

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Integrando técnicas de aprendizado de máquina para
descoberta de conhecimento interpretável de prontuários
eletrônicos de pacientes**

José Maria Clementino Junior

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências
de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

José Maria Clementino Junior

**Integrando técnicas de aprendizado de máquina para
descoberta de conhecimento interpretável de prontuários
eletrônicos de pacientes**

Dissertação apresentada ao Instituto de Ciências
Matemáticas e de Computação – ICMC-USP,
como parte dos requisitos para obtenção do título
de Mestre em Ciências – Ciências de Computação
e Matemática Computacional. *EXEMPLAR DE
DEFESA*

Área de Concentração: Ciências de Computação e
Matemática Computacional

Orientadora: Profa. Dra. Agma Juci Machado Traina
Coorientador: Prof. Dr. Bruno Squizato Faiçal

**USP – São Carlos
Setembro de 2021**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

C626i Clementino Junior, José Maria
Integrando técnicas de aprendizado de máquina
para descoberta de conhecimento interpretável de
prontuários eletrônicos de pacientes / José Maria
Clementino Junior; orientadora Agma Juci Machado
Traina; coorientador Bruno Squizato Faiçal. -- São
Carlos, 2021.
84 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2021.

1. Inteligência Artificial Explicável. 2.
Reconhecimento de Padrões. 3. Registros Médicos
Eletrônicos. 4. Agrupamento. I. Machado Traina, Agma
Juci , orient. II. Squizato Faiçal, Bruno ,
coorient. III. Título.

José Maria Clementino Junior

Towards integration of machine learning techniques and
interpretable knowledge discovery from electronic health
records

Dissertation submitted to the Instituto de Ciências
Matemáticas e de Computação – ICMC-USP – in
accordance with the requirements of the Computer
and Mathematical Sciences Graduate Program, for
the degree of Master in Science. *EXAMINATION
BOARD PRESENTATION COPY*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Profa. Dra. Agma Juci Machado Traina
Co-advisor: Prof. Dr. Bruno Squizato Faiçal

USP – São Carlos
September 2021

*Dedico este trabalho aos meus pais Edilúcia e José Maria
Sem eles nada seria possível.*

AGRADECIMENTOS

Gostaria primeiramente de agradecer a DEUS, pois sem Ele não teria chegado até aqui.

Agradeço minha família, que sempre me apoiou desde o início, proporcionando força para continuar. Em especial, minha mãe Edilúcia, meu pai José Maria, minha irmã Maria Eduarda e minha avó Luzia, sem vocês nada seria possível. Amo vocês família.

A minha orientadora Prof.^a Dra. Agma Traina, pela oportunidade, disponibilidade, atenção e paciência depositada em mim ao longo desse período, que juntamente com Prof. Dr. Caetano me acolheram muito bem, minha gratidão por todos os ensinamentos e conselhos.

Também gostaria de agradecer meu coorientador e amigo Prof. Dr. Bruno Faiçal, por me apresentar o programa de mestrado e por me acolher junto a sua família, além de toda parceria nos trabalhos desenvolvidos.

Agradeço ao Grupo de Bases de Dados e de Imagens (GBdI), por toda estrutura fornecida, aos membros que auxiliaram de maneira direta ou indiretamente nos projetos de pesquisa. Em especial, aos meu amigos que sempre topavam um *happy hour*: Lucas Santiago, André Toshio, Óscar Linares, Christian Bones, Maxwell Sampaio, Willian Dener e Braulio Valentin, tornando a caminhada mais leve, muito obrigado. Agradeço também aos demais amigos externos ao laboratório, mas que acompanharam a minha trajetória: Diego Carneiro, Allan Pires, Erikson Aguiar e Marcos Gôlo.

A Universidade de São Paulo (USP), especificamente ao Instituto de Ciências Matemáticas e de Computação (ICMC), por toda estrutura física e também seu corpo docente.

As instituições de fomento Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) código de financiamento 001 e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) sob número do processo 16/17078-0 e 20/07200-9.

“Conformity is the jailer of freedom and the enemy of growth.”

(John F. Kennedy)

RESUMO

CLEMENTINO JR, M. J. **Integrando técnicas de aprendizado de máquina para descoberta de conhecimento interpretável de prontuários eletrônicos de pacientes.** 2021. 84 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Um desafio a ser superado na área de Aprendizado de Máquina (AM) é a capacidade dos modelos serem explicáveis e apresentarem um alto nível de interpretabilidade e transparência. É possível observar que grande parte dos métodos preocupam-se somente com métricas quantitativas, deixando de explorar a interpretabilidade dos resultados obtidos. Em algumas áreas, tal como a medicina, além da acurácia obtida pelos métodos computacionais é importante que os profissionais compreendam o porquê de uma decisão foi tomada. A importância dada à explicabilidade da decisão do método se justifica por se tratar de áreas críticas, onde uma decisão errada pode causar uma grande perda financeira, a privacidade ou a saúde. Ao encontro desse cenário um campo de pesquisa denominado *Explainable Artificial Intelligence* (XAI) tem surgido com interesse de desenvolver métodos que priorizam explicabilidade, tanto do funcionamento do modelo quanto da explicação das decisões indicadas. Portanto, este trabalho, tem por objetivo a criação de métodos capazes de extrair conhecimento em registros médicos eletrônicos priorizando a explicabilidade dos resultados obtidos. O primeiro método denominado *Bag-of-Attributes Representation* (BOAR) proporciona uma representação capaz de superar o problema de interoperabilidade de bases de dados médicas e servir como entrada para diversos algoritmos de agrupamento. Após o processo de descoberta de padrões por meio dos algoritmos de agrupamento, o segundo método *Multilevel Clustering Explainer* (MCE) tem como ideia principal reunir informações descritivas e visuais sobre a composição dos padrões encontrados, por meio da apresentação das variáveis que mais influenciaram na decisão do algoritmo. O MCE tem a flexibilidade de fornecer explicações em dois níveis: no global apresentando as principais características dos objetos que determinaram a formação de um determinado *cluster* e no local a influência considerando um único objeto. O MCE possui uma grande capacidade de modularização de seus parâmetros em relação a estrutura de funcionamento e de apresentação dos resultados, proporcionando ao profissional de saúde um melhor entendimento das informações e padrões encontrados. O MCE mostrou ser capaz de fornecer informações explicativas que permitem uma análise *top-down* na compreensão da composição, relacionamento e inconsistências dos *clusters*. Neste trabalho, o MCE permitiu a compreensão de quais procedimentos médicos mais influenciaram para a caracterização de diferentes clusters.

Palavras-chave: Inteligência Artificial Explicável, Reconhecimento de padrões, Registros Médicos Eletrônicos, Agrupamento.

ABSTRACT

CLEMENTINO JR, M. J. **Towards integration of machine learning techniques and interpretable knowledge discovery from electronic health records** . 2021. 84 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

An important challenge that is still hard to overcome, is the ability of machine learning models being explicable and having a high level of interpretability and transparency. However, there is a lack of research on models that address the interpretability of knowledge discovery. In areas, such as Medicine, besides the accuracy that computational methods should present, it is essential for medical professionals to understand the reason for a certain decision. The importance of the explicability of the decision suggested by the method is based on the critical nature of these areas. A wrong decision may cause serious consequences on health. In that context, in order to overcome these problems, a new field, Explainable Artificial Intelligence (XAI), has emerged. The goal of XAI is to develop methods that prioritize the explicability. Considering both the model functionality and the explanation of the suggested decisions. In this work, the objective was to create methods capable of extracting knowledge from electronic medical records, prioritizing the results' explanability. The first method called Bag-of-Attributes Representation (BOAR) provides a representation capable of overcoming the problem of interoperability of medical databases and serving as an input to several clustering algorithms. After the process of discovering patterns using clustering algorithms, the second method Multilevel Clustering Explainer (MCE), has as the main idea to gather descriptive and visual information about the composition of the patterns found by presenting the variables that most influenced the decision of the algorithm. Thus, MCE has the flexibility to provide explanations on two levels: globally offering the main characteristics of the objects that determined the formation of a given cluster and on the spot the influence considering a single object. In addition, the MCE has a great capacity to modularize its parameters concerning the operating structure and presentation of results, providing the health professional with a better understanding of the information and patterns found. The MCE proved to provide explanatory information that allows for top-down analysis of clusters' composition, relationship, and inconsistencies. In this work, the MCE allowed understanding of which medical procedures most influenced the characterization of different clusters.

Keywords: Explainable Artificial Intelligence, Pattern Recognition, Electronic Patient Records, Clustering.

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura e hierarquia da IA, supervisionado e não supervisionado	28
Figura 2 – Proposta do conceito XAI	32
Figura 3 – Taxonomia do método de explicabilidade	34
Figura 4 – Interesse em XAI ao longo do tempo	35
Figura 5 – Cenário da interoperabilidade entre as bases de dados	42
Figura 6 – Visão geral do MCE	46
Figura 7 – Visão Geral do BOAR	47
Figura 8 – Exemplo de um dicionário construído pelo módulo construtor.	48
Figura 9 – Funcionamento da etapa 1	50
Figura 10 – Fluxo de construção das informações explicativas nível local	51
Figura 11 – Fluxo de construção das informações explicativas nível global	53
Figura 12 – Apresentação visual	55
Figura 13 – Contexto de aplicação do OMOP	58
Figura 14 – Os 10 primeiros diagnósticos obtidos considerando cada método de agrupamento.	62
Figura 15 – Nível global <i>cluster</i> 45: Informação do <i>cluster</i> 45 e box plot dos procedimentos mais frequentes.	64
Figura 16 – Nível global <i>cluster</i> 45: Valores absolutos dos procedimentos mais influentes.	65
Figura 17 – Nível Global <i>cluster</i> 45: Gráfico Pizza e Gráfico de Barra dos procedimentos mais influentes	65
Figura 18 – Nível Global <i>cluster</i> 45: Gráficos de barras da frequência dos diagnósticos.	66
Figura 19 – Nível Global <i>cluster</i> 45: Gráficos de barras da frequência dos procedimentos	66
Figura 20 – Nível Global <i>cluster</i> 7: Informação do cluster 45 e Box Plot dos procedimentos mais frequentes.	67
Figura 21 – Nível global <i>cluster</i> 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes	68
Figura 22 – Nível global <i>cluster</i> 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes	68
Figura 23 – Nível global <i>cluster</i> 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes	68
Figura 24 – Nível Global <i>cluster</i> 7: Gráficos de Barras da frequência dos diagnósticos.	69
Figura 25 – Nível Global <i>cluster</i> 7: Gráficos de Barras da frequência dos procedimentos.	70

Figura 26 – Nível local <i>cluster</i> 45: Informações do registro e valores absolutos dos procedimentos mais frequentes.	71
Figura 27 – Nível local <i>cluster</i> 45: Informações do registro e valores absolutos dos procedimentos mais frequentes.	71
Figura 28 – Nível local <i>cluster</i> 7: Informações do registro e valores absolutos dos procedimentos mais frequentes.	72
Figura 29 – Nível local <i>cluster</i> 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes.	73

LISTA DE TABELAS

Tabela 1 – Tabela descritiva de comparação dos trabalhos relacionados	40
Tabela 2 – Informações gerais do <i>cohort</i> selecionado e informações básicas obtidas pelos algoritmos.	61

LISTA DE ABREVIATURAS E SIGLAS

AD	Árvore de Decisão
AM	Aprendizado de Máquina
BoCUI	<i>Bag-of-Medical-Concepts</i>
BOW	<i>Bag-of-Words Model</i>
CID	Classificação Internacional de Doenças
DARPA	<i>Defense Advanced Research Projects Agency</i>
DHP	<i>Hashing and Pruning</i>
DIC	<i>Dynamic Itemset Counting</i>
FMLE	<i>Fuzzy Maximum Likelihood Estimates Clustering Algorithm</i>
GSP	<i>Generalized Sequential Pattern</i>
IA	Inteligência Artificial
NLP	Processamento de Linguagem Natural
OHDSI	<i>Observational Health Data Sciences and Informatics</i>
OMOP	<i>Observational Medical Outcomes Partnership</i>
PEP	Prontuário Eletrônico do Paciente
SDAE	<i>Stacked Denoising Autoencoder</i>
SVM	<i>Support Vector Machine</i>
VSM	<i>Vector Space Model</i>
XAI	<i>Explainable Artificial Intelligence</i>

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Motivação	23
1.2	Objetivos do trabalho	24
1.3	Organização do texto	25
2	CONCEITOS E TRABALHOS RELACIONADOS	27
2.1	Considerações iniciais	27
2.2	Detecção de padrões (DP)	27
2.2.1	<i>Trabalhos relacionados - Detecção de padrões</i>	29
2.3	Inteligência artificial explicável (XAI)	31
2.3.1	<i>Problemas em métodos caixas pretas</i>	35
2.3.2	<i>Aspectos éticos no uso de dados</i>	36
2.3.3	<i>Trabalhos relacionados - Inteligência artificial explicável</i>	36
2.4	Discussão dos trabalhos relacionados	37
2.5	Instanciação do problema	39
2.5.1	<i>Interoperabilidade entre as bases de dados</i>	39
2.5.2	<i>Explicabilidade na detecção de padrões</i>	42
2.6	Considerações finais	43
3	O MÉTODO MCE (MULTILEVEL CLUSTERING EXPLAINER)	45
3.1	Visão geral	45
3.2	A representação BOAR	46
3.2.1	<i>Seleção do cohort</i>	46
3.2.2	<i>Construção da representação</i>	47
3.2.3	<i>Resultado Final da Representação</i>	49
3.3	Etapa 1- Preparação dos dados (MCE)	49
3.4	Etapa 2- Construção da arquitetura dos explicadores (MCE)	50
3.4.1	<i>Explicador local</i>	51
3.4.2	<i>Explicador global</i>	52
3.5	Etapa 3- Preparação visual (MCE)	54
3.6	Considerações finais	56
4	RESULTADOS E DISCUSSÃO	57
4.1	Considerações iniciais	57

4.2	OMOP Common Data Model (CDM)	57
4.3	Experimentos BOAR	58
4.3.1	<i>Resultados e discussões</i>	60
4.4	Experimentos MCE	63
4.4.1	<i>Resultados e discussões</i>	63
4.5	Considerações finais	72
5	CONCLUSÃO	75
5.1	Considerações finais	75
5.2	Contribuições	75
5.3	Trabalhos publicados	76
5.4	Outras atividades científicas	76
5.5	Trabalhos futuros	77
	REFERÊNCIAS	79



INTRODUÇÃO

1.1 Motivação

Com os avanços tecnológicos para comunicação e o surgimento de dispositivos inteligentes, a quantidade de dados digitais gerados vem crescendo exponencialmente, colaborando com a criação de grandes bases de dados (NIE *et al.*, 2018). Porém, para que esses dados se transformem em informações relevantes em seu domínio, é necessário o uso de técnicas para o seu gerenciamento e processamento.

No contexto de saúde, também é possível observar o alto crescimento no volume de dados digitais na década atual, estimado em *Exabytes* (10^{18}) (BERGER; DOBAN, 2014). Além do crescimento do volume dos dados, a variabilidade desses dados tem apresentado um crescimento expressivo, devido a novos exames e a necessidade de novas abordagens para interpretação dos dados são continuamente requisitados. Pode-se citar diferentes tipos de dados encontrados na área da saúde, por exemplo prontuário eletrônico do paciente, imagens de exames (2D e 3D), dados estruturados de exames (por exemplo, hemograma), sinais eletrônicos de sensores e registros de monitoramento de pacientes (LIN; LIN; CHEN, 2019; GOMES *et al.*, 2018).

Além da disponibilização dos dados, faz-se necessário a extração de conhecimento e apoio à tomada de decisão, que não são tarefas triviais (KAUSHAL; SHOJANIA; BATES, 2003). A captação, o armazenamento, o gerenciamento, o processamento dos dados e a validação dos conhecimentos extraídos são atividades complexas no contexto computacional. Com isso, existem diversos desafios a serem superados para que o desenvolvimento de métodos para o apoio a decisões médicas, como as distintas modelagens de dados. Um dos desafios, é a diferença entre os modelos de dados que armazenam os registros médicos eletrônicos em diferentes instituições, a diferença dos modelos de dados é resultado da particularidade de cada instituição implementando suas próprias terminologias, códigos e esquemas. Consequentemente, proporcionando uma barreira para divulgação científica, devido à dificuldade na reprodutividade

dos métodos de extração de conhecimento que estão fora do contexto da instituição ([NHSI, 2020](#)).

Outro desafio está relacionado à interpretabilidade dos dados obtidos pelos métodos de AM em tarefas de tomada de decisão. A falta de clareza no funcionamento e resultados apresentados, são encontradas principalmente nos métodos denominados caixas-pretas, pois não possuem a capacidade de interpretabilidade no seu processo de aprendizado e no conhecimento obtido ([ADADI; BERRADA, 2018](#)). Além disso, causam uma lacuna entre o conhecimento obtido e o motivo pelo qual as decisões foram tomadas. Consequentemente, dificultam a compreensão do especialista ao analisar ou validar os resultados obtidos por execuções de métodos de AM. Ambos os desafios são apresentados detalhadamente na [Seção 2.5](#).

Em algumas áreas de aplicações, como na medicina, em sistemas financeiros e judiciais, apenas resultados quantitativos dos métodos não são suficientes. É importante também que o usuário entenda o funcionamento e o motivo pelo qual o método tomou uma determinada decisão ([NEERINCX et al., 2018](#)). A área de estudo XAI (Inteligência artificial explicável), do termo em inglês *Explainable Artificial Intelligence*, objetiva investigar métodos com inferência explicável e formas de explicar o conhecimento adquirido pelos métodos não explicativos.

Alguns trabalhos demonstram a importância da adoção do XAI no contexto médico ([XIE; CHEN; GAO, 2019](#); [MATHEWS, 2019](#)). Entanto, alguns métodos são desenvolvidos para abordagens gerais e acabam entregando informações irrelevantes para o contexto médico, desta forma, dificultando a interpretação dos resultados obtidos. Em alguns trabalhos, como no de [Ribeiro, Singh e Guestrin \(2016a\)](#) oferecem explicabilidade apenas para o nível local, já em [Lundberg e Lee \(2017a\)](#), os autores lidam com explicações locais e global, no entanto não são recomendados para dados esparsos. No trabalho de [Cavaliere et al. \(2020\)](#) lidam com registros médicos eletrônicos especificamente, mas não apresentam explicações visuais. Ambas abordagens citadas trabalham com tarefas supervisionadas. Neste mestrado é proposto um método de explicação para tarefas não supervisionadas. As tarefas não supervisionadas entregam resultados descritivos, no contexto do trabalho são utilizados métodos de agrupamento para encontrar padrões sobre pacientes e diagnósticos similares a partir dos procedimentos realizados. Os detalhes dos objetivos desta pesquisa são apresentados mais detalhadamente na próxima seção.

1.2 Objetivos do trabalho

Este projeto de mestrado teve como objetivo desenvolver métodos pertencentes ao campo de pesquisa de XAI, proporcionando transparência e interpretabilidade dos resultados obtidos no processo de descoberta de padrões, por meio de métodos de agrupamento em registros médicos eletrônicos. Além disso, proporcionar aos especialistas da área médica a capacidade de entender o motivo pela qual as informações foram obtidas, favorecendo a descoberta de possíveis padrões

escondidos e também a validação do conhecimento. Ainda, este trabalho propôs uma nova abordagem de representação vetorial denominada *Bag-of-Attribute Representation* (BOAR). Além disso, proporciona ao método a flexibilidade de atuar em diferentes tipos de *cohorts* em bases de dados distintas. Este trabalho de mestrado contribui para a recuperação e interpretação de informação em bases de dados nos seguintes aspectos:

Considerando o apresentado anterior, é possível sumarizar as principais contribuições deste projeto de mestrado, como:

- Criação de uma nova representação vetorial baseada em atributos extraídos de registros médicos eletrônicos padronizados.
- Desenvolvimento de um método não supervisionado baseado no agrupamento, que proporcione interpretabilidade e transparência nos resultados obtidos, por meio de explicações multinívias visuais.

Neste trabalho os dados de registro médicos eletrônicos utilizados foram disponibilizados pela colaboração que o Grupo de Base de Dados e Imagens (GBdI) do ICMC-USP mantém com o Instituto do Coração (InCor) por meio do Projeto MIVisBD¹.

1.3 Organização do texto

Neste capítulo, foi apresentada a introdução deste projeto, abordando os conceitos iniciais, motivação e objetivos. Este trabalho está dividido em 5 capítulos que são descritos a seguir:

- No [Capítulo 2](#), são apresentados os conceitos fundamentais necessários para compreensão da pesquisa realizada. Primeiramente, os conceitos envolvendo a descoberta de conhecimento, seguido dos trabalhos relacionados da área são sumarizados. Além disso, são apresentados os conceitos relevantes sobre o XAI e os principais trabalhos relacionados que pertencem ao campo de pesquisa. Finalmente, são apresentadas discussões dos trabalhos relacionados de ambas abordagens, apresentando as problemáticas presentes.
- No [Capítulo 3](#), são apresentados os novos métodos desenvolvidos: *Bag-of-Attribute Representation* e *Multilevel Clustering Explainer*.
- No [Capítulo 4](#), são apresentados os experimentos, resultados e discussão dos métodos.
- Finalmente, no [Capítulo 5](#) são apresentados a conclusão, contribuições desta dissertação e os trabalhos futuros almejados.

¹ <<http://gbdi.icmc.usp.br/project/MIVisBD/index.html>>



CONCEITOS E TRABALHOS RELACIONADOS

2.1 Considerações inciais

Neste capítulo, são apresentados os conceitos fundamentais do trabalho desenvolvido. Na [Seção 2.2](#) é discutida a detecção de padrões na tarefas de descoberta de novos conhecimentos, visando as principais abordagens relacionadas à proposta deste trabalho de mestrado e também os trabalhos correlatos à proposta. Na [Seção 2.3](#) são apresentados os principais conceitos e definições sobre o XAI, e também os trabalhos relacionados ao proposto neste trabalho. Na [Seção 2.4](#) é apresentada a sumarização e discussão dos trabalhos relacionados. A [Seção 2.5](#) detalha a definição das problemáticas em que este trabalho objetivou. E por fim, na [Seção 2.6](#) as considerações finais deste capítulo.

2.2 Detecção de padrões (DP)

Detecção, reconhecimento, descrição, classificação ou agrupamento de padrões são focos de pesquisa explorados desde antes dos anos 60, que foram iniciadas em pesquisas teóricas na estatística ([THEODORIDIS; KOUTROUMBAS, 1998](#)). Contudo, com o advento da evolução dos computadores, diferentes campos de pesquisas como: biologia, psicologia, medicina, mercado de ações, visão computacional, estatística, sensoriamento remoto e inteligência artificial (IA) passaram a se preocupar com tais conceitos ([THEODORIDIS; KOUTROUMBAS, 1998; JAIN; DUIN; MAO, 2000](#)).

Em [Watanabe \(1985\)](#), padrão é definido como o contrário de caos, sendo uma entidade incapaz de definir em uma única e concreta explicação. Então, o autor afirma que um padrão pode ser observável, por exemplo, em uma parte da imagem de uma fotografia, uma palavra presente em um texto, uma expressão do rosto humano, sinais sonoros ou qualquer comportamento que

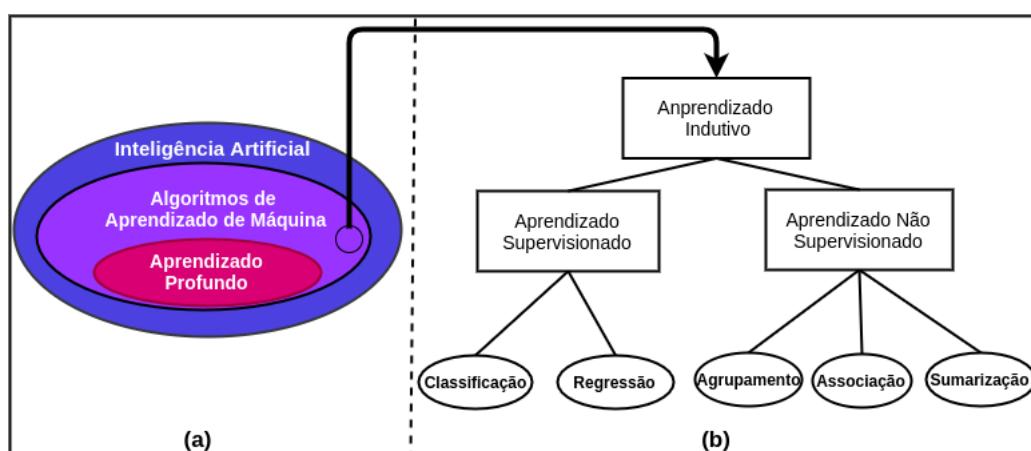
possa ser observável e inferido. Já para [Theodoridis e Koutroumbas \(1998\)](#), o campo de pesquisa de reconhecimento de padrões visa a classificação de objetos (padrões) em uma determinada classe ou categoria.

No contexto da IA, o reconhecimento de padrões é utilizado principalmente em algoritmos de Aprendizado de Máquina (AM), que são constituídos também por métodos estatísticos na mineração e descoberta de conhecimento ([ZHANG; YU; LI, 2019](#)). A [Figura 1](#) (a) apresenta a hierarquia da IA. Já na [Figura 1](#) (b) é apresentada a arquitetura do método indutivo no AM, na qual pode-se observar dois principais paradigmas: o aprendizado supervisionado e não supervisionado e seus respectivos modelos ([FACELI et al., 2011](#)).

No aprendizado supervisionado, após determinar um conjunto de dados (entrada) o algoritmo tem a função de encontrar ou prever uma ou mais classes (*targets*) que já são previamente conhecidas ([BURKOV, 2019](#)). Exemplificando, após observar a ficha de um paciente, a qual contém informações sobre a ocorrência de alguns sintomas, tais como: febre, dor de garganta e tosse (dados de entrada), o algoritmo é responsável por prever se o paciente será ou não diagnosticado com gripe (classe).

Já no paradigma não supervisionado, não existe uma classe pré-definida (*target*), o algoritmo é responsável por buscar por informações ainda não conhecidas em um conjunto de dados (entrada), preocupando-se com a descrição, sumarização e análise de padrões nos dados ([ZHAO; WANG, 2020](#)). Exemplificando, a partir da disponibilização de um conjunto de dados de históricos escolares (entrada), o algoritmo mostra que a frequência de presenças nas aulas influencia exclusivamente na disciplina de matemática.

Figura 1 – Arquitetura e hierarquia da IA, supervisionado e não supervisionado



Fonte: Adaptada de [Faceli et al. \(2011\)](#).

A presença de conceitos de reconhecimento de padrões pode ser encontrada em ambos paradigmas, supervisionado e não supervisionado ([GUO et al., 2015](#)). No paradigma supervisionado, relaciona-se a métodos capazes de mapear dados de entrada em seus respectivos rótulos (ou respostas desejadas - valores discretos ou contínuos), assim, espera-se que os dados

de treinamento sejam conhecidos (NOIA *et al.*, 2019). Por outro lado, no aprendizado não supervisionado, os padrões não são conhecidos e por isso tem o objetivo de descobri-los (KHALID; JUDGE; PINEDO-VILLANUEVA, 2018). No trabalho tem-se como objetivo encontrar padrões desconhecidos e fornecer explicações ao usuário. Sendo assim, é importante destacar algumas tarefas que atuam neste paradigma, tais como: agrupamento, associação e sumarização. Neste trabalho de mestrado, o objetivo é encontrar padrões desconhecidos por meio de métodos de agrupamento, que priorizem a explicabilidade dos resultados. A seguir é apresentado o conceito básico deste tipo de tarefa e alguns de seus principais algoritmos.:

Agrupamento: é a tarefa de identificar exemplos semelhantes em conjuntos de dados não rotulados e criar grupos compostos pelos respectivos elementos semelhantes (OMAR, 2003). Nos métodos propostos nesta dissertação, é possível utilizar algoritmos de agrupamento, por exemplo, a partir da identificação de pacientes que possuem diagnóstico semelhante e seus respectivos grupos de semelhança (HE; QIN; WANG, 2005). Para uma questão de padronização o termo grupo no decorrer do texto será substituído por *cluster*. A seguir são descritos alguns métodos de agrupamento:

- K-Means: considerando um conjunto de registros (R) é estabelecido um parâmetro de entrada (k) que corresponde a quantidade de *clusters* (agrupamentos) a serem formados. Para cada *cluster* (k_i) é determinado um ponto central chamado centróide (C_i), em seguida para cada registro do conjunto de dados (R_i) tem a sua distância calculada entre todos os centróides (C). O registro é atribuído ao centróide com menor distância, e exclusivamente a um centróide. Em cada iteração a posição do centróide é atualizada no sentido de identificar o centro real do grupo de elementos que lhe são atribuídos. O processo é finalizado quando todos os elementos estão atribuídos a algum centróide e quando os centróides não possuem suas posições atualizadas em um novo ciclo, ou quando o número máximo de iterações é alcançado (HAN *et al.*, 2017).
- Agrupamento Hierárquico Aglomerativo: do termo em inglês *Agglomerative Hierarchical clustering*, método que segue a abordagem *bottom-up* (de baixo para cima). Pois dado um conjunto de objetos, cada objeto inicialmente é considerado um *cluster* de um só elemento. Na sequência, é utilizada uma medida de distância definida pelo usuário, cujo objetivo é medir a similaridade entre os objetos. Desta forma, realizando a aglomeração dos objetos em *clusters* aninhados. Uma das vantagens é a flexibilidade do método de lidar com qualquer tipo de medida de distância (FACELI *et al.*, 2011).

2.2.1 Trabalhos relacionados - Detecção de padrões

No trabalho de Simić *et al.* (2018) foi desenvolvida uma estratégia baseada em um modelo híbrido, que combina o método de partição *fuzzy* e estimativas de máxima verossimilhança no diagnóstico de doenças médicas. O algoritmo híbrido tem a capacidade de detectar

agrupamentos de formas, tamanhos e densidades distintas. No primeiro momento, o algoritmo é responsável por extrair as regras dos dados e estimar o número ideal de *clusters*. Em seguida é proposta a combinação do método de partição *fuzzy* e estimativa de uma norma de distância de máxima verossimilhança difusa. Segundo os autores, os resultados obtidos pelo algoritmo proposto *Fuzzy Maximum Likelihood Estimates Clustering Algorithm* (FMLE) são animadores e foram testados em dois conjuntos de dados médicos diferentes, o qual apresentaram valor de acurácia de 78% e 58% respectivamente. Para trabalhos futuros, os autores demonstraram interesse na melhoria da eficiência do método e a possibilidade de realizar os experimentos em um conjunto de dados reais.

Em [Atashi et al. \(2019\)](#) foi proposto um método para encontrar padrões desconhecidos em um conjunto de dados reais de câncer de mama. A busca por padrões foi feita por meio de técnicas de mineração de dados e acompanhamento de profissionais da área da saúde. No primeiro momento, foi aplicado o algoritmo de regra de associação APRIORI para encontrar as regras iniciais. Após a extração, foram encontradas 13100 regras, que posteriormente o ajuste de parâmetros diminuiu para 1206, das quais foram selecionadas as 100 primeiras regras. Em seguida, a equipe de profissionais classificou 10 regras como relevantes no contexto, os padrões encontrados foram considerados para estudos clínicos específicos e pesquisas futuras. A principal limitação do trabalho pode ser atribuída ao alto número de regras descartadas, possibilitando a perda de regras ocultas. Outra limitação que pode ser citada é o baixo número de registros da base de dados.

Já no trabalho de [Garcelon et al. \(2017\)](#), o objetivo foi encontrar pacientes que compartilhavam padrões semelhantes, utilizando medidas de similaridade em um estudo de doenças raras, por meio de dados médicos em prontuários médicos. O modelo proposto foi baseado no *Vector Space Model* (VSM) para calcular a similaridade de um paciente selecionado com todos os demais pacientes do banco de dados. O conjunto de dados utilizado é composto por dados clínicos reais com cerca de 400.000 pacientes. Quanto ao formato dos dados, deve-se enfatizar que foram extraídos de relatórios de texto livre e não estruturados de prontuários eletrônicos. Para a avaliação e comparação dos pacientes foram levadas em consideração 5 tipos de doenças raras. Além do método buscar por pacientes semelhantes, pode-se observar também que foram encontrados pacientes não diagnosticados, mas que apresentavam características semelhantes. Quanto à visualização de similaridade, após selecionar um paciente denominado "índice" o método retorna um grafo com as ligações dos 20 pacientes mais similares. A espessura das arestas representa o grau de similaridade entre os pacientes. Os autores apresentam algumas limitações no trabalho, relacionadas ao baixo nível de precisão, que é atribuído à independência em que o texto livre foi representado e analisado, mas afirmam que é possível entender e melhorar os resultados, por meio da adição de novas informações, como códigos CID e resultados de exames laboratoriais.

O trabalho de [Sushil et al. \(2018\)](#) propôs um método para criação de representações de

pacientes, de forma independente, por meio de anotações de prontuários médicos eletrônicos. O método testou várias combinações de tarefas supervisionadas para prever a mortalidade do paciente, diagnóstico primário e a categoria do procedimento. Para o desenvolvimento da proposta foram utilizados os conceitos de *Bag-of words* (BOW) (ZHANG; JIN; ZHOU, 2010) para sustentar o método desenvolvido, nomeado como *Bag-of-medical-concepts* (BoCUI), que foi implementado importando os conceitos médicos do *Unified Medical Language System* (UMLS). O BoCUI apresentou bons resultados apenas na categoria dos diagnósticos, porém nos demais aspectos o método é prejudicado na tarefa de classificação, devido à presença de conceitos desconhecidos. Uma solução sugerida pelos autores, é a adição de novos dados e adequação para atuação em dados estruturados.

2.3 Inteligência artificial explicável (XAI)

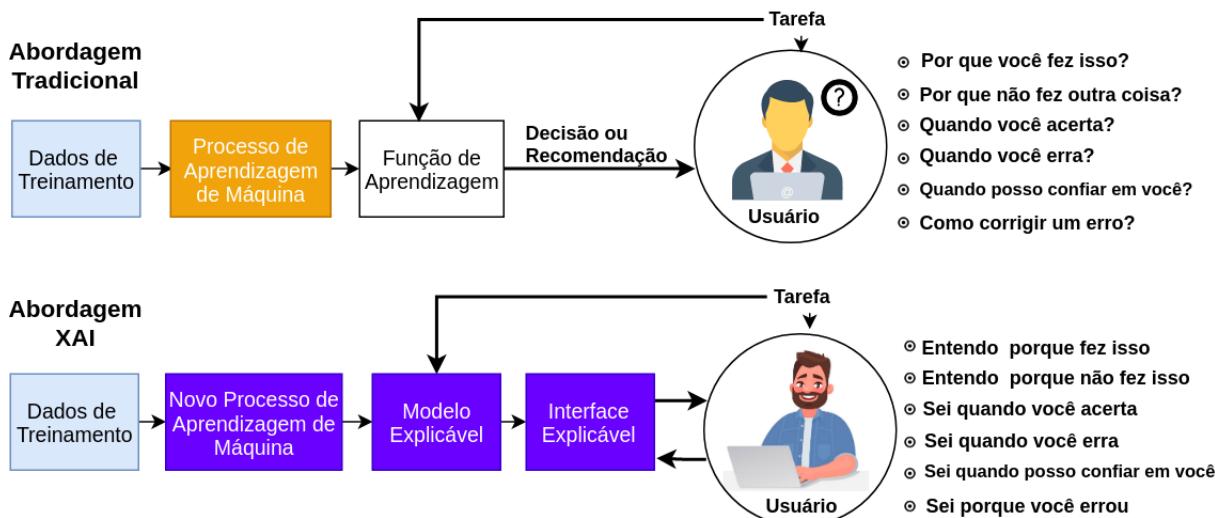
É notório que a IA é cada vez mais utilizada, principalmente em ambientes de tomada de decisões, tais como: auxiliando médicos em diagnósticos, apoio às empresas na auditabilidade de seus clientes, apoio em decisões judiciais e também em controle de veículos (RIBERA; LAPEDRIZA, 2019). No entanto, a popularização dos métodos de IA (principalmente de AM) fez com que outras áreas passem a utilizar tais métodos. Contudo, os profissionais que recebem os resultados obtidos não são especialistas em IA, justificando a necessidade dos métodos serem capazes de ser explicáveis e suas respostas encontradas sejam bem fundamentadas. Desse modo, esses métodos e respostas podem contribuir para um melhor entendimento do usuário e podem colaborar também no processo de tomada de decisão (DOŠILOVIĆ; BRČIĆ; HLUPIĆ, 2018).

Neste contexto, um campo de pesquisa que está emergente na área de IA é a Inteligência Artificial Explicável, do termo em inglês *Explainable Artificial Intelligence* (XAI), que tem como objetivo tornar os resultados disponibilizados por sistemas de IA mais entendíveis para os humanos. Devido a esse rápido crescimento da área de pesquisa, ainda não possível apontar uma definição para o conceito ou termo que seja amplamente consolidado na literatura (MATHEWS, 2019; ADADI; BERRADA, 2018). No entanto, em 2004 o termo explicabilidade já tinha sido utilizado pela primeira vez, no qual foi proposta uma arquitetura para fornecer a capacidade de explicabilidade de agentes em um sistema de simulação de jogos (LENT; FISHER; MANCUSO, 2004). Desde então, o interesse pelo campo de pesquisa vem crescendo e tem sido descrito e nomeado de diferentes maneiras visando oferecer explicação e transparência nas decisões nos sistemas opacos (PÁEZ, 2019). Alguns autores até classificam o XAI como a terceira onda da IA denominada Inteligência Artificial 3.0 (ADADI; BERRADA, 2018). Consequentemente, diferentes nomenclaturas do termo são empregadas além de explicabilidade e interpretabilidade, tais como: compreensibilidade, entendibilidade e inteligibilidade. A maioria dos trabalhos científicos publicados utilizam o termo interpretabilidade para referir-se ao XAI (ADADI; BERRADA, 2018).

Além dos trabalhos científicos, a *Defense Advanced Research Projects Agency* (DARPA-Agência de Projetos de Pesquisa Avançada de Defesa) lançou em 2018 o programa XAI, que tem como objetivo criar um conjunto de técnicas, ferramentas e bibliotecas que produzam modelos de AI mais explicáveis e que simultaneamente não prejudiquem a precisão de respostas dos algoritmos. Outro objetivo do programa é permitir que os usuários humanos entendam, confiem apropriadamente e gerenciam as saídas providas dos modelos de IA ([DARPA, 2018](#)).

Na [Figura 2](#) é apresentada a abordagem do conceito XAI proposto pela DARPA, ilustrando as mudanças que devem ser consideradas em todo processo de descoberta de conhecimento. Além disso, destaca as principais diferenças entre a abordagem tradicional com a do XAI, priorizando o conhecimento do usuário após receber as informações da tarefa realizada.

Figura 2 – Proposta do conceito XAI



Fonte: Adaptada de [DARPA \(2018\)](#).

No trabalho de [Guidotti et al. \(2018\)](#) o principal propósito que envolve o XAI é aplicar explicadores para fornecer informações que facilitam o entendimento do funcionamento de um método. Os autores definem um método caixa-preta, como aquele que não é possível entender o motivo pelo qual suas decisões foram tomadas. [gari et al. \(2019\)](#) completa que um explicador é qualquer procedimento que gera significado ao ser humano, por meio da combinação de: dados de entrada, modelo de aprendizado, saída e as decisões ou previsões realizadas. Além do conceito de caixa-preta, outros conceitos podem ser considerados importantes para a compreensão e contextualização do XAI. Os autores ([ADADI; BERRADA, 2018; MATHEWS, 2019; LENT; FISHER; MANCUSO, 2004](#)) citam alguns dos conceitos:

- **Inteligência Artificial (IA):** um campo de pesquisa que tem por interesse fazer as máquinas aprenderem ou mesmo que imitem a inteligência humana, por meio de algoritmos e experiências.

- **Aprendizado de Máquina (AM):** uma subárea da AI, onde são propostos algoritmos responsáveis por descobrir novos conhecimentos.
- **Aprendizado Profundo:** uma subárea do AM, na qual as redes neurais profundas estão inseridas. Estas, possuem capacidade de lidar com grandes conjuntos de dados, mas não tem alto poder de explicabilidade.
- **Aprendizado Supervisionado e Não Supervisionado:** tipos de paradigmas de algoritmos de AM. O primeiro é baseado na classificação e previsão dos exemplos pré-rotulados. Já no segundo, está relacionada à capacidade de compreensão e descrição de um conjunto de dados sem rótulos.
- **Problemas de modelos caixa-preta:** modelos e algoritmos com alto poder preditivo, porém que não apresentam capacidade de interpretabilidade quanto ao seu funcionamento e as explicações sobre os motivos pelos quais foram apresentados determinados resultados.
- **Acurácia:** mede a capacidade de um algoritmo de AM prever corretamente um resultado esperado. O resultado obtido pela medida está entre 1 e 0. Valores mais próximos de 1 são considerados melhores.
- **IA Responsável:** preocupa-se com os princípios éticos, morais e valores sociais sobre a perspectiva das informações obtidas, levando em consideração três principais conceitos: prestação de conta, responsabilidade e transparência.

Como o campo de pesquisa ainda é recente, a classificação dos modelos para serem considerados aceitáveis no XAI não seguem um critério único. Deste modo, [Adadi e Berrada \(2018\)](#) apresentam uma taxonomia explicativa para auxiliar no processo de comparação entre os métodos existentes. A [Figura 3](#) mostra que o método de explicabilidade pode ser classificado em dois tipos: *Intrínseco* e *Post-hoc*, que são definidos a seguir:

- **Modelos intrínsecos:** nesta classe os modelos já possuem interpretabilidade em sua própria definição e funcionamento. Exemplos destes modelos podem ser citados: árvore de decisão, regras de associação, modelos lineares, simples naïve Bayes ([DOŠILOVIĆ; BRČIĆ; HLUPIĆ, 2018; PÁEZ, 2019](#)).
- **Modelos Post-hoc:** nesta classe estão os modelos denominados caixas pretas, nos quais não apresentam uma linha de raciocínio clara relativa aos resultados obtidos. Alguns modelos podem ser citados: redes neurais profundas, máquina de vetores de suporte e árvores aleatórias ([PÁEZ, 2019; CORTEZ; EMBRECHTS, 2011](#))

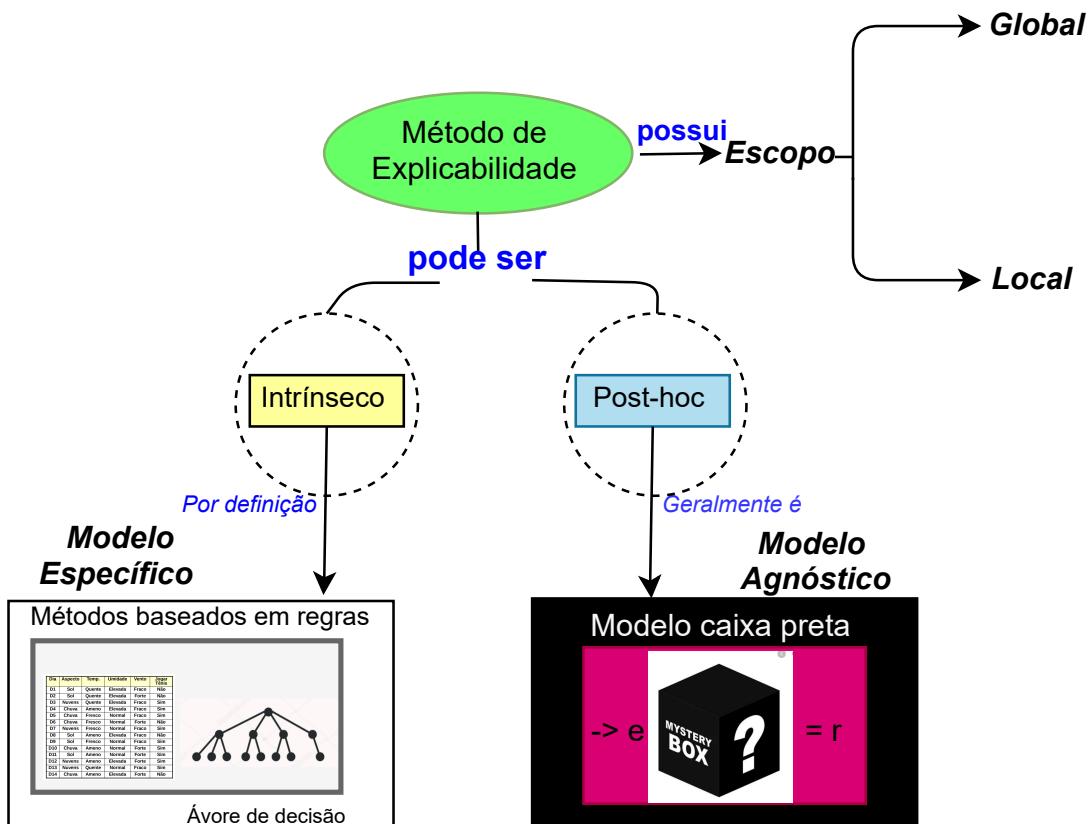
Em sua grande parte os modelos *Post-hoc* são denominados modelos agnósticos, nos quais não é possível determinar o motivo pelo qual uma entrada proporciona um resultado. Já nos modelos intrínsecos que são denominados modelos específicos por sua própria definição, é

possível percorrer um caminho lógico da entrada de dados até as decisões tomadas pelo modelo, facilitando a interpretabilidade das respostas obtidas.

Outro fator importante na taxonomia dos modelos é o seu escopo. A Figura 3 ilustra modelos distintos e seus respectivos escopos, seja ele global ou local. Na sequência são apresentadas as definições:

- **Escopo Global:** refere-se ao entendimento dos resultados em um escopo geral do modelo, ou seja, é capaz de demonstrar todo o percurso lógico que o modelo percorre para encontrar todas as decisões tomadas pelo modelo.
- **Escopo Local:** refere-se ao entendimento de uma única decisão específica, ou seja, analisando localmente cada instância de uma decisão tomada.

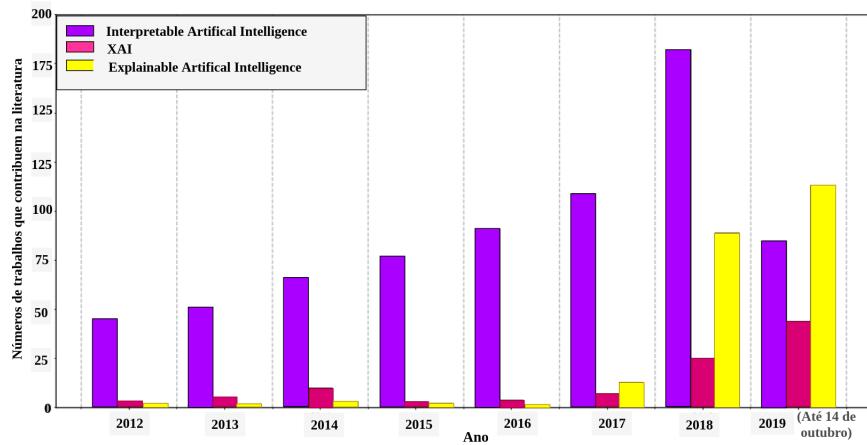
Figura 3 – Taxonomia do método de explicabilidade



Fonte: Adaptada de [Adadi e Berrada \(2018\)](#).

O crescimento do interesse na pesquisa sobre o XAI é notório, conforme ilustra a Figura 4. [Arrieta et al. \(2019\)](#) mostram que nos últimos oito anos vem surgindo um aumento significativo na busca por este novo campo de pesquisa. Os dados encontrados foram obtidos na base de dados da *Scopus* atualizado até o dia 14 de outubro de 2019. Portanto, é possível observar a crescente adesão e contribuições do XAI que vem favorecendo o interesse de instituições governamentais e privadas.

Figura 4 – Interesse em XAI ao longo do tempo



Fonte: Adaptada de [Adadi e Berrada \(2018\)](#).

2.3.1 Problemas em métodos caixas pretas

Na computação termos como caixa-preta, caixa-cinza e caixa-branca são utilizados para expressar diferentes níveis de funcionamento da parte interna de componentes, modelos ou técnicas ([WU; LIU; XIAO, 2011](#)). No cenário da IA, métodos caixas-pretas em sua grande maioria, estão presentes na subárea de AM denominada aprendizado profundo. Os modelos caixas-pretas se destacam entre os métodos de AM pela sua capacidade de muitas vezes encontrarem previsões poderosas. Porém, um dos obstáculos da adoção destes modelos está diretamente ligado à falta de transparência que estes modelos proporcionam.

Grande parte dos métodos caixas-pretas apresentam bons resultados de acurácia, mas existe a dificuldade de compreender o funcionamento e as respostas providas pelo modelo. Isto consequentemente acaba impedindo com que as pessoas confiem absolutamente nos resultados indicados ([GARI et al., 2019](#)). Essa desconfiança faz com que áreas tais como: financeira, saúde, militar e judicial ([ADADI; BERRADA, 2018; MATHEWS, 2019](#)) restrinjam a adoção desses métodos, devido à necessidade de expressar o motivo pelo qual a decisão foi tomada e envolver aspectos éticos e legais e discriminatórios ([GARI et al., 2019; LEPRI et al., 2018; ARRIETA et al., 2019](#)).

Algoritmos de aprendizado profundo são considerados caixas-pretas, em ([DOŠILOVIĆ; BRČIĆ; HLUPIĆ, 2018](#)) eles são criticados pelo fato de seu funcionamento ser considerado como aproximações em que as respostas obtidas nem sempre são confiáveis. Além disso, a falta de confiança e a presença de resultados tendenciosos também são problemas encontrados em reconhecimento de padrões visuais ([SZEGEDY et al., 2014](#)) e também no processamento de linguagem natural ([WOLF; MILLER; GRODZINSKY, 2017](#)).

Um problema clássico que gerou repercussão foi um algoritmo de classificação de

imagens de pessoas, desenvolvido pela Google em 2015, no qual classificou um ser humano com uma *tag* de gorila. No entanto a solução apresentada pela empresa em 2018, na qual apenas removeram a *tag* incorreta do banco de dados, não agradou aos críticos que mais tarde catalogaram como uma alternativa embaracosa (BARR, 2015). Outra pesquisa realizada por Obermeyer *et al.* (2019), que um algoritmo comercial utilizado por organizações de saúde apresentava resultados tendenciosos para pacientes negros.

2.3.2 Aspectos éticos no uso de dados

Com todo o crescimento e sucesso de métodos e algoritmos de AM em diversos setores, assim como toda tecnologia emergente, problemas e dificuldades também são encontradas. Em algumas áreas de aplicações é necessário maior atenção tanto na utilização dos dados quanto nas informações obtidas, principalmente em aplicações que envolvem aspectos éticos e legais dos indivíduos.

Segundo Jobin, Ienca e Vayena (2019), os principais pilares que envolvem os aspectos éticos são: transparência, justiça, equidade, não maleficência, responsabilidade e privacidade. Algumas medidas já vêm sendo tomadas de encontro a esses aspectos, a União Europeia inseriu o direito à explicação na Lei Geral de Proteção de Dados Pessoais. Acredita-se que a explicabilidade é um dos pré-requisitos para alcançar a transparência das informações obtidas, dando a oportunidade de verificar se os critérios de confiança dentro do contexto da aplicação foram atingidos (MATHEWS, 2019).

No Brasil foi proposto o projeto de lei nº 5051 de 2019¹, o qual estabelece os princípios para o uso da IA no Brasil. Ressaltando a importância de garantir o bem estar do ser humano, bem como oferecer: transparência, confiabilidade, o respeito à dignidade humana, à liberdade, à democracia e à igualdade. Um dos pontos mais ressaltados é a obrigatoriedade que todo sistema de IA deverá ter a supervisão de uma pessoa humana de maneira compatível com cada tipo de aplicação.

2.3.3 Trabalhos relacionados - Inteligência artificial explicável

No trabalho de Ribeiro, Singh e Guestrin (2016b) foi proposto o LIME uma técnica que é capaz de atuar em distintos modelos de AM. O trabalho foi dividido em dois momentos principais. No primeiro o LIME atuou na classificação de texto realizado por um algoritmo SVM, em que logo em seguida realiza a transformação para um modelo linear e apresenta ao usuário a linha de corte que determina a decisão tomada. No segundo o LIME trabalha com a classificação de imagens realizado por redes neurais, após a classificação da classe da imagem predita o LIME retorna somente os pixels que mais influenciaram a predição. O método proposto pelos autores tem por objetivo fornecer explicações simples, seja ela uma palavra específica em um documento

¹ <<https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>>

ou uma forma em uma fotografia que estão influenciando as previsões. Para validação do método levando em consideração na classificação de texto, o método atingiu mais que 90% na medida *recall*. Quanto ao processo de classificação em imagens o método atingiu 89% de acurácia em seu modelo.

[Shirataki e Yamaguchi \(2017\)](#) apresentam um método de classificação de texto, em específico em avaliações, positivas ou negativas de comentários. Para classificação do texto foi utilizado o algoritmo SVM, que atingiu até 90% de acurácia combinando com a técnica BOW. Para sustentar o princípio de interpretabilidade, os autores propuseram um método que após classificada a opinião do texto, são retornadas ao usuário as três palavras principais que influenciaram a classificação retornada. Para apresentação das palavras que foram retornadas é utilizado o conceito de atribuição de pesos das palavras principais, por meio da contagem da frequência que aparecem naquela ou nas demais opiniões.

Já no trabalho de [gari et al. \(2019\)](#) foi proposta uma arquitetura para realizar a tradução dos resultados gerados pelo algoritmo de árvore de decisão em uma explicação narrativa priorizando a construção do conhecimento lógico. Os autores afirmaram que a combinação entre técnicas simbólicas e sub-simbólicas podem produzir interpretabilidade para sistemas de IA. O trabalho se divide em três principais objetivos: a) criar uma explicação narrativa a partir de um resultado ou previsão de uma decisão; b) inferir novos conhecimentos; e c) agir em conformidade. Contudo, demonstram que dada uma determinada estrutura de árvore de decisão (simbólica) é possível para cada regra uma explicação textual. De modo que essa fusão das técnicas favorece a interpretabilidade, observabilidade e explicabilidade. O foco da pesquisa foi a criação de explicações narrativas, ocultando o processo anterior de aprendizado de máquina.

Portanto, é visível que a adoção e o uso de modelos interpretativos é uma boa opção para obter o conhecimento, seja em referência à transparência (entender como funciona o modelo) ou de interpretabilidade (entender como uma decisão foi obtida pelo modelo) ([PÁEZ, 2019](#)). A presente seção apresentou os principais conceitos que abrangem o campo de pesquisa do XAI e os trabalhos que estão relacionados com a proposta de pesquisa apresentada. Na sequência, na [Seção 2.4](#) são apresentados alguns modelos e técnicas que auxiliam na descoberta de um novo conhecimento, tema esse que também está ligado à proposta desta dissertação.

2.4 Discussão dos trabalhos relacionados

Após apresentar os principais conceitos fundamentais e os trabalhos relacionados dos dois campos distintos de pesquisa (XAI e detecção de padrões), é necessário discutir sobre estes trabalhos, destacando as principais diferenças envolvendo a proposta de pesquisa. Na [Tabela 1](#) é sumarizado os principais aspectos das diferentes abordagens, considerando nove aspectos de comparações: tipo de dado, formato de dado, abordagens utilizadas, contexto da aplicação, escopo-XAI, definição-XAI e classificação do modelo-XAI.

Os quatro primeiros trabalhos (RIBEIRO; SINGH; GUESTRIN, 2016b; SHIRATAKI; YAMAGUCHI, 2017; GARI *et al.*, 2019; LUNDBERG; LEE, 2017b) foram encontrados considerando a *string* de busca relacionada ao XAI. Já os demais (SIMIĆ *et al.*, 2018; ATASHI *et al.*, 2019; GARCELON *et al.*, 2017; SUSHIL *et al.*, 2018) priorizou-se a descoberta de conhecimento, principalmente aquelas utilizam a representação vetorial. Na sequência é apresentada uma discussão envolvendo cada aspecto discutido nos trabalhos:

Tipo de Tarefa de Aprendizado: Em AM existem duas principais abordagens a supervisionada e a não supervisionada, os conceitos e métodos pertencentes de cada uma já foram apresentados na Seção 2.2. Portanto, os trabalhos de (RIBEIRO; SINGH; GUESTRIN, 2016b; SHIRATAKI; YAMAGUCHI, 2017; LUNDBERG; LEE, 2017b) apresentam apenas tarefas de classificação e em gari *et al.* (2019) não detalham a tarefa, apenas priorizam apresentam as abordagens XAI utilizadas. Já nos trabalhos (ATASHI *et al.*, 2019; GARCELON *et al.*, 2017; SUSHIL *et al.*, 2018) utilizam somente métodos de agrupamento, mas não lidam com os aspectos de explicabilidade, enquanto (SIMIĆ *et al.*, 2018) utilizam uma abordagem híbrida combinando técnicas de classificação e agrupamento para encontrar o melhor número de *clusters*. O trabalho proposto por esta dissertação visa explorar os resultados de tarefas de agrupamento, com intuito de fornecer explicações ao profissional de saúde.

Formato dos Dados e Tipo de Dados: A maioria dos trabalhos são baseados em textos livres, ou seja, dados não estruturados para entrada dos seus respectivos métodos. Apenas em Ribeiro, Singh e Guestrin (2016b), Lundberg e Lee (2017b) que além de texto livre, imagens também são aceitas. No trabalho Shirataki e Yamaguchi (2017) não é especificado qual tipo de dados que foi utilizado pelo método. Enquanto na proposta do trabalho de mestrado é só utilizados apenas dados relacionais, herdados dos conceitos recuperados por meio dos *cohorts*.

Abordagens Utilizadas no Contexto da Aplicação: referem-se às abordagens, métodos ou técnicas utilizadas para solucionar o problema proposto por cada trabalho. Em (RIBEIRO; SINGH; GUESTRIN, 2016b; LUNDBERG; LEE, 2017b) pode ser utilizado qualquer algoritmo de classificação classificação, tanto de textos quanto de imagens. Segundo o autor o método é funcional indiferente do contexto da aplicação. Já em (SHIRATAKI; YAMAGUCHI, 2017) além do *Support Vector Machine* (SVM) foi utilizada também a abordagem de BOW na classificação e análise de satisfação de clientes, referente à compra de mercadorias. Em gari *et al.* (2019) os autores extraem regras narrativas (Prolog) das regras geradas, por meio de árvores de decisão para facilitar a compreensão de diagnósticos médicos. No trabalho de Simić *et al.* (2018) utiliza combinações entre lógica *Fuzzy* para integrar o algoritmo de agrupamento *Fuzzy maximum likelihood* (FMLE), em um conjunto de dados com diagnósticos de doenças. Em Garcelon *et al.* (2017) os autores utilizaram o modelo espaço vetorial (VSM) para buscar similaridades entre pacientes diagnosticados com doenças raras. Em Sushil *et al.* (2018) é utilizada também a combinação entre BOW, *Stacked Denoising Autoencoder* (SDAE) e um nova representação criada por eles para representar conceitos médicos, que foi denominada de *bag-of-concept-medicine*.

para representação de pacientes no contexto médico.

Escopo-XAI: referente ao escopo oferecido pelo métodos, seja ele pertencente à uma visão global ou local de explicabilidade, os trabalhos (RIBEIRO; SINGH; GUESTRIN, 2016b; SHIRATAKI; YAMAGUCHI, 2017) possuem características pertencentes ao escopo local e no trabalho de (LUNDBERG; LEE, 2017b) são considerados ambos os escopos. Já em (SIMIĆ *et al.*, 2018; ATASHI *et al.*, 2019; GARCELON *et al.*, 2017; SUSHIL *et al.*, 2018; GARI *et al.*, 2019) não foi levado em consideração ou não foi especificado este aspecto do XAI. No entanto, no trabalho de (GARI *et al.*, 2019) é possível inferir que o escopo pode ser atribuído a ambos. Já o método proposto nesta dissertação abrange tanto o escopo global quanto o local, devido às características herdadas no modo de representação da árvore de decisão e das regras simbólicas.

Definição e Modelo XAI: Quanto à definição e ao modelos seguindo a taxonomia do XAI, os trabalhos de (RIBEIRO; SINGH; GUESTRIN, 2016b; SHIRATAKI; YAMAGUCHI, 2017; SIMIĆ *et al.*, 2018; LUNDBERG; LEE, 2017b; GARCELON *et al.*, 2017; SUSHIL *et al.*, 2018) apresentam propriedades *post-hoc* característicos de modelos agnósticos. Já em (GARI *et al.*, 2019) a proposta de pesquisa tem por definição o modelo intrínseco por definição própria dos modelos.

A Tabela 1 sumariza os aspectos discutidos, visando facilitar o entendimento das particularidades em que a proposta deste mestrado se insere. Destaca-se que dentre os trabalhos analisados, a proposta deste mestrado é a única que lida com problemas de interoperabilidade de bases de dados médicas, com objetivo de encontrar os padrões desconhecidos por meio de algoritmos de agrupamento. E principalmente por fornecer aos usuários informações explicativas sobre os resultados encontrados, apresentando de maneira visual a importância das variáveis na composição dos *clusters*.

2.5 Instanciação do problema

O trabalho de pesquisa teve como objetivo primordial oferecer explicação para métodos de agrupamento. No entanto, notou-se uma dificuldade na comparação de diferentes métodos com diferentes bases de dados. Desse modo, pensou-se em desenvolver uma representação genérica que permita que o método de explicabilidade atue em diferentes métodos de agrupamento em bases de dados distintas. Desta forma, esta abordagem pode auxiliar o usuário a entender os diferentes métodos de agrupamentos por meio da comparação dos resultados interpretáveis. Na Subseção 2.5.1 é apresentado o problema da interoperabilidade entre bases de dados e na sequência, na Subseção 2.5.2 as limitações relacionadas à explicabilidade.

2.5.1 Interoperabilidade entre as bases de dados

Um dos desafios encontrados na área da saúde em nível mundial é a padronização e compatibilidade entre os diversos padrões de armazenamento de dados de saúde. É comum que

Tabela 1 – Tabela descritiva de comparação dos trabalhos relacionados

Autor	Tipo de Tarefa de Aprendizado	Tipo de Dados	Formato dos Dados	Abordagens utilizadas	Contexto da aplicação	(XAI) Escopo	(XAI) Intrínseco	(XAI) Post-hoc	Modelo Agnóstico	Modelo Específico
Ribeiro, Singh e Guestrin (2016b)	Classificação	Não estruturado	Texto Livre e Imagem	Indiferente	Indiferente	Local	Post-hoc	Agностico		
Lundberg e Lee (2017b)	Classificação	Não estruturado	Texto Livre e Imagem	Indiferente	Indiferente	Local e Global	Post-hoc	Agностico		
Shirataki e Yamaguchi (2017)	Classificação	Não estruturado	Texto Livre	SVM, BOW	Análise de satisfação de produtos: (CD, livros cerveja)	Local	Post-hoc	Agnostico		
gari <i>et al.</i> (2019)	Não se aplica	Não se aplica	Texto Livre	Árvore de decisão e PROLOG	Diagnóstico de doenças:	Local	Intríseco	Específico		
Simić <i>et al.</i> (2018)	Híbrido	Não estruturado	Texto Livre	FMLE, Lógica de Fuzzy	Diagnóstico de doenças: Ecocardiograma, Hepatite e Distúrbios hepáticos	Não se aplica	Post-hoc	Agностico		
Atashri <i>et al.</i> (2019)	Agrupamento	Não estruturado	Texto Livre	APRIORI	Detecção de padrões em pacientes diagnosticados com câncer de mama	Não se aplica	Intríseco	Específico		
Garcelon <i>et al.</i> (2017)	Agrupamento	Não estruturado	Texto Livre	VSM	Similaridade entre pacientes diagnosticos de sete doenças raras	Não se aplica	Post-hoc	Agностico		
Sushil <i>et al.</i> (2018)	Agrupamento	Não estruturado	Texto Livre	BOW, BOW, SDAE	Representação de pacientes quanto à mortalidade, categoria primária de diagnóstico e sexo	Não se aplica	Post-hoc	Agnostico		
Proposta de pesquisa realizada neste mestrado	Agrupamento	Estruturado	Texto estruturado	<i>Bag-of-Attribute Representation</i> , Agnóstico com relação a algoritmos de agrupamento	Detecção de padrões no tratamento sobre qualquer diagnóstico médico presente no banco de dados	Globa e Local	Intríseco	Específico		

Fonte: Dados da pesquisa.

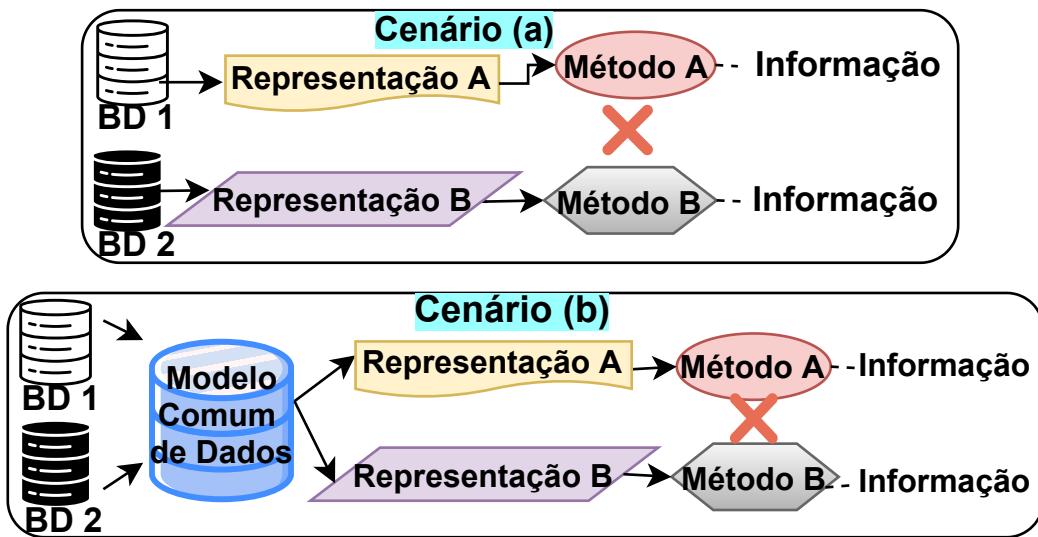
cada instituição possua sua infraestrutura para gerenciar os dados oriundos de suas atividades da maneira que melhor satisfaça suas necessidades (LOUIE *et al.*, 2007; LAVRAS, 2011; RODRIGUES *et al.*, 2020). O desenvolvimento de soluções tecnológicas sob demanda de acordo com a necessidade favorece o desenvolvimento de diversos tipos de sistemas e modelos de dados (QADRI *et al.*, 2020). O desenvolvimento fragmentado de sistemas para prontuário eletrônico tem sido visto como um desastre sanitário e econômico (JALEEL *et al.*, 2020), pois dificulta a recuperação de informações relacionadas a uma ocorrência e o desenvolvimento em partes é mais custoso, porque necessita de um tempo adicional para o treinamento de uma equipe na infraestrutura da organização e também para fazer a integração de um novo sistema aos sistemas já existentes.

Em um cenário de desenvolvimento fragmentado e heterogêneo de sistemas e modelos de dados, as pesquisas direcionadas para avanços dos métodos destinados ao apoio à decisão médica é prejudicada. As pesquisas comumente ocorrem com dados de uma instituição que segue um determinado modelo (NHSI, 2020; HENKE *et al.*, 2018). A comparação entre diferentes métodos ou de um método em dados de diferentes instituições não é trivial e exige um grande esforço, desacelerando possíveis avanços e novas descobertas científicas no ambiente médico (LOUIE *et al.*, 2007).

Muitos métodos para extrair padrões são encontrados na literatura, em especial no contexto médico, o qual é possível encontrar uma gama de trabalhos que lidam com a detecção de padrões em registros médicos eletrônicos (RAZZAK; IMRAN; XU, 2020; FUNKNER; YAKOVLEV; KOVALCHUK, 2017; LISMONT *et al.*, 2016). No entanto, não é possível a comparação entre eles, dificultando a escalabilidade dos métodos.

A [Figura 5](#) detalha os cenários encontrados e suas particularidades. No cenário (a) apresenta dois diferentes modelos de dados (*BD1* e *BD2*) consequentemente tornando também distintas as representações (A e B) e impossibilitando a comparação justa dos resultados providos de dois ou mais métodos. Já o cenário (b) apresenta um Modelo Comum de Dados, o qual é responsável pela padronização dos diferentes modelos de dados (*BD1* e *BD2*). Apesar da busca pelo resultado ser padronizado, a representação ou até mesmo as técnicas de pré-processamento não são iguais, ou seja, a entrada do método também não é semelhante, o que dificulta a comparação entre os métodos. Este trabalho buscou encontrar um modo de que as representações dos dados e os métodos de análise sejam compatíveis, conforme é explicado no [Capítulo 3](#).

Figura 5 – Cenário da interoperabilidade entre as bases de dados



Fonte: Elaborada pelo autor.

2.5.2 Explicabilidade na detecção de padrões

Algoritmos de aprendizado de máquina (AM) vêm se popularizando em diversas áreas de conhecimento, auxiliando tomadas de decisão. No contexto médico, estes métodos podem auxiliar na construção de diagnósticos precoces e precisos (XIE; CHEN; GAO, 2019; ARRIETA *et al.*, 2019). Por outro lado, também é comum que tais métodos exijam dos profissionais que os utilizam um conhecimento específico (relacionado a Aprendizado de Máquina e/ou Ciência de Dados), principalmente quando são algoritmos denominados caixas-pretas, o qual o seu funcionamento não é claro ou conhecido (GUIDOTTI *et al.*, 2018). Caso este conhecimento não seja de domínio do profissional da área da saúde, a compreensão do resultado ou decisão indicada pelo método é comprometida e pode fragilizar a convicção do médico sobre um possível diagnóstico. Com isso, é justificada a necessidade dos métodos sejam capazes de ser explicáveis e suas respostas encontradas sejam bem fundamentadas, contribuindo para o melhor entendimento do usuário e colaborando também para o processo de auxílio à tomada de decisão.

É importante destacar que ao possibilitar a compreensão do usuário sobre o processo realizado e dos respectivos resultados, eleva-se a convicção do usuário em concordar ou discordar com o resultado obtido e contribui para tomadas de decisões mais substanciadas. Porém, existem alguns problemas a serem superados: (i) a maioria dos métodos são propostos para contextos genéricos e acabam fornecendo informações desnecessárias ou excesso de informações, o que pode prejudicar a compreensão de profissional de áreas específicas (assim como os da área da saúde); (ii) a maioria dos métodos estudam a explicabilidade em tarefas de predição. Portanto, é possível observar uma lacuna de trabalhos sobre a explicabilidade dos algoritmos utilizados para a descoberta de conhecimento, por exemplo, a aplicação de técnicas de agrupamento de dados.

2.6 Considerações finais

Esta seção apresentou os principais conceitos que envolvem esta pesquisa, os trabalhos relacionados encontrados na literatura e os problemas que motivaram a criação dos métodos. Foi possível notar que existem dois grandes problemas: (i) a interoperabilidade das bases de dados, o que dificulta a escalabilidade na utilização de diferentes métodos de descoberta de padrões; (ii) a explicabilidade aplicada em diferentes métodos de agrupamento para padrões desconhecidos. Na próxima sessão são apresentados os métodos propostos nesta dissertação para atender às lacunas encontradas na literatura.



O MÉTODO MCE (MULTILEVEL CLUSTERING EXPLAINER)

3.1 Visão geral

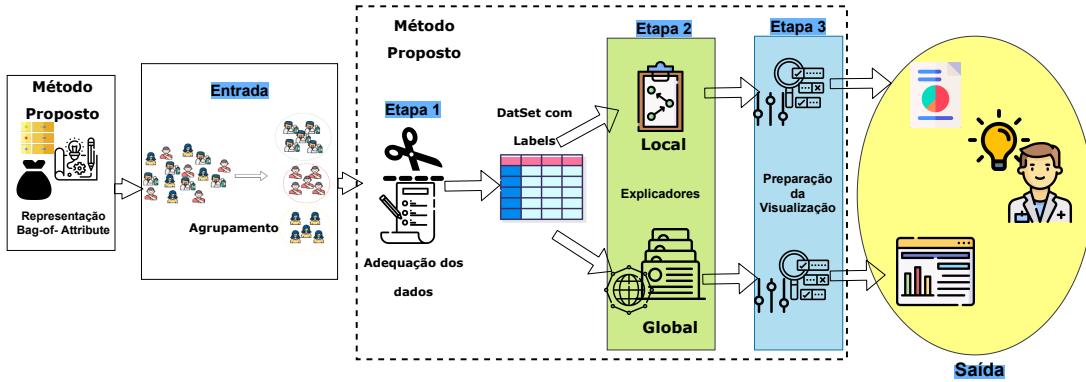
Com o objetivo de colaborar para a descoberta de conhecimento em registros eletrônicos de pacientes, este trabalho propõe um novo método baseado na explicabilidade dos resultados para tarefas não-supervisionadas, permitindo ao usuário explorar a explicabilidade local e global dos dados médicos analisados. Em específico, o método possibilita a flexibilização na granularidade de informação e das explicações providas pelo resultado do método. Além disso, oferece o detalhamento de informações dos registros eletrônicos de pacientes para auxiliar o especialista na análise dos dados de interesse e na sua tomada de decisão.

O método proposto, denominado *Multilevel Clustering Explainer* (MCE), possui 3 etapas principais: (etapa 1) Adequação dos dados, (etapa 2) Construção das informações explicativas (com diferentes níveis de detalhamento, Local ou Global), (etapa 3) Recuperação de conteúdo e apresentação das informações. O fluxo de tarefas do MCE é apresentado na [Figura 6](#) e as etapas que compõem o método proposto são descritas a seguir.

É importante destacar que a entrada do método requer o resultado provido por algoritmos de agrupamento, que seja representado por uma matriz numérica contendo a definição e os respectivos valores dos atributos e também a identificação do *cluster* ao qual o objeto pertence. No entanto, conforme citado na [Subseção 2.5.1](#) o problema da interoperabilidade das bases de dados, estende também no momento de comparar resultados providos de diferentes métodos de agrupamento, dessa maneira dificultando a comparação e análise dos resultados resultantes de métodos distintos.

Portanto, a proposta de desenvolver uma representação única, denominada *Bag-of-Attribute Representation* (BOAR) como entrada do MCE. A utilização do BOAR é motivada

Figura 6 – Visão geral do MCE



Fonte: Elaborada pelo autor.

pelo fato que a representação permite ao profissional de saúde analisar e comparar resultados providos por diferentes métodos de agrupamentos. Desta maneira, dando o método MCE o funcionamento independente da base de dados e dos algoritmos de agrupamento. O desenvolvimento da representação é apresentado na [Seção 3.2](#).

3.2 A representação BOAR

A representação denominada de BOAR (Bag of Attributes Representation), tem como objetivo projetar uma representação única para diferentes tipos de banco de dados e métodos de agrupamento. Desta forma, possibilitando a realização de estudos comparativos entre distintos algoritmos de agrupamento.

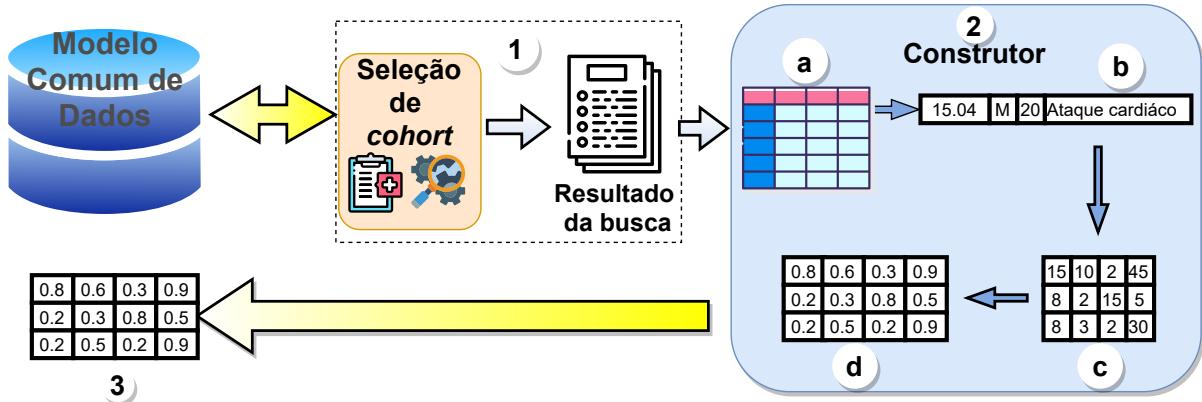
O funcionamento do BOAR consiste em 3 etapas principais: Seleção de *Cohort*, Construtor e Resultado Final. A visão geral da representação é apresentada na [Figura 7](#) e no [Algoritmo 1](#). Nas Subseções [3.2.1, 3.2.2](#) e [3.2.3](#) são apresentados os detalhes de cada etapa da representação.

3.2.1 Seleção do cohort

Nessa etapa, o módulo seleção de *cohort* é responsável por duas tarefas principais: (i) a busca e recuperação das informações em um banco de dados; (ii) o armazenamento dos resultados encontrados. O armazenamento persistente, possibilita guardar os resultados em formatos de arquivos diferentes, evitando assim a necessidade de repetição da etapa. No [Algoritmo 1](#) é apresentado o processamento da base de dados para a seleção de *cohort*.

Primeiramente, o algoritmo necessita como entrada, uma consulta realizada em um banco de dados relacional padronizado. Na linha 1, é realizada a busca pelo *cohort* desejado. A definição da busca é de preferência e demanda do usuário, possibilitando ao usuário recuperar informações de diversos interesses, como por exemplo: informações de diagnósticos, informações de triagem,

Figura 7 – Visão Geral do BOAR



Fonte: Elaborada pelo autor.

Algoritmo 1 – Seleção *cohort* e construtor BOAR**Entrada:** : Base de Dados EHR: *EHR_BD***Entrada:** : Método de busca: *Busca***Entrada:** : Atributos de interesse *AtDeInteresse***Saída:** : Representação Final *VetorF*

```

1: Cohort ← SelecaoCohort(EHR_BD, Busca)
2: para Atributo ∈ AtDeInteresse faz
3:   para Coluna ∈ Cohort faz
4:     se Atributo = Coluna.Nome então
5:       Vcaracteristicas.AddLinha(Coluna.valores)
6:     fim se
7:   fim para
8: fim para
9: Vcaracteristicas.RemoverSimbolosEspeciais()
10: Vcaracteristicas.TransMinusculas()
11: Vcaracteristicas.RemoverEspaos()
12: Vcaracteristicas.AtributosParaPalavra()
13:
14: VetorF ← ComputarFrequenciaAtributo(Vcaracteristicas)
15: VetorF.ComputarTFIAF()
16: retorna VetorF

```

resultados de exames e internações. Este processo pode ser visualizado também na Figura 7 etapa(1).

3.2.2 Construção da representação

O módulo de construção é responsável por criar uma tabela/matriz usando as informações exportadas do módulo de seleção de *cohort*. Após retornadas as informações da exportação, a identificação das colunas do conjunto de dados são tratadas como atributos. Neste momento, o

usuário tem a possibilidade de selecionar quais atributos vão ser utilizados para a construção da representação.

Conforme é apresentado no [Algoritmo 1](#) linhas 2 à 8 e na [Figura 8](#), é apresentada a construção do dicionário de atributos, dado que para cada valor de atributo torna-se uma palavra no dicionário. Baseado no conceito da técnica *bag-of-words*, não é permitida a duplicação das palavras no dicionário, ou seja, os valores dos atributos não se repetem. O principal motivo para considerar atributos como palavras é dado pelo motivo em que os valores de atributos são conhecidos e finitos dentro de uma estrutura comum de dados estruturados. É importante destacar que foram utilizados apenas atributos categóricos para construção do dicionário.

Na sequência, nas linhas 9 à 12 do [Algoritmo 1](#) são realizadas tarefas de pré-processamento com intuito de manter a integridade e padronização dos valores, tais como: remoção de símbolos especiais, padronização dos atributos categóricos para letras minúsculas e remoção de espaços em branco. Logo após é necessário realizar a contagem da frequência dos atributos em cada registro do conjunto de dados e armazená-lo em um vetor n-dimensional n . A [Figura 8](#) apresenta o funcionamento das tarefas de pre-processamento e a contagem de frequência dos atributos de um determinado *cohort* selecionado.

Figura 8 – Exemplo de um dicionário construído pelo módulo construtor.

Cohort Selecionado			
ID	Conceito	Exame	Gênero
01	Hipertensão	ExameA	M
02	Diabetes	ExameB	F
03	Hipertensão	ExameA	F

↓

hipertensao	diabetes	examea	exameb
2	1	2	1

Dicionário

Fonte: Elaborada pelo autor.

Para implementação da contagem de frequências, é utilizada uma adaptação da medida estatística *Frequency-inverse Document Frequency* (*TF-IDF*), a qual os documentos são substituídos por atributos, nomeando assim a nova medida *Term Frequency-inverse Attribute Frequency* (*TF-IAF*). A função do *TF-IAF* é penalizar os atributos que aparecem frequentemente no vetor e indicar o nível de importância de cada variável em relação a todos os registros do conjunto de dados. A computação dos atributos dos registros e a aplicação do *TFIAF* são apresentadas nas linhas 14 e 15 do [Algoritmo 1](#). A saída do módulo é uma matriz composta do vetor *TF-IAF*. Este módulo foi implementado para usar os dados recuperados pelo módulo

seletor de coorte ou armazenados em um arquivo.

3.2.3 Resultado Final da Representação

Ao final do processamento, é gerado o resultado final da representação, uma matriz, na qual as colunas são representadas pelos atributos e as linhas pela frequência que os atributos tiveram sobre aquela instância. É importante destacar que, ao representar os dados desta forma, incluindo a suposição de que todo o conteúdo de cada atributo é uma única palavra, o método BOAR permite a comparação de qualquer número de algoritmos diferentes com bancos de dados, modelos de dados e configurações distintas. Além disso, esta representação de dados permite que o BOAR gere dicionários compactos e evite a necessidade de algumas etapas generalizadas (como a remoção de token) que são necessárias para lidar com texto livre, sem degradar a qualidade dos resultados da análise. O resultado final da representação proposta possibilita ser aplicável em diferentes algoritmos de agrupamento conforme foi apresentado no trabalho (Clementino *et al.*, 2020), o qual são aplicados dois algoritmos distintos para realizar o agrupamento. Sendo que os resultados dos algoritmos de agrupamento serão utilizados na próxima Seção 3.3 do método MCE.

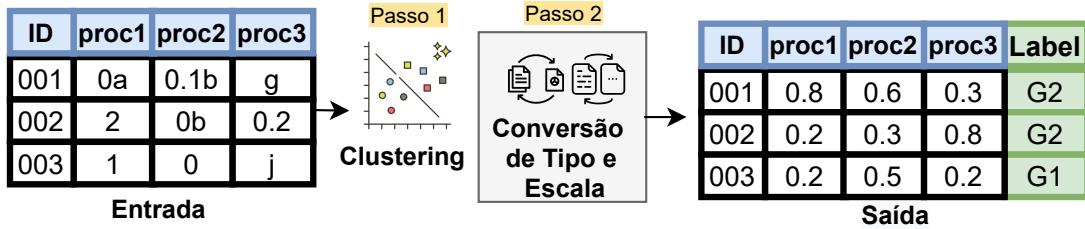
3.3 Etapa 1- Preparação dos dados (MCE)

Nesta etapa é usado o conjunto de dados que o algoritmo de agrupamento utilizou. Especificamente neste trabalho o *cohort* refere-se aos registros de internações e os respectivos Ids dos *clusters* que cada objeto foi inserido, providos da representação do BOAR. No entanto, como o método não é exclusivamente dependente da representação, a atual etapa realiza também uma verificação dos tipos de dados presentes no conjunto de dados que o algoritmo de agrupamento usou. Caso existam dados no formato textual, devem ser convertidos para um formato numérico sem (ou com o mínimo) de perda de informação.

Para atender as especificações básicas de entrada do método, além das colunas dos atributos, outras duas são essenciais: ID dos registros e os *labels* dos *clusters*. Em relação ao ID's dos registros, são inseridos na primeira posição do conjunto de dados, o atributo deve conter os Ids (identificadores) que ligam cada objeto ao seu respectivo registro de internação, permitindo recuperar informações adicionais para futura exibição. Por fim, outro atributo é inserido na última posição do conjunto de dados utilizado pelo algoritmo de agrupamento. Esta coluna contém os Ids dos *cluster* (Label) que cada objeto foi inserido no processo de agrupamento, realizado anteriormente e independente do método proposto. A Figura 9 ilustra o conjunto de dados ao final desta Etapa e o Algoritmo 2 detalha a execução descrita anteriormente.

Na linha 1 do Algoritmo 2 uma função é invocada para realizar a verificação e conversão dos dados (caso necessário) para o formato numérico no conjunto de dados utilizado pelo algoritmo de agrupamento (anterior e independente do método proposto). Na linha 2, uma nova

Figura 9 – Funcionamento da etapa 1



Fonte: Elaborada pelo autor.

coluna contendo o Id do registro de internação ao qual o objeto se refere é inserida na primeira posição do conjunto de dados usado pelo algoritmo de agrupamento. É importante ressaltar que caso o conjunto de dados utilizado pelo algoritmo de agrupamento já contenha o Id do registro de internação, o processo da linha 2 não é necessário. Por fim, na linha 3 uma nova coluna contendo o Id do *cluster* em que o objeto foi agrupado é inserida na última posição do conjunto de dados usado pelo algoritmo de agrupamento.

Algoritmo 2 – Adequação dos dados

Entrada: DataSet_Clust, Id_Clust, EHR

Saída: DataSet_Clust

- 1: *DataSet_Clust* \leftarrow Conversao(*DataSet_Clust*)
 - 2: *DataSet_Clust*[, 0] \leftarrow *EHR*[, *Id_registroInternacoes*]
 - 3: *DataSet_Clust*[, *n*] \leftarrow *Id_Clust*[]
 - 4: **retorna** *DataSet_Clust*
-

Após o processo de adequação dos dados, o *DataSet* original contém adicionalmente o atributo de identificação dos registros e identificação das *labels*. Com isso, este conjunto será processado na Etapa 2, durante a construção dos explicadores.

3.4 Etapa 2- Construção da arquitetura dos explicadores (MCE)

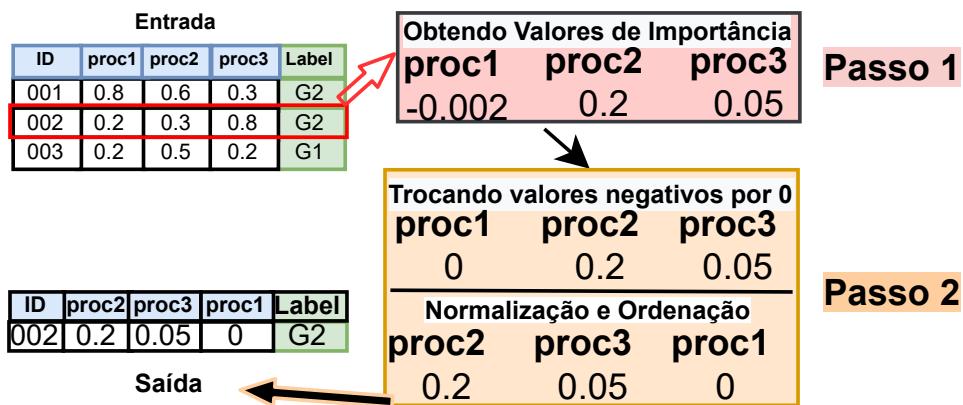
A Etapa 2 possui dois níveis de construção das informações explicativas, Local e Global. No nível Local as informações explicativas são construídas referindo-se a um registro de internação específico. Enquanto que no nível Global as informações explicativas referem-se aos atributos mais influenciadores para a formação de um *cluster* de registros de internações, considerados semelhantes no processo de agrupamento.

3.4.1 Explicador local

Neste nível, as informações explicativas descrevem os atributos que mais influenciaram para que o objeto tenha sido inserido em um determinado *cluster*. Destacamos que para este trabalho, o termo “atributos” refere-se aos procedimentos médicos realizados durante a internação de um paciente. Porém, o termo “atributos” pode referir-se a outra informação que esteja em investigação pelo especialista usada no processo de agrupamento.

A Figura 10 exemplifica o cálculo da influência dos atributos para o objeto indicado pelo especialista. A indicação é realizada pela especificação do número da linha que o objeto está posicionado no conjunto de dados ou pela indicação do Id do registro de internação (Figura 10-Entrada).

Figura 10 – Fluxo de construção das informações explicativas nível local



Fonte: Elaborada pelo autor.

Os valores dos atributos que constituem o referido objeto são recuperados do conjunto de dados para o cálculo das influências de cada atributo. O MCE é proposto para ser flexível com o processo de cálculo das influências, o qual pode ser realizado com uma abordagem específica do especialista ou por métodos existentes na literatura científica (Figura 10-Passo 1).

Neste trabalho, as influências dos atributos são calculadas gerando um novo conjunto de dados que consiste em amostras permutadas e os respectivos Ids do *cluster* que o objeto foi inserido no processo de agrupamento. Nesse novo conjunto de dados, um modelo interpretável (por exemplo, árvores de decisão) é treinado e usado para avaliar o aprendizado, o qual deve ser capaz de realizar uma classificação precisa da amostra utilizada dos dados. A função aprendida pelo modelo de aprendizado de máquina é utilizada para mapear os atributos mais significativos para a predição, estimando as influências. Com isso, o aprendizado alcançado pelo modelo interpretável permite inferir um valor de magnitude da influência que cada atributo apresenta para os objetos serem classificados no *cluster* adequado. Por exemplo, um atributo terá influência “0” se não for utilizado pelo modelo interpretável. Enquanto que um atributo pode ter a magnitude

de sua influência elevada caso este atributo indique a classificação correta do objeto em uma condição específica, como um valor superior a um limiar.

Após o cálculo das influências, os valores são ajustados para serem objetivos e sucintos ao contexto médico. Esse ajuste é necessário porque o processo de cálculo de influência pode gerar grande quantidade de informação tornando a sua aplicação prática demorada e com complexidade desnecessária. Por exemplo, valores negativos referem-se aos atributos que desfavorecem a inserção do objeto ao *cluster* em que foi inserido. Todavia, se o objeto está em um determinado *cluster* é porque os atributos com influência favorável (valores positivos) a inserção do objeto ao referido *cluster* são superiores aos atributos com valores negativos (atributos desfavoráveis). O MCE direciona as informações explicativas construídas para uma abordagem direta que favoreça a sua aplicação prática. Para isso, os valores negativos são convertidos ao valor “0” representando que o atributo não influenciou na inserção do objeto ao referido *cluster* ([Figura 10](#)-Passo 2). Essa conversão evita a grande variedade de valores negativos e concentra-se na explicabilidade da inserção do objeto ao referido *cluster*.

Ainda no (Passo 2), os valores são normalizados e ordenados de maneira decrescente. A normalização mínima e máxima dos valores positivos é realizada para que os valores de influência dos atributos expressem de maneira clara a magnitude de influência dos atributos. A ordenação decrescente é realizada como processo final antes de entregar os valores para a Etapa 3, apresentação das informações. O [Algoritmo 3](#) detalha a implementação da construção das informações explicativas da Etapa 3 no nível Local.

Algoritmo 3 – Explicador local

Entrada: *Id_Objeto, DataSet_Clust*

Saída: *Inf_Local*

- 1: *Objeto* \leftarrow *Recuperacao(Id_Objeto, DataSet_Clust)*
 - 2: *Inf_Local* \leftarrow *CalculoInfluencia(Objeto)*
 - 3: *Inf_Local* \leftarrow *RemocaoValoresNegativos(Inf_Local)*
 - 4: *Inf_Local* \leftarrow *Normalizacao(Inf_Local)*
 - 5: *Inf_Local* \leftarrow *Ordenacao(Inf_Local)*
 - 6: **retorna** *Inf_Local*
-

Inicialmente (linha 1) o objeto indicado pelo especialista é recuperado no conjunto de dados gerados pela Etapa 1. Na linha 2, é invocado o método que calcula a influência dos atributos do objeto. Na linhas 3 e 4, os valores de influência negativos são substituídos por “0” e normalizados. Por fim, os atributos são ordenados de maneira decrescente e o processamento é finalizado.

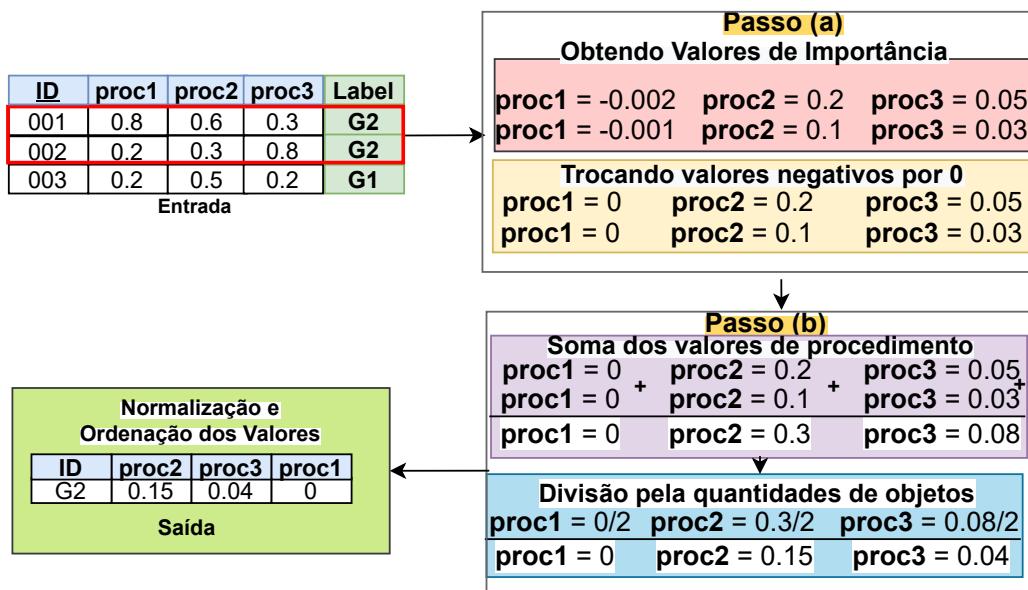
3.4.2 Explicador global

A principal diferença entre os níveis Local e Global está na granularidade das informações estimadas, no qual o nível Global realiza o processamento considerando todos os objetos de um

grupo definido pelo especialista.

A Figura 11 ilustra o processamento realizado para o cálculo dos atributos mais influentes para a inserção de um objeto no respectivo grupo. Inicialmente, o especialista informa o Id do grupo que deseja-se gerar as informações de explicabilidade e todos os elementos que compõem o respectivo grupo são recuperados do conjunto de dados gerado na Etapa 1 (Entrada).

Figura 11 – Fluxo de construção das informações explicativas nível global



Fonte: Elaborada pelo autor.

Na Figura 11-(Passo a), os valores de influência dos atributos de todos os objetos são calculados e os valores negativos substituídos por “0”. Esses processamentos são realizados da mesma forma que o detalhado no nível Local.

Posteriormente, na Figura 11-(Passo b), é calculado o impacto das influências individuais no nível global. O especialista pode escolher qual equação (média, mediana ou moda) utilizar por meio de parâmetros na inicialização do método. Apesar de ter diversas equações implementadas, detalhamos o cálculo utilizando a média aritmética para exemplificar o conceito das equações no método. O cálculo é detalhado na Equação 3.1 onde, G refere-se a um grupo de registros que compõem um *cluster*, n é o total de *clusters* criados pelo algoritmo de agrupamento, R são os objetos do *dataset*, m é a quantidade de objetos do *dataset*, A_i é o i -ésimo atributo calculado, com i percorrendo todos os atributos de uma determinada posição em todos os objetos. Assim, a equação calcula a média das influências de cada atributo considerando todos os objetos de um grupo específico. Como resultado é obtido um vetor da média das influências de cada atributo para um determinado grupo, expressando uma sumarização dos atributos mais influentes para

que um objeto permaneça no respectivo grupo.

$$ExpGlobal = G_n \left(\frac{\sum_{i=1}^m A_i}{m} \right) \quad (3.1)$$

Por fim, é realizada a normalização dos valores de influência média dos atributos para o grupo indicado pelo especialista. Estes cálculos também são semelhantes ao realizado no nível Local. O [Algoritmo 4](#) detalha a implementação para o cálculo das influências dos atributos mais influentes para a formação do respectivo grupo (Global).

Algoritmo 4 – Explicador global

Entrada: *Id_Cluster, DataSet_Clust*

Saída: *Inf_Global*

```

1: DataSet_Ojetos  $\leftarrow$  Recuperacao(Id_Cluster, DataSet_Clust)
2: para Objeto  $\in$  DataSet_Ojetos faz
3:   Objeto  $\leftarrow$  CalculoInfluencia(Objeto)
4:   Objeto  $\leftarrow$  RemocaoValoresNegativos(Objeto)
5:   DataSet_Temp.inserir(Objeto)
6: fim para
7: para Atributo  $\in$  DataSet_Temp faz
8:   Media_Temp  $\leftarrow$  Media(Atributo)
9:   Inf_Global.inserir(Media_Temp)
10: fim para
11: Inf_Global  $\leftarrow$  Normalizacao_Ordenacao(Inf_Global)
12: retorna Inf_Global
```

Na linha 1, os objetos pertencentes ao grupo indicado pelo especialista são recuperados do conjunto de dados gerado pela Etapa 1. Nas linhas 2-6, é calculada a influência dos atributos para cada objeto que compõe o referido grupo, dos quais também são retirados os valores negativos. O *DataSet_Tem* é um conjunto de dados temporário com os valores não negativos das influências de cada atributo de todos os objetos.

Nas linhas 7-10, é calculada a média aritmética de cada atributo do conjunto de dados temporário (*DataSet_Tem*), resultando em um vetor (*Inf_Global*) com os valores médios das influências que os objetos do referido grupo possuem. Por fim, este vetor possui seus valores normalizados e ordenados de maneira decrescente, processamento semelhante ao realizado no nível Local.

3.5 Etapa 3- Preparação visual (MCE)

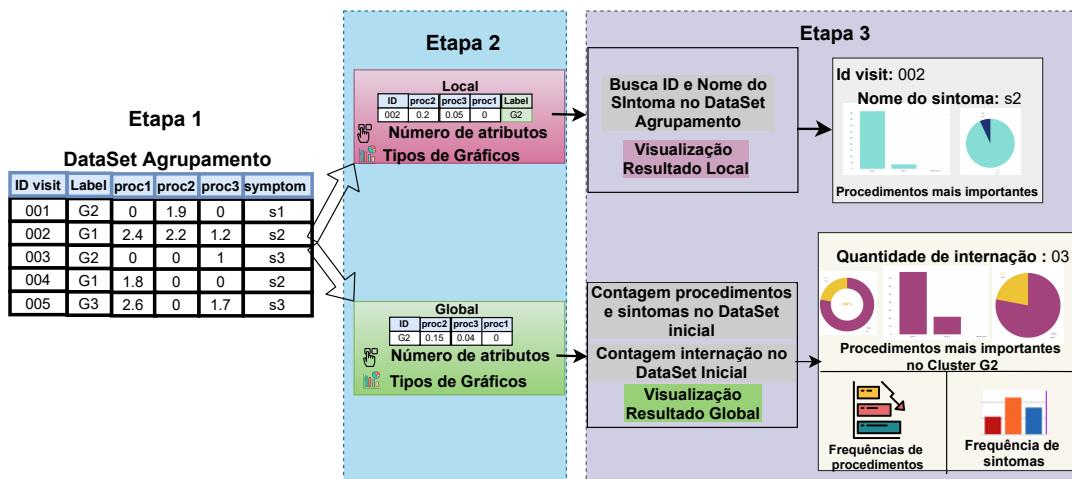
Esta Etapa tem o objetivo de inserir informações complementares (*Local*: Id do registro de internação de paciente e Diagnósticos da internação; *Global*: Total de internações do grupo investigado) nos registros de internação e organizar as informações explicativas geradas na Etapa 2 em gráficos e para apresentá-las de maneira visual. A sumarização destas informações em

um formato visual proporciona ao especialista uma compreensão clara e rápida das informações necessárias para apoiar a tomada de decisões.

É possível personalizar a apresentação visual nas seguintes características: (i) tipo de gráfico, (ii) quantidade de atributos a serem exibidos (atributos de interesse), (iii) valores absolutos ou em porcentagem, (iv) se deve ser exibida a soma dos valores não pertencentes aos atributos de interesse. É importante ressaltar que os parâmetros de personalização suportados são os mesmos para ambos os níveis.

A Figura 12 exemplifica a construção das apresentações visuais para ambos os níveis (Local e Global). É possível observar que as informações adicionais e os gráficos exibidos são diferentes entre os níveis, as quais são direcionadas ao contexto específico do nível que o especialista definiu previamente. Alguns exemplos práticos das apresentações visuais que o MCE gera podem ser vistas na sessão de resultados, apresentada a seguir.

Figura 12 – Apresentação visual



Fonte: Elaborada pelo autor.

O **Algoritmo 5** apresenta os detalhes da execução da Etapa 3. Como explicado anteriormente, esta etapa utiliza a saída da Etapa 2 conforme o nível definido pelo especialista. Nas linhas 1 - 4 são recuperados os dados no registro de internação de paciente (EHR) de uma internação específica indicada pelo “ID_Internacao”. A execução adequada com o nível definido pelo método de recuperação de dados “Recuperar_Dados” é definida pelo parâmetro “Nível”, que pode indicar *Local* ou *Global*.

Nas linhas 5 - 9 os dados recuperados dos “EHR” foram inseridos no grupo identificado pelo “Id_Clust”, também com execução orientada pelo parâmetro “Nível”. O método “Plotar” constrói o gráfico usando as informações de explicabilidade do nível Global (“Inf_Global”), os dados recuperados (“DS_Visu”) e os parâmetros de configuração indicado pelo vetor de parâmetros “Vetor_Parametros” e com execução direcionada pelo parâmetro “Nível”. Por fim, na linha 10, o método “Salvar” gera um arquivo com os gráficos e as informações complementares

recuperadas.

Algoritmo 5 – Construção visual

Entrada: *Inf_Local, Inf_Global, Id_Internaao*

Entrada: *Id_Clust, Nivel, EHR, Vetor_Parametros*

```

1: se Nivel = Local então
2:   DS_Visu  $\leftarrow$  Recuperar_Dados(Id_Internaao, Nivel, EHR)
3:   GF_Visu  $\leftarrow$  Plotar(Inf_Local, DS_Visu, Nivel, Vetor_Parametros)
4: senão
5:   se Nivel = Global então
6:     DS_Visu  $\leftarrow$  Recuperar_Dados(Id_Clust, Nivel, EHR)
7:     GF_Visu  $\leftarrow$  Plotar(Inf_Global, DS_Visu, Nivel, Vetor_Parametros)
8:   fim se
9: fim se
10: Salvar(GF_Visu)

```

3.6 Considerações finais

Este capítulo apresentou a visão geral e o funcionamento do MCE, também detalhou a representação BOAR, qual é utilizada no método MCE. É importante destacar que um método não depende do outro, mas para proposta deste trabalho de mestrado são utilizados de maneira complementar. No Capítulo 4 é apresentado os resultados referentes aos dois métodos propostos.



RESULTADOS E DISCUSSÃO

4.1 Considerações iniciais

Este capítulo tem por objetivo apresentar o conjunto de dados utilizado para desenvolvimento dos métodos propostos. Conforme citado anteriormente, uma das motivações do trabalho é oferecer flexibilidade ao método para ser capaz de atuar em diferentes bases de dados. Nesta direção, no contexto de dados e análise médicas, utilizou-se nesse trabalho o *OMOP Common Data Model* (CDM), um padrão comum de dados que auxilia na padronização das diferentes terminologias e modelos de dados. Portanto, na [Seção 4.2](#) são apresentados os conceitos básicos que envolvem a estrutura do OMOP. A [Seção 4.3](#) detalha os experimentos e resultados do método BOAR e a [Seção 4.4](#) os experimentos e resultados do MCE.

4.2 OMOP Common Data Model (CDM)

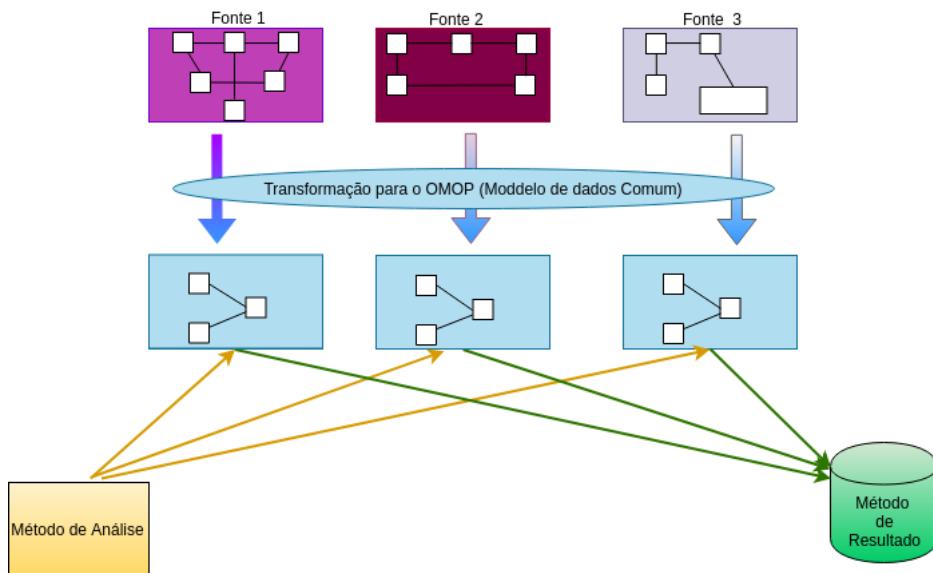
O OMOP Common Data Model (CDM) é fruto de uma pesquisa colaborativa e coordenada pela *Columbia University*, a qual possui uma grande variedade de apoiadores (academia, governo e indústria) que trabalham em áreas multidisciplinares, tais como: estatística, ciência da computação e ciências clínicas e epidemiologia. As pesquisas desenvolvidas direcionam-se a produzir métodos analíticos e ferramentas baseadas em dados padronizados para OMOP ([OMOP, 2020](#)).

A [Figura 13](#) apresenta de maneira visual a proposta do OMOP. Os elementos “Fonte 1”, “Fonte 2” e “Fonte 3” representam diferentes origens de EHR, as quais podem ser instituições diferentes ou modelos de dados diferentes em uma mesma instituição. Os dados de diferentes fontes são transformados (ou adequados)¹ ao modelo de dados comum do OMOP e passam

¹ Instruções para a transformação dos dados para o modelo OMOP são disponibilizados pela empresa OHDSI no endereço <https://www.ohdsi.org/web/wiki/doku.php?id=documentation:etl_best_practices>.

a ter a mesma estrutura. O processo de transformação é representado pela elipse intitulada “*Transformação para o OMOP (Modelo de dados Comum)*”. Após a transformação, os diferentes conjuntos de dados podem ser acessados da mesma forma por um “Método de Análise” (quadrado amarelo no canto esquerdo inferior) ou por um “Método de Resultado” (quadrado verde no canto inferior direito). Ressaltamos que o acesso por diferentes métodos a um mesmo conjunto de dados também é suportado, onde os métodos precisam suportar o acesso a dados organizados com o modelo OMOP.

Figura 13 – Contexto de aplicação do OMOP



Fonte: Adaptada de [Omop \(2020\)](#).

O OMOP pode contribuir significativamente com pesquisas na área da computação na medicina, em especial as pesquisas que utilizam EHR. O uso deste modelo de dados comum permite que novas pesquisas possam ser desenvolvidas utilizando os avanços alcançados neste trabalho e que métodos nele desenvolvidos possam ser aplicados em conjuntos de dados de outras organizações.

4.3 Experimentos BOAR

O objetivo do experimento é apresentar as configurações de implementação da representação, juntamente com a análise de dois algoritmos diferentes de agrupamento aplicados no contexto médico. Contextualizando o cenário da aplicação: sabe-se que sintomas ou doenças semelhantes podem levar o profissional de saúde a escolher os procedimentos clínicos, que são representados pelos *clusters* de pacientes que seguem trajetórias clínicas semelhantes. Esses *clusters* fornecem informações úteis para a melhoria dos serviços de saúde, auxiliando os procedimentos médicos com mais segurança. Em especial, o maior *cluster* apresenta características

preponderantes para caracterizar o *cohort* analisado, pois esse *cluster* é construído a partir da maioria dos registros.

Os experimentos foram realizados sobre uma amostra anônima Lima et al. (2019) de um conjunto de dados de Registros Eletrônicos de Saúde (EHR) fornecido pelo Instituto do Coração (InCor) da Universidade de São Paulo, Brasil. Este estudo foi realizado em conformidade com as leis de proteção de dados aplicáveis. Esta pesquisa foi aprovada pelo Conselho de Revisão do InCor (IRB) sob o número CAEE17146019.0.0000.0068. A amostra de dados inclui EHR de 94.603 pacientes atendidos no InCor no período de 04/08/1998 a 23/09/2018, representados no modelo comum da OMOP. O código-fonte está disponível publicamente ².

Para implementação do módulo de Seleção *cohort*, foram recuperados registros referentes a admissão de internações hospitalares. A *cohort* contém informações, tais como: procedimentos realizados durante a internação e os diagnósticos dos respectivos procedimentos. É importante destacar que o termo “diagnósticos” na estrutura de dados do INCOR também está relacionado aos sintomas.

Para a avaliação são considerados os registros que possuem no mínimo a presença de um diagnóstico. O intervalo dos registros considerados é de 29/05/2007 à 31/05/2008. Caso o paciente tenha mais que uma admissão nesse período, os dados de cada admissão são tratados de forma independente. O resultado obtido pelo módulo é armazenado em um arquivo usando o formato *JavaScript Object Notation (JSON)*. Este armazenamento permite que diferentes experimentos possam ser realizados usando o mesmo *cohort* selecionado. Assim, o *cohort* recuperou 5.530 admissões com 543 atributos. Portanto, o *DataSet* criado, além dos atributos relacionados aos procedimentos e diagnósticos também contém atributos relacionados à data dos procedimentos e a identificação do código de admissão. O identificador é utilizado apenas para o mapeamento entre as informações contidas no vetor de características resultante e o *DataSet* inicial.

Com base no *cohort* selecionado, o módulo de construção da representação constrói um dicionário com n dimensões, correspondendo à quantidade máxima de atributos recuperados na busca do *cohort*. É importante lembrar que as informações referentes aos diagnósticos não são utilizadas na construção do dicionário. Após calculado o TF-IAF, a frequência de cada atributo é adicionada à sua respectiva coluna do vetor de característica. A matriz resultante com os pesos é inserida na dimensão correspondente do vetor representativo. Por fim, na matriz final os registros recebem novamente os atributos das informações de identificação e o diagnósticos correspondentes.

Para validação da representação proposta, faz-se necessário que a matriz final seja capaz de ser utilizada em diferentes algoritmos de agrupamento. Previamente a tarefa de agrupamento, faz-se necessária a definição do número de K/Grupos e também o *setup* dos experimentos. Para

² <https://github.com/clementinojr/Bag-of-Attribute-Representation-BOAR->

isso, é utilizada a técnica de validação cruzada com $KFold = 10$. Foram selecionados 20% e 80% do vetor de representação do BOAR, para estimar o k de cada iteração e avaliar a etapa de agrupamento, respectivamente.

Para a etapa de estimação, sabendo que o número de diagnósticos diferentes da representação é de 750, ou seja, possibilita-se a formação de 750 *clusters* únicos. Logo, para selecionar o número de *clusters*: foram divididos em 10 subconjuntos ($KEstim$) de tamanho 75 cada um, considerando os seguintes intervalos: 75, 150, 225, 300, 375, 450, 525, 550, 625, 700, 750. Para cada interação do valor do intervalo foi computado o coeficiente silhueta (Gupta; Singhal; Sai Sabitha, 2018). Ao final o $KClust$ que obteve o melhor resultado do calculado da silhueta é utilizado como o valor K-selecionado para a etapa de Agrupamento. Este processo é detalhado no [Algoritmo 6](#).

Algoritmo 6 – Configuração de validação cruzada

Entrada: : Vetor Representação BOAR *VetorBOAR*

```

1: para  $KFold \in [1, 10]$  faca                                ▷ Etapa Estimando K
2:    $LSilhouette \leftarrow Zeros(10, 2)$           ▷  $10 \times 2$  matrix  $KEstim \in [75, 150, \dots, 750]$ 
3:    $Labels \leftarrow Clustering(VetorBOAR, KClust, 20\%)$ 
4:    $Silhueta \leftarrow ComputeSilhueta(Labels)$ 
5:    $LSilhouette.Add(Silhueta, KClust)$ 
6: fim para                                         ▷ Etapa Agrupamento
7:
8:  $K \leftarrow MaxSilhueta(LSilhouette).KClust$ 
9:  $Labels \leftarrow Clustering(VetorBOAR, K, 80\%)$ 
10:  $WriteLabels(Labels, KFold)$ 
11:

```

Após o algoritmo de agrupamento realizar o processo de rotulação (*WriteLabels*) dos *clusters*, é selecionado o rótulo mais frequente de cada $KFold$ para que na sequência faça a recuperação dos 10 diagnósticos mais frequentes em cada *cluster*.

4.3.1 Resultados e discussões

Para prova de conceito, nesta [Subseção 4.3.1](#) são apresentados os resultados do uso de BOAR aplicado em dois algoritmos de agrupamento bem conhecidos, *SK-Means* e *Hierarchical Agglomerative Clustering (HAC)*. Os resultados obtidos em ambos os algoritmos de agrupamento foram compatíveis, em relação a consistência dos *clusters* formados e também em ambos os algoritmos, o maior *cluster* é o mesmo. Esse resultado mostra que o BOAR é robusto para destacar características das internações, para ambos métodos avaliados. Esses resultados também estão disponíveis no repositório ³, para análise posterior.

A [Tabela 2](#) apresenta as informações gerais do *cohort* selecionado, o período dos dados coletados referem-se a aproximadamente um ano entre as datas 28/05/2007 à 31/05/2008. O

³ <https://github.com/clementinojr/Bag-of-Attribute-Representation-BOAR-/tree/master/Experiments/>

cohort selecionado contém um total de 5.530 registros de internações contando com a presença total de 12.684 diagnósticos com um total de 12.684 diagnósticos, dos quais 755 são únicos. Ainda na Tabela 2 é possível observar as informações dos resultados obtidos pelos dois algoritmos de agrupamento analisando os diagnósticos encontrados no *cluster* mais populoso em ambos os algoritmos, nota-se que a frequência dos diagnósticos mais frequentes é similar em ambos os algoritmos.

Tabela 2 – Informações gerais do *cohort* selecionado e informações básicas obtidas pelos algoritmos.

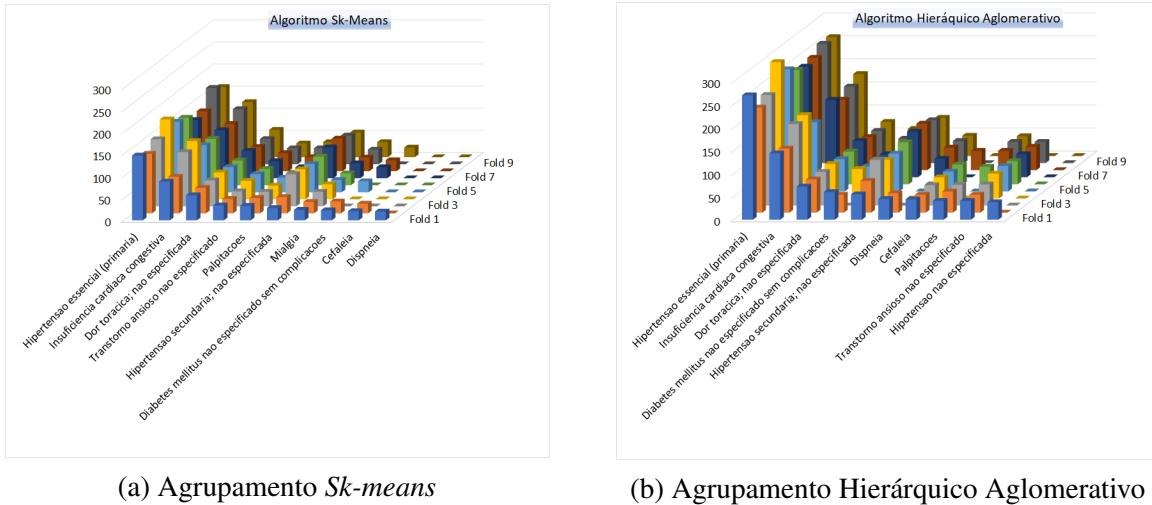
Informações gerais	Quantidade
Registros de internações	5.530
Total de diagnósticos	12.684
Total de diagnósticos únicos	755
Algoritmo SK-Means	Média de diagnósticos no cluster
Maior cluster	1.020 ± 96
diagnósticos mais frequentes	
Hipertensão essencial (primária)	153 ± 16
Insuficiência cardíaca congestiva	110 ± 16
Dor no peito, não especificada	66 ± 6
Algoritmo Hierarchical Agglomerative Clustering	Média de diagnósticos no cluster
Maior cluster	1.885 ± 98
Diagnósticos mais frequentes	
Hipertensão essencial (primária)	255 ± 19
Insuficiência cardíaca congestiva	161 ± 15
Dor no peito, não especificada	88 ± 11

A Figura 14 apresenta a frequência (eixo y) dos diagnósticos que compõem o *cluster* (eixo x) mais populoso para cada iteração da validação cruzada (eixo z) em ambos os algoritmos. Os 10 diagnósticos mais frequentes foram selecionados com base na primeira iteração ($K = 1$) da validação cruzada.

A maioria dos diagnósticos nos registros dos *clusters* encontrados está relacionada a doenças cardíacas. A frequência dos 3 primeiros diagnósticos ocorre na mesma ordem em ambos os *clusters* da Figura 14. Nota-se que 90% dos diagnósticos apareceram em ambos os *clusters*. Ou seja, 18 dos 20 diagnósticos estão presentes em ambos os resultados.

Apesar das diferenças nos valores absolutos apresentados na Figura 14, é possível observar um padrão na quantidade de registros agrupados considerando os diagnósticos “Hipertensão

Figura 14 – Os 10 primeiros diagnósticos obtidos considerando cada método de agrupamento.



(a) Agrupamento *Sk-means*

(b) Agrupamento Hierárquico Aglomerativo

Fonte: Elaborada pelo autor.

essencial (primária)”, “ Insuficiência cardíaca congestiva ” e “ Dor no peito, não fontes ” - os três diagnósticos mais frequentes. Além de estar na mesma ordem de frequência ao comparar os dois gráficos, o mesmo padrão também pode ser visto quando a interação de validação cruzada acontece é realizada em ambos os algoritmos. Esse comportamento semelhante também é observado em registros de outros diagnósticos que não estão na mesma ordem, como: “ Hipertensão secundária, não especificada ” e “ Transtorno de ansiedade, não especificada ”.

É interessante notar que o número de registros diferentes dos mesmos diagnósticos e os padrões de presença seguem um comportamento semelhante. A variação no número de registros é uma consequência do algoritmo de agrupamento usado, pois *SK-Means* usa a distância de cada elemento a um centróide para definir seu *cluster*, e HAC usa a distância média de cada elemento a todos os outros elementos da base para definir seu *cluster*. Por outro lado, o padrão de formação de *clusters* semelhantes indica consistência e desempenho robusto em ambos os casos (BOAR + HAC e BOAR +SK-Means).

Considerando o comportamento de ambos os algoritmos de agrupamento, os gráficos da Figura 14 indicam que o algoritmo SK-Means apresenta um desempenho mais estável, uma vez que os diagnósticos ausentes são os menos frequentes. Embora, certas interações da validação cruzada o algoritmo HAC mostra a ausência de alguns registros, em outros ciclos ele mostra uma frequência maior do que outros diagnósticos (por exemplo, registros com diagnósticos “ Dispneia ”). Além disso, destaca-se a vantagem de possibilitar a comparação entre diferentes algoritmos utilizando o mesmo banco de dados e configurações. Além do progresso feito na literatura científica sobre a representatividade do EHR, o método BOAR proposto permite que os profissionais de saúde avaliem o impacto da descoberta do conhecimento semântico usando diferentes algoritmos, aproveitando uma abordagem de comparação simples, mas robusta.

4.4 Experimentos MCE

Os experimentos estão organizados em duas partes, Local e Global. Assim, a avaliação dos resultados é também realizada em duas partes e busca identificar relações entre as informações exibidas nos níveis Local e Global. Estas relações são entendidas como semelhanças entre os atributos indicados como os mais influentes e predominância em diagnósticos no referido *cluster* (entendido como um padrão). Além disso, considera-se importante a relevância da apresentação dos resultados permitindo ao especialista o entendimento dos resultados obtidos. Neste contexto é apresentado os procedimentos realizados e considerados os mais influentes e importantes para o agrupamento do *cluster* à qual ele pertence.

Os dados utilizados para o agrupamento possuem grande volume de registros de internações, o que dificulta a apresentação das informações descritivas de cada *cluster*. Portanto, foram selecionados os registros de internações que compõem os quatro *clusters* mais populosos construídos no agrupamento realizado pelo algoritmo *SK-Means*. Os *cluster* selecionados possuem os Ids 7, 5, 9 e 45, os quais possuem 2185, 1184, 356 e 330 objetos/registros de internações respectivamente. O direcionamento dos experimentos aos *clusters* mais populosos é motivado pela possibilidade de encontrar padrões mais específicos, porém de difícil capacidade analítica devido ao grande volume de dados.

É importante ressaltar que abordagens qualitativas são amplamente utilizadas na literatura científica (RIBEIRO; SINGH; GUESTRIN, 2016b; LUNDBERG; LEE, 2017c), porque um método de explicabilidade pode ser mais adequado para uma determinada área do que em outras áreas. Também é importante ressaltar que esta premissa é citada como motivação para o trabalho desenvolvido, já que não são encontrados na literatura científica métodos de explicabilidade voltados para o contexto médico. Portanto, os resultados são analisados qualitativamente com o apoio dos profissionais da saúde.

4.4.1 Resultados e discussões

Nesta seção são descritos e detalhados os resultados do método proposto. Os resultados são apresentados em seus respectivos níveis (Local e Global) para uma boa apresentação. Contudo, os resultados permitem serem referenciados entre as subseções e analisados observando características complementares entre os níveis. É importante lembrar que são apresentados os resultados que pertencem ao *cluster* 45 e 7 pois são suficientes para discussão. Porém, os resultados de todos os *cluster* (no nível Global) e dos registros utilizados para esta análise (no nível Local) estão disponíveis em um repositório público⁴.

⁴ <https://github.com/clementinojr/Multilevel-Clustering-Explainer-an-explainable-Approach-to-Hospitalization-Record>

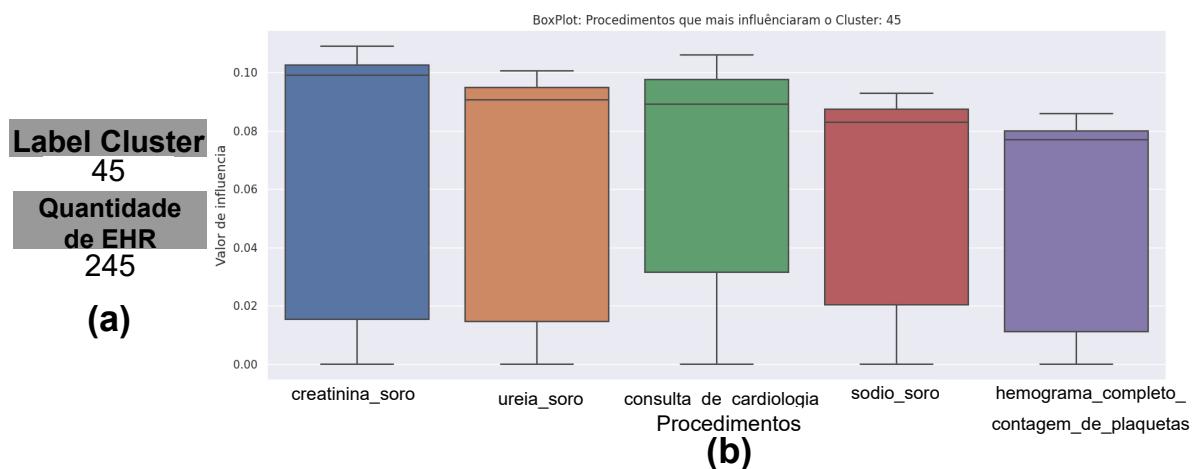
Nível Global

Na sequência serão mostrados os resultados para os *clusters* 45 e 7. Neste nível de explicabilidade o MCE apresenta as informações explicativas de maneira geral de um *cluster*, tal como os atributos mais influentes para que objetos sejam inseridos no referido *cluster*. Na sequência são apresentadas as imagens resultantes que demonstram as informações recebidas pelo especialista neste nível de explicabilidade.

Cluster 45

Na Figura 15(a) é apresentado o *Label* do *cluster* e a quantidade de registros de internações que compõem este *cluster*. Neste experimento os valores apresentados pertencem aos *clusters* 45 com total de 245 registros de internações. Já na Figura 15(b) é apresentado o Gráfico BoxPlot dos cinco procedimentos que mais influenciaram a formação do *cluster* 45, é possível notar que os valores sejam diferentes, todos os procedimentos não apresentam uma distribuição simétrica.

Figura 15 – Nível global *cluster* 45: Informação do *cluster* 45 e box plot dos procedimentos mais frequentes.



Fonte: Elaborada pelo autor.

Já na Figura 16 são apresentados os valores absolutos dos cinco procedimentos mais importantes para a formação do *cluster* 45, sendo também demonstrada na última linha a soma da quantidade dos demais procedimentos que influenciaram na formação do agrupamento.

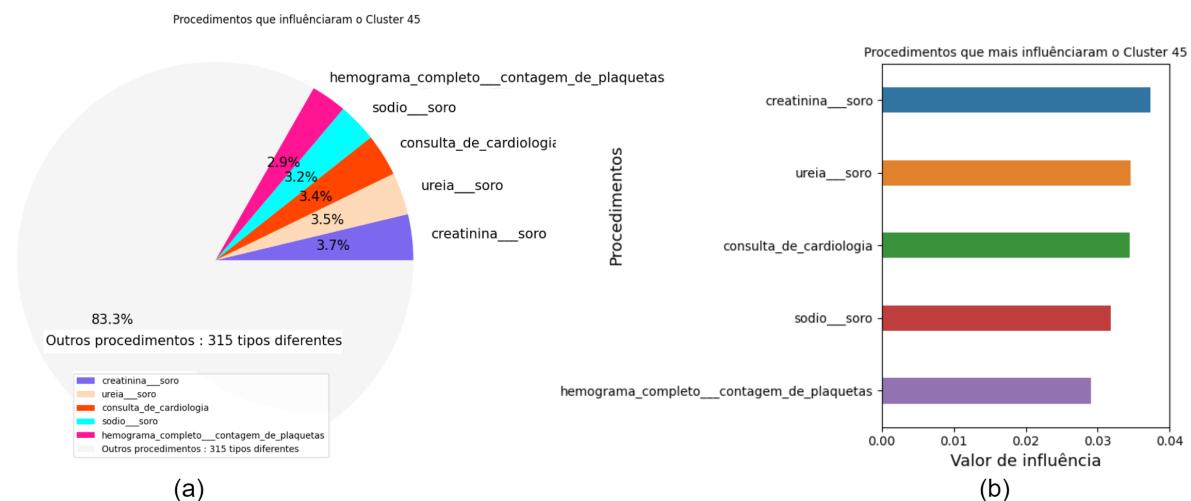
Na Figura 17 são apresentados dois diferentes tipos de representação, em(a) é apresentado utilizando valores em porcentagem(%) e em (b) os valores absolutos são apresentados graficamente. os procedimentos (atributos) considerados mais influentes para que um registro tenha sido inserido no *cluster* com Id 45. Assim, os cinco procedimentos com mais influência para a inserção de um registro no *cluster* 45 foram “Creatinina/Soro” com 3.7%, “Ureia/Soro” com 3.5%, “Consulta de cardiologia” com 3.4%, “Sodio/Soro” com 3.2% e “Hemograma com-

Figura 16 – Nível global *cluster* 45: Valores absolutos dos procedimentos mais influentes.

Nome do Atributo	Valor
creatinina__soro	0.037385952484440335
ureia__soro	0.03456455868036763
consulta_de_cardiologia	0.03445624706507187
sodio__soro	0.03185052071446546
hemograma_completo__contagem_de_plaquetas	0.02911533953283154
Outros procedimentos : 315 tipos diferentes	0.8328625085568883

Fonte: Elaborada pelo autor.

pleto + contagem de plaqueta” com 2.9%. Ainda na região 4 é possível notar que outros 315 procedimentos foram realizados nos 245 registros de internações e dividem os 83% de influência complementares. É importante destacar que a quantidade de procedimentos a serem apresentados é definida inicialmente pelo especialista.

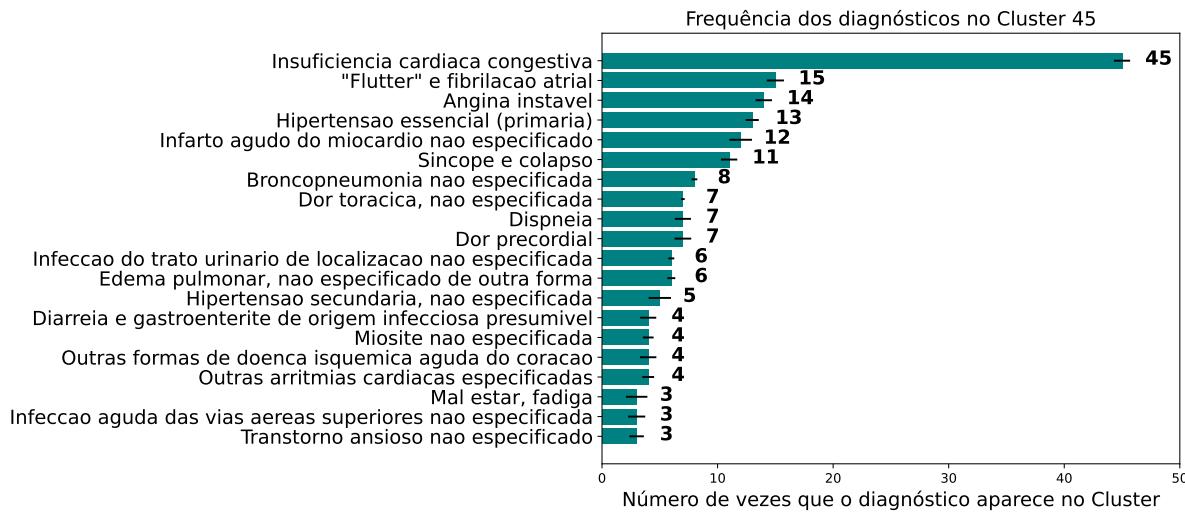
Figura 17 – Nível Global *cluster* 45: Gráfico Pizza e Gráfico de Barra dos procedimentos mais influentes

Fonte: Elaborada pelo autor.

Na Figura 18 é apresentado um novo gráfico de barras horizontal com a frequência dos diagnósticos nos registros analisados. O diagnóstico predominante deste *cluster* é “Insuficiencia cardíaca congestiva” existente em 45 dos 246 registros analisados. Nestes dados é possível que existam mais de um diagnóstico para uma internação, mas não é considerado consistente a repetição de algum diagnóstico em um registro de internação.

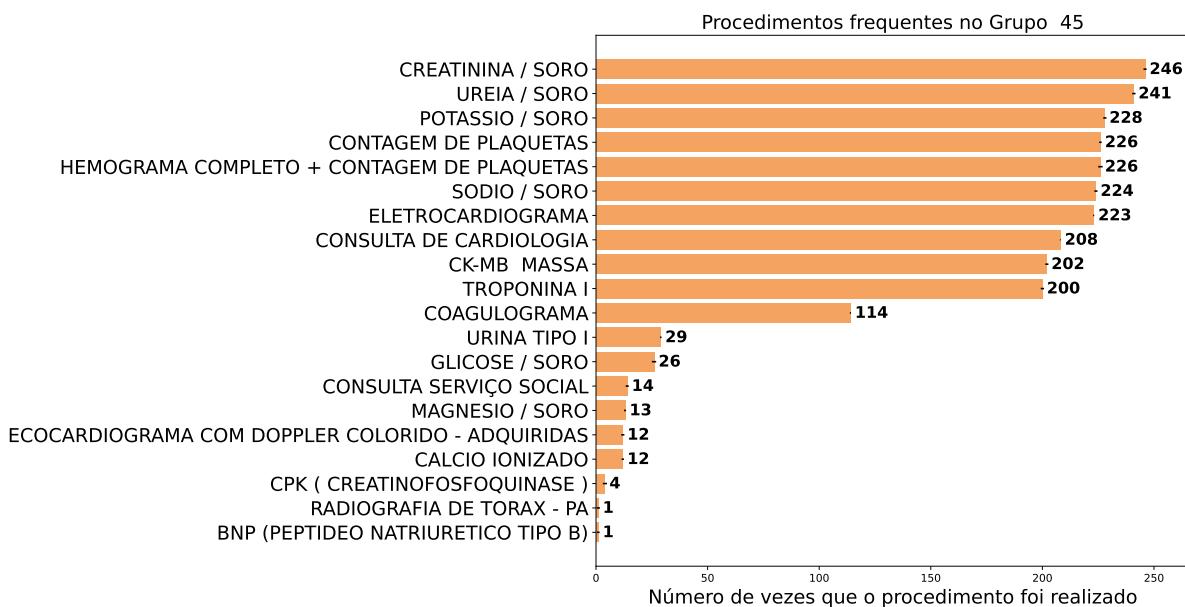
Na Figura 19, um gráfico de barras horizontal apresenta as frequências dos procedimentos realizados nas 245 internações. Os procedimentos predominantes deste *cluster* são “Creatinina/Soro” com 246 realizações e “Ureia/Soro” com 241 realizações. Com esses dados é possível conhecer os procedimentos comumente solicitados para pacientes com queixas semelhantes.

Figura 18 – Nível Global *cluster* 45: Gráficos de barras da frequência dos diagnósticos.



Fonte: Elaborada pelo autor.

Figura 19 – Nível Global *cluster* 45: Gráficos de barras da frequência dos procedimentos .



Fonte: Elaborada pelo autor.

Uma característica importante nesses dados é a independência entre a quantidade de internações e o número de vezes que os procedimentos foram realizados, tal que os procedimentos podem ser realizados mais de uma vez durante uma internação e também é possível que exista internações onde alguns procedimentos não tenham sido realizados. No caso de procedimentos realizados mais de uma vez em uma internação, isso pode ser visto ao comparar a quantidade de registros de internações (245) e a quantidade de vezes que o procedimento “Creatitina/Soro” foi realizado (246). Por outro lado, o caso de existirem procedimentos não realizados em todas as internações pode ser notado nos demais procedimentos que possuem

frequência inferior a quantidade de registros de internações, como exemplo é possível citar o procedimento “Ureia/Soro” que foi realizado 241 vezes em 245 internações.

Por fim, é importante ressaltar que caso ocorra a presença de um procedimento com o mesmo número de internações não é prova que este procedimento tenha sido realizado em todas as internações, podem existir repetições em algumas internações e a não realização em outras internações.

Cluster 7

Na [Figura 20\(a\)](#) também é apresentado o *label* ao qual o *cluster* pertence e a quantidade de registros de internações únicas neste caso as 151 que formam o *cluster* 7. Na [Figura 20\(b\)](#) o gráfico *BoxPlot* aponta a distribuição dos valores dos cinco procedimentos mais importantes na formação do cluster 7, podendo observar um *outlier* no procedimento “hematocrito”.

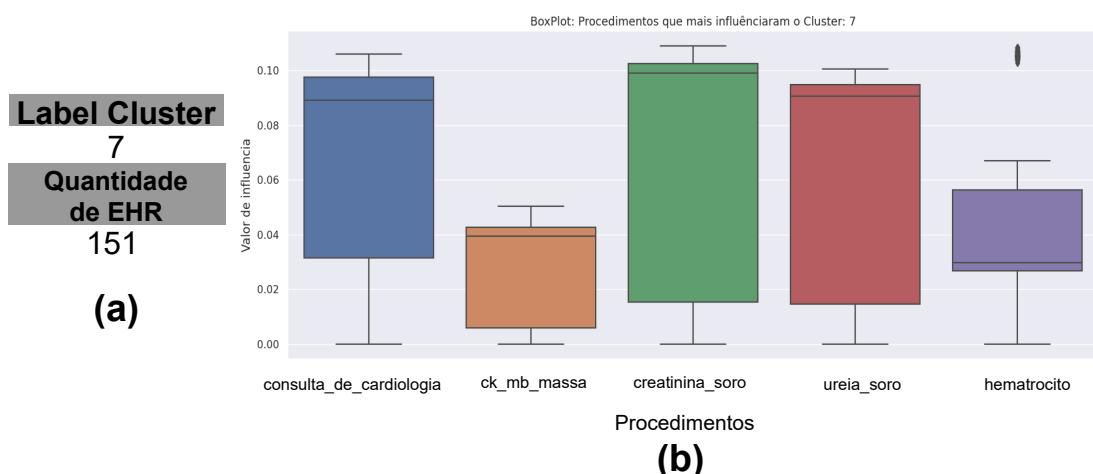


Figura 20 – Nível Global cluster 7: Informação do cluster 45 e Box Plot dos procedimentos mais frequentes.

Fonte: Elaborada pelo autor.

A [Figura 21](#) mostra uma tabela que contém os valores absolutos da influência de cada procedimento em relação à formação do *cluster* 7. Os cinco procedimentos mais importantes são apresentados distintamente, para os 324 demais é realizada a soma dos valores e apresentados em um único procedimento na coluna “Outros procedimentos”.

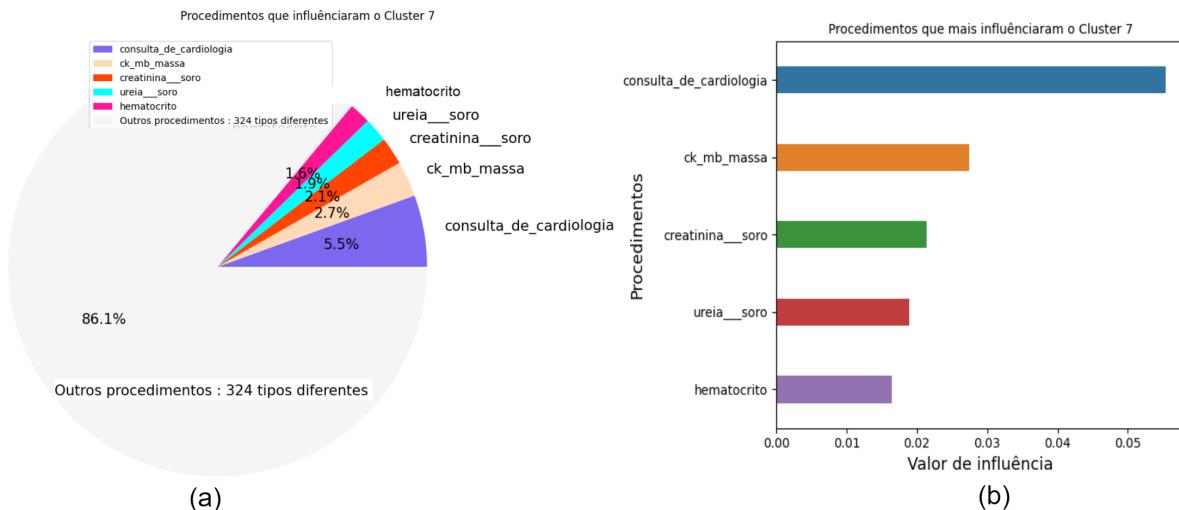
Na [Figura 23\(a\)](#) é apresentado o gráfico de pizza referindo à porcentagem dos procedimentos, enquanto na [Figura 23\(b\)](#) os valores de influência absolutos são apresentados no gráfico de barras horizontal. É possível notar que no *cluster* 7 os procedimentos mais influentes foram “Consulta de cardiologia” com 5,5%, “CK-Mb Massa” com 2,7%, “Creatinina/Soro” com 2,1%, “Ureia/Soro” com 1,9% e “Hematocrito” com 1,6%. Ainda é possível notar que outros 324 procedimentos foram realizados nos 151 registros de internações e dividem os 86% de influência complementares.

Figura 21 – Nível global cluster 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes

Nome do Atributo	Valor
consulta_de_cardiologia	0.05533858562717018
ck_mb_massa	0.027448208023301304
creatina__soro	0.021331845639094724
ureia__soro	0.018880403220863335
hematocrito	0.0164489832880354
Outros procedimentos : 324 tipos diferentes	0.86225301665377

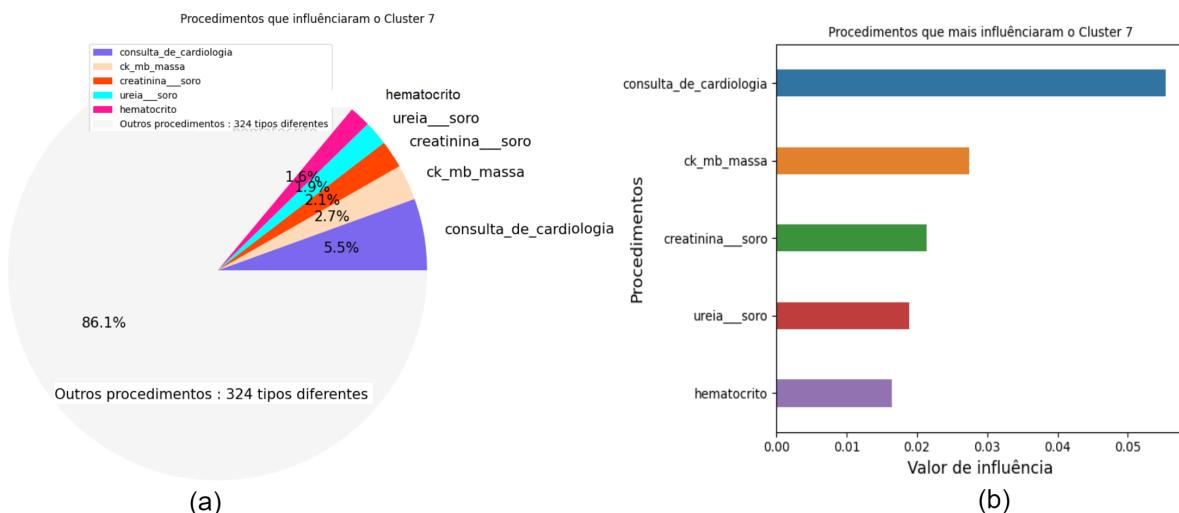
Fonte: Elaborada pelo autor.

Figura 22 – Nível global cluster 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes



Fonte: Elaborada pelo autor.

Figura 23 – Nível global cluster 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes

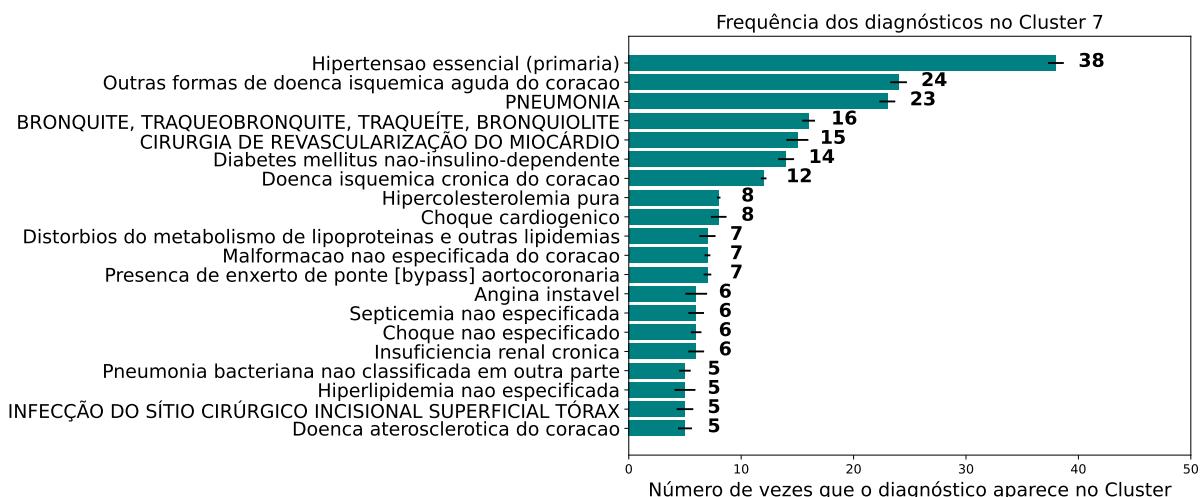


Fonte: Elaborada pelo autor.

Na [Figura 24](#) são apresentadas as frequências dos diferentes diagnósticos presente nos registros do *cluster* 7. O diagnóstico predominante é “Hipertensão essencial (primaria)” presente em 38 registros do total de 151, seguido de “Outras formas de doença isquemica aguda do coração” com 24 e “Pneumonia” com 23 registros.

Na [Figura 18](#) é apresentada a frequência dos procedimentos que foram realizados nos registros de internações que foram agrupados no *cluster* 7. Os quatro procedimentos mais frequentes são “Potassio/Soro”, “Sodio/Soro”, “Lactato(Lab. Centro Cirurgico)“ com os valores 4702, 4602, 3402, 3300 respectivamente. É possível notar algumas características do cluster, mesmo com o baixo número de internações únicas, as internações eram compostas por um grande número de procedimentos até mesmo a repetição de procedimentos em um único registro.

[Figura 24 – Nível Global cluster 7: Gráficos de Barras da frequência dos diagnósticos.](#)



Fonte: Elaborada pelo autor.

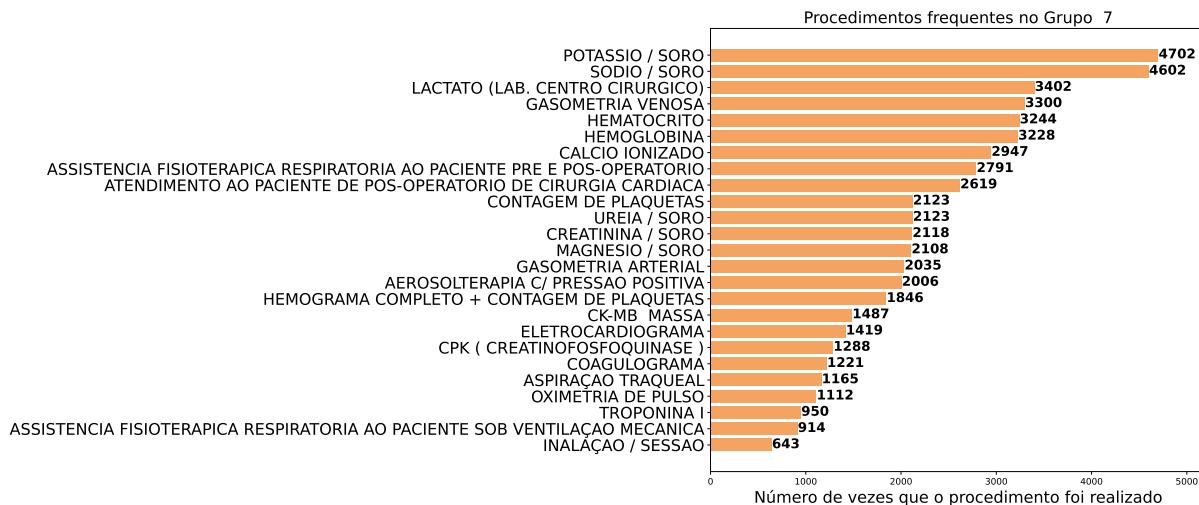
Nível Local

Cluster 45

Para apresentação explicativa no nível local a seleção do registro de internação é realizada de maneira aleatória, mas o método permite também o usuário selecionar o registro desejado por meio da especificação da “ID” de internação. Apesar do registro de internação ter sido selecionado de maneira aleatória, o *cluster* 45 foi escolhido de maneira específica para permitir também uma análise ligada ao nível Global. Essa abordagem permite uma visão *top-down* das informações explicativas.

As informações explicativas são apresentadas na [Figura 26](#). A [Figura 26\(a\)](#) é onde estão localizadas algumas informações descritivas do registro, como o label do *cluster* que o registro foi atribuído, o Id da internação e o nome do diagnóstico. Essas informações apresentam

Figura 25 – Nível Global cluster 7: Gráficos de Barras da frequência dos procedimentos.



Fonte: Elaborada pelo autor.

para o especialista que o registro pertence ao *cluster* 45, o Id desse registro de internação é 8903852871656482.0 e o nome do diagnóstico é “Hipertensao secundaria, nao especificada.”. Cruzando estas informações com as informações do nível Global (Figura 16 e Figura 17) é possível saber que este registro é um dos cinco registros existentes no *cluster* com Id 45 que são ligados a ocorrência “Hipertensao secundaria, nao especificada”.

Já na Figura 26(b), é apresentada a taxa de influência dos cinco procedimentos (atributos) mais influentes. Os procedimentos mais influentes em ordem decrescente são “Creatinina/Soro” com 0.10037, “Consulta de Cardiologia” com 0.09870, “Ureia/Soro” com 0.09737, “Sodio-/Soro” com 0.08826 e “Hemograma Completo + Contagem de plaqueta” com 0.07476. Embora o procedimento “Creatinina/Soro” não esteja entre os procedimentos mais influentes no nível Global para que este registro fosse inserido no *cluster* atual, a soma das taxas dos outros quatro procedimentos (que coincide com os mais influentes no nível Global) totaliza uma taxa de 0.35909. Superando assim, em termos dos procedimentos mais influentes, o primeiro procedimento com um valor de taxa três vezes maior. Com esse tipo de interpretação é possível compreender porque este registro foi inserido no *cluster* 45. Por fim, outros 192 procedimentos dividem uma taxa de influência com valor 2.21322 associado ao termo “Outros procedimentos”, nota-se que os valores individuais de influência são inferiores aos apresentados como mais influentes.

Quatro procedimentos indicados como mais influentes para a inserção deste registro em um *cluster* são também indicados como os procedimentos mais influentes para a construção do *cluster* 45 (nível Global). Apenas o quinto procedimento mais influente em ambos os níveis são iguais. Além disso, os procedimentos na segunda e terceira posição de ambos os níveis são apresentados em posições invertidas no nível oposto. Visualizando a semelhança entre os

Figura 26 – Nível local *cluster* 45: Informações do registro e valores absolutos dos procedimentos mais frequentes.

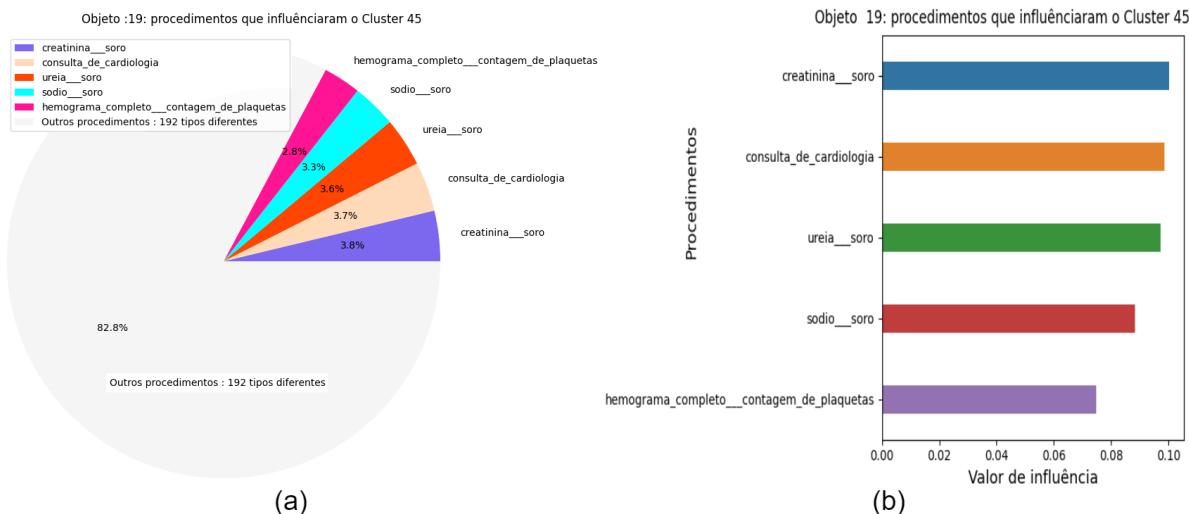
(a)	Label do Cluster	ID da Hospitalização	Nome da Ocorrência
	45	8903852871656482.0	Hipertensao secundaria, nao especificada
(b)	Nome do Atributo	Valor	
	creatinina__soro	0.037385952484440335	
	ureia __ soro	0.03456455868036763	
	consulta_de_cardiologia	0.03445624706507187	
	sodio__soro	0.03185052071446546	
	hemograma_completo__contagem_de_plaquetas	0.02911533953283154	
Outros procedimentos : 315 tipos diferentes			0.8328625085568883

Fonte: Elaborada pelo autor.

procedimentos com maior influência em ambos os níveis, é possível compreender melhor porque o registro com Id 8903852871656482.0 foi inserido no *cluster* 45. Além disso, estas informações explicativas demonstram a coerência entre os dois níveis (Local e Global) para a compreensão do processo de agrupamento.

Na Figura 27(a), é apresentado um gráfico de pizza com valores em porcentagem referente a influência que os procedimentos apresentaram. Em Figura 27(b) apresenta os valores absolutos da influência representado por um gráfico de barras horizontal.

Figura 27 – Nível local *cluster* 45: Informações do registro e valores absolutos dos procedimentos mais frequentes.



Fonte: Elaborada pelo autor.

Cluster 7

Para representação do nível local do *cluster 7*, foi selecionado um registro aleatoriamente, ou seja, para a escolha do registro basta indicar que o registro deve pertencer ao *cluster 7*. Na [Figura 28\(a\)](#) são apresentadas informações referentes a descrição e característica do registro selecionado, tais como: a identificação do *cluster 7*, a identificação do registro de internação 6716956757924323.0 e o nome do diagnóstico do registro “Infarto agudo do miocardio”.

Com os resultados globais do *cluster 7* obtidos e apresentados na [Figura 21](#) e os resultados locais na [Figura 28](#) é possível notar algumas características semelhantes na formação do *cluster 7*. Os cinco procedimentos considerados mais influentes na caracterização do *cluster* no nível local são: “Consulta de cardiologia” 0.07851, “Angiografia com balão e stent” 0.05968, “Ck-mb/massa” 0.03933, “Creatinina/soro” 0.03063 e “Ureia/soro”.

Realizando a comparação dos procedimentos locais quatro deles “Consulta de cardiologia”, “Ck-mb/massa” e “Creatinina/soro”, Ureia/soro também marcaram presença entre os cinco procedimentos mais importantes do nível global. Representando assim, 80% da correspondência dos procedimentos nos dois níveis de explicação.

Figura 28 – Nível local *cluster 7*: Informações do registro e valores absolutos dos procedimentos mais frequentes.

	Label do Cluster	ID da Hospitalização	Nome da Ocorrência
(a)	7	6716956757924323.0	Infarto agudo do miocardio
		Nome do Atributo	Valor
		Outros procedimentos : 157 tipos diferentes	0.9692820801196376
		consulta_de_cardiologia	0.0785145010801352
(b)		angiografia_com_bal_o_e_stent	0.0596838987757039
		ck_mb_massa	0.0393385861143942
		creatinina_soro	0.0306379478048596
		ureia_soro	0.0273253520684429

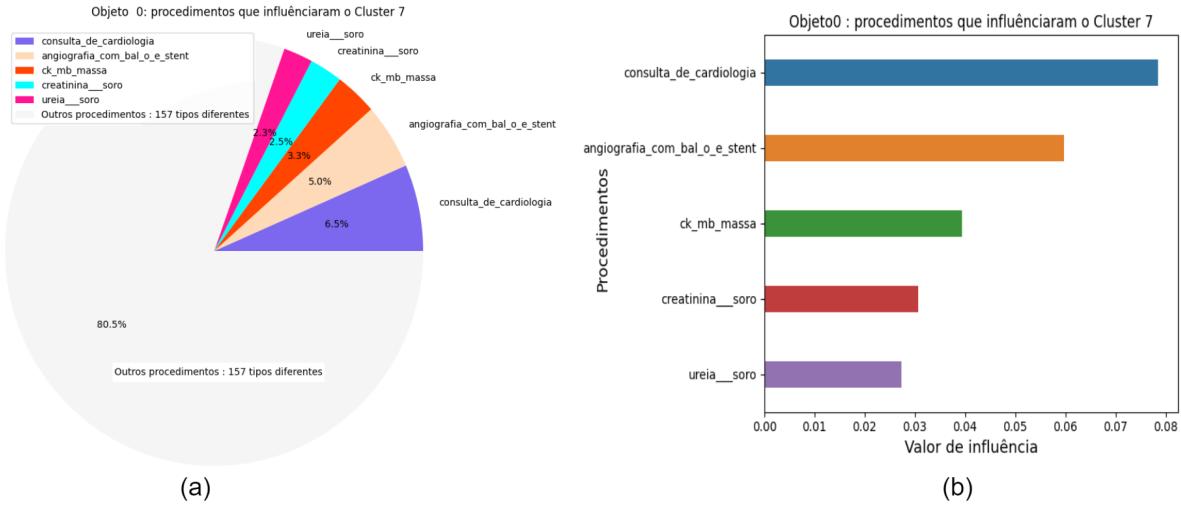
Fonte: Elaborada pelo autor.

A [Figura 29\(a\)](#), apresenta um gráfico de pizza com valores em porcentagem referentes à influência que os procedimentos apresentaram. A [Figura 29\(b\)](#) indica os valores absolutos da influência representado por um gráfico de barras horizontal.

4.5 Considerações finais

Neste capítulo foram apresentados dois métodos que se complementam, primeiramente a Representação *Bag-of-Attributes* (BOAR) para registros médicos eletrônicos de saúde (EHR) que fornecem uma representação única para qualquer algoritmo de agrupamento que utiliza

Figura 29 – Nível local cluster 7: Gráfico pizza e gráfico de barra dos procedimentos mais influentes.



Fonte: Elaborada pelo autor.

o modelo de espaço vetorial (VSM) como entrada. O BOAR se destaca principalmente pela sua capacidade de atuar em diferentes *datasets* apoiando-se em um modelo comum de dados, oferecendo a capacidade de produzir um modelo comum para avaliar, comparar e descobrir informações semelhantes em EHR, independente dos algoritmos de agrupamento.

Uma vez apresentados os resultados do agrupamento faz-se necessária a compreensão das informações obtidas, o que é realizado pelo *Multilevel Clustering Explainer* (MCE). Conforme demonstrado pelos resultados obtidos, o método proposto MCE é capaz de reunir informações explicativas sobre os dados utilizados na detecção de padrões e recuperar conteúdo com informações descritivas para complementar a explicação no contexto médico. O método proposto também permite que o especialista personalize a forma de apresentação dos resultados, com a inserção de parâmetros diferentes aos *default*. Por fim, o método se mostrou robusto na construção de informações explicativas e também ao permitir uma análise *top-down* para compreender possíveis casos de inconsistências.

Por característica dos dados utilizados no processo de agrupamento, é importante detalhar que neste trabalho as informações explicativas referem-se ao quanto um procedimento influenciou no agrupamento por estar presente a sua realização nos respectivos registros de internações. Ou seja, a informação refere-se ao quanto o resultado de um procedimento pode ter agregado informações (favorável ou não favorável) por ter sido realizado em um paciente para apoiar a tomada de decisão do especialista sobre um possível diagnóstico.



CONCLUSÃO

5.1 Considerações finais

Essa dissertação de mestrado visou contribuir com a explicabilidade e utilização de EHR por meio de utilização de *clustering*, que permitiu encontrar padrões significativos em bases de registros eletrônicos de saúde. As novas técnicas apresentadas avançaram no estado da arte relativo à interpretabilidade sobre a formação de *clusters* aplicado ao contexto médico.

A utilização da representação BOAR também proposta neste trabalho, traz ao profissional médico mais facilidade no processo de obtenção do conhecimento, pois o método permite lidar desde o pré-processamento dos dados até a apresentação dos resultados. A estrutura modular do método também permite a alteração dos parâmetros desejados de forma direta e simples.

5.2 Contribuições

Este trabalho apresentou duas contribuições principais. A primeira dela foi *Bag-of-Attributes Representation* (BOAR) que compôs a criação de uma representação genérica baseada no modelo de espaço vetorial, que permite superar a interoperabilidade entre bases de dados e também a comparação justa dos resultados obtidos por diferentes tipos de algoritmos de agrupamento.

A segunda contribuição deste trabalho foi o *Multilevel Clustering Explainer* (MCE), que é um método de construção e apresentação de informações explicativas do processo de detecção de padrões em registros de internações de pacientes. No MCE o objetivo foi de reunir e apresentar informações explicativas em multiníveis (Local e Global), referente ao processo realizado na detecção de padrões (tal como agrupamento) para apoiar a tomada de decisões do especialista da área de saúde. Ou seja, o especialista tem acesso ao grupo de pacientes cujos indicadores são semelhantes ao do paciente que está sendo analisado.

5.3 Trabalhos publicados

No período de desenvolvimento da pesquisa foram elaborados e publicados os seguintes artigos:

- Clementino, J. M.; Bones, C. C.; Faiçal, B. S.; Linares, O. C.; Lima, D. M.; Gutierrez, M. A.; Traina, C.; Traina, A. J. M“ **Bag-of-Attributes Representation: A Vector Space Model for Electronic Health Records Analysis in OMOP**”, 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), 2020, pp. 197-202, doi: 10.1109/CBMS49503.2020.00045. (**Qualis A3**)
- Clementino, J. M.; Faiçal, B. S.; Bones, C. C.; Gutierrez, M. A.; Traina, C.; Traina, A. J. M“, “ **Multilevel Clustering Explainer: An Explainable Approach to Electronic Health Records**”, 2021 IEEE 34rd International Symposium on Computer-Based Medical Systems (CBMS), 2021, pp. 253-258, doi: 10.1109/CBMS52027.2021.00073. (**Qualis A3**)

5.4 Outras atividades científicas

Outras atividades acadêmicas foram realizadas no período do curso de mestrado, sendo as seguintes:

- Orientações de TCC
 - BALBONI, Camila de Bernardi; GOBETTI, Claudio Gomes; VIEIRA, Edson de Oliveira. **Explainable Artificial Intelligence: Técnica de explicação profunda para detecção de fraudes em finanças**. Orientador: José Maria Clementino Junior. 2021. 24 f. TCC (Graduação) – Engenharia da Computação, Universidade Virtual de São Paulo (UNIVESP), São Paulo. 2021.
 - REZENDE, Diego.H. Cornélio; FURTADO, José .A. Alves. **Transfer Learning para detecção de Manchas de Bitot em crianças de 6 a 59 meses para diagnóstico de deficiência de vitamina A e prevenção da desnutrição infantil**. Orientador: José Maria Clementino Junior. 2021. 24 f. TCC (Graduação) – Engenharia da Computação, Universidade Virtual de São Paulo (UNIVESP), São Paulo. 2021.
 - OLIVEIRA, Cristiano C. Oliveira; SILVA, Jorge Mauro. **Aprendizagem de Máquina aplicada em base de dados de saúde: Uma análise preditiva do risco de doenças cardíacas**. Orientador: José Maria Clementino Junior. 2021. 24 f. TCC (Graduação) – Engenharia da Computação, Universidade Virtual de São Paulo (UNIVESP), São Paulo. 2021.

- MIYAMOTO, Douglas Massao; GUITZEL, Elizabete; CARVALHO, Gerson Borges. **Particionamento e armazenamento seguro na nuvem: Uma metodologia para utilização de volume de dados superior a capacidade de armazenamento em disco do servidor pela divisão do banco de dados e uso escalonado das partes.** Orientador: José Maria Clementino Junior. 2021. 24 f. TCC (Graduação) – Engenharia da Computação, Universidade Virtual de São Paulo (UNIVESP), São Paulo. 2021.
- SOUZA, Marcos Lázaro; MOTA, Reginaldo da Silva; OLIVEIRA, Rogério Misias. **Desenvolvimento de um sistema inteligente IoT para monitoramento de temperatura de alimentos.** Orientador: José Maria Clementino Junior. 2021. 30 f. TCC (Graduação) – Engenharia da Computação, Universidade Virtual de São Paulo (UNIVESP), São Paulo. 2021.
- GRACELLI, Ricardo Alexandre; ROSÁRIO, Rodrigo Dias; DUARTE, Roberto Carlos. **Sistema automático para leitura e controle de temperatura corporal.** Orientador: José Maria Clementino Junior. 2021. 24 f. TCC (Graduação) – Engenharia da Computação, Universidade Virtual de São Paulo (UNIVESP), São Paulo. 2021.

5.5 Trabalhos futuros

Existem várias maneiras de explorar trabalhos futuros relacionados às propostas do trabalho. Dentre elas pode-se citar:

- Acrescentar ao BOAR uma abordagem semântica possibilitando assim trabalhar com EHR descritivos.
- Desenvolver um estudo de caso com profissionais, com intuito de analisar e validar os resultados visuais obtidos pelo MCE.
- Estudar o comportamento do MCE não apenas em relação a formação de um único *cluster*, mas também a formação de sub-*clusters*, visando entender possíveis hierarquias intra-*cluster*.
- Criar uma ferramenta para integrar o método MCE, com intuito de permitir ao profissional de saúde uma interface para alteração da parametrização de algoritmos, funções de distância e gráficos dos resultados.

REFERÊNCIAS

- ADADI, A.; BERRADA, M. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). **IEEE Access**, v. 6, p. 52138–52160, 2018. Citado nas páginas [24](#), [31](#), [32](#), [33](#), [34](#) e [35](#).
- ARRIETA, A. B.; DÍAZ-RODRÍGUEZ, N.; SER, J. D.; BENNETOT, A.; TABIK, S.; BARBADO, A.; GARCÍA, S.; GIL-LÓPEZ, S.; MOLINA, D.; BENJAMINS, R. *et al.* Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. **arXiv preprint arXiv:1910.10045**, 2019. Citado nas páginas [34](#), [35](#) e [42](#).
- ATASHI, A.; TOHIDINEZHAD, F.; DORRI, S.; NAZERI, N.; GHOUSI, R.; MARASHI, S.; HAJIALIASGARI, F. Discovery of hidden patterns in breast cancer patients, using data mining on a real data set. **Studies in health technology and informatics**, v. 262, p. 142–145, 2019. Citado nas páginas [30](#), [38](#), [39](#) e [40](#).
- BARR, A. Google mistakenly tags black people as ‘gorillas,’ showing limits of algorithms. **The Wall Street Journal**, v. 1, p. 2015, 2015. Citado na página [36](#).
- BERGER, M. L.; DOBAN, V. Big data, advanced analytics and the future of comparative effectiveness research. **Journal of Comparative Effectiveness Research**, v. 3, n. 2, p. 167–176, 2014. Disponível em: <<https://doi.org/10.2217/cer.14.2>>. Citado na página [23](#).
- BURKOV, A. **The Hundred-Page Machine Learning Book**. [S.l.]: Andriy Burkov, 2019. Citado na página [28](#).
- Cavaliere, F. *et al.* Parkinson’s disease diagnosis: Towards grammar-based explainable artificial intelligence. In: **2020 IEEE Symposium on Computers and Communications (ISCC)**. [S.l.: s.n.], 2020. p. 1–6. Citado na página [24](#).
- Clementino, J. M.; Bones, C. C.; Faiçal, B. S.; Linares, O. C.; Lima, D. M.; Gutierrez, M. A.; Traina, C.; Traina, A. J. M. Bag-of-attributes representation: A vector space model for electronic health records analysis in omop. In: **2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.: s.n.], 2020. p. 197–202. Citado na página [49](#).
- CORTEZ, P.; EMBRECHTS, M. J. Opening black box data mining models using sensitivity analysis. In: **2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)**. [S.l.: s.n.], 2011. p. 341–348. Citado na página [33](#).
- DARPA. 2018. Disponível em: <<https://www.darpa.mil/program/explainable-artificial-intelligence>>. Acesso em: 01/02/2012. Citado na página [32](#).
- DOŠILOVIĆ, F. K.; BRČIĆ, M.; HLUPIĆ, N. Explainable artificial intelligence: A survey. In: **2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)**. [S.l.: s.n.], 2018. p. 0210–0215. Citado nas páginas [31](#), [33](#) e [35](#).

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. *et al.* Inteligência artificial: Uma abordagem de aprendizado de máquina. 2011. Citado nas páginas [28](#) e [29](#).

FUNKNER, A. A.; YAKOVLEV, A. N.; KOVALCHUK, S. V. Towards evolutionary discovery of typical clinical pathways in electronic health records. **Procedia Computer Science**, v. 119, p. 234 – 244, 2017. ISSN 1877-0509. 6th Intl. Young Scientist Conference on Computational Science, YSC 2017, 01-03 November 2017, Kotka, Finland. Citado na página [41](#).

GARCELON, N.; NEURAZ, A.; BENOIT, V.; SALOMON, R.; KRACKER, S.; SUAREZ, F.; BAHI-BUISSON, N.; HADJ-RABIA, S.; FISCHER, A.; MUNNICH, A.; BURGUN, A. Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse: Dr. warehouse and the needle in the needle stack. **Journal of Biomedical Informatics**, v. 73, p. 51 – 61, 2017. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046417301764>>. Citado nas páginas [30](#), [38](#), [39](#) e [40](#).

GARI, R. C.; CIATTO, G.; DELLA LUCE, J.; OMICINI, A. Interpretable narrative explanation for ML predictors with LP: A case study for XAI. In: BERGENTI, F.; MONICA, S. (Ed.). **WOA 2019 – 20th Workshop ‘From Objects to Agents’**. Sun SITE Central Europe, RWTH Aachen University, 2019, (CEUR Workshop Proceedings, v. 2404). p. 105–112. Disponível em: <<http://ceur-ws.org/Vol-2404/paper16.pdf>>. Citado nas páginas [32](#), [35](#), [37](#), [38](#), [39](#) e [40](#).

GOMES, J. B.; RODRIGUES, J. J.; AL-MUHTADI, J.; ARUNKUMAR, N.; RABÊLO, R. A.; FURTADO, V. An iot-based smart solution for preventing domestic co and lpg gas accidents. In: IEEE. **2018 IEEE 10th Latin-American Conference on Communications (LATINCOM)**. [S.I.], 2018. p. 1–6. Citado na página [23](#).

GUIDOTTI, R.; MONREAL, A.; RUGGIERI, S.; TURINI, F.; GIANNOTTI, F.; PEDRESCHI, D. A survey of methods for explaining black box models. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 51, n. 5, p. 93:1–93:42, ago. 2018. ISSN 0360-0300. Disponível em: <<http://doi.acm.org/10.1145/3236009>>. Citado nas páginas [32](#) e [42](#).

GUO, J.; CHEN, Q.; WANG, C.; QIU, H.; LIU, B.; JIANG, Z.-H.; ZHANG, W. Comparison of two exploratory data analysis methods for classification of phyllanthus chemical fingerprint: unsupervised vs. supervised pattern recognition technologies. **Analytical and Bioanalytical Chemistry**, v. 407, n. 5, p. 1389–1401, Feb 2015. ISSN 1618-2650. Disponível em: <<https://doi.org/10.1007/s00216-014-8371-x>>. Citado na página [28](#).

Gupta, R.; Singhal, A.; Sai Sabitha, A. Comparative study of clustering algorithms by conducting a district level analysis of malnutrition. In: **2018 8th Intl. Conference on Cloud Computing, Data Science Engineering (Confluence)**. [S.I.: s.n.], 2018. p. 280–286. ISSN null. Citado na página [60](#).

HAN, L.; LUO, S.; WANG, H.; PAN, L.; MA, X.; ZHANG, T. An intelligible risk stratification model based on pairwise and size constrained kmeans. **IEEE J. Biomedical and Health Informatics**, v. 21, n. 5, p. 1288–1296, 2017. Disponível em: <<https://doi.org/10.1109/JBHI.2016.2633403>>. Citado na página [29](#).

HE, Q. P.; QIN, S. J.; WANG, J. A new fault diagnosis method using fault directions in fisher discriminant analysis. **AIChE Journal**, v. 51, n. 2, p. 555–571, 2005. Disponível em: <<https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.10325>>. Citado na página [29](#).

- HENKE, R. M.; KARACA, Z.; MOORE, B.; CUTLER, E.; LIU, H.; MARDER, W. D.; WONG, H. S. Impact of health system affiliation on hospital resource use intensity and quality of care. **Health Services Research**, v. 53, n. 1, p. 63–86, 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6773.12631>>. Citado na página 41.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. **IEEE Transactions on pattern analysis and machine intelligence**, Ieee, v. 22, n. 1, p. 4–37, 2000. Citado na página 27.
- JALEEL, A.; MAHMOOD, T.; HASSAN, M. A.; BANO, G.; KHURSHID, S. K. Towards medical data interoperability through collaboration of healthcare devices. **IEEE Access**, IEEE, v. 8, p. 132302–132319, 2020. Citado na página 41.
- JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of ai ethics guidelines. **Nature Machine Intelligence**, Nature Publishing Group, p. 1–11, 2019. Citado na página 36.
- KAUSHAL, R.; SHOJANIA, K. G.; BATES, D. W. Effects of Computerized Physician Order Entry and Clinical Decision Support Systems on Medication Safety: A Systematic Review. **JAMA Internal Medicine**, v. 163, n. 12, p. 1409–1416, 06 2003. ISSN 2168-6106. Disponível em: <<https://doi.org/10.1001/archinte.163.12.1409>>. Citado na página 23.
- KHALID, S.; JUDGE, A.; PINEDO-VILLANUEVA, R. An unsupervised learning model for pattern recognition in routinely collected healthcare data. Scitepress, 2018. Citado na página 29.
- LAVRAS, C. Atenção primária à saúde e a organização de redes regionais de atenção à saúde no brasil. **Saúde e Sociedade**, scielo, v. 20, p. 867 – 874, 12 2011. ISSN 0104-1290. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-12902011000400005&nrm=iso>. Citado na página 41.
- LENT, M. V.; FISHER, W.; MANCUSO, M. An explainable artificial intelligence system for small-unit tactical behavior. In: MENLO PARK, CA; CAMBRIDGE, MA; LONDON; AAAI PRESS; MIT PRESS; 1999. **Proceedings of the national conference on artificial intelligence**. [S.l.], 2004. p. 900–907. Citado nas páginas 31 e 32.
- LEPRI, B.; OLIVER, N.; LETOUZÉ, E.; PENTLAND, A.; VINCK, P. Fair, transparent, and accountable algorithmic decision-making processes. **Philosophy & Technology**, v. 31, n. 4, p. 611–627, Dec 2018. ISSN 2210-5441. Disponível em: <<https://doi.org/10.1007/s13347-017-0279-x>>. Citado na página 35.
- LIMA, D. M.; RODRIGUES-JR, J. F.; TRAINA, A. J.; PIRES, F. A.; GUTIERREZ, M. A. Transforming two decades of epr data to omop cdm for clinical research. **Stud Health Technol Inform**, v. 264, p. 233–237, 2019. Citado na página 59.
- LIN, Y.-K.; LIN, M.; CHEN, H. Do electronic health records affect quality of care? evidence from the hitech act. **Information Systems Research**, v. 30, n. 1, p. 306–318, 2019. Disponível em: <<https://doi.org/10.1287/isre.2018.0813>>. Citado na página 23.
- LISMONT, J.; JANSSENS, A.-S.; ODNOLETKOVA, I.; BROUCKE, S. vanden; CARON, F.; VANTHIENEN, J. A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways. **Computers in Biology and Medicine**, v. 77, p. 125 – 134, 2016. ISSN 0010-4825. Citado na página 41.

LOUIE, B.; MORK, P.; MARTIN-SANCHEZ, F.; HALEVY, A.; TARCZY-HORNOCH, P. Data integration and genomic medicine. **Journal of Biomedical Informatics**, v. 40, n. 1, p. 5 – 16, 2007. ISSN 1532-0464. Bio*Medical Informatics. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046406000244>>. Citado na página 41.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 4768–4777. ISBN 9781510860964. Citado na página 24.

_____. A unified approach to interpreting model predictions. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems 30**. Curran Associates, Inc., 2017. p. 4765–4774. Disponível em: <<http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>>. Citado nas páginas 38, 39 e 40.

_____. A unified approach to interpreting model predictions. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 4768–4777. ISBN 9781510860964. Citado na página 63.

MATHEWS, S. M. Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review. In: ARAI, K.; BHATIA, R.; KAPOOR, S. (Ed.). **Intelligent Computing**. Cham: Springer International Publishing, 2019. p. 1269–1292. ISBN 978-3-030-22868-2. Citado nas páginas 24, 31, 32, 35 e 36.

NEERINCX, M. A.; WAA, J. van der; KAPTEIN, F.; DIGGELEN, J. van. Using perceptual and cognitive explanations for enhanced human-agent team performance. In: HARRIS, D. (Ed.). **Engineering Psychology and Cognitive Ergonomics**. Cham: Springer International Publishing, 2018. p. 204–214. ISBN 978-3-319-91122-9. Citado na página 24.

NHSI. **National Health Interview Survey**. 2020. Disponível em: <<https://medical-data-models.org/?lang=en>>. Acesso em: 27/01/2020. Citado nas páginas 24 e 41.

NIE, R.; HE, M.; CAO, J.; ZHOU, D.; LIANG, Z. Pulse coupled neural network based mri image enhancement using classical visual receptive field for smarter mobile healthcare. **Journal of Ambient Intelligence and Humanized Computing**, Springer, p. 1–12, 2018. Citado na página 23.

NOIA, A. D.; MARTINO, A.; MONTANARI, P.; RIZZI, A. Supervised machine learning techniques and genetic optimization for occupational diseases risk prediction. **Soft Computing**, Jul 2019. ISSN 1433-7479. Disponível em: <<https://doi.org/10.1007/s00500-019-04200-2>>. Citado na página 29.

OBERMEYER, Z.; POWERS, B.; VOGELI, C.; MULLAINATHAN, S. Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, American Association for the Advancement of Science, v. 366, n. 6464, p. 447–453, 2019. Citado na página 36.

OMAR, E. P. e Vilma França e N. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 1, n. 1, p. 495–504, 2003. ISSN

2316-6533. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/article/view/280>>. Citado na página 29.

OMOP. **OMOP Common Data Model**. 2020. Disponível em: <<https://www.ohdsi.org/data-standardization/the-common-data-model/>>. Acesso em: 27/01/2020. Citado nas páginas 57 e 58.

PÁEZ, A. The pragmatic turn in explainable artificial intelligence (xai). **Minds and Machines**, v. 29, n. 3, p. 441–459, Sep 2019. ISSN 1572-8641. Disponível em: <<https://doi.org/10.1007/s11023-019-09502-w>>. Citado nas páginas 31, 33 e 37.

QADRI, Y. A.; NAUMAN, A.; ZIKRIA, Y. B.; VASILAKOS, A. V.; KIM, S. W. The future of healthcare internet of things: a survey of emerging technologies. **IEEE Communications Surveys & Tutorials**, IEEE, v. 22, n. 2, p. 1121–1167, 2020. Citado na página 41.

RAZZAK, M. I.; IMRAN, M.; XU, G. Big data analytics for preventive medicine. **Neural Computing and Applications**, Springer, v. 32, n. 9, p. 4417–4451, 2020. Citado na página 41.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should I trust you?": Explaining the predictions of any classifier. In: **Proceedings of the 22nd ACM SIGKDD, San Francisco, CA, USA, August 13-17, 2016**. [S.l.: s.n.], 2016. p. 1135–1144. Citado na página 24.

_____. Why should i trust you?: Explaining the predictions of any classifier. In: **ACM. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**. [S.l.], 2016. p. 1135–1144. Citado nas páginas 36, 38, 39, 40 e 63.

RIBERA, M.; LAPEDRIZA, À. Can we do better explanations? a proposal of user-centered explainable ai. In: **IUI Workshops**. [S.l.: s.n.], 2019. Citado na página 31.

RODRIGUES, L. S.; CAZZOLATO, M. T.; TRAINA, A. J. M.; TRAINA, C. Taking advantage of highly-correlated attributes in similarity queries with missing values. In: **SPRINGER. International Conference on Similarity Search and Applications**. [S.l.], 2020. p. 168–176. Citado na página 41.

SHIRATAKI, S.; YAMAGUCHI, S. A study on interpretability of decision of machine learning. In: **2017 IEEE International Conference on Big Data (Big Data)**. [S.l.: s.n.], 2017. p. 4830–4831. Citado nas páginas 37, 38, 39 e 40.

SIMIĆ, S.; BANKOVIĆ, Z.; SIMIĆ, D.; SIMIĆ, S. D. A hybrid clustering approach for diagnosing medical diseases. In: **SPRINGER. International Conference on Hybrid Artificial Intelligence Systems**. [S.l.], 2018. p. 741–752. Citado nas páginas 29, 38, 39 e 40.

SUSHIL, M.; ŠUSTER, S.; LUYCKX, K.; DAELEMANS, W. Patient representation learning and interpretable evaluation using clinical notes. **Journal of Biomedical Informatics**, v. 84, p. 103 – 113, 2018. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046418301266>>. Citado nas páginas 30, 38, 39 e 40.

SZEGEDY, C.; ZAREMBA, W.; SUTSKEVER, I.; Bruna Estrach, J.; ERHAN, D.; GOODFELLOW, I.; FERGUS, R. Intriguing properties of neural networks. In: . [S.l.: s.n.], 2014. 2nd International Conference on Learning Representations, ICLR 2014 ; Conference date: 14-04-2014 Through 16-04-2014. Citado na página 35.

THEODORIDIS, S.; KOUTROUMBAS, K. Pattern recognition. Academic press, 1998. Citado nas páginas 27 e 28.

WATANABE, S. **Pattern recognition: human and mechanical.** [S.l.]: John Wiley & Sons, Inc., 1985. Citado na página 27.

WOLF, M. J.; MILLER, K.; GRODZINSKY, F. S. Why we should have seen that coming: Comments on microsoft's tay "experiment,"and wider implications. **SIGCAS Comput. Soc.**, ACM, New York, NY, USA, v. 47, n. 3, p. 54–64, set. 2017. ISSN 0095-2737. Disponível em: <<http://doi.acm.org/10.1145/3144592.3144598>>. Citado na página 35.

WU, R. M.; LIU, D. A.; XIAO, Y. Z. Scheme design for machine based on black box-gray box theory. In: **Advanced Materials Science and Technology, ICMST 2010.** [S.l.]: Trans Tech Publications Ltd, 2011. (Advanced Materials Research, v. 181), p. 247–250. Citado na página 35.

XIE, Y.; CHEN, X. A.; GAO, G. Outlining the design space of explainable intelligent systems for medical diagnosis. In: **Joint Proceedings of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces (ACM IUI 2019), Los Angeles, USA, March 20, 2019.** [s.n.], 2019. Disponível em: <<http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-18.pdf>>. Citado nas páginas 24 e 42.

ZHANG, W.; YU, Y.-C.; LI, J.-S. Dynamics reconstruction and classification via koopman features. **Data Mining and Knowledge Discovery**, v. 33, n. 6, p. 1710–1735, Nov 2019. ISSN 1573-756X. Disponível em: <<https://doi.org/10.1007/s10618-019-00639-x>>. Citado na página 28.

ZHANG, Y.; JIN, R.; ZHOU, Z.-H. Understanding bag-of-words model: a statistical framework. **International Journal of Machine Learning and Cybernetics**, Springer, v. 1, n. 1-4, p. 43–52, 2010. Citado na página 31.

ZHAO, T.; WANG, Y. Interpolation and stratification of multilayer soil property profile from sparse measurements using machine learning methods. **Engineering Geology**, v. 265, p. 105430, 2020. ISSN 0013-7952. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0013795219313420>>. Citado na página 28.

