problemática desta pesquisa está em analisar como o processamento e a agregação de dados – oriundos de diferentes bases que possuam relação com Órteses, Próteses e Materiais Especiais (OPME), em uma arquitetura de referência para Big Data organizada com diferentes ferramentas – podem impactar o processo de auditoria de insumos médico-hospitalares, como no trabalho de auditores do Sistema Único de Saúde do Brasil, tendo como objeto principal o fornecedor desses insumos. A partir desse raciocínio, é possível dividir os aspectos discutidos anteriormente em dois níveis de problemática, quais sejam: a corrupção em saúde e as informações fragmentadas. Com base nisso, será possível constituir as problemáticas centrais desta dissertação de mestrado.

1.2.1 Problemática em corrupção em saúde

Um dos pilares que justifica a problemática desta dissertação diz respeito às irregularidades, à incidência de corrupção no campo da saúde e aos prejuízos que isso acarreta para a população brasileira e o Sistema Único de Saúde (SUS). Por isso, é importante analisar aspectos da realidade do Brasil e do SUS como uma forma de compreender os malefícios que problemas associados a fraudes e corrupção em saúde podem desencadear.

O trabalho desenvolvido por Machoski e Araujo (2020) constroem uma avaliação a respeito de como a corrupção em saúde pública causa prejuízos ao desenvolvimento econômico de municípios brasileiros. Para isso, os autores adotaram abordagens estatísticas a fim de mensurar o impacto desse problema por meio de auditorias. O processo consistiu em duas etapas de trabalho: a primeira seria a análise de vínculos entre a auditoria e o crescimento econômico; e, a segunda, o impacto da corrupção no desenvolvimento da saúde pública do município auditado. Finalmente, os pesquisadores identificaram que as ocorrências de corrupção possuem consequências duradouras, podendo alastrar-se como prejuízo durante anos.

Assim, é possível exemplificar a problemática de corrupção em saúde e identificar como ela pode prejudicar o desenvolvimento do país. Isso se constitui uma motivação para a discussão de abordagens que possam contribuir para o combate a irregularidades e fraudes na saúde, o que também faz parte do protocolo de trabalho das trilhas de auditoria do projeto FiscalizaSUS mediante o contexto de Órteses, Próteses e Materiais Especiais (OPME). A próxima subseção apresentará outro aspecto pertinente para o projeto: a presença de informações fragmentadas.

1.2.2 Problemática em informações fragmentadas

Um dos tópicos de interesse desta dissertação de mestrado é abordar meios para abordar informações distribuídas em diferentes bases de dados. Para isso, é necessário desenvolver uma sistemática de integração entre as bases de dados que serão utilizadas por esta vertente do projeto FiscalizaSUS.

A pesquisa publicada por Bispo *et al.* (2020) recorreu a fontes de dados públicas a respeito de cirurgia bariátrica, obtidas através da plataforma do Departamento de Informação e Informática do SUS (DataSUS), e baseou-se na ontologia do tópico estudado para aplicar a integração desses dados. Os pesquisadores envolvidos aplicaram um fluxo de trabalho a fim de compreender o assunto tratado, selecionando as fontes de dados pertinentes e executando as

consultas. Assim, o objetivo foi construir, a partir das bases coletadas, uma base geral para realizar consultas.

Ademais, a lógica apresentada pelo artigo é pertinente para os processos de integração de diferentes bases de dados no contexto da saúde, e demonstra a importância desse tipo de problemática, além de oferecer um parâmetro comparativo para esta dissertação. A esse respeito, um dos principais pontos deste trabalho é partir para um processo de agregação entre diferentes conjuntos de dados, obtidos ao longo do projeto FiscalizaSUS. Em seguida, os objetivos da dissertação discutem a respeito do direcionamento desses elementos discutidos como problemática.

1.3 OBJETIVOS

A seguir, serão apresentados os objetivos desta pesquisa.

1.3.1 Objetivo Geral

O objetivo geral desta dissertação de mestrado é desenvolver a integração de diversas bases de dados, com diferentes perfis, para suportar o trabalho de auditores do Departamento Nacional de Auditoria do SUS (DenaSUS). O processo visa a qualificar os dados, bem como a dicionarizar as bases utilizadas e agregá-las a um modelo relacional estratificado.

1.3.2 Objetivos específicos

Para atingir o objetivo geral desta pesquisa será necessário realizar os seguintes objetivos específicos:

- Realizar um mapeamento sistemático da literatura com o intuito de buscar embasamento teórico sobre ferramentas de Big Data e Inteligência aplicadas à auditoria em saúde.
- Implementar processos de pré-processamento, análise exploratória, processamento e estruturação dos dados na plataforma FiscalizaSUS, focando em bases de dados relacionadas a OPMEs, programas de auxílio como o Bolsa Família, fornecedores de OPMEs e notas fiscais eletrônicas.

- Dicionarizar e desenvolver modelos relacionais de bancos de dados capazes de cruzar informações entre as bases de dados coletadas.
- Mapear e auxiliar no desenvolvimento de trilhas de auditoria associadas a inconsistências e anomalias relacionadas à temática de Órteses, Próteses e Materiais Especiais (OPME).

1.3.3 Hipótese da Pesquisa

O processamento e a estruturação de dados dirigidos pela arquitetura *Big Data* da Plataforma FiscalizaSUS possibilitam a criação de trilhas de auditoria para o Sistema Único de Saúde do Brasil (SUS) capaz de prover suporte aos auditores do Departamento Nacional de Auditoria do Sistema Único de Saúde (DenaSUS) do Ministério da Saúde, no contexto das Órteses, Próteses e Materiais Especiais (OPME).

1.3.4 Questões de Pesquisa

Q1: Como os métodos e técnicas de *Big Data* e os métodos computacionais podem ser empregados em aplicações de auditoria em saúde?

Q2: É possível integrar bases de diversas fontes de dados públicos para desenvolver trilhas de auditoria no SUS?

Q3: Como a utilização de trilhas de auditoria baseadas em dados e métodos computacionais pode auxiliar no processo de tomada de decisões dos auditores em relação a Órteses, Próteses e Materiais Especiais (OPME)?

Esta dissertação foi dividida em oito seções. Além desta seção introdutória, na segunda seção, apresentamos o referencial teórico. Na terceira seção, serão discutidos os trabalhos correlatos. Na quarta seção, será realizado o mapeamento sistemático da literatura. A quinta seção traz a metodologia. A sexta seção apresenta os resultados deste estudo. Na sétima seção, faremos uma discussão sobre os dados encontrados. Na oitava seção, traremos as considerações finais desta pesquisa.

2 REFERENCIAL TEÓRICO

O intuito desta seção é apresentar a fundamentação teórica para a construção das trilhas de auditoria desenvolvidas ao longo do projeto FiscalizaSUS, no âmbito da OPME.

2.1 DIMENSÕES DE PESQUISA

Nesta subseção, detalharemos as dimensões de alcance da pesquisa, discorrendo sobre: auditoria, OPME, Big Data, valores anômalos, entre outras.

2.1.1 Auditoria

A auditoria consiste em uma metodologia para avaliar a conformidade dos requisitos estabelecidos de modo a realizar a implementação de um sistema, de um cronograma de atividades ou conjunto de ações, considerando os requisitos norteadores estabelecidos previamente, de modo a verificar a integridade desses repertórios em dado cenário, podendo apresentar atores como: o auditor, profissional encarregado de desempenhar as tarefas de análise dos fatos, partes responsáveis e usuários (Brasil, 1988; Cordeiro, 2025; Rodrigues, 2013). Por sua vez, as análises são constituídas para disponibilizar transparência em relação a métodos, diretrizes, resultados e tomadas de decisões de fatores de qualidade e efetividade (Brasil, 1988).

É preciso também discutir a respeito dos tipos de controle que um processo de auditoria pode desempenhar, sendo estes de perfil externo e interno. Uma auditoria interna pode ser descrita como um processo desempenhado pela própria organização, visando a avaliar o controle interno e o comprometimento com as diretrizes definidas para execução de tarefas (Amorim, 2022). Já a auditoria externa remete a uma análise conduzida por agentes externos, independentes da instituição, empresa ou órgão sendo auditado. Esse fluxo de trabalho visa a disponibilizar resultados referentes aos procedimentos desempenhados, considerando as operações desempenhadas ou até mesmo o âmbito financeiro (Satka, 2017; Zhou *et al.*, 2023). Portanto, essas duas vertentes no campo de auditoria consistem em dois caminhos distintos para a apuração e o controle dos processos e resultados alcançados na esfera de uma instituição ou de um ente jurídico, proporcionando, assim, uma maior transparência e resiliência acerca das ações conduzidas no meio em que está sendo explorado.

O meio financeiro também destaca essa área do conhecimento como uma ferramenta propícia para assinalar o grau de confiança do ambiente auditado, com base no parecer do profissional responsável pela análise (Cordeiro, 2025). Por meio disso, pode-se estabelecer um panorama de possibilidades para avaliação acerca de gastos e investimentos, visando à execução de um objetivo com base em uma lógica de transparência e resiliência mediante a aplicação de recursos para um propósito final.

Além disso, esse ramo do conhecimento pode ser inserido na ótica clínica como um processo de trabalho cíclico. Ou seja, desenvolver uma lógica de trabalho baseada em etapas que buscam coletar e analisar dados, além de avaliar a qualidade dos possíveis impactos que as informações adquiridas podem desempenhar e seu real impacto para o cenário estudado. Partindo do processo de análise, as mudanças são aplicadas, podendo passar novamente por todo o processo até que se atinja o objetivo final proposto no planejamento (Limb *et al.*, 2017; Ojeda-Thies *et al.*, 2021). Busca-se, portanto, organizar uma atividade responsável por inferir tópicos de interesse para a melhoria na eficiência da oferta de serviços de saúde de modo que seja possível mensurar a relevância das informações obtidas.

Não obstante, as auditorias podem ser empregadas como uma ferramenta de análise qualitativa para um contexto desejado. O estudo apresentado por Ravn *et al.* (2022) demonstra a utilização de uma auditoria qualitativa, tomando como base prontuários médicos a respeito de reabilitação cardíaca de pacientes, com o intuito de investigar por que os pacientes não estavam aderindo como esperado ao tratamento. Tendo isso em vista, a análise dos dados obtidos levou à identificação de alguns padrões, como o desinteresse em continuar o tratamento cardíaco, administração de medicamentos por conta própria, sensação de não estar progredindo, entre outros fatores listados no estudo. Desse modo, pode-se observar que a análise qualitativa dos documentos médicos obtidos demonstra as adversidades presentes no ambiente de estudo, proporcionando, assim, uma melhor compreensão do cenário, podendo facilitar a elaboração de uma medida efetiva para solução do problema.

Outro modelo de auditoria que pode ser discutido é referente a auditorias operacionais. Essa abordagem constitui-se de processos comparativos entre documentos e normas que demonstrem uma padronização em relação a uma atividade (Aydogdu, 2025). O Tribunal de Contas da União atribui o conceito de auditoria operacional, no escopo brasileiro, a uma metodologia de controle externo a fim de averiguar os resultados das atividades sendo estas executadas em determinada repartição pública. O processo visa a elucidar métricas referentes à eficiência, eficácia e economia de recursos em atividades e programas referentes à administração pública. Desse modo, pode-se compreender esse método como um caminho para

aperfeiçoar e mensurar a aplicação de determinada política pública ou regra de negócio, baseando-se em diretrizes para orientação dos objetivos finais estabelecidos.

Além das metodologias de auditoria citadas anteriormente, pode-se acrescentar a esse conjunto as auditorias orientadas por conformidade, ou *compliance*. Esse método preza pelo cumprimento das normativas previamente estabelecidas conforme um planejamento elaborado anteriormente, ou seja, visa à implementação do processo de trabalho por meio de diretrizes (Aragão; Garbaccio, 2020; Rodrigues, 2013; Vedanabhatla; Gupta, 2013). Portanto, a auditoria realiza um conjunto de atividades basilares para a fundamentação teórica deste trabalho, considerando o escopo de atuação do projeto FiscalizaSUS, sobretudo para o campo de Órteses, Próteses e Materiais Especiais, que é o objeto central deste trabalho de dissertação.

A discussão a respeito dos tipos de auditoria e perfis marca um tópico fundamental para esta dissertação de mestrado, sobretudo do ponto de vista da seção 4, que discute a execução e os resultados de um Mapeamento Sistemático da Literatura (MSL) a respeito de auditoria em saúde intermediada por Big Data. Isso facilita o entendimento a respeito dos conceitos selecionados a partir da massa de artigos analisados para o mapeamento desenvolvido. Desse modo, faz-se necessário compreender como esse processo metodológico existe e é abordado no âmbito do Sistema Único de Saúde e quais órgãos competentes auxiliam na execução desse tipo de atividade.

2.1.2 Auditoria no SUS

Tendo em vista o conceito de auditoria, o Sistema Único de Saúde (SUS) conta com o Sistema Nacional de Auditoria do SUS (SNA), estabelecido pela Lei nº 8.689, de 27 de julho de 1993 (Brasil, 1993). O propósito do SNA é realizar avaliações técnicas, contábeis, financeiras e patrimoniais em relação ao SUS (Unasus, 2021).

Segundo o Manual de auditoria do Ministério da Saúde do Brasil (1998), algumas das finalidades de auditorias em saúde podem ser listadas como:

- 1) coleta de dados e identificação de padrões que permitam ao SNA identificar a qualidade, despesas, investimentos e quantidade de gastos na saúde;
- 2) avaliação objetiva dos elementos que compõem o objeto auditado, buscando a melhoria dos serviços;
- 3) avaliação qualitativa, quantitativa e efetiva dos serviços de saúde ofertados;
- 4) angariamento de dados que possibilitem a melhoria do Sistema Único de Saúde e atendimento adequado ao usuário.

Além disso, esse órgão conta atualmente com o Departamento Nacional de Auditoria do SUS (DenaSUS), encarregado de abordar diretamente a logística de auditoria interna no SUS. As atividades que competem ao DenaSUS giram em torno de metodologias voltadas para: realizar auditoria interna, políticas públicas e aplicação de recursos federais no escopo do SUS; propor melhorias nos processos de otimização e gerenciamento de riscos; coordenar a governança em políticas públicas; coordenar e orientar atividades de auditoria; sistematizar e padronizar os procedimentos de auditoria (Diário Oficial da União, 2023).

O DenaSUS tem um papel de relevância no combate a fraudes e desvios de verbas por meio de processos de auditoria em saúde no âmbito do Sistema Único de Saúde do Brasil. Nesse sentido, o objeto de discussão dessa dimensão compõe uma das bases de motivação e atuação para o projeto FiscalizaSUS. Trata-se de uma iniciativa colaborativa entre diversas instituições parceiras para auxiliar auditores do Sistema Único de Saúde (SUS), buscando compreender as necessidades dos auditores mediante o contexto de aplicação de grandes quantidades de dados relativos aos processos de auditoria, de modo a identificar como o uso de técnicas de processamento, armazenamento e análise pode otimizar o fluxo de informações que esses profissionais obtêm em seu cotidiano.

No ano de 2015, foi deflagrada uma operação de combate à corrupção contra irregularidades envolvendo Órteses, Próteses e Materiais Especiais (OPME) pelo Ministério Público Federal do estado de Minas Gerais. O esquema conhecido como Máfia das Próteses oferecia propinas para médicos mineiros aumentarem a venda de seus produtos, e, conseqüentemente, obterem mais lucros sem se responsabilizar pelo estado de saúde real dos pacientes, vítimas desses procedimentos. As ações irregulares envolviam médicos e clínicas, gerando um mercado paralelo que movimentava cerca de R\$ 20 bilhões por ano no país (Máfia [...], 2024).

Nesse processo, é possível perceber que as auditorias constituem um mecanismo de avaliação propositiva. O objetivo é averiguar o andamento de atividades, identificar possíveis irregularidades e anomalias, assim como promover suporte a órgãos de controle, visando a uma melhor transparência e eficiência dos processos, sobretudo na saúde pública. O projeto FiscalizaSUS envolve discussões e propostas para o enfrentamento a essas ocorrências no Sistema Único de Saúde (SUS), em parceria com o DenaSUS e outras instituições parceiras do projeto.

Por meio de situações como as destacadas anteriormente, fica claro como o SUS no Brasil se encaixa na temática de auditoria em saúde, e quais elementos são responsáveis por alertar a ocorrência de possíveis irregularidades e atividades suspeitas no âmbito de OPME. Por

isso, é importante que os processos de auditoria na esfera do SUS investiguem comportamentos inadequados a respeito desses insumos. A discussão desta subseção serve como motivação para as temáticas exploradas do ponto de vista prático do trabalho, a fim de provocar a discussão de métricas que podem ser adotadas para auxiliar os profissionais da área de auditoria no SUS.

2.1.3 Órteses, Próteses e Materiais Especiais (OPME)

Para o entendimento a respeito de objeto de pesquisa abordado neste trabalho, é importante compreender e conceituar do que se trata Órteses, Próteses e Materiais Especiais (OPME). Em primeiro lugar, OPME pode ser compreendido como um conjunto de ferramentas voltado para fins médicos e odontológicos com a finalidade de auxiliar pacientes em processos de diagnóstico ou reabilitação (Brasil, 2016; Rufino; Reis; Ribeiro, 2024). Ou seja, por ser um elemento presente no meio médico, é importante analisar como esses artifícios podem ser avaliados mediante um processo de auditoria em saúde.

Também é importante discutir a respeito dos gastos que a esfera federal tem com esse tipo de produto. No período de 2015 a 2019, houve um gasto de mais de R\$ 6 milhões em compras desse insumo, ou seja, uma taxa de aproximadamente R\$ 1.3 milhão por ano (Cruz *et al.*, 2024). Portanto, este setor movimenta valores substanciais em dinheiro, e acaba demonstrando seu impacto econômico mediante a esfera da saúde pública.

Um ponto importante a esse respeito trata da heterogeneidade que essa área possui no Brasil. O fornecimento desse tipo de produto é conduzido, em sua maioria, por empresas multinacionais estrangeiras, e, em menor escala, por empresas brasileiras de médio a pequeno porte, gerando uma concentração de atividades em grupos minoritários, dificultando a padronização em relação à classificação das OPMEs (Cruz *et al.*, 2024). Esse ponto em particular pode ser trabalhado de forma minuciosa a fim de compreender possíveis inconformidades presentes em processos de aquisição desses materiais.

Além disso, como citado na subseção anterior, essa área está à mercê de enfrentar problemas ligados à corrupção de profissionais da saúde e eventuais prejuízos à oferta de serviços de saúde à população. Assim, no desenvolvimento das trilhas ligadas a Órteses, Próteses e Materiais Especiais (OPME) do projeto FiscalizaSUS foram adotadas bases de dados referentes a fornecedores desse tipo de insumo e notas fiscais eletrônicas referentes a transações envolvendo esses produtos, com o objetivo de averiguar possíveis irregularidades mediante esse contexto. Diante disso, é possível refletir a respeito da necessidade de fomentar atividades de

auditoria em saúde, buscando uma melhor transparência de dados e ações desempenhadas pelos profissionais de saúde.

2.1.4 Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)

Como discutido anteriormente, as Órteses, Próteses e Materiais Especiais (OPME) constituem o elemento central de pesquisa deste trabalho, também compõem diversas trilhas de auditoria inseridas no projeto FiscalizaSUS. Nesse sentido, faz-se necessário discutir a respeito da padronização em relação às ações desempenhadas pelo SUS.

Para isso, existe o Sistema de Gerenciamento da tabela de procedimentos, medicamentos e OPM do SUS (SIGTAP). Esse sistema diz respeito a um conjunto de dados organizados e ordenados de forma tabular no qual são inseridas informações pertinentes a respeito dos serviços de saúde prestados pelo Sistema Único de Saúde (Miranda *et al.*, 2022; Santos *et al.*, 2023). Ou seja, consiste em uma padronização adotada para averiguar a confiabilidade e mapear o desenvolvimento de ações desenvolvidas pelo SUS. Além disso, é pertinente discutir acerca da formatação que um registro nesse sistema apresenta, e quais classes estão dispostas em sua numeração. A Figura 2, a seguir, demonstra a configuração dessa classe de documentos.

Figura 2 – Exemplo de procedimento SIGTAP registrado

■ Procedimento

Procedimento: 07.02.04.028-2 - ENDOPROTESE AORTICA BIFURCADA																					
Grupo:		07 - Órteses, próteses e materiais especiais																			
Sub-Grupo:		02 - Órteses, próteses e materiais especiais relacionados ao ato cirúrgico																			
Forma de Organização:		04 - OPM em assistência cardiovascular																			
Competência:		04/2025		 Histórico de alterações																	
Modalidade de Atendimento: Hospitalar Complexidade: Não se Aplica Financiamento: Média e Alta Complexidade (MAC) Sub-Tipo de Financiamento: Instrumento de Registro: AIH (Proc. Especial) Sexo: Ambos Média de Permanência: Tempo de Permanência: Quantidade Máxima: 1 Idade Mínima: Não se aplica Idade Máxima: Não se aplica Pontos: Atributos Complementares:																					
Valores <table border="1"> <tr> <td>Serviço Ambulatorial:</td> <td>R\$ 0,00</td> <td>Serviço Hospitalar:</td> <td>R\$ 17.218,60</td> </tr> <tr> <td>Total Ambulatorial:</td> <td>R\$ 0,00</td> <td>Serviço Profissional:</td> <td>R\$ 0,00</td> </tr> <tr> <td></td> <td></td> <td>Total Hospitalar:</td> <td>R\$ 17.218,60</td> </tr> </table>										Serviço Ambulatorial:	R\$ 0,00	Serviço Hospitalar:	R\$ 17.218,60	Total Ambulatorial:	R\$ 0,00	Serviço Profissional:	R\$ 0,00			Total Hospitalar:	R\$ 17.218,60
Serviço Ambulatorial:	R\$ 0,00	Serviço Hospitalar:	R\$ 17.218,60																		
Total Ambulatorial:	R\$ 0,00	Serviço Profissional:	R\$ 0,00																		
		Total Hospitalar:	R\$ 17.218,60																		
Descrição	CID	CBO	Leito	Serviço Classificação	Habilitação	Redes	Origem	Regra Condicionada	Renases	TUSS											

Fonte: <http://sigtap.datasus.gov.br/tabela-unificada/app/sec/procedimento/exibir/0702040282/04/2025>. Acesso em: 15 abr. 2025

Com base na Figura 2, pode-se observar que o procedimento descrito possui um número e um identificador, que é composto por nomenclaturas como: grupo, subgrupo e forma de organização. Além disso, ele contém outras informações de caráter descritivo, como modalidade do atendimento, financiamento, entre outros elementos em destaque. A lógica empregada por trás dos procedimentos SIGTAP foi de fundamental valor para a construção das trilhas de auditoria em OPMEs relativas a esse tópico, pois proporciona um melhor entendimento.

2.1.5 Big Data

O conceito de Big Data compõe um dos principais pilares para a fundamentação teórica deste trabalho, uma vez que as trilhas de auditoria estão diretamente relacionadas a essa lógica. Esse conceito foi introduzido inicialmente por Laney *et al.* (2001), baseando-se em três heurísticas principais: volume, velocidade e variedade. No primeiro ponto, o autor discute acerca do conceito de volume de dados, exemplificando, por meio da logística firmada em *e-commerces*, o armazenamento de uma maior quantidade de dados referentes a operações transacionais, destacando como as empresas assimilam informações a respeito de seus ativos. Seguindo essa lógica, compreende-se que o volume de Big Data está atrelado a uma quantidade de informações considerada elevada para os mais variados nichos de atuação, impactando as possibilidades de interação com esses elementos.

No segundo ponto, relativo à velocidade, é preciso refletir a respeito de métodos capazes de facilitar a interação com esses materiais. Algumas dessas abordagens podem girar em torno da reorganização periódica dos dados, buscando um controle interno apurado e a facilidade em adquirir informações de registros já obtidos. Somado a esses componentes está a construção de uma arquitetura de Big Data compatível com o volume alvo de operações (Laney *et al.*, 2001). Nesse processo, a velocidade de interação com grandes quantidades de dados consiste em uma sistemática que varia conforme a metodologia aplicada para utilização dos conjuntos desejados.

O terceiro ponto de interesse para a fundamentação do Big Data é o fator de variedade. Os dados podem possuir os mais distintos formatos e estarem disponíveis de maneira não estruturada, dificultando o processo de entendimento e sua utilização para um propósito final (Laney *et al.*, 2001). A esse respeito, algumas estratégias foram sugeridas por Laney *et al.* (2001) com o objetivo de tratar dessas informações. Inicialmente, foi proposta a descoberta por padrões que permitissem identificar a relação entre os dados, a adoção de um formato universal para trabalhar com esses componentes, técnicas de indexação de dados e gerenciamento de metadados. Com o passar do tempo e as inovações tecnológicas, novos eixos foram propostos para o conceito de Big Data, como veracidade e valor.

A veracidade consiste em avaliar a qualidade dos dados obtidos mediante um cenário de Big Data. Desse modo, é preciso definir quais elementos presentes em uma amostra serão de fato selecionados para a análise final, tendo em vista que dados mal formatados podem não ser úteis para a construção de um produto final, sendo necessário prezar pela condição em que os dados se encontram (Ishwarappa; Anuradha, 2015).

Quanto ao valor, pode-se afirmar que ele é tudo o que pode ser extraído como útil em um contexto Big Data para atender a determinado cenário. Sem apreciação do valor, todo esse volume de dados não terá serventia (Ishwarappa; Anuradha, 2015).

Ademais, todos os conceitos de Big Data comentados anteriormente compõem uma base para a compreensão do aspecto técnico deste trabalho e a tomada de decisões ao longo do processo de manipulação de dados referentes às trilhas de auditoria envolvendo Órteses, Próteses e Materiais Especiais (OPME). A relevância de cada conceito descrito nesta seção pode ser identificada no Quadro 2, a seguir.

Quadro 2 – Aplicação dos conceitos de Big Data - Fonte: Elaborada pelo autor, 2025

Conceito	Aplicação
Volume	Quantidade de dados utilizada para integração e aplicação nas trilhas de auditoria do projeto FiscalizaSUS
Veracidade	Atestado de resiliência que uma base de dados pode demonstrar, verificando como ela pode se relacionar com os objetivos do projeto
Variedade	A variedade de dados está explicitamente relacionada às diversas fontes utilizadas para captação de informações pertinentes ao projeto, e, sobretudo, todo o escopo desta dissertação.
Velocidade	Para velocidade, é possível relacionar com a arquitetura de Big Data disponibilizada pelo projeto FiscalizaSUS, tendo em vista a capacidade de interação entre múltiplas ferramentas e trabalho com os dados adquiridos.

Fonte: produzido pelo autor, 2025

2.1.6 Arquitetura de referência para Big Data

Desde o século passado, a quantidade de dados em saúde vem aumentando cada vez mais, desafiando a capacidade desse setor de armazenar e trabalhar com esses vastos volumes de informação (Agarwal *et al.*, 2025). Mediante essa problemática, o uso de tecnologias associado à Big Data pode desempenhar um papel-chave para a abordagem de cenários que envolvam grandes volumes de dados na área da saúde. Tendo em vista esse cenário, é

importante refletir a respeito de como os dados podem ser armazenados e utilizados simultaneamente para o desenvolvimento de demandas nos mais diversos cenários e para diferentes propósitos. Portanto, é necessário discutir a respeito do conceito de arquitetura de referência para a Big Data.

Nessa perspectiva, o trabalho de Ataei e Litchfield (2022) desenvolve uma revisão do estado da arte em relação a como as arquiteturas de referência para Big Data têm sido desenvolvidas e planejadas ao longo do tempo. Segundo os resultados apontados pelos autores, um percentual significativo de arquiteturas não é projetado de forma distribuída, o que acaba não sendo vantajoso para operações que envolvam a integração de diferentes grupos de dados para um mesmo propósito.

No mesmo trabalho, foram identificados alguns padrões interessantes em relação ao ciclo de vida dessas arquiteturas de referência. Os processos basicamente consistem em uma logística de Extração, Transformação e Carregamento (ETL), que acaba sendo relevante para cenários em que dados não estruturados e de fontes variadas estejam empregados para consolidação da arquitetura (Ataei; Litchfield, 2022; Fan; Lu, 2024).

Em uma revisão sistemática da literatura levantada por Ataei e Litchfield (2020), algumas questões de pesquisa foram feitas com o intuito de elucidar o conceito de Arquitetura de Referência. A esse respeito, os autores pontuaram alguns tópicos principais, quais sejam:

- 1) são ambientes que possuem alto grau de abstração;
- 2) descrevem aspectos qualitativos;
- 3) fomentam a padronização e a formatação de outras arquiteturas, visando a facilitar o entendimento e a implementação de novas versões.

Tendo em vista a descrição dessa linha de raciocínio a respeito do conceito de arquitetura de referência, essa metodologia também foi empregada no projeto FiscalizaSUS com o objetivo de reunir dados de diferentes origens e criar mecanismos de interação entre diferentes tabelas, fazendo com que haja sinergia entre esses conjuntos nos processos de trilhas de auditoria que englobam múltiplas fontes de dados dos mais diversos formatos e origens.

Assim, faz-se basilar discutir a respeito de arquitetura de referência, uma vez que esse raciocínio foi necessário para a construção da arquitetura responsável por comportar as ferramentas utilizadas ao longo do projeto, assim como pela integração das diferentes bases de dados, de origem tanto pública quanto privada. Logo, foi um fator fundamental para a integração entre diversos conjuntos de dados, de diferentes origens, em um único ambiente integrado. Assim, por meio da discussão desse conceito, é possível partir para os próximos

aspectos técnicos inseridos nas trilhas de auditoria com foco em Órteses, Próteses e Materiais Especiais.

2.1.7 Detecção de valores anômalos

A detecção de valores anômalos em amostras de dados é um dos principais fatores presentes na construção metodológica deste trabalho, sobretudo para a análise de irregularidades em relação a valores monetários que envolvem Órteses, Próteses e Materiais Especiais (OPME). Diante disso, dados que geralmente eram reconhecidos como fora do padrão, apresentando comportamento anômalo em comparação com o restante da amostra, podem representar valor em questões relacionadas à veracidade dos dados coletados, e servirem como base para processos que envolvam a detecção de irregularidades (Singh; Upadhyaya, 2012). Com isso, os valores que excedem seu comportamento em relação à amostra absoluta de dados podem ser avaliados mediante sua relevância no cenário.

A respeito das trilhas de auditoria de Órteses, Próteses e Materiais Especiais (OPME) do projeto FiscalizaSUS, o objetivo é identificar a presença de comportamentos indevidos em relação a operações que envolvem o uso desses insumos. Mediante a necessidade desse tipo de avaliação, foi empregado o método tradicional de Tukey, que consiste em uma abordagem gráfica para visualização de componentes estatísticos presentes nos conjuntos de dados estudados, como os quartis e a mediana. Esse método comporta-se com menos sensibilidade em relação a valores extremos, uma vez que adota como pontos focais os quartis da amostra em análise (SEO, 2006). Logo, essa técnica permite uma análise voltada para valores médios que apresentam comportamento anômalo em relação às médias amostrais.

Na literatura, foi possível identificar algumas publicações que indicam o uso do método de Tukey para abordar a presença de valores discrepantes em relação à amostra empregada. O trabalho, desenvolvido por Huyghues-Beaufond *et al.* (2020), propõe um *framework* para detectar *outliers* em grandes conjuntos de dados, por meio da agregação de métodos estatísticos como a técnica de Tukey. Os autores descreveram esse processo, por meio da análise dos 1º e 3º quartis, a fim de estimar o alcance referente à amostra central de dados e identificar os valores descritos além dos limites impostos pelo método. Os autores também destacam a possibilidade de essa analogia ser utilizada para destacar falsos negativos e falsos positivos. Nesse sentido, é importante ilustrar como o método pode ser descrito em relação à detecção dos *outliers* identificados por meio das seguintes equações de modo a detectar os

limites do 1º (Q1) e 3º (Q3) quartis, e Intervalo Interquartilico (IQR) (Dutoit; Steyn; Stumpf, 2012; Tukey *et al.*, 1977; Zijlstra; Van-Der Ark; Sijtsma, 2007):

Limite Inferior:

$$Q1 - 1,5 \times IQR$$

Limite Superior:

$$Q3 + 1,5 \times IQR$$

Intervalo Interquartilico:

$$IQR = Q3 - Q1$$

Com essas equações, é possível compreender que os valores situados fora dos limites superior e inferior são classificados como *outliers*, indicando um valor discrepante na amostra (Tukey *et al.*, 1977). Além disso, a publicação de Costelloe *et al.* (2015) traz uma aplicação de valores discrepantes para escassez de medicamentos em um ambiente de farmácia de comunidade. Assim como a publicação de Huyghues-Beaufond *et al.* (2020), os autores empregaram o método de Tukey como uma alternativa para abordar a temática discutida em seu trabalho. Assim, foi feita uma comparação de médias em relação aos dados utilizados. Com isso, o estudo apresenta uma aplicação desse artifício estatístico para o campo da saúde.

Esses conceitos abordados pelos artigos, podem ser aproveitados para a construção de trilhas de auditoria referentes a Órteses, Próteses e Materiais Especiais (OPME) no âmbito do projeto FiscalizaSUS, uma vez que os principais objetivos das trilhas remetem à identificação de valores anômalos e a irregularidades identificadas nos diversos conjuntos de dados referentes à OPME. O entendimento estatístico a respeito da detecção de valores classificados como *outliers* foi um dos elementos relevantes para a identificação de anomalias nas trilhas de OPME, sobretudo quando se trata de procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP).

Nesta seção, apresentamos o referencial teórico deste estudo. Na próxima seção, faremos a exposição dos trabalhos relacionados com o objeto de estudo desta pesquisa.

3 TRABALHOS CORRELATOS

A auditoria em saúde compõe um elemento imprescindível para a condução deste trabalho. Desse modo, foi feita uma busca por trabalhos que tratassem dessa temática. Uma parte dos trabalhos possui foco de pesquisa na atividade de auditoria em saúde em si, enquanto outros desenvolvem ferramentas mediadas por métodos computacionais para explorar esse cenário. Assim, serão discutidas quatro publicações com metodologias singulares e diferentes abordagens para o tema de interesse. Por fim, será estabelecida uma comparação entre os trabalhos discutidos e a pesquisa desta dissertação de mestrado.

Em primeiro lugar, o estudo desenvolvido por Hillan *et al.* (2024) apresenta um exemplo de auditoria referente à vacinação na Austrália. Os autores adotaram uma metodologia responsável por auxiliar os gestores de saúde em pontos críticos, como manuseio e armazenamento de insumos médicos, que norteia a eficácia na aplicação de vacinas em seu campo de estudo. Para atingir esse objetivo, os pesquisadores seguiram métodos, como a implementação de um questionário eletrônico com o intuito de avaliar as melhores práticas para o armazenamento de vacinas e estabelecer uma dinâmica de acompanhamento com os gestores de saúde.

Essas medidas resultaram em um caminho capaz de auxiliar as unidades responsáveis pelo armazenamento das vacinas a compreender mais como avaliar a qualidade da gestão de cadeia de frio e alertar sobre a importância do processo de auditoria como um elemento capaz de ressaltar tópicos pertinentes em seu respectivo cenário. Com isso, essa pesquisa destaca o uso de uma iniciativa metodológica envolvendo processos de auditoria em saúde. Apesar de Hillan *et al.* (2024) terem alcançado seu objetivo mediante o processo de auditoria em saúde, o estudo não abordou o uso de ferramentas de processamento de Big Data para seu contexto de uso, que é um dos focos principais para o desenvolvimento das trilhas de auditoria vinculadas a Órteses, Próteses e Materiais Especiais.

Em contrapartida, o estudo apresentado por Joudaki *et al.* (2015) traz uma abordagem de aplicação de processamento inteligente da informação mediante um escopo de auditoria em saúde, inclusive retomando conceitos explorados na fundamentação teórica do trabalho, iniciando o processo pela identificação da natureza do problema e os objetivos definidos nos níveis de planejamento. Desse modo, o foco do trabalho está em otimizar a detecção de fraudes e irregularidades em reivindicações de médicos. Seguindo essa linha de raciocínio, os autores da pesquisa desenvolveram uma sistemática de mineração de dados em uma base

disponibilizada por uma empresa privada focada no ramo de seguros de saúde, aplicando conceitos de auditoria para abranger uma análise concisa da problemática.

Nesse processo, os dados foram devidamente organizados e pré-processados para desenvolver uma avaliação coesa para as próximas etapas de inferências. Assim, indicadores também foram selecionados, como a taxa de pacientes que foram consultados mais de uma vez no mesmo mês, o custo médio de medicamentos, a taxa de antibióticos prescritos por médicos, entre outros indicadores que mostravam as ações desses profissionais de acordo com sua atuação profissional (Joudaki *et al.*, 2015).

Ademais, os indicadores foram empregados como uma forma de separar os dados em diferentes grupos para facilitar o processo de identificação de anomalias. Nesse processo, foram aplicadas análise de clusters e análise de discriminantes, baseadas em combinações lineares das variáveis de interesse adotadas para avaliar a eficácia dos indicadores ao longo do processo. Tendo isso em vista, esse trabalho consiste em uma análise de auditoria em saúde intermediada pela aplicação de técnicas estatísticas em um contexto de grandes quantidades de dados, uma vez foi utilizada uma superfície de mais de 400 mil registros de reclamações médicas para análise (Joudaki *et al.*, 2015).

Outro trabalho que desenvolve uma temática relacionada à identificação de fraudes no contexto médico foi o estudo publicado por Herland, Khoshgoftaar e Bauder (2018), que toma como escopo de análise o Medicare, programa estabelecido pelo governo estadunidense, como uma forma de prestar suporte a populações de risco que precisem de atendimento. Nesse contexto, ainda é possível destacar a menção de insumos médicos, como órteses e próteses presentes nesse âmbito de atividades fraudulentas, e os prejuízos que acabam reduzindo a disponibilidade de investimentos para esse tipo de programa assistencial.

Partindo dessa problemática, os pesquisadores utilizaram dados relacionados a pagamentos e cobranças relacionados ao *Medicare*. O estudo ainda utilizou três algoritmos principais para análise dos dados, quais sejam: Regressão logística, *Random Forest* e *Gradient True Boosting*. Por ser um contexto de Big Data, foram utilizados o Hadoop e o Apache Spark para auxiliarem na organização dos dados. Das técnicas listadas, o primeiro apresentou o melhor desempenho em comparação com os demais, tendo em vista o conjunto de dados completo, apresentando um score de 0.816 (Herland; Khoshgoftaar; Bauder, 2018).

As principais contribuições do trabalho estão inseridas no escopo de uso de ferramentas de Big Data para o processamento de grandes quantidades de dados em saúde, além de propor esse modo de identificação de irregularidades como uma alternativa a auditorias manuais. Apesar de esse trabalho aplicar o processamento de Big Data para identificação de

fraudes no contexto da saúde, ele não discute a respeito de uma arquitetura de Big Data integrada, e de trilhas de auditorias que estejam agregadas a esse tipo de ação (Herland; Khoshgoftaar; Bauder, 2018).

Já o trabalho desenvolvido por Gomes *et al.* (2025) trata de uma problemática envolvendo um dos objetos centrais desta pesquisa, que são as OPMEs. Inicialmente, os autores do trabalho introduzem o cenário relativo a nomenclaturas de insumos não padronizados, dificultando a identificação desses materiais em processos de auditoria. O intuito do estudo foi desenvolver uma metodologia para classificação de OPMEs, permitindo identificar qual tipo de material estava sendo utilizado em determinado procedimento médico. O treinamento foi realizado por meio de algoritmos de aprendizado de máquina, buscando identificar, por meio de descrições textuais compostas por diversas bases de dados, qual tipo de OPME foi aplicado em determinado caso de uso.

Nesse mesmo trabalho, a ferramenta OPMInEr, que faz parte da arquitetura de Big Data do FiscalizaSUS, foi um artifício fundamental para a execução dos experimentos de classificação dos materiais presentes, tendo em vista o armazenamento de bases de dados adquiridas, oriundas da Agência Nacional de Vigilância Sanitária (Anvisa), tendo como foco notas fiscais do estado do Rio Grande do Norte e o Cadastro Oficial de Materiais do Governo Federal (CATMAT). Desse modo, também as técnicas de Inteligência Artificial foram aplicadas, sendo a *Linear Support Vector* a que apresentou melhor desempenho em comparação com outras listadas no trabalho, como *Random Forest*, *Decision Tree*, *Gradient Boost*, entre outras técnicas de classificação aplicadas ao longo da pesquisa (Gomes *et al.*, 2025).

Assim, o artigo “*A smart classifier of orthoses, prostheses and special materials (OPMEs) in invoices*” conclui que é possível automatizar a classificação e a identificação de OPMEs em notas fiscais. Esse estudo promove, com isso, a possibilidade de identificação de valores irregulares e de quantidades anômalas de OPMEs sendo compradas por parte das autoridades de auditoria competentes (Gomes *et al.*, 2025).

Os trabalhos discutidos nesta seção trazem contribuições para o estado da arte desta pesquisa em diferentes competências. Primeiro, descrevem como um processo de auditoria pode impactar na qualidade de armazenamento de insumos médicos; segundo, expondo exemplos de auditorias automatizadas por meio do uso de tecnologias de *Big Data* em cenários diferentes do da saúde. Apesar disso, fazem uso de técnicas e abordagem de conceitos que podem ser pertinentes também para a pesquisa desenvolvida neste trabalho de dissertação. Por fim, o trabalho apresentando por Gomes *et al.* (2025) expõe um cenário mais convergente com o objeto de pesquisa presente nesta dissertação ao tratar de Órteses, Próteses e Materiais

Especiais (OPME) no contexto de auditoria em saúde, utilizando, para isso, processamento de grandes volumes de dados para o treinamento do algoritmo de classificação. O Quadro 3, a seguir, apresenta as contribuições e lacunas demonstradas pelos trabalhos citados nesta seção.

Quadro 3 – Trabalho Correlatos

Autores	Fez uso de Big Data para auditoria?	Foco em integrar diversos conjuntos de dados, de diferentes perfis para interação com modelo relacional?	Usou técnicas de <i>Machine Learning</i>?	Tem foco temático em saúde?
(Hillan <i>et al.</i> , 2024)	Não	Não	Não	Sim
(Joudaki <i>et al.</i> , 2015)	Sim	Não	Sim	Sim
(Herland <i>et al.</i> , 2018)	Sim	Não	Sim	Sim
(Gomes <i>et al.</i> , 2025)	Sim	Não	Sim	Sim

Fonte: produzido pelo autor, 2025

Por meio dessa síntese do conteúdo dos trabalhos correlatos, é possível traçar um comparativo a respeito da singularidade presente nesta dissertação. É importante destacar que o processo de integração, estruturação e qualificação de dados são basilares para as trilhas de auditoria do projeto FiscalizaSUS. Nenhum dos artigos abordados, por mais que recorram a mais de uma base de dados, demonstra foco nesses processos de agregação de dados. As soluções discutidas, em sua maioria, apontam para o uso de técnicas computacionais como suporte à área de auditoria. Portanto, o diferencial deste trabalho de mestrado é estar focado na agregação de múltiplas bases de dados, com os mais diversos perfis e utilidades, para viabilizar e facilitar a implementação de trilhas de auditoria focadas em OPME.

Nesta seção, expomos os trabalhos que estabelecem relação com esta pesquisa e trouxeram contribuições para o estudo. Na próxima seção, serão descritas as estratégias para mapeamento da literatura sobre o tema em tela.

4 MAPEAMENTO SISTEMÁTICO DA LITERATURA

Nesta seção, abordaremos a temática de auditoria em saúde e as estratégias de mapeamento da literatura sobre o tema. Este processo busca coletar artigos que tratem do tema de auditoria em saúde, com a mediação de métodos inteligentes. Ao longo da seção, serão apresentadas as estratégias utilizadas para extração e filtragem de material, assim como os resultados adquiridos. Sendo assim, também será discutida uma síntese narrativa a respeito do que foi feito e as ameaças à validade encontradas.

4.1 OBJETIVOS

O objetivo é analisar artigos relacionados à temática de auditoria em saúde e verificar como o tema é tratado ao redor do mundo. Para isso, o primeiro passo foi adotar o Mapeamento Sistemático da Literatura (MSL). A estratégia foi selecionar artigos publicados em bases de dados pertinentes para adquirir um paradigma referente ao estado da arte em estudos que tratam de abordagens semelhantes ao foco desta dissertação.

Um dos pontos para este mapeamento foi a adoção do protocolo PICO (População, Intervenção, Controle e Resultado) como um caminho para a construção de questões de pesquisa (Riera-Mestre *et al.*, 2024; Roever *et al.*, 2021). De acordo com ROEVER *et al.* (2021), o protocolo possibilita um processo transparente em relação à seleção de artigos que estejam relacionados ao assunto focal da pesquisa e ao desenvolvimento de trabalhos futuros. Assim, o Quadro 4 contempla a divisão do protocolo PICO adotado com base em temáticas pertinentes ao tópico de auditoria em saúde no escopo do projeto FiscalizaSUS.

Quadro 4 – Protocolo PICO

Acrônimo	Categoria	Descrição
P	População	Publicações que abordam aplicações de auditoria em saúde mediadas por métodos inteligentes e lidam com grandes volumes de dados para identificação de anomalias e/ou inconformidades.
I	Intervenção	Contextos de uso de inteligência automatizada de negócios para análise de grandes quantidades de dados.
C	Controle	Identificação de inconformidades/anomalias.
O	Resultados	Eficácia, efetividade, eficiência, indícios de fraude, irregularidades, não conformidades, estatísticas descritivas, classificação inteligente, relatórios gerenciais, descoberta de dados.

Fonte: produzido pelo autor, 2025

Tendo em vista o conteúdo distribuído no Quadro 4, foram formuladas as seguintes questões de pesquisa:

- Q1: quais os principais algoritmos e métodos estatísticos utilizados?
- Q2: quais os tipos de dados ou qual a área da saúde dos estudos analisados?
- Q3: quais os principais desafios e limitações identificados?
- Q4: quais as contribuições desenvolvidas pelas pesquisas?
- Q5: em quais anos foram publicados mais artigos?
- Q6: quais os meios de publicação mais populares?
- Q7: quais países têm mais publicações nessa área?

Pode-se compreender o item População como publicações que abordam aplicações de auditoria em saúde mediadas por métodos inteligentes e lidam com grandes volumes de dados para identificação de anomalias e/ou inconformidades. Já Intervenção seria o conjunto de abordagens identificadas para abordar a problemática de auditoria em saúde. O Controle traria a identificação de inconformidades/anomalias. Os Resultados se configurariam como uma lista de parâmetros relativos à eficácia, efetividade, eficiência, indícios de fraude, irregularidades, não conformidades, estatísticas descritivas, classificação inteligente, relatórios gerenciais, descoberta de dados. As questões de pesquisa propostas podem se relacionar com o protocolo PICO por meio das seguintes diretrizes:

- Q1 e Q2: buscam identificar o uso de técnicas computacionais e em quais áreas da saúde o estudo está inserido – População (P) e Intervenção (I).

- Q3 e Q4: baseiam-se nos resultados adquiridos e nas limitações identificadas ao longo da intervenção apresentada na publicação – Intervenção (I), Controle (C) e Resultados (O).
- Q5, Q6 e Q7: remetem a quais países mais contribuíram para a área de estudo, além de verificar períodos de publicação e as bases mais populares para divulgação – População (P), Intervenção (I), Controle (C), Resultados (O).

4.2 ESTRATÉGIAS DE BUSCA E SELEÇÃO

Para executar a busca em periódicos que pudessem contribuir para esta dissertação, foram adotadas as principais bases de busca para o ramo da computação. Essas fontes incluem SCOPUS, ACM, IEEE Xplore e Web of Science. O processo de busca foi feito por meio de ferramentas que recebem padrões de texto em um formato próprio para cada uma das bases. Essa fase do mapeamento também partiu do proposto pelo protocolo PICO adotado, dando origem, em primeiro lugar, a uma *string* de busca genérica adaptada para cada uma das bases utilizadas, haja vista que existem peculiaridades nos mecanismos de buscas, que serão apresentadas a seguir:

- String Genérica: (“health*” AND “audit*”) AND (“Big Data” OR “Business Intelligence” OR “Data Science” OR “AI” OR “Mining” OR “Artificial Intelligence”) AND (“Efficacy” OR “Effectiveness” OR “Efficiency” OR “Irregularit*” OR “Fraud*” OR “Non-conformitie*” OR “cheat*” OR “Statistics” OR “Descriptive” OR “Smart Classification” OR “management report*” OR “Data Discovery”).

SCOPUS: TITLE-ABS-KEY (health AND audit) AND (“big data” OR “business intelligence” OR “data science” OR “AI” OR “mining” OR “artificial intelligence”) AND (“Efficacy” OR “Effectiveness” OR “Efficiency” OR “Irregularities” OR “Irregularity” OR “Fraud” OR “Non-conformities” OR “cheats” OR “Statistics” OR “Descriptive” OR “Smart Classification” OR “management reports” OR “Data Discovery”).

IEEE Xplore: (“health” AND “audit”) AND (“Big Data” OR “Business Intelligence” OR “Data Science” OR “AI” OR “Mining” OR “Artificial Intelligence”) AND (“Efficacy” OR “Effectiveness” OR “Efficiency” OR “Irregularities” OR “Irregularity” OR “Fraud” OR “Non-conformities” OR “cheats” OR “Statistics” OR “Descriptive” OR “Smart Classification” OR “management reports” OR “Data Discovery”).

ACM: (“health*” AND “audit*”) AND (“Big Data” OR “Business Intelligence” OR “Data Science” OR “AI” OR “Mining” OR “Artificial Intelligence”) AND (“Efficacy” OR

“Effectiveness” OR “Efficiency” OR “Irregularities” OR “Irregularity” OR “Fraud” OR “Non-conformities” OR “cheats” OR “Statistics” OR “Descriptive” OR “Smart Classification” OR “management reports” OR “Data Discovery”).

Web of Science: (“health” AND “audit”) AND (“big data” OR “business intelligence” OR “data science” OR “AI” OR “Mining” OR “Artificial Intelligence”) AND (“Efficacy” OR “Effectiveness” OR “Efficiency” OR “Irregularities” OR “Irregularity” OR “Fraud” OR “Non-conformities” OR “Cheats” OR “Statistics” OR “Descriptive” OR “Smart Classification” OR “management reports” OR “Data Discovery”).

4.3 CRITÉRIOS DE SELEÇÃO DE FONTES

Com o intuito de obter um espaço amostral de artigos filtrados e coerentes com o campo de pesquisa desta dissertação, foram utilizados Critérios de Inclusão (CI) e Critérios de Exclusão (CE) para determinar a relevância dos materiais para o presente contexto. Assim, esses critérios seguem listados a seguir:

Critérios de Inclusão:

- CI1 – Artigos que estejam disponíveis de forma online em bases de dados digitais
- CI2 – Artigos que possuam texto da pesquisa no título, no resumo ou nas palavras-chave
- CI3 – Trabalhos publicados entre 2015 e 2025
- CI4 – Artigos escritos em inglês
- CI5 – Artigos que utilizem métodos, abordagens ou aplicações de auditorias em saúde baseadas em *Big Data*.
- CI6 – Artigos publicados em Journals.

Critérios de Exclusão:

- CE1 – Estudos duplicados
- CE2 – Surveys
- CE3 – Revisões sistemáticas, e outros tipos de revisão de literatura
- CE4 – Artigos que proponham métodos, abordagens ou aplicações de auditorias em saúde não baseadas em *Big Data*
- CE5 – Artigos não relacionados à auditoria em saúde
- CE6 – Artigos indisponíveis para download.

Esses critérios foram de fundamental importância para a aquisição de trabalhos que realmente contemplassem a temática desejada, filtrando e descartando materiais fora do contexto desejado.

4.4 ESTRATÉGIA DE EXTRAÇÃO DE DADOS

Para avaliar e adquirir informações a respeito da massa de artigos selecionados para a construção do mapeamento sistemático, foi elaborado um formulário de extração de dados a respeito de tópicos que podem tanto contribuir para um entendimento a respeito do estado da arte dos artigos, como auxiliar nas respostas das questões de pesquisa adotadas mediante Protocolo PICO. O objetivo é adquirir um método eficiente para extração de informações pertinentes nos artigos. O Quadro 5 destaca cada uma das informações pertinentes para análise e construção da síntese narrativa.

Quadro 5 – Formulário de Extração de dados

Bloco de análise	Descrição do bloco	Conteúdo esperado
Tipo Auditoria	Segundo a metodologia e desenvolvimento do estudo, com quais tipos de auditoria o artigo mais se aproxima?	Clínica, Compliance, Financeira, Qualitativa
Algoritmos/Técnicas Utilizados	Quais algoritmos e técnicas foram implementados para a condução da pesquisa e experimentos práticos	Campo aberto
Área da saúde	Área da saúde utilizada como escopo	Campo aberto
Periodicidade/Origem temporal do dado	Periodicidade da auditoria, e quando não encontrada remete ao período de dados utilizados	Campo aberto
Indicadores-chave de performance	Indicadores relevantes para a construção da pesquisa	Campo aberto
Métricas quantitativas	Métricas quantitativas adotadas pelos autores no decorrer das atividades	Campo aberto
Métricas qualitativas	Métricas qualitativas adotadas pelos autores no decorrer das atividades	Campo aberto
Desafios/Limitações	Desafios e limitações abordados no artigo	Campo aberto
Resultados	Resultados encontrados no artigo	Campo aberto
Ano	Ano de publicação	Campo aberto
País	País de origem da publicação	Campo aberto

Fonte: produzido pelo autor, 2025

O formulário de extração foi elaborado com base em padrões presentes no protocolo PRISMA. A metodologia por trás dessa técnica de extração de dados, presente em processos de Revisões Sistemáticas da Literatura (RSL), diz respeito a extrair dados de artigos publicados num escopo delimitado por pesquisadores por meio do preenchimento de uma série de campos como métricas, técnicas, resultados, entre outros valores (Marcondes; Silva, 2023; Page *et al.*,

2021). Assim, o formulário presente no Quadro 5 foi adotado para viabilizar a identificação dos processos conduzidos nos artigos e responder às questões de pesquisa.

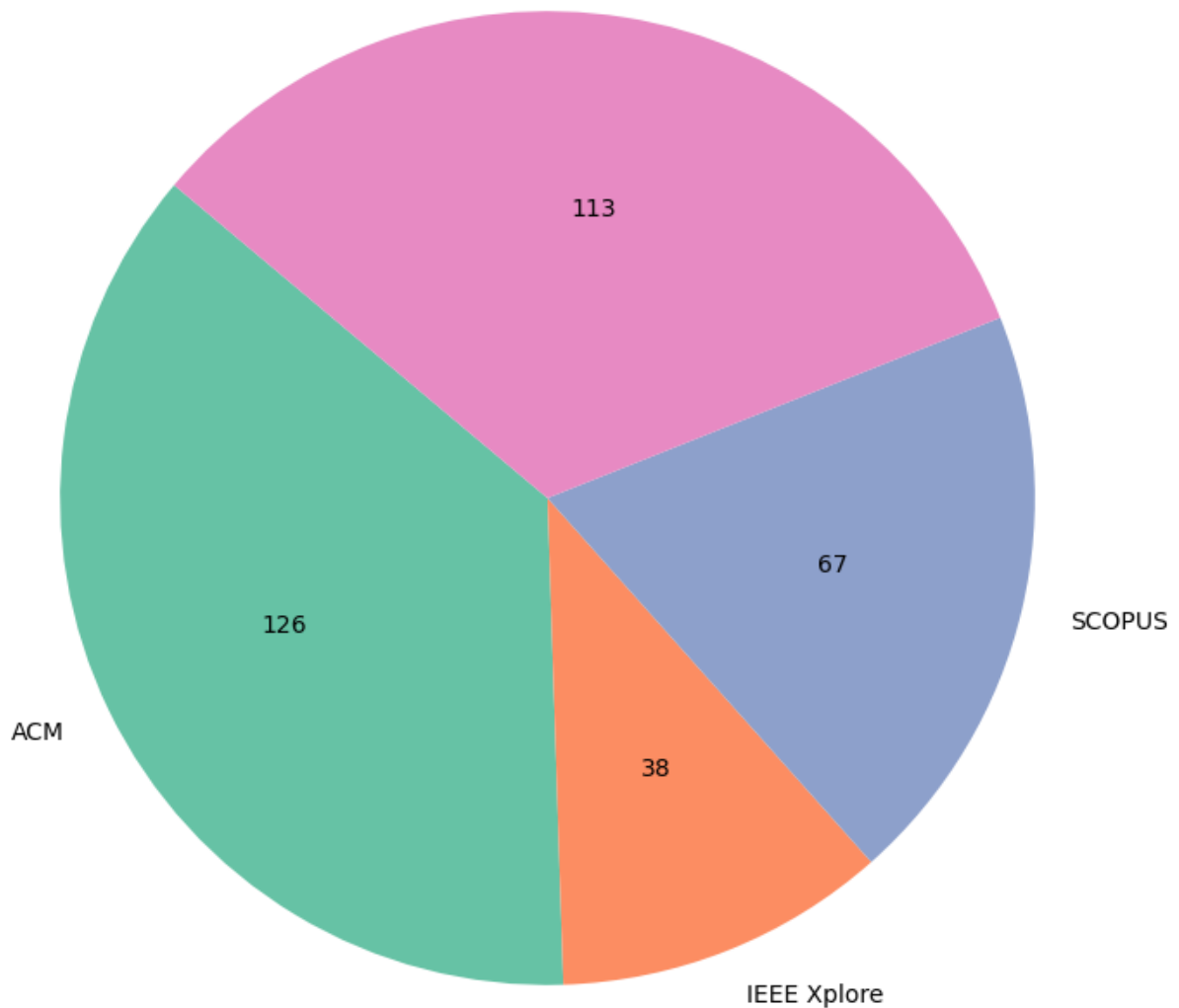
4.5 CONDUÇÃO DO MAPEAMENTO SISTEMÁTICO

A condução do Mapeamento Sistemático da Literatura (MSL) pode ser detalhada segundo os passos a seguir:

1. acesso à base de dados e utilização da *string* de busca específica;
2. aplicação dos filtros de inclusão e exclusão;
3. avaliação de títulos, resumos, palavras-chave e metodologia dos artigos adquiridos;
4. intermediação da plataforma Rayann para performar um processo conjunto com outros pesquisadores, permitindo que cada um exerça a seleção ou a exclusão de artigos com base nos critérios propostos nesta seção;
5. avaliação conjunta para verificação dos artigos selecionados ou em situação para desempate, se são descartados ou não, e;
6. submissão dos artigos inclusos ao formulário de extração de dados presente no Quadro 5.

Por meio desse passo a passo adotado, é possível compreender o fluxo de atividades que levaram à execução do MSL, à aplicação dos critérios de inclusão e exclusão e à extração de informações pertinentes a cada um dos artigos incluídos para a etapa final do trabalho. Ademais, esse fluxo auxilia na apresentação de resultados em seus níveis qualitativos e quantitativos, assim como na construção de uma síntese narrativa do que foi analisado. A partir disso, pode-se visualizar, na Figura 3, o percentual de artigos adquiridos nas bases utilizadas.

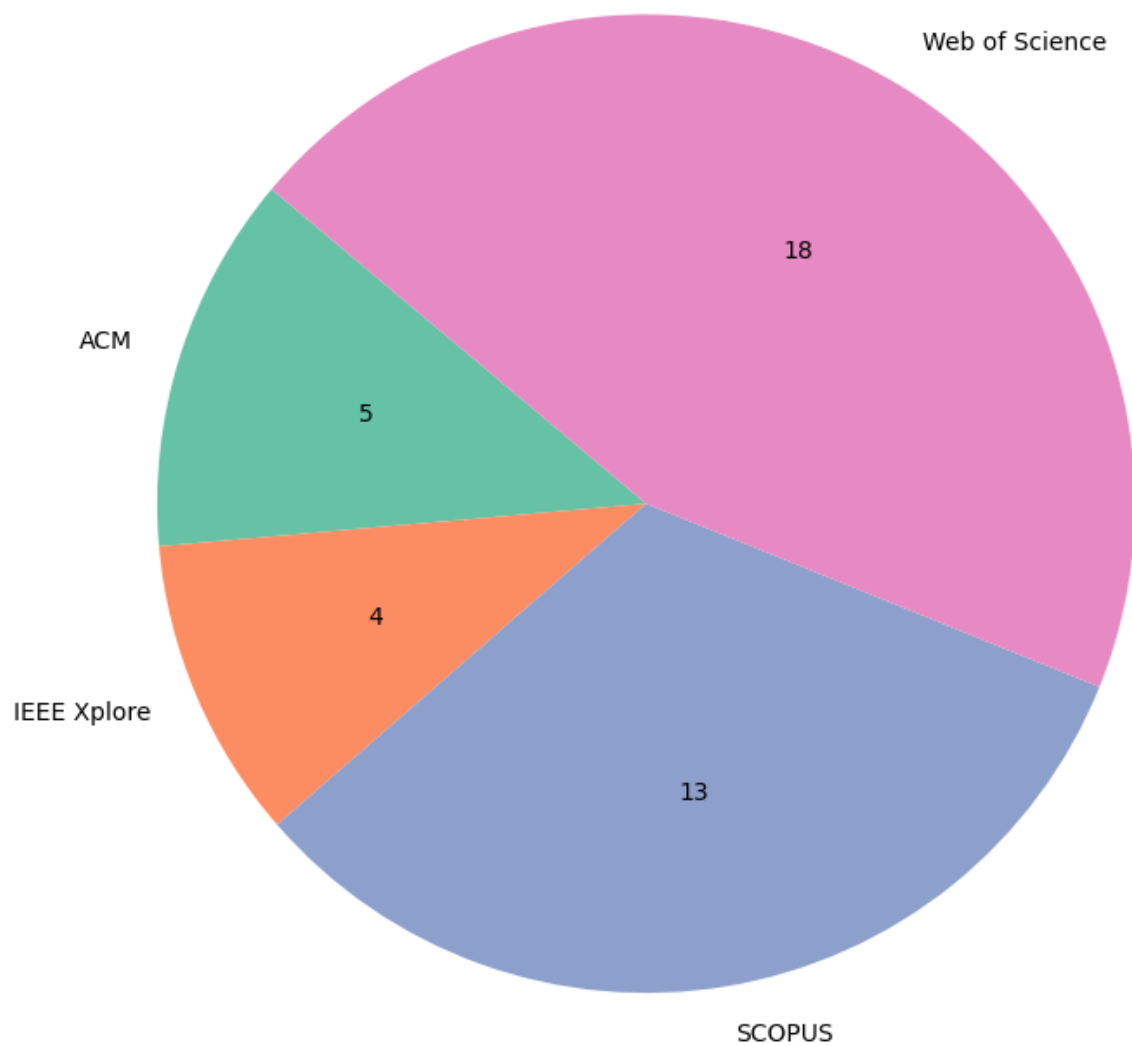
Figura 3 – Quantitativo de Artigos publicados em *Journals* coletados por base
Distribuição das bases de dados pesquisadas
 Web of Science



Fonte: produzida pelo autor com linguagem python3, 2025.

Com base na Figura 3, é possível visualizar um quantitativo de 344 artigos adquiridos. Em primeiro lugar, esse material foi exportado em formato RIS ou bib para a ferramenta de revisão sistemática Rayann, na qual foi possível identificar de imediato a presença de 26 artigos duplicados entre as bases de dados, sobrando então 318 para aplicação dos filtros. Logo após a aplicação dos critérios de inclusão e exclusão e discussão entre pares, o resultado foi uma amostra de 40 artigos para análise. A Figura 4 ilustra esse novo quantitativo.

Figura 4 – Quantitativo dos artigos restantes para análise
Distribuição das bases de dados restantes

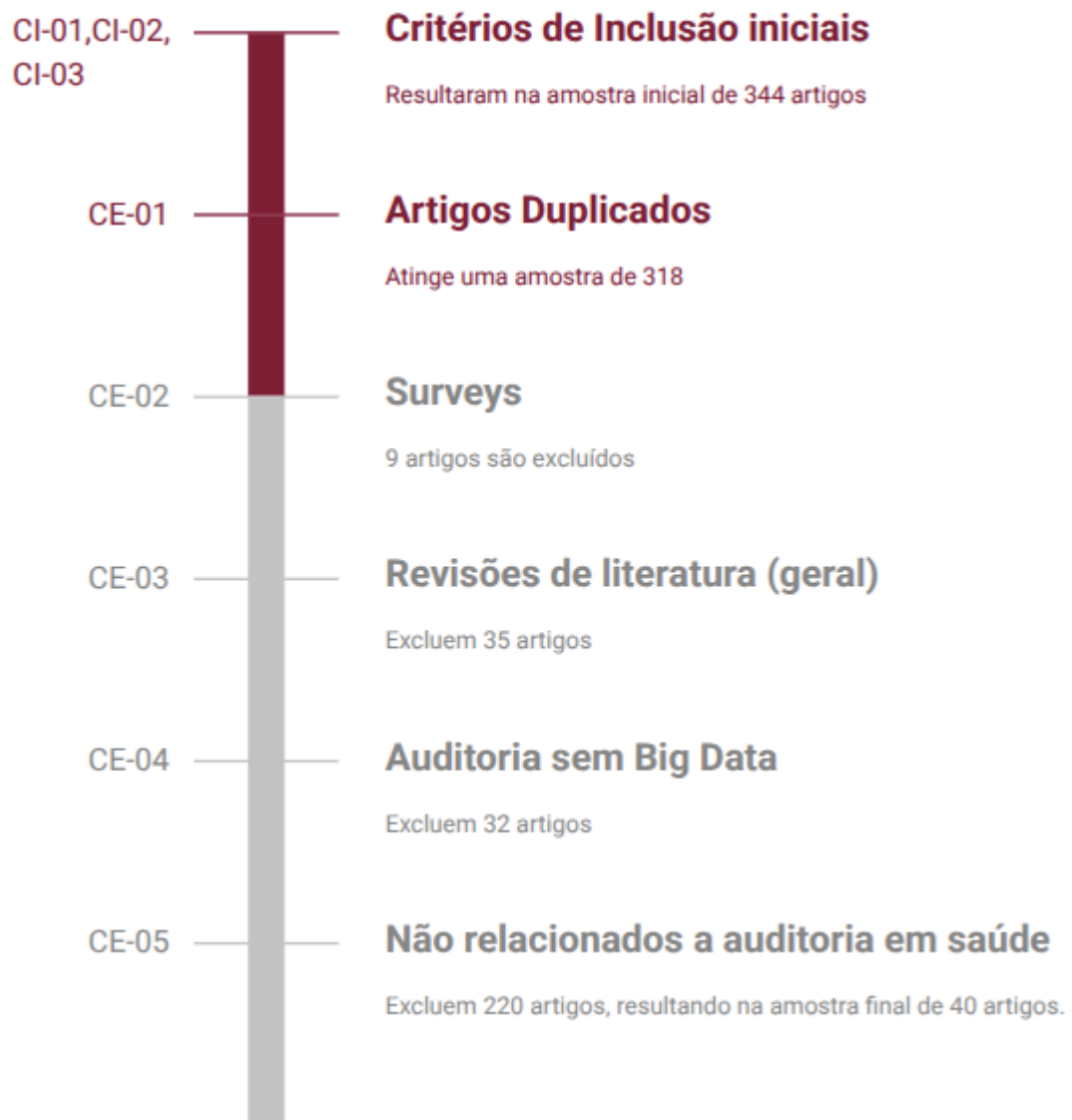


Fonte: produzida pelo autor com linguagem python3, 2025.

4.6 RESULTADOS E DISCUSSÃO DO MAPEAMENTO

Após aplicação do fluxo metodológico para o desenvolvimento do Mapeamento Sistemático da Literatura (MSL), será possível discutir a respeito das questões de pesquisa abordadas nesta seção, tendo como base a amostra final de materiais coletados para análise. O fluxograma pode ser compreendido pela Figura 5, que explica de forma direta a aplicação dos critérios de inclusão iniciais para aquisição da amostra total de 344 artigos, assim como os critérios de exclusão adotados e o quantitativo descartado em cada um dos níveis de execução.

Figura 5 – Fluxograma para exclusão de materiais não pertinentes



Fonte: desenvolvida pelo autor, 2025.

4.7.1 Quais os principais algoritmos e métodos estatísticos utilizados?

Para responder à primeira questão de pesquisa desenvolvida a partir do protocolo PICO, foi utilizado o formulário presente no Quadro 5 como caminho para estruturar esses dados em um formato de planilha. Desse modo, foi desenvolvida uma nuvem de palavras a fim de verificar quais métodos foram mais utilizados nos artigos separados para análise. A Figura 6 expõe essa nuvem de palavras.

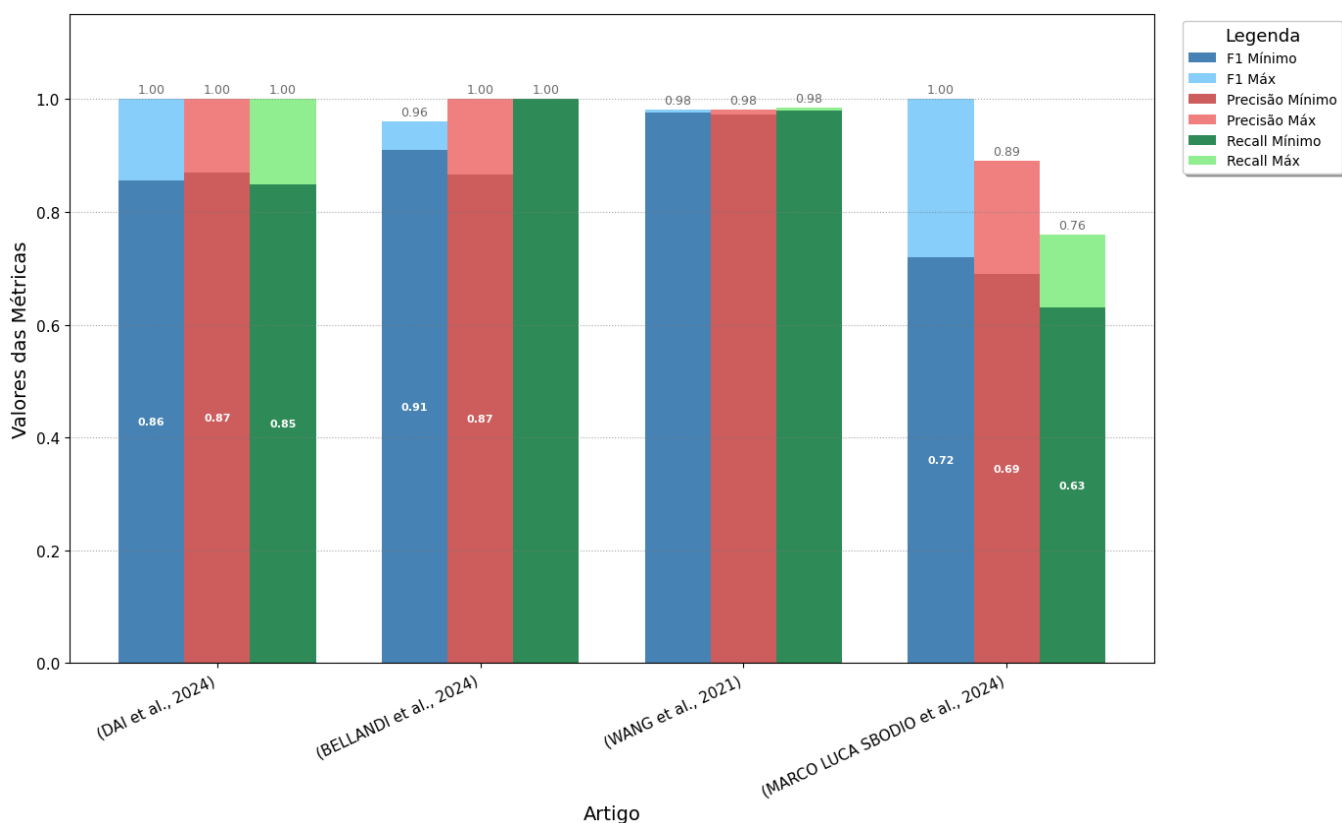
Tabela 1 – Eficiência de algoritmos e métodos

Citação	Eficiência
(Dai <i>et al.</i> , 2024)	F1-Score = 0.856 ~1.0; Precisão = 0.870 ~1; <i>Recall</i> =0.848~1.0;
(Bellandi <i>et al.</i> , 2024)	F1-Score = 0.909 ~0.960; Precisão =0.866 ~1; <i>Recall</i> = 1;
(Wang <i>et al.</i> , 2021)	F1-Score = 0.981 ~ 0.976; Precisão = 0.981 ~ 0.972; <i>Recall</i> = 0.979 ~ 0.985
(Sbodio <i>et al.</i> , 2024)	F1-Score = 0.72 ~1.0 ; <i>Recall</i> = 0.63 ~0.76; Precisão = 0.69 ~0.89

Fonte: desenvolvida pelo autor, 2025.

Já com base na Figura 7, a seguir, pode-se observar uma comparação mais direta em relação às métricas de desempenho dos algoritmos utilizados nas publicações.

Figura 7– Ranqueamento por publicação
Comparativo de Métricas de Desempenho (Mínimo e Máximo)



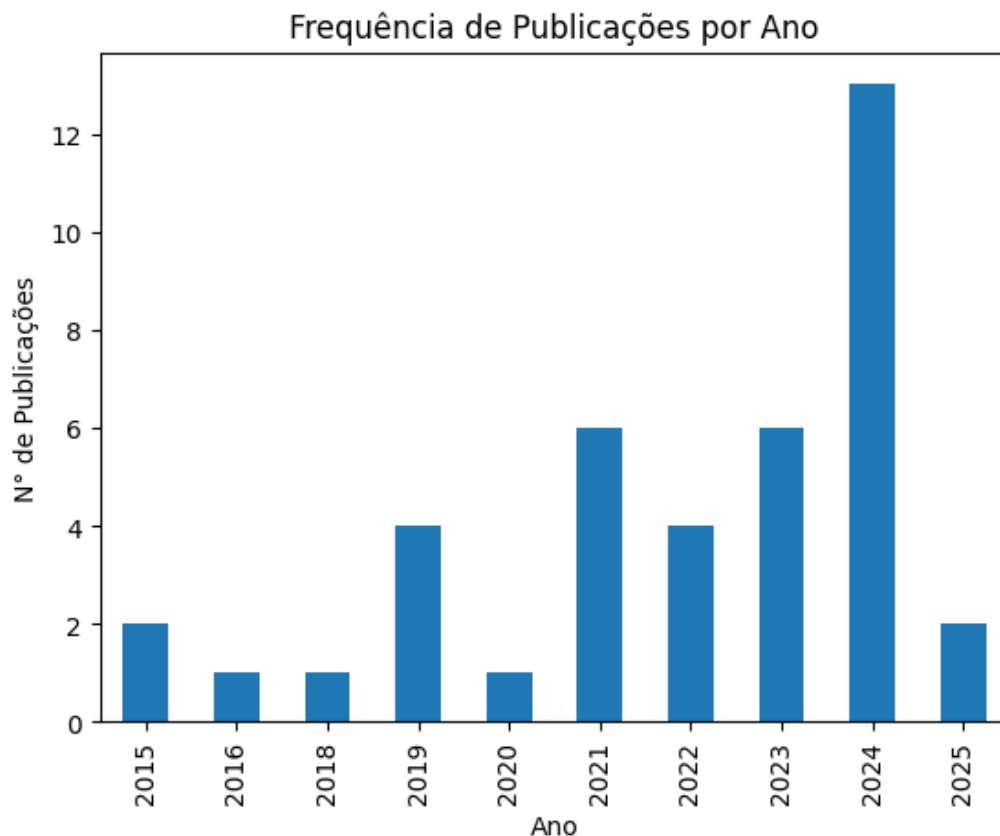
Fonte: elaborada pelo autor com linguagem python3, 2025

É importante ressaltar que os pesquisadores executaram diversos experimentos para avaliação de suas soluções, portanto, há um intervalo entre valores mínimos e máximos sendo

4.7.3 Em quais anos foram publicados mais artigos?

Para responder à pergunta de pesquisa, foram coletados dados a respeito dos anos de publicação em cada um dos artigos separados para amostragem final. Esses dados foram estruturados em formato de planilha .csv e separados por coluna e artigo, fazendo com que o gráfico da Figura 9 demonstrasse a distribuição do volume de artigos produzidos considerando o intervalo de 2015 a 2025. Isso ajuda a entender quais períodos mais se destacaram em relação a publicações relacionadas à temática de auditoria em saúde que utilizassem abordagens baseadas em Big Data.

Figura 9 – Quantitativo de publicações por ano



Fonte: produzida pelo autor com linguagem python3, 2025

Por meio da Figura 9, pode-se perceber um aumento significativo de publicações entre 2023 e 2024. Como o processo de extração e análise de artigos foi feito no primeiro semestre de 2025, ainda não se pode identificar qual seria o montante completo de materiais. Porém, nota-se um valor comparável ao ano de 2015 e superior aos anos de 2016, 2018 e 2020. O propósito final foi destacar o interesse pelo tema nas publicações coletadas ao longo do tempo. Isso também pode ser um dos fatores de provocação para condução de pesquisas na área.

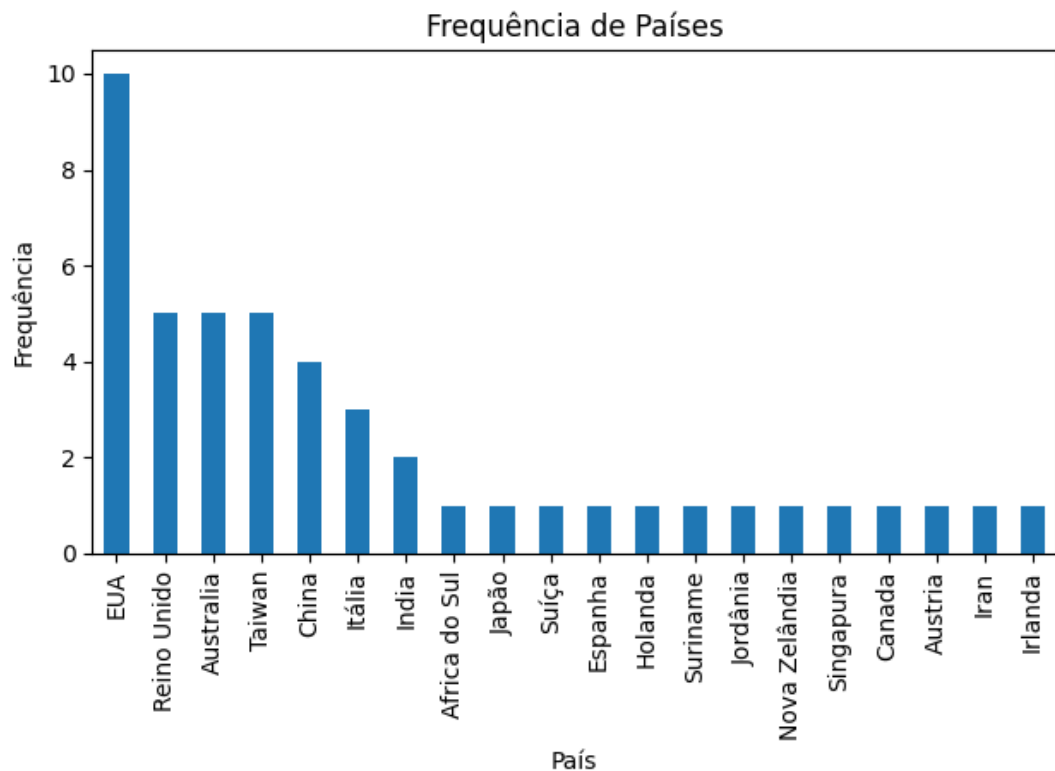
4.7.4 Quais os meios de publicação mais populares?

Uma questão importante a ser respondida remete-se às etapas iniciais de coleta e extração de dados e à utilização da *string* de busca desenvolvida a partir do protocolo PICO. Cada uma das fontes, ACM, IEEE Xplore, SCOPUS e Web of Science gerou sua respectiva amostra de materiais, cujas maiores incidências estão dispostas nas Figuras 3 e 4. Portanto, o quantitativo final de artigos para cada uma das bases de periódicos foi respectivamente: 18 para Web of Science, 14 para SCOPUS, 7 para ACM e 4 para IEEE Xplore.

4.7.5 Quais países têm mais publicações nessa área?

Tendo em vista os objetivos de Controle (C), População (P) e Resultados (O) estabelecidos a partir do protocolo PICO, é preciso apresentar a distribuição de países envolvidos nas publicações adquiridas a fim de compreender em quais regiões o tema está sendo mais incidente. Assim, a Figura 10, a seguir, expõe um gráfico de barras com essas informações.

Figura 10 – Frequência de países na amostra selecionada



Fonte: Imagem produzida pelo autor com linguagem python3, 2025

Com base na Figura 10, é possível notar uma distribuição de publicações superiores considerando a ordem Estados Unidos da América (EUA), Reino Unido e Austrália, com demais países apresentando cerca da metade do quantitativo desenvolvido pelos EUA.

4.7.6 Quais os principais desafios e limitações identificados?

Outra questão de pesquisa pertinente visa a elucidar como os autores abordaram em seus artigos a temática de desafios e limitações destacados ao longo do processo metodológico. Esses tópicos foram encontrados de forma mais específica para cada um dos materiais analisados favorecendo o entendimento a respeito do papel de cada aplicação proposta. Assim, foi feito um recorte acerca do espaço amostral completo, de modo que, dos 40 artigos, 5 foram selecionados para exposição devido a seu grau de semelhança e ao potencial de contribuição para esta dissertação. O Quadro 6 demonstra esses resultados com base no formulário de extração presente no Quadro 5. As limitações demonstram questões relativas ao comportamento das bases de dados utilizadas, à segurança e à otimização dos algoritmos utilizados.

Quadro 6 – Amostra dos artigos selecionados e seus desafios/limitações

Artigo	Citação	Desafios/Limitação
An unsupervised Bayesian hierarchical method for medical fraud assessment	(Ekin, Lakomski e Musal, 2019)	Heterogeneidade nas fontes de dados adotadas. Foi feito um recorte para uma carga volumétrica focada no estado de Vermont (EUA), onde o desvio padrão das amostras é menor.
Density-Based Outlier Detection for Safeguarding Electronic Patient Record Systems	(Boddy <i>et al.</i> , 2019)	Densidade de dados, e complexidade nas redes hospitalares
MRDACE: An Intelligent Architecture For Secure Sharing And Traceability Of The Medical Images And Patients' Records	(Verma Tripathi, Pant e Mrdace, 2025)	Assegurar a integridade dos dados mediante tentativas de acesso maliciosas.
A Blockchain-Based E-Healthcare System With Provenance Awareness	(Sun <i>et al.</i> , 2024)	Manutenção da integridade e segurança dos dados a longo prazo.
FraudAuditor: A Visual Analytics Approach for Collusive Fraud in Health Insurance	(Zhou <i>et al.</i> , 2023)	Otimização do algoritmo utilizado, tratamento de outliers e falsos positivos e variedade de dados.

Fonte: produzido pelo autor, 2025

Além do recorte apresentado, foi possível identificar alguns pontos-chave ao longo da leitura dos artigos no geral. A qualidade e a disponibilidade dos dados se mostraram um tema recorrente para as publicações de Bellandi *et al.* (2024), Dafaalla *et al.* (2023) e Groh *et al.* (2022). A pesquisa de Navarrete, Gallego e Padilla (2021) apresenta também alguns tópicos interessantes para essa parte do estudo: a avaliação de fraudes financeiras além de abordagens tradicionais e a dificuldade dos modelos em classificar o objeto de estudo como uma empresa em estado de falência. Esses fatores destacados podem ser alertas para a formulação de metodologias capazes de se adaptar a um cenário de auditoria em saúde.

4.7.7 Quais as contribuições desenvolvidas pelas pesquisas?

Os mesmos artigos utilizados como exemplos para citação de desafios e limitações também apresentaram contribuições interessantes para a temática de auditoria em saúde, intermediada por Big Data. Essas informações foram separadas e agrupadas no Quadro 7 a fim

de expor a contribuição particular de cada um dos artigos já previamente avaliados na seção anterior. Desse modo, os trabalhos giram em torno da preservação da integridade das informações adquiridas e da identificação de anomalias por meio de métodos estatísticos e uso de *blockchain*.

Quadro 7 – Amostra de contribuições dos artigos selecionados

Artigo	Citação	Contribuições
An unsupervised Bayesian hierarchical method for medical fraud assessment	(Ekin, Lakomski e Musal, 2019)	Identificação de fraudes médicas por meio da aplicação de métodos estatísticos
Density-Based Outlier Detection for Safeguarding Electronic Patient Record Systems	(Boddy <i>et al.</i> , 2019)	Proteção dos Registros Médicos Eletrônicos de pacientes.
MRDACE: An Intelligent Architecture For Secure Sharing And Traceability Of The Medical Images And Patients' Records	(Verma Tripathi, Pant e Mrdace, 2025)	Arquitetura para preservar a integridade e segurança de dados médicos
A Blockchain-Based E-Healthcare System With Provenance Awareness	(Sun <i>et al.</i> , 2024)	Mecanismo de auditoria intermediado pelo uso de <i>blockchain</i> para averiguar a confiabilidade de processos que envolvam Registros Médicos Eletrônicos
FraudAuditor: A Visual Analytics Approach for Collusive Fraud in Health Insurance	(Zhou <i>et al.</i> , 2023)	Método capaz de abordar problemas que envolvam fraudes em seguros de saúde, por meio de técnicas computacionais, como detecção de comunidades, cálculo de semelhança e <i>Principal Component Analysis</i> (PCA)

Fonte: produzido pelo autor, 2025

Os outros artigos citados na seção 4.7.6 também apresentaram contribuições interessantes para o trabalho. A publicação de Navarrete, Gallego e Padilla (2021), por exemplo, destaca a importância pela busca de parâmetros mais profundos para identificação de irregularidades. Isso significa que os autores não reduziram o escopo a indicadores financeiros. Ademais, os outros trabalhos contribuíram com a extração de indicadores pertinentes para os problemas apresentados em cada um deles, promovendo a identificação de lacunas e medidas propositivas para resolução dos problemas. A partir disso, pode-se refletir a respeito da

existência de uma solução definitiva para problemas desse perfil. Cada cenário pode demandar abordagens distintas e singulares, em virtude de assegurar resultados mais consistentes.

4.8 SÍNTESE NARRATIVA

Nesta subseção, serão discutidos os principais aspectos notados ao longo do processo de resposta às questões de pesquisa abordadas, mediante uma síntese narrativa. Para Rabinowitz (1998), esse processo é um mecanismo capaz de agregar as informações dispersas para transformá-las em algo relevante para o entendimento do processo. Nesse caso, o objetivo da síntese narrativa é apresentar os dados adquiridos resumidamente, com o intuito de provocar novas questões de pesquisa e propostas para trabalhos futuros.

Na seção 2 desta dissertação, foram definidos os conceitos relativos à auditoria e seus tipos, como clínica, financeira, *compliance*, operacional e qualitativa. Desse modo, foi estabelecido um paradigma para a classificação dos artigos coletados por meio desses padrões. Um exemplo disso é o trabalho de Zhou *et al.* (2023), que apresenta evidências mais claras para um processo de auditoria financeira, tendo em vista seu objeto de estudo relacionado a seguros em saúde. Todavia, ao longo da leitura do material, percebeu-se que algumas publicações não possuíam tendências tão explícitas. Portanto, esse processo tornou-se uma análise qualitativa mediante os conteúdos apresentados nos trabalhos.

Acrescem-se à discussão as análises referentes a métodos utilizados e áreas/tipo de dado em saúde presentes nos artigos. Foi observada uma vasta variedade de áreas, tipos de dados e métodos diferentes. Identificou-se a presença de estudos envolvidos com oncologia, cardiologia, que usaram registros médicos eletrônicos como fontes para análise, além da aplicação de diferentes abordagens, desde métodos estatísticos tradicionais até algoritmos de *machine learning* e até mesmo aplicações com Inteligência Artificial Generativa. Por meio disso, observou-se uma grande diversidade de informações para esses dois campos abordados.

Também foi feita uma análise a respeito dos anos com mais publicações, destacando os países que mais contribuíram para o escopo, com base na *string* de busca empregada e nos meios de publicação mais utilizados. Por meio disso, os resultados apontaram uma vasta gama de artigos publicados no ano de 2024. Considerando que 2025 segue em sua metade, existe a possibilidade de mais publicações relativas à auditoria em saúde, inseridas em um contexto de Big Data, serem cada vez mais presentes. Além disso, os Estados Unidos da América foram o país que mais trouxe publicações que se encaixam no escopo desta pesquisa, seguido por Reino Unido e Austrália. Por fim, a plataforma que mais contribuiu quantitativamente para o

levantamento de dados foi Web of Science, com 18 artigos, enquanto as demais não passaram de 13 artigos.

Outra questão importante a respeito desse processo de pesquisa na literatura remete a trabalhos que tenham envolvimento direto com Órteses, Próteses e Materiais (OPME). Ao longo da extração, não foram identificadas publicações que estivessem direcionadas a essa área, demonstrando um contraste com a proposta estabelecida nesta dissertação de mestrado, que tem como um de seus objetos de estudo OPME. Esse também pode ser um fator de incentivo e provocação para o desenvolvimento de novos trabalhos que busquem investigar esse nicho e propor soluções para contextos de auditoria em OPME.

O Quadro 20, presente no apêndice A desta pesquisa, contém alguns artigos da amostra total que foram submetidos ao formulário de extração do Quadro 5. Assim, foram selecionados os tópicos mais pertinentes como: tipo de auditoria, técnicas/métodos utilizados, indicadores e métricas. Isso também proporciona uma análise comparativa entre as publicações e suas características singulares. Foram identificadas alternativas para atuação em auditoria em saúde englobando diversos raciocínios distintos. Além disso, cada métrica e cada indicador variavam conforme o tipo de problema abordado na pesquisa. Os próximos passos giraram em torno dos desafios e resultados encontrados nesses materiais.

Finalmente, limitações e resultados também se comportaram de forma diversa e específica para cada um dos artigos analisados, tendo poucos pontos em comum. Por isso, foi feita uma apresentação estratificada de 5 artigos do conjunto geral a fim de expor esses tópicos de interesse para as questões de pesquisa propostas. Os desafios podem ser resumidos em assuntos relativos ao volume e à heterogeneidade das fontes de dados usadas até as questões acerca de segurança e melhoria dos algoritmos. Já os resultados giram em torno da identificação de anomalias, segurança de dados, aplicação de modelos capazes de classificar irregularidades, e uso de *blockchain* a fim de atestar a confiabilidade dos dados médicos.

4.9 AMEAÇAS À VALIDADE

Para aquisição de uma base de dados que seguisse adequadamente os critérios de inclusão e exclusão, o processo de seleção de artigos foi feito com a participação de 3 pesquisadores familiarizados com a temática. Além disso, a participação desses pesquisadores foi, sobretudo, focada na exclusão ou inclusão de artigos marcados como incertos, ou “*maybe*” na plataforma *Rayann*. Assim, serão discutidos tópicos como validação da *string* de busca

utilizada, variedade dos métodos e tipo de dados/áreas da saúde utilizados, vieses de inclusão e exclusão e a validade temporal.

4.9.1 Validação da string de busca

A *string* de busca foi validada por meio da análise de um artigo de controle capaz de atender às demandas propostas. O estudo adotado para isso foi o “FraudAuditor: A Visual Analytics Approach for Collusive Fraud in Health Insurance” (Zhou *et al.*, 2023), que tem como escopo de atuação a análise de dados referentes a seguros de saúde mediante fraudes. Desse modo, a *string* de busca do protocolo PICO apresentou a capacidade de adquirir materiais coerentes com a temática de auditoria em saúde. Apesar disso, é importante ressaltar que alguns materiais relacionados ao assunto podem acabar não sendo coletados. Ademais, houve dificuldade para identificar publicações que abordassem auditoria em saúde com intermédio de técnicas de Big Data no ramo de Órteses, Próteses e Materiais Especiais (OPME), destacando um alerta em relação às possibilidades e à importância de pesquisa a respeito desse tema. Desse modo, a pesquisa atual, utilizando a *string* de busca proposta, não identificou publicações relacionadas ao tópico de OPME.

4.9.2 Variedade de Algoritmos, técnicas e tipos de dado em saúde

Ao longo da extração de dados referentes às técnicas abordadas nos artigos selecionados, observou-se uma ampla variedade de métodos utilizados para identificação de padrões e eventuais anomalias nos dados avaliados. Assim, foi feita uma análise a partir de nuvem de palavras para apresentação das heurísticas empregadas. O mesmo procedimento foi utilizado para os tipos de dados e áreas da saúde, apresentando uma ampla variedade.

4.9.3 Viés de Inclusão e Exclusão

Os materiais adquiridos por meio desse processo de mapeamento sistemático não representam a totalidade do estado da arte a respeito da área de auditoria em saúde, sendo somente um fragmento de dados coletados por meio de um conjunto de Critérios de Exclusão (CE) e Critérios de Inclusão (CI) que auxiliaram na formação do conjunto de artigos utilizados.

4.9.4 Viés de Seleção Temporal

Em relação à periodicidade das publicações avaliadas, foi observado que, em todos os anos, houve artigos pertinentes para contribuir com esta dissertação de mestrado. Nesse caso, o ano de 2024 foi o que retornou mais materiais, o que pode indicar um crescimento do interesse de pesquisadores a respeito do tema até esse período.

4.10 CONSIDERAÇÕES FINAIS

Esta subseção contempla o Mapeamento Sistemático da Literatura (MSL), com o intuito de explorar técnicas utilizadas, dificuldades e contribuições dos artigos adquiridos, identificar a periodicidade das publicações e os países que mais contribuíram para o cenário. Desse modo, do conjunto inicial de 344 artigos publicados em *Journals*, 40 surgiram como resultados após a aplicação dos Critérios de Inclusão (CI) e Critérios de Exclusão (CE). A escolha dos artigos, delimitada a *Journals*, deve-se ao fato de esse tipo de publicação passar por critérios mais rigorosos para sua publicação, resultando em materiais desenvolvidos com mais rigor técnico.

Um ponto importante observado ao longo da leitura dos artigos coletados diz respeito ao aspecto subjetivo da área. Conforme descrito na seção 2 do trabalho, os conceitos de auditoria abordados demonstraram diversidade, e, em alguns casos, até mesmo a presença de raciocínios mútuos. Isso significa que alguns artigos foram interpretados a partir de mais de um tipo de auditoria. O artigo de Dafaalla *et al.* (2023), por exemplo, foi concebido ao longo da leitura como um conteúdo que apresentava características qualitativas e clínicas. Isso significa que, em alguns casos, mais de um tema de auditoria foi abordado, também tendo em vista a finalidade do estudo analisado.

Em suma, os artigos apresentaram diversas soluções e casos de uso diferentes para abordar auditorias em saúde com uso de Big Data. Isso também revelou possibilidades de aplicação para técnicas estatísticas tradicionais, algoritmos de aprendizado de máquina, *blockchain* e até mesmo uso de inteligência artificial generativa. Portanto, este Mapeamento Sistemático da Literatura auxilia no entendimento de como outros autores estão enxergando e abordando essa área do conhecimento, além de destacar contribuições e desafios enfrentados pelos autores. Na próxima seção, detalharemos a metodologia adotada para este estudo.

5 METODOLOGIA

Esta seção detalha a metodologia adotada para coleta e análise dos dados desta pesquisa.

5.1 COLETA DA BASE DE DADOS

O primeiro passo para a construção prática deste trabalho foi a coleta e o consumo de múltiplas bases de dados. Os dados abertos utilizados para a construção das trilhas de auditoria direcionadas a Órteses, Próteses e Materiais Especiais (OPME), no âmbito de pessoa jurídica e auxílio do governo federal, encontram-se disponíveis no portal gov.br, por meio dos endereços eletrônicos presentes no Quadro 8.

Quadro 8 – Dados referentes a Bolsa Família e Cadastro da Pessoa Jurídica

Link	Base	Perfil
https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica--cnpj	Cadastro Nacional da Pessoa Jurídica	Dado Público
https://dados.gov.br/dados/conjuntos-dados/bolsa-familia---pagamento	Base de dados de pagamentos para pessoas que recebem Bolsa Família	Dado Público

Fonte: elaborado pelo autor, 2025

Esses dados foram coletados manualmente, baixando planilhas em formato “.csv” presentes no portal de acesso a dados aberto do governo federal. Essas duas bases de dados serão utilizadas para a identificação de vínculos entre outros componentes assimilados na arquitetura de *Big Data* do projeto FiscalizaSUS. Esse raciocínio também será aproveitado mais à frente para o desenvolvimento de um modelo de dados integrado.

Os dados referentes ao programa Bolsa Família foram coletados considerando os períodos de 2019 a 2021, tendo em vista a mudança na logística e o nome por trás do benefício a partir do período de 2023, que passa a ser identificado como “Novo Bolsa Família”. Isto pode ser verificado por meio dos links utilizados para baixar as planilhas. Foi observado, também, que a estrutura de dados a partir da nova nomenclatura se mantém similar à anterior. Esse fator é importante para a modelagem de um quadro que armazena o nome dos auxílios governamentais e um identificador para cada um.

Além das bases citadas anteriormente, o projeto FiscalizaSUS já contava com dados referentes a fornecedores de OPMEs, notas fiscais de OPME oriundas dos estados do Rio Grande do Norte e do Ceará, tabela de médicos que atendem no Sistema Único de Saúde do Brasil e informações relacionadas às malhas de cidades e estados brasileiros para facilitar a identificação geográfica de um estabelecimento de saúde onde esses insumos sejam utilizados.

Um ponto importante gira em torno da não divulgação de dados pessoais e informações sensíveis que podem estar presentes nas bases de pagamentos do Bolsa Família e de pessoa jurídica, uma vez que um dos intuitos presentes neste trabalho é respeitar as diretrizes mencionadas na Lei nº 13.709, de 14 de agosto de 2018, Lei Geral de Proteção de Dados (LGPD). Todas as bases destacadas anteriormente são utilizadas de maneira conjunta para a pavimentação das trilhas de auditoria mencionadas na introdução do trabalho. Em primeiro lugar, será discutido o tratamento dos novos dados referentes ao Bolsa Família e aos sócios da empresa. Ademais, há o tratamento e a ingestão de dados referentes às trilhas de auditoria para ocorrências com *Stents* Recorrentes e procedimentos com valores anômalos para procedimentos registrados pelo Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP).

Neste sentido, como já foram apresentadas as bases principais relacionadas à trilha de Bolsa Família e sócios, vale destacar a origem dos dados para procedimentos SIGTAP e *Stents* Recorrentes. Essas informações remetem a Autorização de Internação Hospitalar (AIH), Cadastro Nacional de Estabelecimentos de Saúde (CNES) e Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP), que são bases oriundas do Departamento de Informação e Informática do SUS (DataSUS). Esse é o órgão encarregado de atuar como um elemento essencial para a transformação digital do Sistema Único de Saúde (SUS), e, consequentemente, para a disponibilização de bases de dados para livre acesso ao público. O Quadro 9 apresenta a origem dos dados, a descrição e a trilha de auditoria envolvida.

Quadro 9 – Origem das trilhas de *Stents* recorrentes e procedimentos SIGTAP

Origem	Descrição da origem	Trilha de Auditoria
AIH	DataSUS	<i>Stents</i> recorrentes
SIGTAP	DataSUS	<i>Stents</i> recorrentes
CNES	DataSUS	<i>Stents</i> recorrentes
API IBGE	Instituto Brasileiro de Geografia e Estatística	<i>Stents</i> recorrentes
Notas fiscais Rio Grande do Norte e Ceará	Acordo cooperação entre Ministério Público Federal (MPF), Secretaria Estadual de Tributação (SET/RN) e LAIS/UFRN	Procedimentos SIGTAP
SIGTAP	DataSUS	Procedimentos SIGTAP
API IBGE	Instituto Brasileiro de Geografia e Estatística (IBGE)	Procedimentos SIGTAP

Fonte: elaborado pelo autor, 2025

Não obstante esses quadros estarem organizados, pré-processados e atestados em relação à veracidade dos dados identificados, o objetivo é viabilizar a construção de trilhas de auditoria relativas a Órteses, Próteses e Materiais Especiais do projeto FiscalizaSUS. Todo o processo relacionado à coleta, tratamento e carregamento dos dados é suportado pelas ferramentas instaladas na arquitetura de Big Data do projeto, viabilizando a escalabilidade e a construção de novos conjuntos de dados devidamente estruturados e tratados para consumo ao longo das trilhas de auditoria.

A seguir, serão discutidos os processos de construção relativos às trilhas de auditoria de *Stents* recorrentes e Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP). Assim, destacando quais métricas foram aplicadas para seu desenvolvimento, como os dados foram tratados, o relacionamento de vínculos entre as trilhas de auditoria de OPME, estratificação do modelo de dados geral, carregamento no banco de dados do projeto e conexão entre as trilhas.

5.1.1 Trilha de *Stents* recorrentes

A trilha de *Stents* recorrentes, cadastrada na plataforma do projeto FiscalizaSUS, considera a arquitetura de Big Data orientada para a condução das atividades do projeto. Assim como a trilha discutida anteriormente, ela conta com o uso de dados referentes a fornecedores e notas fiscais de Órteses, Próteses e Materiais Especiais (OPME). A partir do raciocínio apresentado anteriormente, serão discutidos os paradigmas dessa trilha de auditoria.

A proposta da trilha é disponibilizar um mecanismo de auxílio em atividades de vigilância e controle para o uso de procedimentos que envolvam esse tipo de material, considerando eventos cíclicos, por exemplo, um mesmo paciente ter realizado o procedimento a cada 30 dias. Esse intervalo temporal é justificado por meio de uma experiência empírica do conjunto de auditores que participam do projeto FiscalizaSUS, por se tratar de um período cuja reutilização de um OPME seria limitada ou menos provável.

Para verificar essa periodicidade, foi aplicado o uso integrado das bases de dados referentes a procedimentos médicos envolvendo OPME, informações de estabelecimentos de saúde, base de dados referente à Autorização de Internação Hospitalar (AIH) e a componentes geográficos disponibilizados por meio de consumo de dados do Instituto Brasileiro de Geografia e Estatística (IBGE). Para tanto, foram cruzados valores referentes aos períodos de aplicação dos *stents* buscando realizar uma análise de frequência referente a essa atividade.

Assim, as principais métricas adotadas para identificação de indícios presentes na trilha podem ser identificadas como:

- 1) periodicidade e frequência entre os procedimentos;
- 2) total de procedimentos com *Stents* por paciente;
- 3) frequência de repetições dos procedimentos com base em unidades de saúde e profissional de saúde;
- 4) eventos reincidentes no recorte temporal de análise;
- 5) identificação de indícios por meio de CNPJ do fornecedor, e comparação com a média.

As métricas destacadas estão cadastradas no ambiente de homologação do projeto FiscalizaSUS². O acesso ao ambiente de homologação é feito por meio da ferramenta de autenticação Sabiá, desenvolvida pelo Laboratório de Inovação Tecnológica em Saúde. Por meio desse acesso, os auditores do Departamento Nacional de Auditoria do SUS (DenaSUS)

² Podem ser identificadas por meio do link: <https://fiscalizasus.homolog.lais.ufrn.br/pergunta/10>.

podem colaborar e visualizar as camadas visuais desenvolvidas pelos pesquisadores do projeto FiscalizaSUS e sugerir mudanças para melhores resultados. Desse modo, por meio da elaboração de métricas de avaliação, será possível identificar situações que apresentam inconformidades quanto aos períodos de realização dos procedimentos com *Stents*, proporcionando a visualização desses indícios em diferentes níveis de complexidade, na esfera municipal, estadual ou na unidade de saúde envolvida, por meio do vínculo com o CNES identificado via integração das bases de dados adotadas.

5.1.2 Trilha de procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)

Essa trilha de auditoria tem como objeto de estudo a realização de Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP) que apresentem informações inconsistentes ou anômalas, além de considerar fatores como prazos de validade e valores de Órteses, Próteses e Materiais Especiais empregados nessas atividades. Mediante isso, poderão ser apresentados os meios para desenvolver essa trilha de auditoria.

Para tanto, utiliza análise de *outliers*, valores anômalos e fora do padrão da amostra, por meio do método tradicional de *Tukey*, que foi empregado por apresentar menor sensibilidade a valores extremos, e por estar mais focado na análise dos *quartis* e mediana do conjunto de dados adotado. Esse método foi um facilitador para identificar a concentração de valores monetários referentes aos procedimentos SIGTAP. Com base nesse raciocínio, foram elaboradas métricas para o desenvolvimento do processo analítico e identificação de eventos irregulares. O fluxo foi construído seguindo os seguintes elementos estatísticos:

- 1) média de valor por procedimento;
- 2) mediana das médias de valor por procedimento;
- 3) primeiro quartil das médias;
- 4) terceiro quartil das médias;
- 5) desvio padrão e variações interquartil.

As métricas referentes a essa trilha de auditoria foram cadastradas na plataforma de homologação do projeto FiscalizaSUS, seguindo as diretrizes firmadas e o trabalho conjunto com auditores do SUS. Ao final da análise, é calculada uma média geral que destaca o comportamento dos dados avaliados, viabilizando a identificação de irregularidades e anomalias no presente contexto. Em seguida, também será possível orientar a trilha para a

identificação gráfica das irregularidades, destacando a identificação por meio da posição geográfica, com os desvios de valores nos procedimentos realizados.

Ademais, será debatido o processo de tratamento e curadoria dos dados utilizados para o processo de análise das trilhas referentes a Órteses, Próteses e Materiais Especiais no âmbito do projeto FiscalizaSUS. As próximas etapas foram fundamentais para realizar a agregação das bases de dados e apresentar as métricas das trilhas de auditoria desenvolvidas nesta dissertação de mestrado.

5.2 TRATAMENTO DOS DADOS DA TRILHA BOLSA FAMÍLIA E MÉDICO FORNECEDOR

Nesta seção do trabalho, será discutido o percurso para o tratamento de dados e as ferramentas utilizadas. No momento da coleta dos dados oriundos do portal www.gov.br, foi possível perceber que os arquivos adquiridos estavam formatados como “.csv”. Em seguida, foram armazenados em uma camada primária da ferramenta *MinIO* presente na arquitetura de Big Data do projeto FiscalizaSUS. A partir disso, foi possível elaborar as próximas etapas de tratamento de dados, contando também com outras ferramentas disponibilizadas na arquitetura do projeto.

O segundo passo consistiu em realizar a leitura dos dados no ambiente *Jupyter Lab* a fim de verificar fatores de relevância para a análise, como sua veracidade. Inicialmente, foi observado que as planilhas de pessoas jurídicas possuíam asteriscos nas colunas de CPF e CNPJ como uma forma de anonimizar essa informação. Para testar o cruzamento entre as bases de dados adquiridas, foi necessário realizar o tratamento dessa coluna, removendo esses asteriscos e, assim, aproveitando os 6 números centrais do CPF e os 8 primeiros números do CNPJ. Também foi verificada a coluna de nome, um elemento-chave para o cruzamento com as demais tabela que compõem as trilhas de auditoria, sendo padronizada adequadamente, passando por pré-processamento e formatação devida.

A mesma lógica foi aplicada para a tabela de beneficiários do auxílio do governo federal, garantindo assim uma organização regularizada por meio do pré-processamento de caracteres indesejáveis e de texto nas colunas de CPF e NOME. Desse modo, foi testado um cruzamento dessas variáveis utilizando a linguagem de programação python com intermédio da biblioteca *pyspark*, sendo possível validar a formatação dos dados adquiridos entre as duas bases oriundas da fonte original.

Uma base de dados importante para as trilhas de auditoria diz respeito ao conjunto de

dados referentes a médicos que atendem no sistema único de saúde. Essa base foi lida por meio da ferramenta *OpenSearch*, utilizando a biblioteca *Pandas* para criação de um *dataframe* de modo a analisar a forma dos dados obtidos. Foi feita uma edição nas colunas de CPF e CNPJ para que ambas estivessem nos mesmos moldes que as respectivas colunas presentes na tabela de pessoas jurídicas.

Em seguida, a base referente a Fornecedores de Órteses, Próteses e Materiais Especiais (OPME) foi agregada ao cruzamento com as demais bases. Por meio do campo de CNPJ, foi possível aplicar um teste, de modo a identificar se um usuário que possui vínculo societário e recebe o benefício social está envolvido em processos transacionais envolvendo OPME. Apesar disso, não foi identificado qualquer indício real entre essas bases com os dados disponíveis até o momento. A mesma situação pôde ser identificada quando a base de funcionários públicos foi apreciada para os cruzamentos.

Por fim, a base de notas fiscais eletrônicas de OPME dos estados do Rio Grande do Norte e Ceará foram agregadas ao conjunto de dados. Essa base possui o parâmetro de *idfornecedor*, que ajuda na identificação dos valores monetários adquiridos por um fornecedor que possua vínculo no programa de transferência de renda ou em caso de um médico apresentar vínculo societário com o fornecedor do insumo. Nesse caso, vale destacar que as notas fiscais adquiridas passaram por um processo de classificação por meio da ferramenta OPMINER, desenvolvida pelos pesquisadores do projeto FiscalizaSUS a fim de tratar e classificar notas fiscais eletrônicas em grupos adequados conforme apresentado na publicação de Gomes *et al.* (2025), visando a adquirir uma classificação adequada para o seu respectivo material.

Assim, os dados de pré-processamento, limpeza e formatação foram passados para um Grafo Acíclico Direcionado (DAG) no ambiente *Airflow* disponível na arquitetura do projeto FiscalizaSUS, viabilizando o carregamento das tabelas no banco de dados fornecido. A partir dos tópicos explorados nos procedimentos dessas trilhas, também foi possível considerar uma logística similar para o tratamento das trilhas de auditoria relativas aos *Stents* recorrentes e Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP).

5.2.1 Tratamento dos dados para trilha de *Stents* recorrentes

Assim como as trilhas descritas anteriormente, o processo de implementação consiste no desenvolvimento de um ciclo de Extração, Transformação e Carga (ETL). Em primeiro lugar, os dados originais, descritos no Quadro 9, também são armazenados em uma camada

stage disponível na arquitetura de Big Data do projeto FiscalizaSUS, tendo sido adaptados para o formato *parquet*. Com isso, as bases de dados são consumidas por meio do Apache Spark e colocadas em um processamento de rotina por meio de um Grafo Acíclico Direcionado (*DAG*), com intermédio do *Apache Airflow*. Ademais, são utilizados como referência identificadores a respeito dos *Stents* e da categoria de angioplastia para filtrar procedimentos que estejam inseridos nesse tópico, permitindo o descarte de valores duplicados.

Desse modo, os próximos passos visam a coletar a quantidade de procedimentos e de valores relacionados às Autorizações de Internações Hospitalares (AIH) e armazená-los em um novo conjunto de dados orientado ao procedimento de angioplastia. Além disso, informações relativas a datas são transformadas para o formato *date*, sendo mais adequado para etapas futuras, como visualização. Não obstante, ainda são calculados os indícios por meio de análise temporal decorrente em relação ao tempo de internações recorrentes de um paciente em períodos de 15 a 30 dias, cada um representando um tipo de anomalia. Também é importante identificar médicos solicitados e autorizados, assim como pacientes que não possuem uma identificação.

Nessa trilha, são selecionadas somente as colunas pertinentes para a trilha de auditoria de *Stents* recorrentes, sendo aplicada uma agregação com a tabela de Cadastro Nacional de Estabelecimentos de Saúde (CNES), e as tabelas de malha de cidades e estados. Finalmente, após a agregação no conjunto geral da trilha, os dados resultantes são carregados na camada de consumo do *MinIO* e enviados para o banco de dados disponibilizado pelo projeto.

5.2.2 Tratamento dos dados para trilha de Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)

A trilha referente a Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP) utiliza bases de dados relativas a notas fiscais eletrônicas, procedimentos SIGTAP realizados e Fornecedores. A ideia por trás disso é possibilitar a visualização dos indícios indicados nas métricas, por meio da integração com as bases de geolocalização do Instituto Brasileiro de Geografia e Estatística (IBGE).

A metodologia de pré-processamento e análise exploratória dos dados seguiu a mesma lógica presente na trilha de *Stents* recorrentes, buscando identificar valores que pudessem ser cruzados entre as bases de dados adotadas para desenvolvimento da trilha respeitando seu propósito final. Com os dados devidamente formatados e com a identificação de colunas de comportamento semelhante entre os conjuntos de dados explorados, tornou-se viável partir para

o carregamento das bases devidamente formatadas no banco de dados disponibilizado pelo projeto FiscalizaSUS. Na Figura 11, a seguir, está disponibilizado o modelo de fluxograma adotado para o tratamento de dados nas bases de referentes às trilhas de auditoria de Órteses, Próteses e Materiais Especiais (OPME).

Figura 11 – Fluxograma de trabalho para estruturação de dados



Fonte: elaborada pelo autor, 2025.

Por meio desse modelo, é possível fornecer uma lógica de trabalho baseada em etapas cíclicas. Isso também foi viabilizado por meio das ferramentas disponíveis na arquitetura de *Big Data* do projeto FiscalizaSUS. Inicialmente, é feita a coleta e o armazenamento dos dados no *MinIO*. O próximo passo consiste em conduzir uma análise exploratória e qualitativa do material adquirido. Consequentemente, é verificado se os elementos já estão corretamente alinhados e depois são transferidos para a camada do banco de dados por meio das *DAGs* presentes na camada de processamento do *Apache Airflow*. Em seguida, os dados são carregados no banco de dados, e, assim, é possível partir para as próximas etapas de estratificação e dicionarização do modelo de dados.

As tabelas podem ser consumidas no ecossistema Apache Superset, em que são desenvolvidas as visualizações após a aplicação de métricas e análises de vínculo. Um ponto importante é a questão da dicionarização do banco de dados, que será discutida na seção 6, que também constitui um fator importante para esta dissertação do ponto de vista organizacional de tabelas adquiridas e tratadas.

5.3 ESTRATIFICAÇÃO DOS MODELOS DE DADOS

Em decorrência das etapas finalizadas anteriormente, foi preciso avaliar a qualificação dos dados para um modelo relacional de dados. Por meio dessa lógica, abre-se a possibilidade de integrar diferentes elementos para uma mesma trilha de auditoria. Baseando-se nisso, foi preciso refletir a respeito da disposição das bases de dados por meio de um modelo de dados estratificado. Assim, foi adotado o esquema estrela como metodologia para integração entre as bases. Uma tabela de fato pode ser identificada como o foco central presente para a regra de negócio, enquanto as tabelas de dimensões são dispositivos auxiliares que contemplam registros complementares para o fato avaliado (Colaço Júnior, 2018).

Nesse contexto, dispondo do ambiente pgAdmin na arquitetura do projeto FiscalizaSUS, foram estabelecidos *schemas* para particionar as informações adquiridas conforme sua utilidade nas trilhas de auditoria. É importante mencionar também que as tabelas estão divididas em diferentes *schemas*, pois algumas pertencem ao uso geral de outras trilhas do FiscalizaSUS e as demais são referentes a tarefas específicas. Assim, foram construídas divisões a partir de duas nomenclaturas, *public* e *sociedade*, salvo a exceção da tabela referente a Notas Fiscais Eletrônicas (NFE) que pertence ao schema de *opme* e está agregada à base de fornecedores que pertence às trilhas de auditoria discutidas anteriormente.

O modelo de dados utilizado consiste em dados já adquiridos pelo projeto FiscalizaSUS, além dos conjuntos presentes nas tabelas de pessoa jurídica e pagamentos de Bolsa Família. A partir da qualificação dos dados utilizados para as trilhas de auditoria, foi possível desenvolver uma estratificação dos dados coletados e organizá-los orientados a um esquema estrela. Isso possibilitou a comunicação entre diferentes tabelas a fim de promover análise de vínculos e conexões entre as partes. Dessa forma, torna-se viável identificar quanto um fornecedor teria recebido em uma transação envolvendo Órteses, Próteses e Materiais Especiais (OPME) devido à conexão com a tabela responsável por armazenar esses dados. Finalmente, isso proporciona a ligação entre diferentes fontes de dados, mais liberdade para a identificação de vínculos entre as bases e identificação de anomalias.

5.4 DICIONARIZAÇÃO DE DADOS

Um ponto pertinente para o processo metodológico desta dissertação gira em torno da dicionarização dos dados coletados ao longo da implementação do projeto FiscalizaSUS. Isso foi um fator importante para a verificação de valores pertinentes nas bases analisadas,

identificando qual classificação seria atribuída a cada um dos valores. Além disso, serve como uma base documental estratificada em comparação com os dados adquiridos das fontes originais.

O primeiro passo foi avaliar a tipificação de cada elemento presente nas colunas. Para dados descritivos, como CPF, CNPJ e nomes, de maneira geral, foram atribuídos o tipo *character varying*, enquanto elementos voltados para identificadores e valores monetários receberam, respectivamente, o tipo *integer* e *decimal*. Nas fontes de dados do IBGE, observaram-se componentes de geolocalização, além dos nomes de municípios e estados. Essa coluna, em específico, consiste em um atributo em formato *json* contendo coordenadas geográficas da localidade. Em alguns casos, foram utilizadas *flags* do tipo *bit* para verificação de vínculos e questões relacionadas a sociedades vigentes dos fornecedores de OPME, e de pessoas jurídicas no geral. Esses passos foram fundamentais para o processo de documentação das bases de dados, após todo o trabalho de qualificação, limpeza e carregamento.

Em seguida, é importante comentar que os dados referentes ao Bolsa Família foram distribuídos para outras tabelas de interesse, com dados referentes ao tipo de auxílio, ao sócio e ao tempo. Durante o tratamento e a análise, foram feitos cruzamentos entre essas informações para alimentar as tabelas, de fato, no banco de dados. Ademais, serão observados os resultados oriundos do processo metodológico apresentado nesta seção, e a construção das tabelas presentes no dicionário de dados. A qualificação e a estratificação dos dados possibilitaram diversas perspectivas para o projeto FiscalizaSUS no campo de OPME. Os dados foram agregados e digitalizados viabilizando sua consulta por meio de um esquema estrela e da visualização de indícios pelo ambiente *Apache Superset*. Finalmente, toda essa integração também é refletida nas trilhas de OPME abordadas nesta dissertação. Na seção a seguir, apresentaremos os resultados desta pesquisa.

5.5 ESTUDO DE CASO

Nesta subseção, será feita a descrição das atividades técnicas para o processo de qualificação das bases de dados utilizadas. Desse modo, analisou-se qualitativamente a proposta metodológica adotada. Esta etapa do trabalho consistiu em um Estudo de Caso (Yin, 2015) para avaliar as trilhas de auditoria do projeto FiscalizaSUS, associadas a vínculos entre fornecedores de OPME, médicos que atuam no SUS e beneficiários do programa Bolsa Família. O intuito

principal foi discutir a metodologia empregada para avaliar a construção das trilhas e a apresentação dos indícios ao nível visual. Ademais, foram convidados 3 colaboradores do projeto FiscalizaSUS com experiência na área de auditoria em saúde para avaliar as trilhas apresentadas.

5.5.1 Elementos a ser criticados

Para identificar e avaliar a assertividade das trilhas de auditoria mencionadas anteriormente nesta seção, foram separados alguns tópicos basilares, quais sejam:

- A trilha atende ao requisito principal (identificação de vínculo)?
- A camada visual está compatível com os indícios apresentados?
- O que pode ser melhorado na trilha de auditoria?

Foi imprescindível receber a devolutiva dos colaboradores mediante a clareza do requisito principal desses mecanismos orientados mediante a arquitetura de dados do projeto FiscalizaSUS. Para isso, avaliou-se o modelo relacional proposto e se ele seria capaz de identificar o vínculo de um mesmo indivíduo presente nas bases de interesse para as trilhas. A partir disso, este modelo seria encaminhado para consumo na camada visual do projeto, *Apache Superset*, de modo que o destaque do indício encontrado pudesse ser avaliado. Dessa forma, tanto o processo de integração de bases de dados como a visualização de irregularidades foram focos para melhorias futuras do projeto. Além disso, a visão desses colaboradores pode contribuir para o aperfeiçoamento desta pesquisa, a melhoria dos processos desencadeados ao longo do projeto FiscalizaSUS e sugestões de abordagens que possam proporcionar resultados mais otimizados que os atuais.

5.5.2 Resultados do Estudo de Caso Qualitativo

Conforme discutido, foi implementado um Estudo de Caso qualitativo para avaliar a construção das trilhas de auditoria associadas a vínculos anômalos, envolvendo fornecedores de OPME, médicos e pessoas cadastradas no programa Bolsa Família. A partir disso, alguns pontos puderam ser destacados:

- Os painéis são capazes de cumprir seu propósito basilar – eles foram construídos com o intuito de destacar vínculos irregulares entre diferentes bases de dados, por meio de elementos-chave entre as tabelas.

- Os painéis podem aprofundar seu grau de complexidade – o primeiro avaliador sugeriu que fossem apresentados os tipos de OPME que o fornecedor em questão comercializa de forma individual.
- O segundo avaliador sugeriu a padronização das nomenclaturas utilizadas nas tabelas que apresentavam os indícios de teste, e o mesmo procedimento para os gráficos referentes à localização geográfica. Também foi solicitada a apresentação de datas referentes a períodos de recebimento e início de vínculos societários.
- O terceiro avaliador sugeriu a utilização de mais indícios geográficos. Além disso, novas bases de dados podem ser agregadas ao modelo de dados relacional atual a fim de obterem-se novas perspectivas para análise mediante questões de localização.
- Outra sugestão foi apresentar mais indícios de teste para replicação de filtros no *Apache Superset*. Ademais, foi sugerida a apresentação dos tipos de OPME representados no teste.

5.5.3 Discussão do Estudo de Caso

As colocações dos colaboradores convidados foram pertinentes para reflexão acerca do desenvolvimento do trabalho e abertura para perspectivas futuras. Tendo em vista os retornos obtidos a partir da linha de raciocínio dos avaliadores, alguns pontos podem ser destacados como considerações:

- Adaptação das trilhas de auditoria para apresentação específica das OPME sendo comercializadas por um fornecedor identificado como anômalo.
- A padronização dos campos presentes em gráficos e tabelas foi acatada.
- A representação de mais indícios simulados pode não ser tão pertinente para a validação do modelo de dados. Em contrapartida, filtros associados ao tipo de produto apresentado no vínculo podem ser úteis.
- A apresentação de mais componentes gráficos, com informações geográficas, pode não ser tão adequada para um cenário de testes, como é o atual.
- A necessidade de avaliação da adição de campos de interesse para a identificação dos vínculos.

Em suma, o Estudo de Caso qualitativo descrito nesta etapa do trabalho serviu para melhorias futuras e aperfeiçoamento das trilhas de auditorias integradas ao projeto FiscalizaSUS.

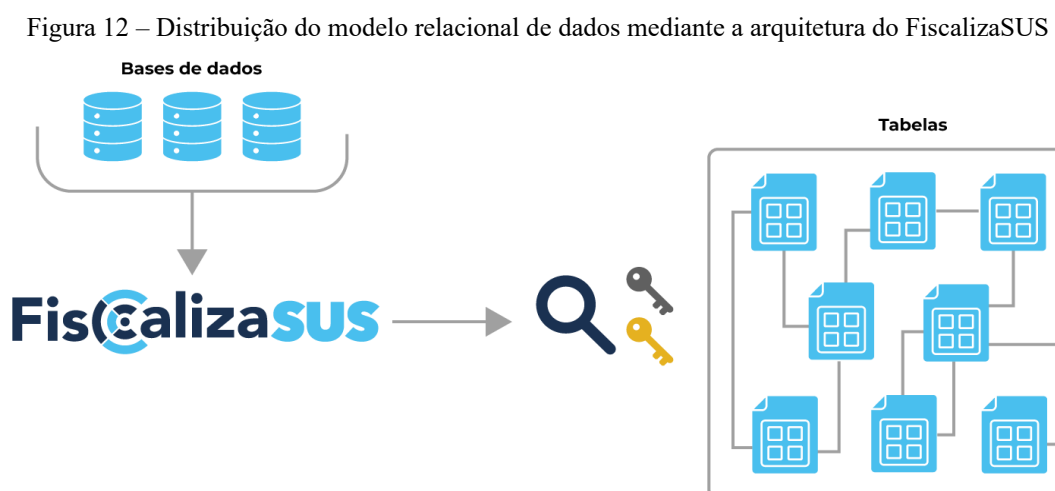
6 RESULTADOS

O intuito desta seção é apresentar quais resultados foram alcançados ao longo da construção das trilhas de auditoria referentes a Órteses, Próteses e Materiais Especiais (OPME) conduzidas pelo projeto FiscalizaSUS. Nessa direção, serão apontadas as repercussões que essas atividades desempenharam para o projeto, assim como o atendimento às questões de pesquisa propostas no trabalho.

6.1 AGREGAÇÃO DE DIVERSAS BASES DE DADOS EM ESQUEMA ESTRELA

As trilhas de auditoria do projeto FiscalizaSUS, no âmbito de OPMEs, contemplam diversos elementos de interesse, como fornecedores, pessoas jurídicas, notas fiscais, médicos que atendem no Sistema Único de Saúde, beneficiários do programa Bolsa Família, entre outras bases de dados pertinentes para o cumprimento dos objetivos firmados. Uma das questões técnicas previstas para a consulta das tabelas de forma agregada é justamente a eficiência mediante os quesitos de armazenamento e consulta posterior dos dados na camada visual do projeto disponibilizada por meio do ecossistema Apache Superset.

Desse modo, foi elaborado um esquema estrela no sistema de banco de dados do FiscalizaSUS, consistindo na presença de tabelas de fato e tabelas de dimensão. O primeiro grupo contém dados de nível numérico, como os identificadores de tabelas de dimensão que precisam estar conectados a determinada tabela de fato; e no âmbito de dimensão estão os registros descritivos contidos nas bases de dados, como nome, CNPJ e outras informações. A Figura 12 ilustra a lógica a respeito das relações entre os entes selecionados.



Fonte: elaborada pelo autor, 2025.

Conforme apresentado na metodologia desta dissertação, os dados são coletados e passados para a arquitetura do FiscalizaSUS, na qual ocorre todo o processo de tratamento e carregamento. Após esse ciclo de etapas, esses materiais são transformados em tabelas, fato ou dimensão, estabelecendo conexões com chaves identificadoras de outras tabelas pertinentes para o mesmo conjunto de trilhas de auditoria. A lógica por trás das conexões pode ser compreendida por meio de situações pertinentes, como casos em que um mesmo fornecedor possua um vínculo societário e de funcionário público que atende no Sistema Único de Saúde (SUS), ou esteja cadastrado como beneficiário do programa Bolsa Família. Com base nessas possibilidades, as tabelas utilizadas no modelo foram desenvolvidas. Por meio disso, é estabelecida a integração entre as tabelas utilizadas pelo projeto no âmbito de OPME.

Com base nisso, também foi importante partir para a dicionarização do conjunto de dados utilizado, estabelecendo a função e o tipo de cada coluna. Além disso, foi executado um teste volumétrico com alguns registros a fim de validar a funcionalidade do modelo relacional construído a partir da integração das bases de dados. Em seguida, serão também apresentados, de forma gráfica, os resultados obtidos no ambiente de homologação do projeto FiscalizaSUS para as trilhas de auditoria referentes a procedimentos SIGTAP e *Stents* recorrentes. Por fim, o modelo de dados pode ser visualizado na Figura 19 do Apêndice A, de forma mais detalhada.

6.2 DICIONARIZAÇÃO E VALIDAÇÃO DO MODELO ESTRATIFICADO PARA AS TRILHAS QUE ENVOLVEM AS BASES DE BOLSA FAMÍLIA E PESSOA JURÍDICA

Com todos os dados adquiridos e adicionados ao modelo estratificado, torna-se necessário dicionarizar os dados utilizados para a condução das trilhas de auditoria mediante a problemática de Órteses, Próteses e Materiais Especiais (OPME). Para os dados adquiridos por meio do portal www.gov.br, o dicionário de dados presentes para as bases foi utilizado como referência para a atribuição de alguns elementos que foram preservados para o dicionário final nesse escopo de trilhas voltadas à OPME. Desse modo, seguem listados, no Quadro 10, os dados referentes a cada componente das trilhas de auditoria, contendo seus respectivos elementos e descrições.

Quadro 10 – Dimensão sócio

Coluna	Descrição	Tipo
id	Identificador da tabela de dimensão sócio	INTEIRO
CNPJ básico	Número base de inscrição do Cadastro de Pessoa Jurídica (CNPJ)	CHARACTER VARYING
CPF	Número base do Cadastro de Pessoa Física (CPF)	CHARACTER VARYING
nome	Nome da pessoa física cadastrada como sócio	CHARACTER VARYING
flag_prog_auxilio	Usada para auxiliar na identificação de sócios que possuem vínculo com o programa de transferência de renda do governo federal.	INTEIRO
flag_func_publico	Usada para auxiliar na identificação de sócios que são cadastrados como funcionários públicos	INTEIRO

Fonte: elaborado pelo autor, 2025

A Dimensão sócio armazena dados descritivos a respeito de uma pessoa jurídica oriunda das bases de dados públicas utilizadas. Os principais dados podem ser compreendidos como o CNPJ básico, CPF e nome. Esses campos são fundamentais para o processo de identificação de vínculos com outras bases de dados utilizadas neste trabalho, como médicos, funcionários públicos, fornecedores de OPME e beneficiários do programa Bolsa Família.

Outro componente importante está na dimensão de auxílio. Esse aspecto do trabalho armazena informações utilizadas para identificação de qual tipo de auxílio está vinculado a um indivíduo. Conforme descrito na metodologia, foram utilizados dados do Bolsa Família, mas essa estrutura pode receber dados referentes a outros auxílios também. Isso pode provocar o incremento de novas bases de dados, como a estrutura referente ao “Novo Bolsa Família” para identificação de novos vínculos.

Quadro 11 – Dimensão Auxílio

Coluna	Descrição	Tipo
id	id referente a tabela	INTEIRO
nome_auxilio	Nome do tipo de auxílio que um indivíduo pode receber	CHARACTER VARYING

Fonte: elaborado pelo autor, 2025

Já a Dimensão Funcionário Público diz respeito a dados descritivos a respeito de médicos que atendem no Sistema Único de Saúde. O Quadro 12 apresenta elementos pertinentes ao cruzamento com as demais partes do banco de dados. Observa-se nele a presença de nome, CPF e CNPJ dos funcionários armazenados. Portanto, essa é mais uma dimensão qualificada pelo processo de limpeza e transformação de dados que se torna útil para o projeto.

Quadro 12 – Dimensão Funcionário Público

Coluna	Descrição	Tipo
id	id referente a tabela	INTEIRO
nome_funcionario	Nome do funcionário público	CHARACTER VARYING
cpf_funcionario	Número base do Cadastro de Pessoa Física (CPF), seguindo o formato aplicado nas tabelas de sócio e Bolsa Família	CHARACTER VARYING
cnpj_funcionario	Número base de inscrição do Cadastro de Pessoa Jurídica (CNPJ), seguindo o formato aplicado nas tabelas de sócio e bolsa família	CHARACTER VARYING

Fonte: elaborado pelo autor, 2025

O Quadro 13, Fato Sociedade, é resultado do cruzamento entre sócios, fornecedores e funcionários públicos. Após o cruzamento entre as 3 dimensões citadas, os identificadores são armazenados no fato. Além disso, há também a análise temporal para as datas de início e fim, assim como uma flag indicando se a sociedade ainda está vigente.

Quadro 13 – Fato Sociedade

Coluna	Descrição	Tipo
id_socio	Identificador de sócio	INTEIRO
id_fornecedor	Identificador de fornecedor	INTEIRO
id_funcionario público	Identificador de funcionário público	INTEIRO
data_inicio	Data de início da sociedade	DATE
data_fim	Data de fim da sociedade	DATE
flag_corrente	Sociedade vigente	INTEIRO

Fonte: elaborado pelo autor, 2025

Ao mesmo tempo, o Quadro 14, Fato Bolsa Família, gira em torno de um raciocínio parecido. Ele armazena o valor de parcela que um indivíduo teria recebido em um momento também identificado no cruzamento entre os dados pertinentes. Além disso, guarda o identificador de pessoa jurídica e o tipo de auxílio que aquele sócio recebeu.

Quadro 14 – Fato Bolsa Família

Coluna	Descrição	Tipo
valor parcela	Valor da parcela do programa social recebida	DECIMAL
id_auxilio	Identificador do auxílio recebido	INTEIRO
id_socio	Identificador do sócio	INTEIRO
id_tempo	Identificador de tempo	INTEIRO

Fonte: elaborado pelo autor, 2025

A Dimensão Fornecedor (Quadro 15) foi um dado de suma importância para este trabalho. A partir disso, foi possível identificar dados-chave para os cruzamentos entre os demais componentes, com o campo de CNPJ. Além disso, ela possui outros dados descritivos referentes ao indivíduo em processo de análise, como: endereço, latitude, longitude, CEP e nome fantasia.

Quadro 15 – Dimensão Fornecedor

Coluna	Descrição	Tipo
idfornecedor	identificador da tabela	INTEIRO
CNPJ	CNPJ do fornecedor	CHARACTER VARYING
nome	nome do fornecedor	CHARACTER VARYING
endereço	endereço do fornecedor	CHARACTER VARYING
data_inicio	data inicial	TIPO DATA
data_fim	data final	TIPO DATA
flag_endereco_suspeito	identificador para endereço suspeito	BIT
flag_corrente	identificador para fornecedor ativo	BIT
nível	nível do fornecedor	CHARACTER
CEP	CEP do fornecedor	INTEIRO
lat	latitude do fornecedor	DOUBLE PRECISION
lng	longitude do fornecedor	DOUBLE PRECISION
idestado	chave estrangeira com a dimensão estado	INTEIRO
idcidade	chave estrangeira com a dimensão de cidade	INTEIRO
nome_fantasia	nome fantasia do fornecedor	CHARACTER VARYING

Fonte: elaborado pelo autor, 2025

Por fim, é possível sinalizar o vínculo com as dimensões de malha estado e malha cidade, que também viabilizam o processo de localização geográfica nos cruzamentos entre as bases. A seguir, observam-se os Quadros 16 e 17, respectivamente, de Dimensão Cidade e Dimensão Estado.

Quadro 16 – Dimensão Cidade

Coluna	Descrição	Tipo
idcidade	identificador da tabela	INTEIRO
idestado	chave estrangeira da tabela de dimensão estado	INTEIRO
cod_municipio	código do município	CHARACTER VARYING
nome	nome do município	CHARACTER VARYING
geojson	coluna responsável por armazenar o localizador geográfico do município	JSON
populacao_cidade	população do município	INTEIRO
populacao_estado	população do estado	INTEIRO
multiplicador_estado		DOUBLE PRECISION

Fonte: elaborado pelo autor, 2025

Quadro 17 – Dimensão Estado

Coluna	Descrição	Tipo
idestado	identificador da tabela estado	INTEIRO
nome	nome do estado	CHARACTER VARYING
cod_uf	código do estado	CHARACTER VARYING
uf	sigla do estado	CHARACTER VARYING
regiao	região em que o estado se encontra	CHARACTER VARYING
geojson	geojson referente a posição geográfica do estado	JSON
populacao	população do estado	INTEIRO
multiplicador_pais		DOUBLE PRECISION

Fonte: elaborado pelo autor, 2025

Esses dados servem como auxiliares na construção de outros quadros de dimensão e, posteriormente, para a localização dos indícios detectados em nível geográfico. Por meio deles, é possível adquirir pelo menos duas perspectivas: a primeira seria de nível municipal e a segunda de nível estadual. Os nomes de cada cidade e cada município são cruzados com as

tabelas contidas no projeto e seus respectivos identificadores armazenados, facilitando o processo de integração entre elementos de diferentes funções.

Para armazenamento de dados temporais, foi feita uma dimensão própria para isso. Na Dimensão Tempo (Quadro 18) são inseridos valores desde o dia da semana ao ano. O objetivo é facilitar a interação com múltiplas(os) tabelas/quadros que contêm informações temporais e o vínculo desses valores em tabelas/quadros de fato. Por meio desse quadro, são identificadas as relações de datas analisadas nas trilhas de auditoria, sobretudo as focadas em Órteses, Próteses e Materiais Especiais.

Quadro 18 – Dimensão tempo

Coluna	Descrição	Tipo
idtempo	identificador da tabela tempo	INTEIRO
diadasemana	Dia da semana escrito por extenso	CHARACTER VARYING
numero_do_dia	número do dia da semana	INTEIRO
data	data referenciada	DATE
diautil	identificador de dia útil	BIT
periodo	período referente a mês e ano	CHARACTER VARYING
ano	ano descrito na tabela	INTEIRO
desc_mes	descrição do mês	CHARACTER VARYING
mes_abreviado	mês utilizado de forma abreviada	CHARACTER VARYING
numero_do_mes	número do mês	INTEIRO
desc_semestre	descrição do semestre	CHARACTER VARYING
nível	nível da relação temporal	CHARACTER VARYING

Fonte: elaborado pelo autor, 2025

Por último, o Quadro 19 (de Fato Notas), composta por elementos de diversas bases contidas no projeto FiscalizaSUS. Assim como os demais quadros de fato utilizados, ele recebe identificadores de outras partes do banco de dados e os armazena. O intuito desse quadro no trabalho é vincular uma nota fiscal eletrônica emitida a um fornecedor e procedimento. Essa lógica viabiliza a construção das trilhas de auditoria apresentadas na metodologia desta dissertação.

Quadro 19 – Fato Notas

Coluna	Descrição	Tipo
id_fatonotas	identificador da tabela	INTEIRO
idprocedimento	identificador do procedimento [chave estrangeira]	INTEIRO
idunidade	identificador da unidade de saúde [chave estrangeira]	INTEIRO
iddata_emissao	identificador da data de emissão [chave estrangeira]	INTEIRO
idopme	identificador do opme [chave estrangeira]	INTEIRO
idfornecedor	identificador do fornecedor [chave estrangeira]	INTEIRO
idorigemnota	identificador de origem da nota fiscal [chave estrangeira]	INTEIRO
nrnota	número da nota	CHARACTER VARYING
ean	identificador de código de barras	CHARACTER VARYING
nem	código do produto	CHARACTER VARYING
data_de_validade	data de validade do produto	DATE
descricao	descrição do produto	CHARACTER VARYING
qtd	quantidade do produto vendida	INTEIRO
unidade_de_medida	unidade de medida adotada	STRING
valor_unitario	valor da produto individualmente	DECIMAL
valor_total_produto	valor total dos produtos vendidos	DECIMAL
valor_total_nota	valor total presente na nota	DECIMAL
flag_anomalia_valor	identificador de anomalia no valor perante classificação	INTEIRO
qtd_dias_vencidos	cálculo de dias vencidos	INTEIRO

flag_anomalia_vencimento	identificador de vencimento	INTEIRO
codigo_produto	código identificador de produto	INTEIRO
data_de_fabricacao	data de fabricação do produto	DATE
qtd_lote	quantidade de OPMEs no lote	FLOAT
desvio_valor	cálculo de desvio padrão no valor em comparação com a amostra	FLOAT
tipo_opme	tipo de opme	CHARACTER VARYING
tipo_classe	tipo de classe	CHARACTER VARYING

Fonte: elaborado pelo autor, 2025

Com os dados dicionarizados e orientados a um esquema estrela, foi possível agregar os diversos conjuntos de tabelas em um único modelo relacional. Isso permitiu averiguar a veracidade do modelo de dados construído para a realização de consultas nas bases de dados, e, conseqüentemente, para a criação de visualizações por meio do ecossistema Apache Superset, disponibilizado pelo projeto FiscalizaSUS.

Além disso, é possível acrescentar à discussão o teste volumétrico realizado com o modelo relacional de dados referente à trilha de auditoria envolvendo o programa de transferência de renda Bolsa Família. Esse teste foi utilizado como prova do funcionamento do modelo de dados com base em um esquema estrela. A aplicação foi desenvolvida por meio de um indício teste que simulava o vínculo entre as bases de fornecedor, o Bolsa Família e os sócios. Assim, foi possível constatar que a conjuntura do modelo de dados para a trilha de auditoria encontrava-se em estado adequado para a identificação de anomalias reais mediante a presença de vínculos entre os atores destacados para a pesquisa. O resultado do modelo de dados pode ser observado na Figura 13, a seguir.

Figura 13 – Teste do modelo relacional para Bolsa Família e OPME



Fonte: elaborada pelos pesquisadores do FiscalizaSUS, 2025.

6.3 IDENTIFICAÇÃO DE INDÍCIOS NA TRILHA DE PROCEDIMENTOS SIGTAP

Partindo do processo de agregação das bases de dados referentes à trilha de auditoria para Procedimentos Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP), foi possível desenvolver gráficos no ecossistema Apache Superset que demonstram a localização geográfica, graças também à integração com as bases oriundas da API do Instituto Brasileiro de Geografia e Estatística (IBGE), de locais onde foram encontrados inconformidades.

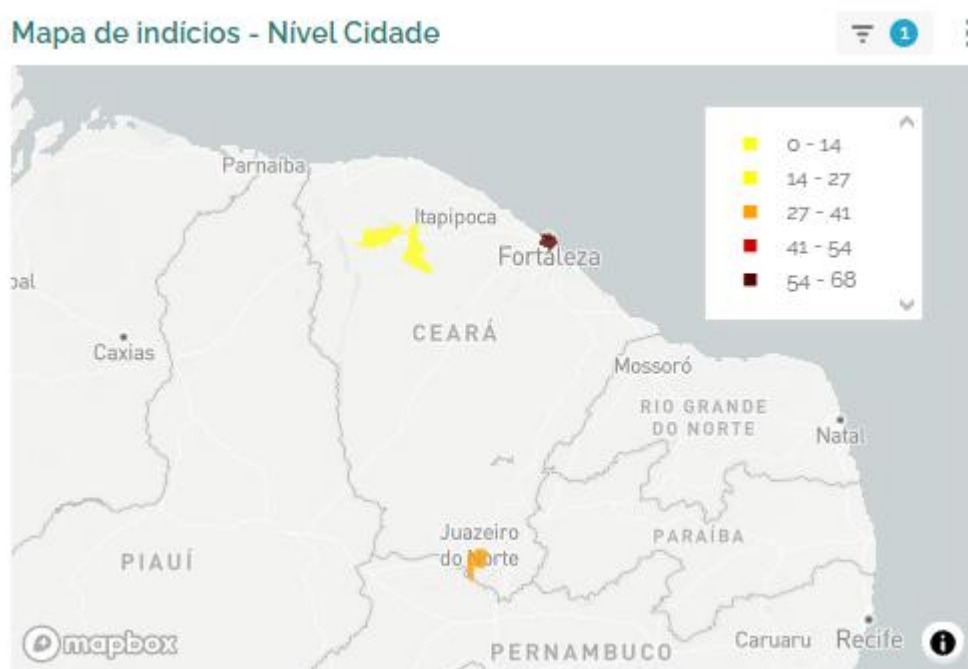
Respectivamente, as Figuras 14 e 15 demonstram essa identificação em dois níveis de complexidade diferentes – estadual e municipal. Nas Figuras, estão em destaque o estado do Ceará e alguns de seus municípios.

Figura 14 – Índícios em procedimentos SIGTAP (Nível-Estadual)



Fonte: elaborada pelos pesquisadores do projeto FiscalizaSUS, 2025

Figura 15 – Índícios em procedimentos SIGTAP (Nível-Municipal)



Fonte: desenvolvida pelos pesquisadores do projeto FiscalizaSUS, 2025

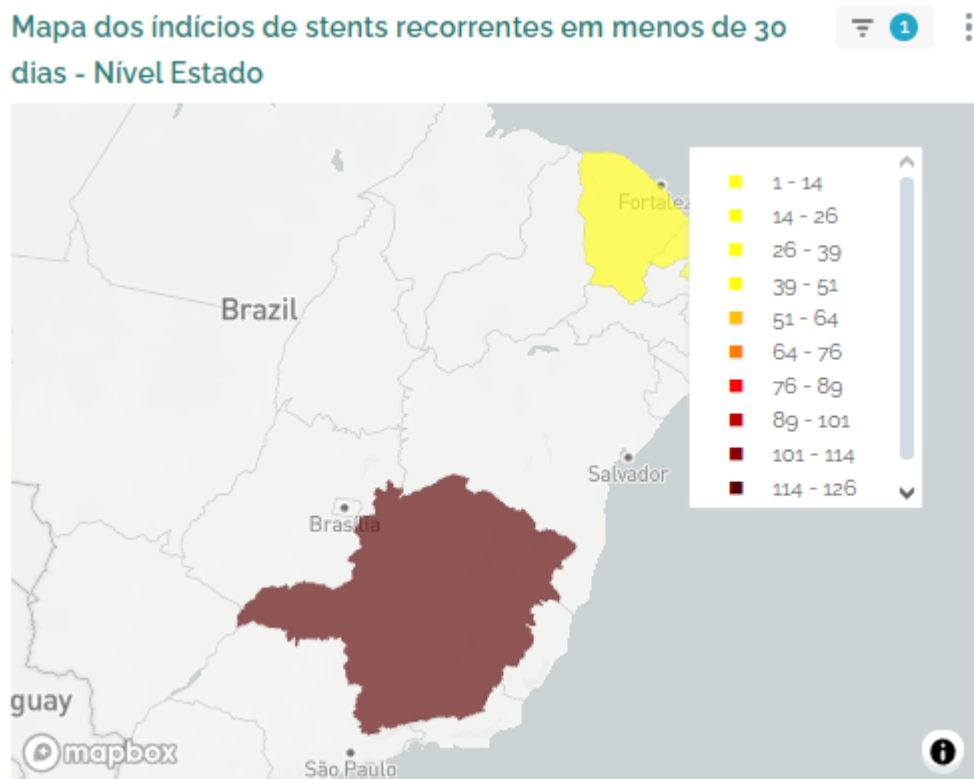
Com base nessas figuras, pode-se dizer que a integração de diferentes fontes de dados e a aplicação de métricas que envolvem o uso de técnicas estatísticas para verificação de valores discrepantes, como foi o caso da utilização do método de Tukey, possibilita a identificação de indícios na trilha de auditoria inserida no âmbito de Órteses, Próteses e Materiais Especiais

(OPME). Partindo destas premissas, também foi pertinente para a pesquisa aplicar um processo de identificação de indícios similar para procedimentos relacionados à trilha de auditoria em *Stents* recorrentes.

6.4 IDENTIFICAÇÃO DE INDÍCIOS NA TRILHA DE *STENTS* RECORRENTES

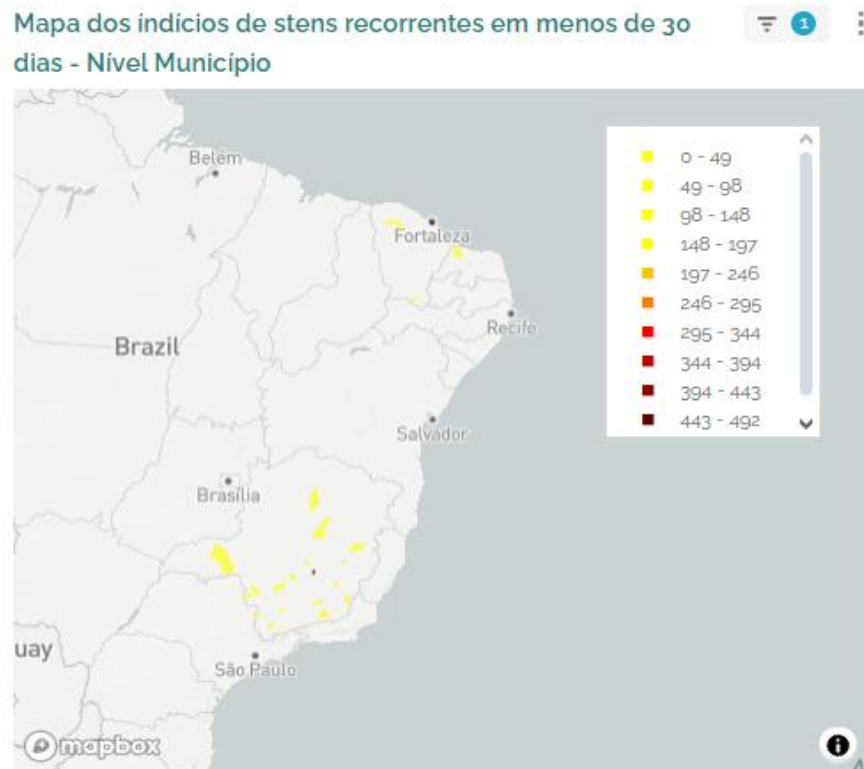
Nas trilhas de *Stents* recorrentes, foram empregadas bases relativas à Autorização de Internação Hospitalar (AIH), Cadastro Nacional de Estabelecimentos de Saúde (CNES) e Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP) para o mapeamento de ocorrências anômalas referentes à aplicação indevida desses procedimentos em períodos de 30 dias. As Figuras 16,17 e 18, a seguir, demonstram situações suspeitas respectivamente, em nível estadual, municipal e de estabelecimento de saúde.

Figura 16 – Indícios em procedimentos com *Stents* recorrentes (nível Estadual)



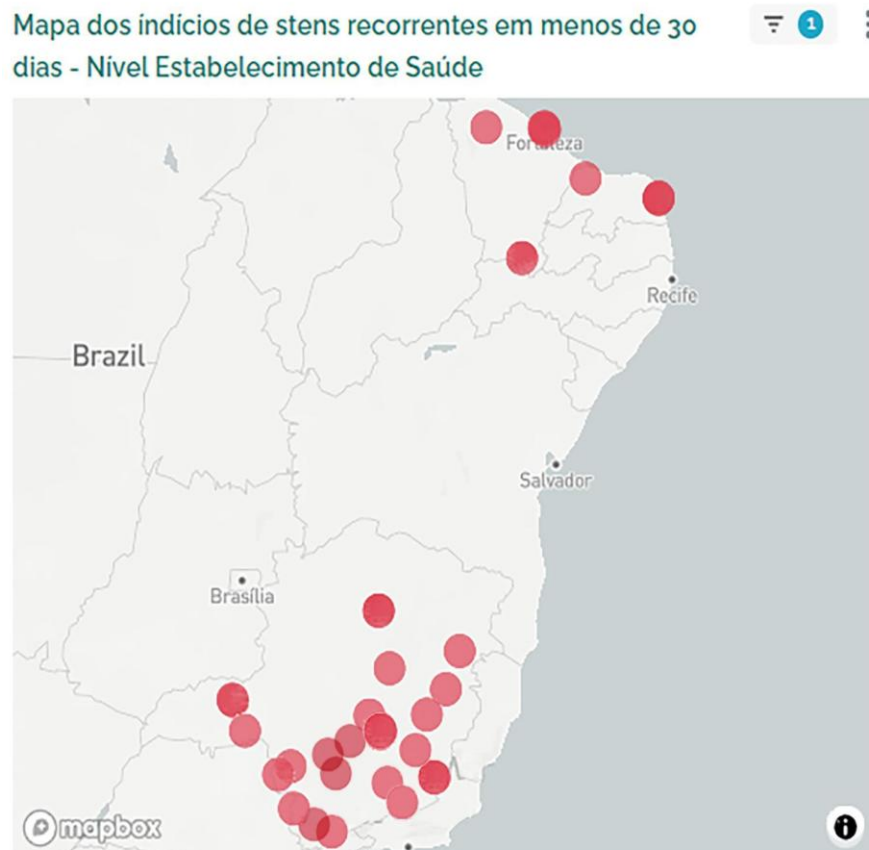
Fonte: desenvolvida pelos pesquisadores do projeto FiscalizaSUS, 2025

Figura 17 – Índios em procedimentos com *Stents* recorrentes (nível Municipal)



Fonte: desenvolvida pelos pesquisadores do projeto FiscalizaSUS, 2025

Figura 18 – Índios em procedimentos com *Stents* recorrentes (Nível Estabelecimento de Saúde) -



Fonte: desenvolvida pelos pesquisadores do projeto FiscalizaSUS, 2025

Como observado para a trilha de auditoria anterior, foi possível construir, por meio da integração de diferentes bases de dados, uma visualização capaz de apresentar, de modo geográfico, a presença de irregularidades para a regra de negócio vigente, tendo em vista o contexto de procedimentos irregulares com *Stents*. Assim, vale discutir o processo de integração entre as trilhas de auditoria desenvolvidas neste trabalho, sobretudo do ponto de vista de vínculos entre indivíduos que podem ser localizados em mais de um cenário.

6.5 RELACIONAMENTO ENTRE AS TRILHAS DE AUDITORIA DE ÓRTESES, PRÓTESES E MATERIAIS ESPECIAIS

Como observado anteriormente, as trilhas de auditoria relativas a Órteses, Próteses e Materiais Especiais (OPME) do projeto FiscalizaSUS possuem objetivos singulares e diferentes aspectos metodológicos para o desenvolvimento da apresentação dos dados. Porém, ao mesmo tempo que são particulares em até certo nível, também é possível identificar a relação entre cada uma delas. Conforme apresentado na metodologia do trabalho, as trilhas de auditoria em OPME consomem algumas bases que são comuns entre si, sobretudo do ponto de vista de fornecedores e notas fiscais, que são objetos de interesse para análise.

Isso significa que um mesmo fornecedor, que está sob suspeita de ter cometido uma irregularidade em um procedimento envolvendo *Stents*, pode também ser identificado como beneficiário do programa de transferência de renda, Bolsa Família, e/ou até mesmo como um médico que atende no Sistema Único de Saúde (SUS). Tal situação pode ocorrer devido aos modelos relacionais criados entre os quadros de Dimensão e Fato, responsáveis por cruzar informações de diversas bases de dados e identificar esses possíveis vínculos em diversas instâncias. Portanto, esse é mais um elemento que reforça a integração entre as diversas bases de dados utilizadas ao longo da construção das trilhas de auditoria de OPME no âmbito do projeto FiscalizaSUS. Na próxima seção, procederemos à análise dos dados.

7 DISCUSSÃO

Nesta seção, discutiremos os dados referentes ao mapeamento da literatura, à agregação de informações, à construção das trilhas do projeto, de modo a apontar seus limites e perspectivas.

7.1 CONSIDERAÇÕES DO MAPEAMENTO SISTEMÁTICO DA LITERATURA

Um dos objetivos propostos neste trabalho foi a condução de um Mapeamento Sistemático da Literatura baseado no protocolo PICO. A meta principal da seção 4 visava a identificar métodos, soluções, abordagens e contribuições para a área de auditoria em saúde. O processo metodológico para análise contou com etapas de inclusão e exclusão, observando fatores relacionados ao conteúdo apresentado e ao perfil das publicações coletadas. A partir disso, foi adquirido um espaço amostral reduzido em relação ao original, de 344 artigos, dos quais foram selecionados 40. Foi feito ainda um formulário de extração para a construção da seção, que visava a identificar questões, como: tipo de auditoria, métodos utilizados, contribuições, entre outros fatores. Por meio disso, foram identificados os seguintes fatores:

1. heterogeneidade de soluções e abordagens;
2. importância do tratamento de dados;
3. viabilidade da implementação de métodos estatísticos;
4. países e anos com mais publicações na área;
5. constatação de ser uma área que pode ser expandida e cada vez mais fomentada.

Nesta seção, também foi possível agregar conceitos pertinentes para o desenvolvimento metodológico desta dissertação de mestrado. Em primeiro lugar, observou-se a diversidade de metodologias para estudos relacionados à auditoria em saúde, bem como o uso de técnicas estatísticas em alguns casos. Além disso, tratamento, limpeza e formatação de dados foram um elemento-chave para o desenvolvimento de parte dos estudos. Tendo isso em vista, os próximos componentes do trabalho foram também baseados nos conteúdos avaliados ao longo do mapeamento e sobretudo na comparação com as publicações filtradas.

7.2 ANÁLISE DE BIG DATA

Ao longo do desenvolvimento deste trabalho, foi apresentado o contexto de Órteses, Próteses e Materiais Especiais (OPME) como um pilar do projeto FiscalizaSUS para

identificação de irregularidades e anomalias, sobretudo considerando relações de vínculo e relações transnacionais envolvendo esses insumos. Desse modo, os resultados para as trilhas de auditoria de OPME apresentados na seção anterior demonstram a possibilidade de integração de diversas bases de dados, de origem pública ou privada, para o reconhecimento de comportamentos anômalos mediante as informações adquiridas.

Nesse processo, a organização das soluções observadas depende integralmente do uso de ferramentas direcionadas para armazenamento, processamento e integração de grandes volumes de dados, e da aplicação de métodos estatísticos, como a técnica de Tukey, para interação com valores discrepantes nas amostras de dados. Esse raciocínio também permite realizar inferências a respeito da frequência com que os procedimentos envolvendo *Stents* ocorrem de maneira irregular e com valores transacionais exorbitantes para procedimentos ligados ao Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP).

Além disso, deve-se evidenciar o valor que o uso de uma arquitetura de referência para Big Data, fornecida pelo projeto FiscalizaSUS e as ferramentas nele presentes – como o *MinIO*, *Apache Airflow*, *Jupyter Lab* e *pgAdmin* – viabilizaram o tratamento, a transformação e o carregamento das bases de dados coletadas de forma escalável e resiliente. Esse fator também assegurou a modelagem das tabelas processadas seguindo a lógica de um esquema estrela, que, como descrito na metodologia desta dissertação, facilita o processo de consulta e interação com todas as bases necessárias devidamente agregadas por meio dos Quadros Fato e Dimensão. A aplicação dessas heurísticas, envolvendo grandes volumes de dados e análise estatística, foi fundamental para a elaboração das trilhas de auditoria em saúde ligadas à OPME.

7.3 AGREGAÇÃO DE BASES DE DADOS

Um fator primordial para a disponibilidade das trilhas de auditoria de Órteses, Próteses e Materiais Especiais (OPME) foi a agregação conjunta dos mais variados tipos de base de dados, sendo adquiridos por meio cooperações firmadas pelo projeto FiscalizaSUS com parceiros como o Ministério Público Federal do Rio Grande do Norte, e a aquisição de bases liberadas para acesso público, como é o caso para as tabelas referentes a beneficiários do programa de transferência de renda do governo federal, o Bolsa Família.

Ao longo da análise exploratória dos dados adquiridos, foi possível observar a presença de elementos em comum entre as tabelas, como colunas referentes a CNPJ e CPF. A partir disso, foi possível elaborar uma metodologia de integração entre as diversas tabelas

adquiridas pelo projeto FiscalizaSUS, reconhecendo a lógica de criação de tabela de fato e de dimensão, com seus respectivos campos de conexão.

Esses fatores permitem o entendimento a respeito de outro passo relevante para o processo de agregação de bases de dados com diferentes perfis. Em primeiro lugar, foram elaboradas tabelas de fato, constituindo valores numéricos, de cunho monetário ou identificadores das tabelas dimensões. Isso permite que as tabelas adotadas para a construção de uma trilha de auditoria em OPME estejam devidamente vinculadas ao propósito de identificação de múltiplos vínculos, anomalias mediante valores e até mesmo a frequência com que um procedimento médico está ocorrendo.

Desse modo, todos esses fatores proporcionam um ambiente escalável para a identificação de anomalias e irregularidades contidas no conjunto de dados resultantes a partir da agregação de várias(os) tabelas/quadros que compõem os elementos necessários para atender às regras de negócio vinculadas à trilha de auditoria específica.

7.4 ANÁLISE DAS TRILHAS DE AUDITORIA DO PROJETO

Neste trabalho, as trilhas de auditoria em OPME ligadas ao projeto FiscalizaSUS foram desenvolvidas em parceria com os auditores do DenaSUS. As motivações para desenvolver as atividades também possuem relação com notícias veiculadas na mídia tradicional, conforme destacadas na introdução desta dissertação, referente às trilhas envolvendo tanto o programa de transferência de renda do governo federal, o Bolsa Família, quanto procedimentos irregulares. Dessa forma, esse raciocínio possibilitou o desenvolvimento do percurso metodológico do trabalho, considerando todo o processo de limpeza e qualificação dos dados, assim como análises mais robustas com métodos estatísticos.

Nas trilhas focadas na identificação de vínculos entre fornecedores, médicos e funcionários públicos que atendem ao SUS, beneficiários do Bolsa Família e cadastro de pessoa jurídica, foi conduzido um processo de análise exploratória inicialmente, para, em seguida, identificar a pertinência dos dados. A partir disso, verificou-se a presença de valores que poderiam ser cruzados, como CPF, nome e CNPJ.

Já para as trilhas de procedimentos SIGTAP e *Stents*, a parte de vínculo também foi levada em consideração. Porém, o tópico principal da discussão atinge valores e periodicidade dos procedimentos médicos desempenhados. Nessa etapa, foi aplicado o método tradicional de Tukey, com o objetivo de identificar elementos discrepantes em relação à amostra total de

valores, utilizando um intervalo de confiança de 95%, como pôde ser observado na seção 2 desta dissertação.

Outro ponto relevante diz respeito à trilha de auditoria do projeto FiscalizaSUS que consome dados da base de médicos que atendem no SUS. A lógica por trás da construção desse mecanismo assemelha-se à trilha de Bolsa Família. Toda a sistemática de limpeza, análise exploratória dos dados e cruzamento de informações em comum entre as bases foi aproveitada. O tipo de indício que pode ser apresentado por essa trilha também foi ilustrado por meio da Figura 13 dos resultados, de modo que há um destaque para a posição geográfica do evento. Assim, foi possível perceber um aproveitamento de raciocínios e metodologias para agregação de dados a uma trilha de auditoria.

Além disso, é imprescindível destacar que todas essas trilhas de auditoria podem ser agregadas entre si. Mediante os elementos disponibilizados no dicionário de dados, pode-se avaliar a importância e as caracterizações das colunas. Observam-se a separação dos dados em elementos pertinentes para análise de vínculo, como CNPJ, CPF e Nome. Não obstante, também existem valores para facilitar as localizações geográficas dos fornecedores, a identificação de períodos temporais e os valores monetários irregulares obtidos a partir das notas fiscais de OPME. Dessa forma, as trilhas de auditoria mencionadas nesta dissertação de mestrado podem trabalhar de forma conjunta, considerando o papel dos vínculos e os cálculos de anomalias em períodos de execução de procedimentos médicos, considerando os custos discrepantes da amostra coletada para análise.

7.5 LIMITAÇÕES E PERSPECTIVAS FUTURAS

Ao longo da modelagem de dados e do processo de desenvolvimento das trilhas de auditoria, foram encontrados alguns impedimentos em relação ao acesso à infraestrutura fornecida pelo projeto devido a questões de volumetria de dados e acesso estável à rede. Desse modo, uma trilha de auditoria que estivesse lidando com uma volumetria maior do que aquilo que a estrutura fornecida anualmente pelo projeto FiscalizaSUS comportava, prejudicava o funcionamento e o acesso ao banco de dados do projeto. Devido a essa questão, foi feito um teste volumétrico com 200 registros e um elemento de teste para validar e testar o modelo relacional de dados desenvolvido para a trilha que envolve indivíduos que recebem o auxílio Bolsa Família. Além disso, também há a possibilidade de aquisição de novos dados a respeito de quadros societários, fornecedores, e demais componentes pertinentes para o desenvolvimento das trilhas de auditoria. Assim, as atividades do projeto no que diz respeito a

OPME poderão ser expandidas e aperfeiçoadas.

O projeto FiscalizaSUS está transferindo suas trilhas de auditoria para um novo ambiente fornecido pelo Ministério da Saúde (MS). Esse novo ambiente foi intitulado Plataforma de Auditoria e Monitoramento de Dados em Saúde (PAMDAS), que oferece uma nova arquitetura de *Big Data* para os pesquisadores. Por esse motivo, no presente momento, a camada de visualização ainda não se encontra disponível para acesso. Essa transferência de dados qualificados continua em execução, e, futuramente, poderão ser utilizados neste novo cenário. Portanto, isso também se mostra um fator que pode provocar novas possibilidades de pesquisa e aperfeiçoamento em relação ao que já foi produzido.

8 CONCLUSÃO

Mediante o contexto da integração de diversas bases de dados relacionadas à temática de Órteses, Próteses e Materiais Especiais (OPME), foi possível dicionarizar os quadros pertinentes adquiridos, e partir para a construção de um modelo relacional de dados respeitando os paradigmas de um esquema estrela para viabilizar a consulta e a filtragem de vínculos entre os dados analisados para o escopo da dissertação. Além disso, foram levados em consideração os resultados para o Mapeamento Sistemático da Literatura (MSL), com representação gráfica dos indícios e a contribuição que uma dissertação como esta pode representar para o campo de auditoria em saúde no SUS.

Os resultados do Mapeamento conduzido auxiliam na demonstração de como as pesquisas no ramo de auditoria em saúde estão sendo conduzidas nos últimos dez anos. Foram identificados trabalhos que abordaram soluções mediadas por algoritmos de *machine learning*, métodos estatísticos tradicionais, uso de *blockchain* e até mesmo inteligência artificial generativa. Além disso, os estudos recorreram a grandes volumes de dados, o que faz com que apresentem cenários convergentes com o que foi abordado nesta dissertação.

Apesar dos achados, é importante reforçar que não foram identificados estudos que tratassem especificamente de Órteses, Próteses e Materiais Especiais (OPME), o que, como discutido anteriormente, serve como provocação para o desenvolvimento de pesquisa nesse nicho. Por isso, o Mapeamento Sistemático da Literatura ajudou a compreender o andamento de pesquisas na área de auditoria em saúde, com possibilidades para novos caminhos de pesquisa.

Outra contribuição proposta nesta dissertação foi a agregação de diversas bases de dados para formulação das trilhas de auditoria de Órteses, Próteses e Materiais Especiais (OPME). Para tanto, foram adquiridas bases relacionadas à pessoa jurídica, ao Bolsa Família, a fornecedores de OPME, às notas fiscais eletrônicas dos estados do Ceará e Rio Grande do Norte, assim como outras bases relacionadas a procedimentos médicos, internações e estabelecimentos de saúde. A partir disso, foram identificados elementos que poderiam ser cruzados entre todos esses componentes para agregação em um único modelo estratificado de dados. Por fim, os dados foram separados em quadros de fato e dimensão com o objetivo de estabelecer análise de vínculos e viabilizar a identificação de indícios por meio da camada visual do projeto.

Os modelos relacionais também colaboraram com o consumo e a consulta de dados por meio do ecossistema *Apache Superset*. A partir desse ambiente, foi possível criar gráficos

que identificassem indícios por meio de vínculos entre as bases de dados. Além disso, a plataforma possibilitou a aplicação das métricas elaboradas na metodologia do trabalho. Outro ponto importante deve-se à transferência das trilhas de auditoria para a plataforma Pandas. Após o fim desse processo, os auditores puderam acompanhar de forma integral o andamento e o aperfeiçoamento das trilhas de auditoria, conforme ocorreu no ambiente de homologação do projeto FiscalizaSUS.

Os resultados apresentados ao longo desta dissertação apresentam a importância das etapas metodológicas adotadas anteriormente. Foi possível assinalar o valor que a agregação e a documentação dos conjuntos de dados proporcionaram para o projeto FiscalizaSUS nas trilhas de OPME. Além disso, o Mapeamento Sistemático da Literatura auxilia na compreensão de abordagens adotadas em auditorias em saúde. Outro ponto importante é considerar a execução de todo esse processo prático em uma arquitetura de Big Data que conta com a presença de diversas ferramentas facilitadoras para o desenvolvimento das trilhas de auditoria do projeto. Assim, a pesquisa desenvolvida nesta dissertação favorece a implementação de trilhas de auditoria em OPME por meio da qualificação e da integração de diversos conjuntos de dados. Esses fatores também estão à disposição dos auditores envolvidos nas cooperações firmadas pelo projeto FiscalizaSUS.

8.1 IMPACTOS SOCIAIS

Esta dissertação pode ser correlacionada com os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), presentes na Agenda 2030, que incluem o Brasil como um dos países parceiros. Assim, o trabalho pode ser inserido nas seguintes objetivos e metas:

1) Objetivo 3: Assegurar uma vida saudável e promover o bem-estar para todos, em todas as idades, por meio das metas 3.8: Atingir a cobertura universal de saúde, incluindo a proteção do risco financeiro, o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais seguros, eficazes, de qualidade e a preços acessíveis para todos e 3.d: Reforçar a capacidade de todos os países, particularmente os países em desenvolvimento, para o alerta precoce, redução de riscos e gerenciamento de riscos nacionais e globais de saúde (Agenda 2030, 2024);

2) Objetivo 9: Construir infraestruturas resilientes promover a industrialização inclusiva e sustentável e fomentar a inovação, assim como a proposta da meta 9.4: Até 2030, modernizar a infraestrutura e reabilitar as indústrias para torná-las sustentáveis, com eficiência

aumentada no uso de recursos e maior adoção de tecnologias e processos industriais limpos e ambientalmente corretos; com todos os países atuando de acordo com suas respectivas capacidades (Agenda 2030, 2024);

3) Objetivo 12: Assegurar padrões de produção e consumo sustentáveis, e diretrizes da meta 12.5 Até 2030, reduzir substancialmente a geração de resíduos por meio da prevenção, redução, reciclagem e reuso (Agenda 2030, 2024);

4) Objetivo 16: Promover Sociedades pacíficas e inclusivas para o desenvolvimento sustentável, proporcionar o acesso à justiça para todos e construir instituições eficazes, responsáveis e inclusivas em todos os níveis, além de agregar os pontos discutidos nas metas 16.5: Reduzir substancialmente a corrupção e o suborno em todas as suas formas e 16.6: Desenvolver instituições eficazes, responsáveis e transparentes em todos os níveis (Agenda 2030, 2024);

O alinhamento com esses ODS se deve às metas firmadas pelo projeto FiscalizaSUS com instituições parceiras, no intuito de abordar e discutir de forma eficiente irregularidades em saúde, e suportar o trabalho de auditores do Sistema Único de Saúde (SUS), como demonstrado a seguir, por meio de sete exemplos de atuação desse projeto, evidenciando suas contribuições para os ODS.

1) Atingir a cobertura universal de saúde, incluindo a proteção do risco financeiro, o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais seguros, eficazes, de qualidade e a preços acessíveis para todos, o que pode ser atribuído também ao uso correto de Órteses, Próteses e Materiais Especiais (OPME) em procedimentos hospitalares.

2) Fortalecer o SUS por meio da prática de vigilância e controle em auditoria na saúde.

3) Atingir a cobertura universal de saúde, incluindo a proteção do risco financeiro, o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais seguros, eficazes, de qualidade e a preços acessíveis para todos.

4) Otimizar a sustentabilidade e a eficiência das indústrias para a produção consciente de insumos médicos, como *Stents* para procedimentos cardíacos.

5) Reduzir resíduos em decorrência de uma produção de componentes biomédicos de forma responsável e sustentável em decorrência da vigilância e controle em saúde.

6) Reduzir substancialmente a corrupção e o suborno em todas as suas formas, que, para este trabalho, relaciona-se com a questão dos vínculos entre fornecedores para a realidade do projeto, e demais anomalias destacadas nos resultados.

7) Desenvolver instituições eficazes, responsáveis e transparentes em todos os níveis, o que pode se relacionar com a promoção de mecanismos de controle e vigilância, adentrando o campo de auditoria em saúde.

Portanto, esta dissertação de mestrado, pertencente ao escopo de atuação do projeto FiscalizaSUS na área de OPME, também se constitui a partir de um conjunto de relações com problemáticas atuais e discutidas em nível global. Ademais, esses fatores indicam a importância do investimento em pesquisas para inovação em saúde e enfrentamento às adversidades observadas na sociedade.

REFERÊNCIAS

- AGARWAL, A. *et al.* Beyond boundaries: charting the frontier of healthcare with big data and ai advancements in pharmacovigilance. **Health Sciences Review**, [s. l.], v. 14, p. 100214, mar. 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772632025000066>. Acesso em: 6 ago. 2025.
- AGÊNCIA SENADO. Orçamento 2025: quase R\$ 1 trilhão para Previdência e R\$ 245 bilhões para saúde. **Senado Notícias**, 14 abr. 2025. Disponível em: <https://www12.senado.leg.br/noticias/materias/2025/04/14/orcamento-2025-quase-r-1-trilhao-para-previdencia-e-r-245-bilhoes-para-saude>. Acesso em: 6 ago. 2025.
- AMAZON. O que é OpenSearch? **AWS**, c2024. Disponível em: <https://aws.amazon.com/pt/what-is/opensearch>. Acesso em: 29 maio. 2025.
- AMORIM, G. Monitoring Transfers to Public Health: Evidence from Randomized Audits in Brazil. **SSRN Electronic Journal**, [s. l.], fev. 2025. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5081625. Acesso em: 6 ago. 2025.
- AMORIM, R. M. O papel da auditoria interna na melhoria contínua do sistema de integridade corporativa. **Revista Latino-Americana De Governança**, [s. l.], v. 2, n. 1, p. e038, 2022. Disponível em: <https://www.revistaregov.org/index.php/revista/article/view/38>. Acesso em: 6 ago. 2025.
- APACHE. What Is Apache Spark? **Apache Spark**, c2018. Disponível em: <https://spark.apache.org>. Acesso em: 6 ago. 2025.
- ARAGÃO, A.; GARBACCIO, G. L. **Compliance and Sustainability: Brazilian and portuguese perspectives**. Coimbra: Faculdade de Direito da Universidade de Coimbra, 2020.
- ASSAYE, B. T. *et al.* Readiness of Big Health Data Analytics by Technology-Organization-Environment (TOE) Framework in Ethiopian Health Sectors. **Heliyon**, [s. l.], v. 10, n. 19, p. e38570-e38570, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405844024146014>. Acesso em: 6 ago. 2025.
- ATAEI, P.; LITCHFIELD, A. T. Big data reference architectures, a systematic literature review. 2020.
- ATAEI, P.; LITCHFIELD, A. The state of big data reference architectures: a systematic literature review. **IEEE**, [s. l.], v. 10, p. 113.789-113.807, 2022. Disponível em: <https://ieeexplore.ieee.org/document/9931012>. Acesso em: 6 ago. 2025.

AYDOGDU, A. L. F. Viewpoints of nurse auditors regarding the profession: a qualitative study. **Journal of Health Organization and Management**, [s. l.], 2025. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/39828602/>. Acesso em: 6 ago. 2025.

BELLANDI, V. *et al.* Data management for continuous learning in EHR systems. **ACM transactions on Internet technology**, [s. l.], 2024. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3660634>. Acesso em: 6 ago. 2025.

BENÍTEZ-HIDALGO, A. *et al.* TITAN: a knowledge-based platform for big data workflow management. **Knowledge-Based Systems**, [s. l.], v. 232, p. 107489, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950705121007516>. Acesso em: 6 ago. 2025.

BISPO, S. L. B. *et al.* Integração de dados abertos em saúde com o modelo OBDA: um estudo de caso na área de cirurgia bariátrica. In: **Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)**. SBC, 2022. p. 401-412.

BODDY, A. J.; HURST, W.; MACKAY, M.; RHALIBI, A. Density-Based Outlier Detection for Safeguarding Electronic Patient Record Systems. **IEEE Access**, [s. l.], v. 7, p. 40.285-40.294, 2019.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2016]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 6 ago. 2025.

BRASIL. Decreto n. 11.798, de 28 de novembro de 2023. Aprova a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções de Confiança do Ministério da Saúde e remaneja e transforma cargos em comissão e funções de confiança. **Diário Oficial da União**: seção 1 – extra B, Brasília, DF, p. 1, 28 nov. 2023. Disponível em: <https://www.in.gov.br/en/web/dou/-/decreto-n-11.798-de-28-de-novembro-de-2023-526282363>. Acesso em: 15 ago. 2025.

BRASIL. **Decreto nº 11.798, de 28 de novembro de 2023**. Aprova a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções de Confiança do Ministério da Saúde e remaneja e transforma cargos em comissão e funções de confiança. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/d11798.htm. Acesso em: 22 abr. 2025.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 6 ago. 2025.

BRASIL. **Lei nº 8.689, de 27 de julho de 1993**. Dispõe sobre a extinção do Instituto Nacional de Assistência Médica da Previdência Social (Inamps) e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/L8689.htm. Acesso em: 6 ago. 2025.

BRASIL. **Manual de normas de auditoria**. Brasília, DF: Ministério da Saúde, 1998.

BRASIL. Ministério da Saúde. **Manual de boas práticas de gestão das Órteses, Próteses e Materiais Especiais (OPME)**. Brasília, DF: Ministério da Saúde, 2016.

BRASIL. Tribunal de Contas. **Manual de Auditoria e demais fiscalizações – parte geral**. Brasília, DF: Tribunal de Contas, 2020.

CARCARY, M. The research audit trail: methodological guidance for application in practice. **Electronic Journal of Business Research Methods**, [s. l.], v. 18, n. 2, p. 166-177, 2020. Disponível em: <https://academic-publishing.org/index.php/ejbrm/article/view/2033>. Acesso em: 6 ago. 2025.

CIRURGIAS sem necessidade, superfaturamento: veja fraudes em procedimentos ortopédicos em hospitais públicos. **G1**, 18 fev. 2024. Disponível em: <https://g1.globo.com/fantastico/noticia/2024/02/18/cirurgias-sem-necessidade-superfaturamento-veja-fraudes-em-procedimentos-ortopedicos-em-hospitais-publicos.ghtml>. Acesso em: 6 ago. 2025.

COLAÇO JÚNIOR, M. **IA para a galera toda**: agentes e inovação experimental sem código. [S. l.]: Methanias Colaço Júnior, PhD, 2025. *E-book*.

COLAÇO JÚNIOR, M. **Projetando sistemas de apoio à decisão baseados em data warehouse**. Rio de Janeiro: Axcel Books, 2018.

CORDEIRO, A. C. **Auditoria financeira**: a definição da materialidade e o seu impacto no trabalho de auditoria. Coimbra: [s. n.], 2025.

COSTELLOE, E. M. *et al.* An audit of drug shortages in a community pharmacy practice. **Irish Journal of Medical Science**, [s. l.], v. 184, n. 2, p. 435-440, 2015. Disponível em: <https://link.springer.com/article/10.1007/s11845-014-1139-7>. Acesso em: 6 ago. 2025.

CRUZ, R. C. S. *et al.* Análise do impacto do Banco de Preços em Saúde (BPS) para redução das assimetrias de informação dos preços de compras de Órteses, Prótese e Materiais Especiais (OPME). **Journal of Management & Primary Health Care**, [s. l.], v. 14, n. spec, p. e006, 2022. Disponível em: <https://jmphec.com.br/jmphec/article/view/1200>. Acesso em: 6 ago. 2025.

CRUZ, R. C. S.; MARZULLO, M. P.; RAMOS, M. C. B.; PETERS, J. R.; CAVALCANTI, I. T. N. Análise do impacto do Banco de Preços em Saúde (BPS) para redução das assimetrias de informação dos preços de compras de Órteses, Prótese e Materiais Especiais (OPME). **JMPHC. Journal of Management & Primary Health Care**, [s. l.], v. 14, n. spec, p. e006, 2022. Disponível em: <https://jmphec.com.br/jmphec/article/view/1200>. Acesso em: 3 fev. 2025.

DAFAALLA, M. *et al.* Outcomes of ST elevation myocardial infarction in patients with cancer: a nationwide study. **European Heart Journal-Quality of Care and Clinical Outcomes**, [s. l.], v. 9, n. 8, p. 806-817, 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/36921979/>. Acesso em: 6 ago. 2025.

DAI, H.-J. *et al.* Integrating predictive coding and a user-centric interface for enhanced auditing and quality in cancer registry data. **Computational and Structural Biotechnology Journal**, [s. l.], v. 24, p. 322-333, 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/38690549/>. Acesso em: 6 ago. 2025.

DUTOIT, S. H. C.; STEYN, A. G. W.; STUMPF, R. H. **Graphical exploratory data analysis**. New York: Springer, 2012.

EKIN, T.; LAKOMSKI, G.; MUSAL, R. M. An unsupervised Bayesian hierarchical method for medical fraud assessment. **Statistical Analysis and Data Mining: the asa data science journal**, [s. l.], v. 12, n. 2, p. 116-124, 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.11408>. Acesso em: 6 ago. 2025.

EXASCALE OBJECT Store for AI Data, Agentic Computing, and Analytics. **MINIO**, c2025. Disponível em: <https://min.io>. Acesso em: 6 ago. 2025.

FALCÃO, M. Z.; RACHID, R.; FORNAZIN, M. AI innovation in healthcare and state platforms under a rights-based perspective: the case of Brazilian RNDs. **Data & Policy**, [s. l.], v. 6, p. e70, 2024. Disponível em: <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/EC3FFDD671CCE1D663EFF0279842DD48/S2632324924000701a.pdf/ai-innovation-in-healthcare-and-state-platforms-under-a-rights-based-perspective-the-case-of-brazilian-rnds.pdf>. Acesso em: 6 ago. 2025.

FAN, X.; LU, J. Enterprise Level Data Warehouse System Based on Hive in Big Data Environment. **Procedia Computer Science**, [s. l.], v. 243, p. 67-75, 2024. Disponível em: <https://dl.acm.org/doi/10.1016/j.procs.2024.09.010>. Acesso em: 6 ago. 2025.

FERREIRA, A. O. M.; SILVINO, Z. R. Auditoria interna para a estratificação do problema da qualidade das equipes de enfermagem na COVID-19. **Research, Society and Development**, [s. l.], v. 9, n. 10, p. e2789108381, 2020. Disponível em: <https://rsdjournal.org/rsd/article/view/8381>. Acesso em: 6 ago. 2025.

FONTES, R. S. *et al.* Sussurro-Detecção na Web de eventos auditáveis que representam riscos à saúde pública. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE (SBCAS), 23., 2023, São Paulo. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 211-217.

GOMES, W. *et al.* A Smart Classifier of Orthoses, Prostheses and Special Materials (OPMEs) in Invoices. **Frontiers in the Internet of Things**, [s. l.], v. 4, p. 1436757, 2025. Disponível em: <https://www.frontiersin.org/journals/the-internet-of-things/articles/10.3389/friot.2025.1436757/full>. Acesso em: 6 ago. 2025.

GOMES, W. *et al.* Prova de Conceito de um Classificador de OPMes em Notas Fiscais. *In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE (SBCAS)*, 23., 2023, São Paulo. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 198-204.

GROH, M. *et al.* Towards Transparency in Dermatology Image Datasets with Skin Tone Annotations by Experts, Crowds, and an Algorithm. **Proceedings of the ACM on human-computer interaction**, [s. l.], v. 6, n. 521, p. 1-26, 7 nov. 2022. Disponível em: <https://dl.acm.org/doi/10.1145/3555634>. Acesso em: 6 ago. 2025.

HERLAND, M.; KHOSHGOFTAAR, T. M.; BAUDER, R. A. Big data fraud detection using multiple medicare data sources. **J Big Data**, [s. l.], v. 5, n. 29, 2018. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0138-3>. Acesso em: 6 ago. 2025.

HILLAN, A. *et al.* Public health audit of vaccine cold chain management in general practice and community pharmacy in Western Australia. **Australian and New Zealand Journal of Public Health**, [s. l.], v. 48, n. 4, p. 100.168, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1326020024000438>. Acesso em: 6 ago. 2025.

HUSSAIN, M. K. *et al.* Big Data in Healthcare. **IJRTE**, [s. l.], v. 8, n. 6, p. 2.127-2.131, 2020. Disponível em: https://www.researchgate.net/publication/345984483_Big_Data_in_Healthcare. Acesso em: 6 ago. 2025.

HUYGHUES-BEAUFOND, N. *et al.* Robust and automatic data cleansing method for short-term load forecasting of distribution feeders. **Applied Energy**, [s. l.], v. 261, p. 114.405, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0306261919320926>. Acesso em: 6 ago. 2025.

IAASB. **Basis for Conclusions: ISA 200 (Revised and Redrafted)**, Overall Objectives of the Independent Auditor and the Conduct of an Audit in Accordance with International Standards on Auditing. New York: International Federation Accountants, 2010.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Agenda 2030**: objetivos de desenvolvimento sustentável: avaliação do progresso das principais metas globais para o Brasil: ODS 3: assegurar uma vida saudável e promover o bem-estar para todas e todos, em todas as idades. Brasília, DF: IPEA, 2024. (Cadernos ODS, 3). Disponível em: <http://dx.doi.org/10.38116/ri2024ODS3>. Acesso em: 6 ago. 2025.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Agenda 2030**: objetivos de desenvolvimento sustentável: avaliação do progresso das principais metas globais para o Brasil: ODS 9: construir infraestruturas resilientes, promover a industrialização inclusiva e sustentável, e fomentar a inovação. Brasília, DF: IPEA, 2024. (Cadernos ODS, 9). Disponível em: <http://dx.doi.org/10.38116/ri2024ODS9>. Acesso em: 6 ago. 2025.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Agenda 2030**: objetivos de desenvolvimento sustentável: avaliação do progresso das principais metas globais para o Brasil: ODS 12: consumo responsável – assegurar padrões de consumo e produção sustentável. Brasília, DF: IPEA, 2024. (Cadernos ODS, 12). Disponível em: <http://dx.doi.org/10.38116/ri2024ODS12>. Acesso em: 6 ago. 2025.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Agenda 2030**: objetivos de desenvolvimento sustentável: avaliação do progresso das principais metas globais para o Brasil: ODS 16: promover sociedades pacíficas e inclusivas para o desenvolvimento sustentável, proporcionar o acesso à justiça para todos e construir instituições eficazes, responsáveis e inclusivas em todos os níveis. Brasília, DF: IPEA, 2024. (Cadernos ODS, 16). Disponível em: <http://dx.doi.org/10.38116/ri2024ODS16>. Acesso em: 6 ago. 2025.

ISHWARAPPA, K.; ANURADHA, J. A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. **Procedia Computer Science**, [s. l.], v. 48, p. 319-324, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050915006973>. Acesso em: 6 ago. 2025.

JOHNSON, J. L.; ADKINS, D.; CHAUVIN, S. A review of the quality indicators of rigor in qualitative research. **American journal of pharmaceutical education**, [s. l.], v. 84, n. 1, p. 7.120, 2020. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/32292186/>. Acesso em: 6 ago. 2025.

JOUDAKI, H. *et al.* Improving fraud and abuse detection in general physician claims: a data mining study. **International journal of health policy and management**, [s. l.], v. 5, n. 3, p. 165-172, 2015. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26927587/>. Acesso em: 6 ago. 2025.

JUPYTERLAB: a next-generation notebook interface. **Jupyter**, c2025. Disponível em: <https://jupyter.org>. Acesso em: 6 ago. 2025.

LANEY, D. *et al.* 3D data management: Controlling data volume, velocity and variety. **META group research note**, [s. l.], v. 949, 2001. Disponível em: <https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Acesso em: 6 ago. 2025.

LIMA, B. C. *et al.* Perfil da prescrição de órteses, próteses e materiais especiais no âmbito do Sistema Único de Saúde por fisioterapeutas no estado do Maranhão. **Brazilian Journal of Health Review**, [s. l.], v. 7, n. 5, p. e73734, 2024. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/73734>. Acesso em: 6 ago. 2025.

LIMA, H. S. C. *et al.* SUS, saúde e democracia: desafios para o Brasil Manifesto de seis ex-ministros da saúde a propósito da 16ª Conferência Nacional De Saúde. **Ciência & saúde coletiva**, [s. l.], v. 24, n. 10, p. 3.713-3.716, 2019. Disponível em: <https://www.scielo.br/j/csc/a/YkZScFP3VBsMbRhyqH9xL3R/>. Acesso em: 6 ago. 2025.

LIMB, C. *et al.* How to conduct a clinical audit and quality improvement project. **Int. J. Surg. Oncol.**, [s. l.], v. 2, n. 6, p. e24, 2017. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29177218/>. Acesso em: 6 ago. 2025.

LOUREIRO, L. H. *et al.* Como a auditoria de enfermagem pode influenciar na qualidade assistencial. **Revista Práxis**, [s. l.], v. 10, n. 19, 2018. Disponível em: <https://revistas.unifoa.edu.br/praxis/article/view/698>. Acesso em: 6 ago. 2025.

MACHOSKI, E.; ARAUJO, J. M. Corruption in public health and its effects on the economic growth of Brazilian municipalities. **The European Journal of Health Economics**, [s. l.], v. 21, n. 5, p. 669–687, 17 fev. 2020.

MÁFIA DAS PRÓTESES: empresa é condenada a pagar R\$ 15 milhões por danos morais coletivos: Multinacional alemã pagava propina a médicos mineiros para aumentar suas vendas de órteses e próteses. **MPF**, 18 set. 2024. Disponível em: <https://www.mpf.mp.br/mg/sala-de-imprensa/noticias-mg/mafia-das-proteses-empresa-e-condenada-a-pagar-r-15-mi-por-danos-morais-coletivos#:~:text=Ainda%20segundo%20a%20ação%2C%20o,R%24%201%2C25%20bilhão>. Acesso em: 3 fev. 2025.

MAFORT, E. M. B. F.; MELECHE, M. A. O papel da auditoria durante a pandemia. **REVISTA FOCO**, [s. l.], v. 17, n. 1, p. e4224, 2024. Disponível em: <https://ojs.focopublicacoes.com.br/foco/article/view/4224>. Acesso em: 6 ago. 2025.

MARCONDES, R.; SILVA, S. L. R. O protocolo prisma 2020 como uma possibilidade de roteiro para revisão sistemática em ensino de ciências. **Revista Brasileira de Pós-Graduação**, [s. l.], v. 18, n. 39, p. 1-19, 2023. Disponível em: <https://rbpg.capes.gov.br/rbpg/article/view/1894>. Acesso em: 6 ago. 2025.

MARQUES, M. C. C. Saúde e poder: a emergência política da Aids/HIV no Brasil. **História, Ciências, Saúde-Manguinhos**, [s. l.], v. 9, p. 41-65, 2002. Disponível em: <https://www.scielo.br/j/hcsm/a/SJHgNdc3WBMKgNGfjKQvqfM/>. Acesso em: 6 ago. 2025.

MENDES, E. V. 25 anos do Sistema Único de Saúde: resultados e desafios. **Estudos Avançados**, [s. l.], v. 27, n. 78, p. 27-34, 2013. Disponível em: <https://www.scielo.br/j/ea/a/gzYFsDyxzXPjJK8WvWvG8th>. Acesso em: 6 ago. 2025.

MENDES, H. Preso em operação contra fraudes na saúde usava “laranja” do Bolsa Família. **G1**, 22 jul. 2016. Disponível em: <https://g1.globo.com/bahia/noticia/2016/07/preso-em-operacao-contra-fraudes-na-saude-usava-laranjas-do-bolsa-familia.html>. Acesso em: 6 ago. 2025.

MINIO. **IBM**, 6 abr. 2021. Disponível em: <https://www.ibm.com/docs/pt-br/cloud-private/3.2.x?topic=private-minio>. Acesso em: 29 maio 2025.

MIRANDA, E. B. M. *et al.* Consolidação dos códigos do sistema de gerenciamento da tabela de procedimentos, medicamentos e OPM do SUS (SIGTAP) para Odontologia. **RECISATEC-REVISTA CIENTÍFICA SAÚDE E TECNOLOGIA**, [s. l.], v. 2, n. 1, p. e2179, 2022. Disponível em: <https://recisatec.com.br/index.php/recisatec/article/view/79>. Acesso em: 6 ago. 2025.

NAKAMURA, J. Brasil cai e encerra 2024 como 10^a maior economia do mundo; veja ranking. **CNN**, 10 mar. 2025. Disponível em: <https://www.cnnbrasil.com.br/economia/macroeconomia/brasil-cai-e-encerra-2024-como-10a-maior-economia-do-mundo-veja-ranking/>. Acesso em: 6 ago. 2025.

NAVARRETE, A.; GALLEGO, A.; PADILLA, P. Are the score ratios conclusive in detecting financial fraud? the case of clínica las condes in chile. **Journal of Legal, Ethical and Regulatory Issues**, [s. l.], v. 24, n. 1, 2021. Disponível em: <https://www.abacademies.org/articles/are-the-score-ratios-conclusive-in-detecting-financial-fraud-the-case-of-clinica-las-condes-in-chile-12457.html>. Acesso em: 6 ago. 2025.

OJEDA-THIES, C. *et al.* Fragility fracture audit. In: FALASCHI, P.; MARSH, D. (ed.). **Orthogeriatrics: the management of older patients with fragility fractures**. 2. ed. Cham: Springer, 2021.

PAGE, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, [s. l.], v. 372, 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33782057/>. Acesso em: 6 ago. 2025.

PANIGUTTI, C. *et al.* FairLens: auditing black-box clinical decision support systems. **Information Processing & Management**, [s. l.], v. 58, n. 5, p. 102657, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S030645732100145X>. Acesso em: 6 ago. 2025.

PINHEIRO, S. D.; BUSATO, I. M. S.; CAVEIÃO, C. Inflação na saúde-OPME em tempos de Covid-19. Importância da auditoria, Amazonas-Brasil. **Revista Saúde e Desenvolvimento**, [s. l.], v. 15, n. 23, p. 96-104, 2021.

PINTO, K.; MOURA, E.; SILVA, M. ROBERTO, J.; SERRA, M.; LOPES, N. Auditoria externa como ferramenta de melhoria na gestão empresarial do setor financeiro. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, [s. l.], p. 153-176, 2021.

RABINOWITZ, P. J. **Before reading: narrative conventions and the politics of interpretation**. Ohio: The Ohio State University Press, 1998.

RAVN, M. B. *et al.* Why do patients with ischaemic heart disease drop out from cardiac rehabilitation in primary health settings. A qualitative audit of patient charts. **Frontiers in rehabilitation sciences**, [s. l.], v. 3, p. 837174, 2022. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9397838/>. Acesso em: 6 ago. 2025.

RIERA-MESTRE, A. *et al.* PICO questions and DELPHI methodology for improving the management of patients with acute hepatic porphyria. **Revista Clínica Española (English Edition)**, [s. l.], v. 224, n. 5, p. 272-280, 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/38642893/>. Acesso em: 6 ago. 2025.

RODRIGUES, C. A. **Aspectos relevantes da auditoria de compliance nas Micro e Pequenas Empresas**. 2013. Trabalho de Conclusão de Curso (Bacharelado em Ciências Contábeis) – Faculdade de Tecnologia e Ciências Sociais Aplicadas, Brasília, DF, 2013.

ROEVER, L. *et al.* Compreendendo o GRADE: PICO e qualidade dos estudos. **Revista da Sociedade Brasileira de Clínica Médica**, [s. l.], v. 19, n. 1, p. 54-61, 2021. Disponível em: <https://docs.bvsalud.org/biblioref/2022/03/1361752/54-61.pdf>. Acesso em: 6 ago. 2025.

RUFINO, C. E. V.; REIS, M. C.; RIBEIRO, M. B. Distribuição das órteses, próteses e materiais especiais no sistema único de saúde do piauí. **Jornal de Ciências da Saúde**, [s. l.], v. 7, n. 2, p. 34-45, 2024. Disponível em: <https://periodicos.ufpi.br/index.php/rehu/article/download/5706/4804/25263>. Acesso em: 6 ago. 2025.

SANTOS, M. A. B. *et al.* **A Remuneração de internações e a tabela de procedimentos do SUS: uma análise a partir do SIH e do SIGTAP**. Brasília, DF: IPEA, 2023. (Texto para Discussão, n. 2925).

SATKA, E. Internal and External Audit in the Function of the Management of the Trade Companies. **Journal of US-China Public Administration**, [s. l.], v. 14, n. 6, p. 330-338, 2017. Disponível em: <https://davidpublisher.com/Public/uploads/Contribute/5ac48743caba7.pdf>. Acesso em: 6 ago. 2025.

SBODIO, M. L. *et al.* Collaborative artificial intelligence system for investigation of healthcare claims compliance. **Scientific Reports**, [s. l.], v. 14, n. 11884, 2024. Disponível em: <https://www.nature.com/articles/s41598-024-62665-0>. Acesso em: 6 ago. 2025.

SEO, S. **A review and comparison of methods for detecting outliers in univariate data sets**. 2006. Tese (Doutorado em Ciências) – University of Pittsburgh, Pittsburgh 2006.

SINGH, K.; UPADHYAYA, S. Outlier detection: applications and techniques. **International Journal of Computer Science Issues (IJCSI)**, [s. l.], v. 9, n. 3, p. 307, 2012. Disponível em: <https://www.studocu.id/id/document/universitas-bina-nusantara/algorithm-programming/outlier-detection-applications-and-techniques/64432241>. Acesso em: 6 ago. 2025.

SIQUEIRA, A. S. E.; SIQUEIRA-FILHO, A. G.; LAND, M. G. P. Análise do impacto econômico das doenças cardiovasculares nos últimos cinco anos no Brasil. **Arquivos Brasileiros de Cardiologia**, [s. l.], v. 109, p. 39-46, 2017.

SOUZA, L. *et al.* Data science to identify crimes against public administration. **Annals of Computer Science and Information Systems**, [s. l.], v. 37, n. 1, p. 281-287, 2023. Disponível em: <https://journals.indexcopernicus.com/search/article?articleId=4553669>. Acesso em: 6 ago. 2025.

SUPERSET. **Superset**, c2024. Disponível em: <https://superset.apache.org/docs/intro/>. Acesso em: 4 jan. 2025.

TEIXEIRA, M. G. *et al.* Vigilância em Saúde no SUS-construção, efeitos e perspectivas. **Ciência & Saúde Coletiva**, [s. l.], v. 23, n. 6, p. 1.811-1.818, 2018. Disponível em: <https://www.scielo.br/j/csc/a/FxcSJBQq8G7CNSxhTyT7Qbn/>. Acesso em: 6 ago. 2025.

TUKEY, J. W. *et al.* **Exploratory data analysis**. Reading: Addison-wesley, 1977.

TUNGARE, S. *et al.* Designing audit and feedback dashboards for dentists to monitor their opioid prescribing. **International Journal of Medical Informatics**, [s. l.], v. 176, p. 105092, 2023. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1386505623001107>. Acesso em: 6 ago. 2025.

UNASUS. **O sistema nacional de auditoria do sus – sna foi criado pela lei 8689 de 27 de julho de 1993**. Brasília, DF: UNASUS, 2025. Disponível em: <https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-permanentes/cssf/arquivos/21-10-2021-fortalecimento-e-reestruturacao-do-denasus-solimar-vieira-mendes>. Acesso em: 6 ago. 2025.

VEDANABHATLA, S.; GUPTA, N. V. A review on audits and compliance management. **Asian Journal of Pharmaceutical and Clinical Research**, [s. l.], v. 6, n. 2, p. 43-45, 2013.

VERMA, P.; TRIPATHI, V.; PANT, B. MRDACE: An Intelligent Architecture For Secure Sharing And Traceability Of The Medical Images And Patients' Records. **ACM Transactions on Computing for Healthcare**, [s. l.], v. 6, n. 3, p. 1-21, 2025. Disponível em: <https://dl.acm.org/doi/10.1145/3712708>. Acesso em: 6 ago. 2025.

WANG, H. *et al.* A Weakly-Supervised Named Entity Recognition Machine Learning Approach for Emergency Medical Services Clinical Audit. **International Journal of Environmental Research and Public Health**, [s. l.], v. 18, n. 15, p. 7.776, 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34360065/>. Acesso em: 6 ago. 2025.

WANG, H. F. *et al.* Quality improvements in decreasing medication administration errors made by nursing staff in an academic medical center hospital: a trend analysis during the journey to Joint Commission International accreditation and in the post-accreditation era. **Therapeutics and clinical risk management**, [s. l.], v. 11, p. 393-406, 2015. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25767393/>. Acesso em: 6 ago. 2025.

YIN, R. K. **Estudo de Caso: Planejamento e métodos**. Porto Alegre: Bookman, 2015.

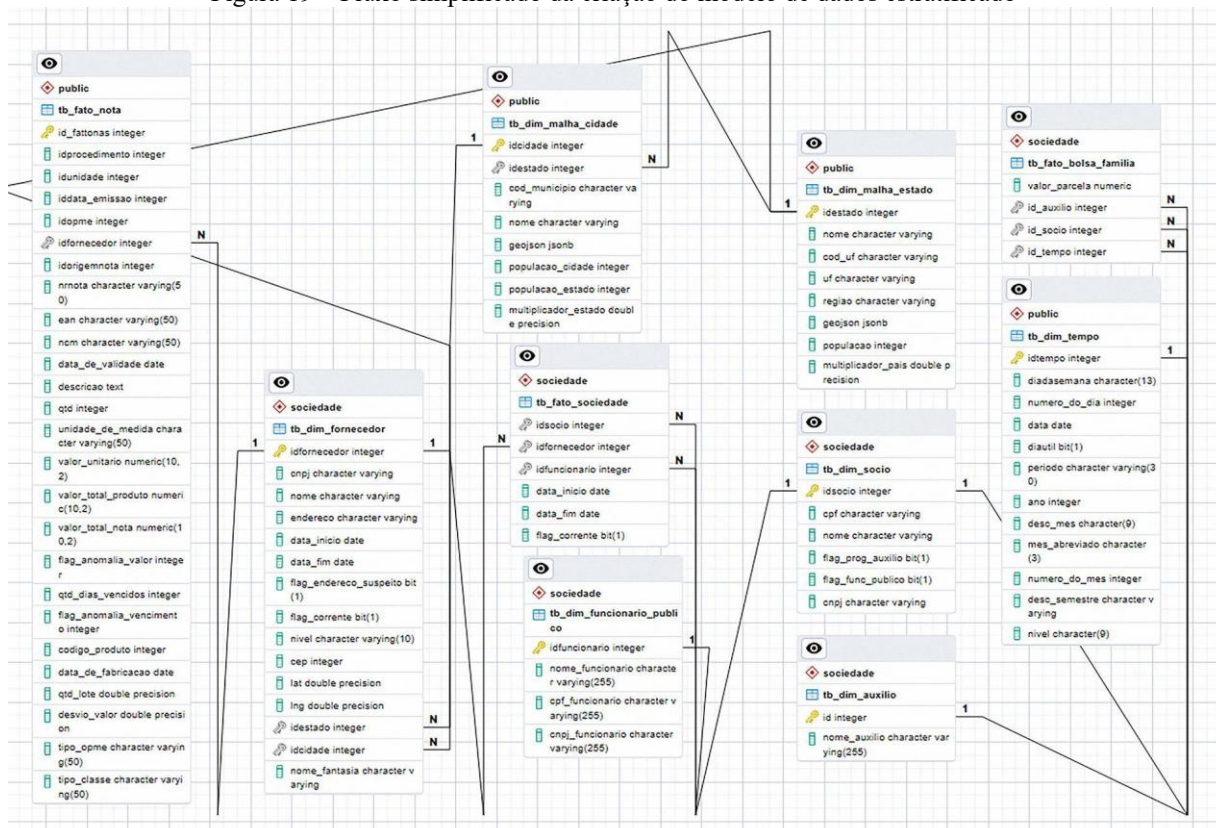
ZHENG, X.; HAMID, M. A. A.; HOU, Y. Data mining algorithm in the identification of accounting fraud by smart city information technology. **Heliyon**, [s. l.], v. 10, n. 9, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405844024060791>. Acesso em: 6 ago. 2025.

ZHOU, J. *et al.* FraudAuditor: A visual analytics approach for collusive fraud in health insurance. **IEEE Transactions on Visualization and Computer Graphics**, [s. l.], v. 29, n. 6, p. 2.849-2.861, 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/37030774/>. Acesso em: 6 ago. 2025.

ZIJLSTRA, W. P.; VAN-DER ARK, L. A.; SIJTSMA, K. Outlier detection in test and questionnaire data. **Multivariate Behavioral Research**, [s. l.], v. 42, n. 3, p. 531-555, 2007. Disponível em: <https://psycnet.apa.org/record/2007-16556-005>. Acesso em: 6 ago. 2025.

APÊNDICE A – MODELO DE DADOS ESTRATIFICADO COMPLETO, UTILIZANDO AS PRINCIPAIS TABELAS PARA AS TRILHAS DE AUDITORIA DE OPME

Figura 19 – Fluxo simplificado da criação do modelo de dados estratificado



Fonte: elaborada pelo autor, 2025

Quadro 20 – Exemplo de artigos que passaram por extração de dados

Autor	Tipo de auditoria	Algoritmos/Métodos	Indicadores	Métricas Quantitativas	Métricas Qualitativas
Tungar <i>et al.</i> , 2023	Clínica	Tableau/Operações com SQL	Taxa de prescrição de opioides; Dados qualitativos mediante os atendimentos	Cálculo das taxas de prescrição de opioides;	Análise qualitativa dos feedback dos atendimentos
Panigutti <i>et al.</i> , 2021	Clínica	Técnicas de Machine learning e Análise estatística	Indicadores demográficos e de eficiência do modelo	F1-score; N de pacientes, N de admissões, média de admissões por paciente, máximo de admissões por paciente, N de códigos ICD-9 Únicos, média de códigos por admissão	Descrições e regras que implicam em erros de classificação; Regras de decisão local e global
Verma Tripathi, Pant e Mrdace, 2025	Conformidade	Técnicas de Blockchain (Proof-of-Authority); Elliptic Curve Digital Signature Algorithm (ECDSA)	Modelagem do Sistema; Modelo de Ameaça; Requisitos Funcionais; Requisitos de Segurança	Tempo de execução; Precisão e Curva AUC; Redução no tempo de execução	Confiança; Auditabilidade; Privacidade;
Zhou <i>et al.</i> , 2023	Financeira	algoritmo de detecção de comunidades [Algoritmo de Louvain]; cálculo de semelhança baseado na correspondência mais próxima. PCA	Visualização da comparação de pacientes; visualização do comportamento do paciente; avaliação estatística	Número de pacientes; Total de pagamentos per capita; Quantas vezes um paciente visitou o mesmo estabelecimento de saúde; Frequência	Similaridade de doenças; Similaridade no comportamento dos pacientes

				das visitas ao longo do tempo; Menor intervalo entre as visitas; Quartis ; Distância Interquartil	
Dafaalla <i>et al.</i> , 2023	Clínica/ Qualitativa	Indicadores de Qualidade da Sociedade Europeia de Cardiologia (ESC) e da Associação para Cuidados Cardiovasculares Agudos (ACVC)	Prevalência de câncer; Taxa de revascularização; Mortalidades em 1 ano e 30 dias; Proporção de pacientes por dados demográficos	Prevalência de Câncer; taxas de revascularização; Mortalidade em ambiente hospitalar; mortalidade em 30 dias; Mortalidade em 1 ano; taxas de realização de angiografia; Proporção dos pacientes por dados demográficos	Uso de estatinas de alta intensidade na alta hospitalar; uso de betabloqueadores na alta; tempo porta-balão

Fonte: produzida pelo autor, 2025