Universidade do Estado do Rio de Janeiro
Centro de Tecnologia e Ciências
Instituto de Matemática e Estatística

Carla Cristina Passos Cruz

Modelagem *Fuzzy* Intuicionista no Reconhecimento de Padrões em Registros no Prontuário Eletrônico de Pacientes atendidos no Hospital Universitário durante a Pandemia da Covid-19

Rio de Janeiro
2024

Carla Cristina Passos Cruz

Modelagem *Fuzzy* Intuicionista no Reconhecimento de Padrões em Registros no Prontuário Eletrônico de Pacientes atendidos no Hospital Universitário durante a Pandemia da Covid-19

Tese apresentada como requisito para obtenção do título de Doutora, ao Programa de Pós-graduação em Ciências Computacionais e Modelagem Matemática, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.ª Dra. Regina Serrão Lanzillotti

Rio de Janeiro

2024

CATALOGAÇÃO NA FONTE

UERJ/REDE SIRIUS/BIBLIOTECA CTC/A

C957 Cruz, Carla Cristina Passos.

Modelagem Fuzzy intuicionista no reconhecimento de padrões em registros no prontuário eletrônico de pacientes atendidos no hospital universitário durante a pandemia da Covid-19 / Carla Cristina Passos Cruz. – 2024.

149 f.: il.

Orientadora: Regina Serrão Lanzillotti

Tese (Doutorado em Ciências Computacionais e Modelagem Matemática) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Processamento eletrônico de dados - Medicina - Teses. 2. Sistemas difusos - Teses. 3. Arquivos médicos - Teses. 4. COVID-19 - Teses. I. Lanzillotti, Regina Serrão. II. Universidade do Estado do Rio de Janeiro. Instituto de Matemática e Estatística. III. Título.

CDU 004:61

Patricia Bello Meijinhos CRB7/5217 - Bibliotecária responsável pela elaboração da ficha catalográfica

_____ _____

Assinatura Data

Carla Cristina Passos Cruz

Modelagem *Fuzzy* Intuicionista no Reconhecimento de Padrões em Registros no Prontuário Eletrônico de Pacientes atendidos no Hospital Universitário durante a Pandemia da Covid-19

Tese apresentada como requisito para obtenção do título de Doutora, ao Programa de Pós-graduação em Ciências Computacionais e Modelagem Matemática, da Universidade do Estado do Rio de Janeiro.

Aprovada em 05 de novembro de 2024.
Banca Examinadora :

_____
Prof.ª Dra. Regina Serrão Lanzillotti (Orientadora)
Instituto de Matemática e Estatística - UERJ

_____
Prof. Dr. Agnaldo Lopes
Faculdade de Ciências Médicas - UERJ

_____
Prof.ª Dra. Euzeli da Silva Brandão
Universidade Federal Fluminense - UFF

_____
Prof.ª Dra. Karla Tereza Figueiredo Leite
Instituto de Matemática e Estatística - UERJ

_____
Prof.ª Dra. Michelle Marcia Viana Martins
Universidade Federal de Viçosa - UFV

_____
Prof.ª Dra. Zochil González Arenas
Instituto de Matemática e Estatística - UERJ

Rio de Janeiro

2024

DEDICATÓRIA

Dedico esta tese com muito amor e carinho à minha mãe, Margarida.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, que me deu forças, resiliência, paciência e coragem para prosseguir nesta longa jornada. A Ele, toda honra e toda glória!

Aos meus pais Margarida e Carlos, e ao meu irmão Rodrigo, pelo amor, carinho, paciência, compreensão e apoio nos momentos mais difíceis ao longo do mestrado, doutorado e da vida.

As amigas de longa data, Aline, Pattrícia, Raissa e Rayza, pela compreensão nos momentos em que me fiz distante em função dos estudos.

A esta universidade, Universidade do Estado do Rio de Janeiro (UERJ), em especial ao Programa de Pós-graduação em Ciências Computacionais e Modelagem Matemática (CcompMat/UERJ) do Instituto de Matemática e Estatística (IME/UERJ), à Coordenação da pós-graduação e seus funcionários, pelo empenho em atender e resolver as solicitações e dúvidas ao longo do doutorado.

Ao Comitê de Ética, pela autorização e aprovação CAAE: 3592920.2.0000.5282 e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, edital n° 12/2020, bolsa CAPES-EPIDEMIAS, pelo apoio financeiro em parte do doutorado.

À minha orientadora, Profª Dra. Regina Lanzillotti, pelo conteúdo ensinado ao longo do mestrado e do doutorado, pela orientação, pela ajuda na escolha do tema deste estudo, pelo carinho dedicado a mim, e pela paciência e apoio dedicados não só neste trabalho como também em assuntos relacionados à vida.

À Profª Dra. Karla Figueiredo, coordenadora do projeto Modelo Hierárquico para Diagnóstico e Desfecho de Pacientes com COVID-19, Utilizando Comitês Baseados em Modelos de Inteligência Artificial a partir de Imagens e Dados Clínicos/Laboratoriais, financiado pela CAPES, por autorizar o uso dos dados em minha tese.

Aos meus colegas de Doutorado, Alisson, Andres, Ighor, João, Luciana e Walisson, pelas trocas de experiências, companheirismo e apoio ao longo desta jornada.

Enfim, agradeço a todos que, direta ou indiretamente, contribuíram de alguma forma ao longo da trajetória na pós-graduação.

RESUMO

CRUZ, Carla Cristina Passos. *Modelagem Fuzzy Intuicionista no Reconhecimento de Padrões em Registros no Prontuário Eletrônico de Pacientes atendidos no Hospital Universitário durante a Pandemia da Covid-19.* 2024. 149 f. Tese (Doutorado em Ciências Computacionais e Modelagem Matemática) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

A Mineração dos Textos corresponde ao pré-processamento para o uso de algoritmos, pois organiza, vincula e transforma os conteúdos da linguagem escrita em informações fidedignas que antecipam o uso a *posteriori* de modelagem que busca por reconhecimento de padrões. Destaca-se que 80% dos dados gerados são não-estruturados e destes, 80% estão no formato textual que lida com a incerteza e ambiguidade. A Recuperação da Informação e o Processamento da Linguagem Natural contabiliza a frequência de ocorrência das palavras, o que permite a abordagem da Estatística Inferencial e Inteligência Computacional. **Objetivo**: aplicar a modelagem Fuzzy Intuicionista para obter padrões segundo termos linguísticos em registros no prontuário eletrônico de pacientes atendidos em Unidade Hospitalar Pública Universitária durante a pandemia da Covid-19. **Método**: Estudo de caso, optou-se por Agrupamento Intuicionista *Fuzzy* que considera o grau de pertinência, não-pertinência e hesitação para descrever cenários. Para aplicação desta modelagem que foi utilizada para obter padrões segundo termos linguísticos, utilizou-se os registros realizados nos prontuários eletrônicos dos pacientes do Hospital Universitário Pedro Ernesto no período da pandemia do Covid-19. As informações não-estruturadas, escritas de forma incorreta que não atendiam a Classificação Internacional de Doenças, passaram pela etapa de Pré-Processamento, onde foram criadas variáveis para a entrada no Sistema *Fuzzy*. Aplicou-se o método Elbow para indicar o provável número de *clusters* e ratificação pelo modelo *Fuzzy* Intuicionista C-*Means*. **Resultados**: O pré-processamento organizou e ajustou os dados e possibilitou a construção de uma base de dados textual associado à incidência de Covid-19 nos casos atendidos no HUPE, que poderá ser utilizado em outras pesquisas. Foram estabelecidas duas Classes Padrões para permitir a classificação de um paciente acometido de Covid-19 em função da Cosseno dos Conjuntos *Fuzzy* Intuicionista segundo os sintomas: febre, falta de ar e cansaço; e comorbidades: hipertensão, doença cerebrovascular e doença cardiovascular. **Conclusão**: Os resultados confirmam que um paciente pode ser enquadrado em determinada classe padrão mediante a estimativa das pertinências do Sistema Lógico Intuicionista *Fuzzy* levando em consideração a relevância da similaridade.

Palavras-chave: *Knowledge Discovery in Text*. Similaridade Cosseno. *Fuzzy* Intuicionista. Covid-19. Uniformização de Protocolos.

ABSTRACT

CRUZ, Carla Cristina Passos. *Intuitionist Fuzzy Modeling in Pattern Recognition in Records in the Electronic Medical Record of Patients treated at the University Hospital during the Covid-19 Pandemic.* 2024. 149 f. Tese (Doutorado em Ciências Computacionais e Modelagem Matemática) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

Text Mining corresponds to pre-processing for the use of algorithms, as it organizes, links and transforms the contents of written language into reliable information that anticipates the subsequent use of modeling that seeks pattern recognition. It is noteworthy that 80% of the data generated is unstructured and of these, 80% are in textual format that deals with uncertainty and ambiguity. Information Retrieval and Natural Language Processing account for the frequency of occurrence of words, which allows the approach of Inferential Statistics and Computational Intelligence. **Objective:** Applying Fuzzy Intuitionist modeling to obtain patterns according to linguistic terms in records in the electronic medical record of patients treated in a Public University Hospital Unit during the Covid-19 pandemic. **Method:** Case study, Fuzzy Intuitionist Grouping was chosen, which considers the degree of relevance, non-pertinence and hesitation to describe scenarios. Applying this modeling, which was used to obtain patterns according to linguistic terms, we used the records made in the electronic medical records of patients at the Pedro Ernesto University Hospital during the period of the Covid-19 pandemic. The unstructured information, written incorrectly and not complying with the International Classification of Diseases, went through the Pre-Processing stage, where variables were created for entry into the Fuzzy System. The Elbow method was applied to indicate the probable number of clusters and ratification by the Fuzzy Intuitionist C-Means model. **Results:** Pre-processing organized and adjusted the data and enabled the construction of a textual database associated with the incidence of Covid-19 in cases treated at HUPE, which can be used in other research. Two Standard Classes were established to allow the classification of a patient suffering from Covid-19 based on the Cosine of the Intuitionist Fuzzy Sets according to the symptoms: fever, shortness of breath and fatigue; and comorbidities: hypertension, cerebrovascular disease and cardiovascular disease. **Conclusion:** The results confirm that a patient can be classified into a certain standard class by estimating the pertinences of the Fuzzy Intuitionist Logic System, taking into account the relevance of the similarity.

Keywords: Knowledge Discovery in Text. Cosine Similarity. Intuitionistic Fuzzy. Covid-19. Protocol Uniformization.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
CF	Conjuntos *Fuzzy*
CFI	Conjuntos *Fuzzy* Intuicionista
CID	Classificação Internacional de Doenças
DE	Distância Euclidiana
DP	Desvio Padrão
FANNY	*Fuzzy Analysis Clustering*
FCM	*Fuzzy C-Means*
FCMdd	*Fuzzy C-Medoids*
FICM	*Fuzzy* Intuicionista *C-Means*
HUPE	Hospital Universitario Pedro Ernesto
IA	Inteligência Artificial
IC	Inteligência Computacional
IFCM	*Intuitionistic Fuzzy C-Means*
KDD	*Knowledge Discovery in Database*
KDT	*Knowledge Discovery in Text*
LC	Lógica Clássica
LC	Lógica *Crisp*
LF	Lógica *Fuzzy*
LFI	Lógica *Fuzzy* Intuicionista
LGPD	Lei Geral de Proteção de Dados
MT	Mineração de Texto
MTD	Matriz Termo-Documento
NP	Não-Pertinências
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana de Saúde
P	Pertinências
PEP	Prontuário Eletrônico do Paciente
PNO	Plano Nacional de Operacionalização
RI	Recuperação da Informação

RSLP	Removedor de Sufixos em Língua Portuguesa
Sars-CoV-2	*Severe Acute Respiratory Syndrome Coronavirus-2*
SIG	Sistema de Informação Gerencial
SIS	Sistema de Informação em Saúde
TCF	Teoria dos Conjuntos *Fuzzy*
TCFI	Teoria dos Conjuntos *Fuzzy* Intuicionistas
TF	Teoria *Fuzzy*
TFI	Teoria *Fuzzy* Intuicionista
TM	*Text Mining*
TP	*Term Presence*
UERJ	Universidade do Estado do Rio de Janeiro
UTI	Unidade de Terapia Intensiva
VOC	Variante de Preocupação
VOI	Variante de Interesse
VP	Valor de Pertinência
WHO	World Health Organization

LISTA DE SÍMBOLOS

μ	Conjunto pertinência
ν	Conjunto não-pertinência
$\neq$	Diferente
u	Elemento do conjunto de pertinência
v	Elemento do conjunto de não-pertinência
$\subset$	Está contido
J	Função objetivo
π	Grau de hesitação (letra grega pi)
$=$	Igual
$\cap$	Interseção
$>$	Maior
$\leq$	Menor ou igual
$\nexists$	Não existe
$\notin$	Não pertence
$\wedge$	Operador lógico “E”
$\vee$	Operador lógico “OU”
z	Padronização
$\forall$	Para todo
$\in$	Pertence
$\leftrightarrow$	Se e somente se
$\rightarrow$	Se..., então
Σ	Somatório (letra grega sigma em maiúsculo)
$\mid$	Tal que
$\cup$	União
ϕ	Valor do ângulo

SUMÁRIO

INTRODUÇÃO

A tecnologia está presente em praticamente todas as nossas atividades, incluindo-se pagamento de contas, videochamadas, dentre outros. O aumento dessa interação entre o homem e a máquina tem gerado um grande fluxo de dados, o que torna o tratamento das informações mais necessárias. Atualmente, decisões são tomadas por sistemas autônomos dotados de Inteligência Artificial (IA), integrada em uma variedade crescente de setores, transformando profundamente o cenário econômico, social e tecnológico (Da Silva; Domingues, 2014).

A mineração dos dados se faz necessária e, para tal utiliza-se o processo denominado *Knowledge Discovery in Database* (KDD), processo de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados (Fayyad; Piatetsky-Shapiro; Smyth, 1996).

Está inserida na Mineração de Dados, capaz de fornecer informação fidedigna a partir da exploração de um grande volume de informações, propiciando relações úteis e compreensíveis para o pesquisador segundo a identificação de padrões (Hand; Mannila; Smyth, 2001).

Os dados gerados podem se apresentar como: estruturados, que são organizados e representados por uma estrutura rígida previamente projetada de forma a armazená-los (Li *et al.*, 2008); não-estruturados, que não possuem uma estrutura definida, pois seus dados são apresentados como vídeos, imagens, áudios e textos (Li *et al.*, 2008); e semi-estruturados, etapa intermediária entre as duas anteriores (Buneman, 1997).

Os dados não-estruturados, por envolverem dados do tipo texto, utiliza-se de técnicas multidisciplinares como a recuperação da informação, aprendizado de máquina, a estatística, a linguística computacional e a mineração de dados (Santos *et al.*, 2014).

O processo da mineração de textos consiste em um conjunto de procedimentos para extrair e recuperar dados textuais considerados relevantes lidando com a imprecisão, incerteza e abreviação de termos, além do sentido e da semântica das palavras (Aranha; Passos, 2006). Tan (1999) e Ferreira (2020) indicam que entre 80% a 90% das informações vigentes no mundo são não-estruturadas, e 80% delas se encontram no formato textual (Chen, 2001).

Baseia-se em passos do *Knowledge Discovery in Text* (KDT) como Coleta, Pré-processamento, Processamento e Pós-Processamento (Tan, 1999), além de métodos como Recuperação da Informação e Processamento da Linguagem Natural (PLN) que envolvem as

etapas consecutivas de análise: morfológica, sintática, semântica e pragmática (Bulegon; Moro, 2010).

As informações clínicas estão sob a forma não-estruturada e encontram-se no Prontuário Eletrônico do Paciente (PEP) em instituição hospitalar. Esta é uma fonte considerada importante no Sistema de Informação em Saúde (SIS), mas apresentam uma variedade terminológica que por vezes não corresponde à orientação constante na Classificação Internacional de Doenças (CID).

Wang *et al*. (2012) indicam que tal fato pode atrasar na Recuperação de Informações, uma vez que as anotações podem ser realizadas no PEP por equipe multidisciplinar, em linguagem natural, inclusive com uso de jargões da Medicina (Rector, 1999; Baud *et al*., 2007; Souza; De Almeida, 2019).

De acordo com Faleiros Júnior, Nogaroli e Cavet (2020) as anotações diárias referentes aos fatos relacionados tanto com a doença quanto aos pacientes redigidas em caderno, porém sem a individualização dos pacientes, tendo evolução para o dossiê que abrange a conduta médica. Entre esses dados, destacam-se os da área da saúde em que as informações clínicas são registradas no prontuário eletrônico do paciente (PEP) de instituições hospitalares, consideradas fontes importantes no Sistema de Informação em Saúde (SIS) para subsidiar pesquisas.

Estas informações podem se apresentar de forma não-estruturada, e por isso, apresentam uma variedade terminológica que por vezes não corresponde a orientação constante na Classificação Internacional de Doenças (CID) (Wang *et al*., 2012). Tal fato indica atraso na recuperação de informações, uma vez que as anotações podem ser realizadas no PEP por equipe multidisciplinar, em linguagem natural, inclusive com uso de jargões da Medicina (Rector, 1999; Baud *et al*., 2007; Souza; De Almeida, 2019).

De França (2017) refere-se ao prontuário não apenas quanto ao registro da *anamnese*[1], mas considera todo acervo documental padronizado, organizado e conciso, referente ao registro dos cuidados médicos prestados. De acordo com o autor, não se pode afirmar que o prontuário é mera peça burocrática destinada ao controle das despesas hospitalares e dos procedimentos, presta-se a múltiplos objetivos, como a análise da evolução de doenças.

De acordo com Dagliati *et al.* (2021), um dos aspectos mais difíceis é a reutilização e compartilhamento de dados coletados via Prontuário Eletrônico do Paciente (PEP). Os autores

[1] Na medicina, a *anamnese* é o diálogo estabelecido entre profissional de saúde e paciente com o objetivo de ajudá-lo a lembrar de situações e fatos que podem estar relacionados a sua doença (Portal Telemedicina, 2022).

ainda ressaltam que os dados de PEP não são apenas essenciais para apoiar as atividades cotidianas, mas também podem subsidiar as pesquisas e apoiar decisões referentes aos medicamentos e as estratégias terapêuticas.

Entre as abordagens utilizadas em Mineração de Texto destacam-se a recuperação da informação (RI), o aprendizado de máquina (AM), a inteligência computacional (IC), o agrupamento ou *cluster*, dentre outros. Este último, é uma técnica estatística multivariada que objetiva dividir uma amostra (ou população) em grupos, de acordo com as características medidas (Mingoti, 2013), que avalia a dimensionalidade da estrutura dos dados, identifica *outliers* e avalia hipóteses de associação (Johnson; Whichern, 2007).

Os agrupamentos são designados em *hard*, quando os dados se enquadram apenas em um agrupamento; ou *soft*, quando se enquadra em mais de um. Este último conduz a metodologia *fuzzy*, uma abordagem que se baseia em uma lógica que permite expressar a ideia de verdade parcial.

Desenvolvida pelo professor Lofti Zadeh, da Universidade da Califórnia, em 1965, a partir da combinação de conceitos da lógica clássica e dos conjuntos de Lukasiewicz. Nesta metodologia, cada elemento é associado a um grau de pertinência e não-pertinência no agrupamento e ambos são complementares (Faceli; Carvalho; Souto, 2005). Entre os métodos de agrupamentos *fuzzy* encontra-se o *Fuzzy* Intuicionista C-*Means* (FICM) que agrupa os documentos da entrada do sistema lógico com base no regime de adesão.

Além do grau de pertinência e não-pertinência, nesta modelagem intuicionista considera-se a hesitação, obtida pela diferença entre um e a soma da pertinência e não pertinência. Este método serve para propiciar o reconhecimento de padrões com o propósito de procurar, detectar estruturas associadas às regularidades ou propriedades, descrever cenários nas áreas do conhecimento tecnológico, biomédico e social (Mamdani; Assilian, 1975).

Outra questão importante é a classificação dos pacientes em grupos, a partir de determinadas características e níveis de complexidade. Para Fantinelli (2022), a classificação dos pacientes em níveis de complexidade pode ser considerada um recurso valioso para a gestão da assistência, pois é possível estimar as necessidades de cuidados para cada paciente.

Muitos hospitais utilizam a escala de Fugulin (2005), que classifica os pacientes de acordo com o grau de dependência da enfermagem, sendo útil para o aperfeiçoamento dos parâmetros oficiais relacionados à temática do dimensionamento de pessoal de enfermagem em instituições hospitalares (Fantinelli, 2022).

Entre os métodos e técnicas de classificação e agrupamentos existentes, destaca-se a modelagem *fuzzy* com similaridade cosseno que classifica indivíduos e/ou objetos em determinada "Classe Padrão", de acordo com características específicas apresentadas (Ye, 2011; Intarapaiboon, 2016).

Nesse contexto, importa ressaltar que em 7 de janeiro de 2020, as autoridades chinesas confirmaram que haviam identificado um novo tipo de coronavírus denominado de *Severe Acute Respiratory Syndrome Coronavirus-2*, conhecido pela abreviação *Sars-Cov*-2 (Meirelles *et al.*, 2020; Pereira *et al.*, 2022), o sétimo coronavírus humano identificado. No Brasil, o primeiro caso da Covid-19 foi registrado em 26 de fevereiro de 2020 com a primeira morte anunciada em 17 de março de 2020.

Até 1° de setembro de 2020, a pandemia causou 122.596 óbitos, considerando apenas os notificados ao Ministério da Saúde do Brasil (Brasil, 2020). Em 28 de agosto de 2020, o Brasil foi o segundo país do mundo em número de mortes e casos da Covid-19 (WHO, 2020). A pandemia da Covid-19 mostrou que os dados e suas análises são componentes importantes para tomada de decisão.

Justificativa

O advento da pandemia da Covid-19 reafirmou a importância dos dados e suas análises como componentes cruciais para tomada de decisão. No contexto de saúde, o uso de técnicas e ferramentas de Processamento da Linguagem Natural (PLN) aumentaram de forma considerável, sobretudo em resposta a pandemia da Covid-19. Recentemente, pesquisadores da Fiocruz desenvolveram Pesquisadores desenvolvem técnica de mineração de texto para análise da Covid longa (Fiocruz, 2024).

Logo, a presente tese propõe estudo de caso a partir de informações de prontuários eletrônicos dos pacientes internados com Covid-19 nas enfermarias e UTI do Hospital Universitário Pedro Ernesto (HUPE), pertencente a Universidade do Estado do Rio de Janeiro (UERJ), de dezembro 2019 até meados de 2021. Fernandes *et. al*. (2019) concluíram que a ilegibilidade e preenchimento errôneo demonstram o preenchimento inadequado e, portanto, uma possível incapacidade de estabelecimento de vínculo efetivo.

A partir das fragilidades identificadas percebe-se a necessidade de estratégias de educação permanente para melhoraria da qualidade de preenchimento dos prontuários médicos. Durante o levantamento prévio dos prontuários referentes aos pacientes do Hospital Universitário Pedro Ernesto (HUPE) observou-se ausência de uniformidade nos registros de profissionais de distintas formações acadêmicas, fato que prejudica a qualidade do Sistema de Informação Gerencial (SIG).

No levantamento foram detectadas as seguintes inconsistências: duplicidade do tipo do registro, grafia incorreta, termos não correspondendo ao CID-10, uso de linguagem popular. Assim, o banco foi elaborado com dados não-estruturados, fato que dificultou a obtenção da frequência dos termos do prontuário.

As informações posteriormente adquiridas, propor uma metodologia *Fuzzy* Intuicionista para agrupamento e classificação dos pacientes em grupos, denominados Classes Padrão, a partir dos sintomas e comorbidades que mais apareceram entre os pacientes.

Garcia-Vidal *et al*. (2024), aplicaram o agrupamento *K-Means* para agrupar pacientes dos Hospitais de Universitari Mutua de Terrassa (Terrassa) e do Universitari La Fe (Valência), de acordo com os valores do limiar do ciclo da reação em cadeia do rRT-PCR e contagens de linfócitos no momento do diagnóstico e duração dos sintomas pré-teste.

Allem *et al*. (2022), analisaram a evolução da pandemia da Covid-19 no Rio Grande do Sul com técnicas de agrupamento espectral, em gráficos ponderados definidos no conjunto de 167 municípios do estado com população de 10.000 ou mais, com dados fornecidos por diversas fontes.

Objetivos

Diante do exposto, foi elaborado o objetivo desta tese: aplicar o Método *Fuzzy* Intuicionista *C-Means* segundo Similaridade Cosseno, para reconhecer Classes Padrões de risco em sintomas e comorbidades, a partir dos termos linguísticos em registros no prontuário eletrônico de pacientes internados nas enfermarias e na UTI no período de dezembro de 2019 até setembro de 2021 de acordo com sintomas e comorbidades da Covid-19 mais recorrentes.

Os objetivos específicos são:

- Elaborar o Pré-processamento nos prontuários do HUPE, para a construção da base estruturada de informações;
- Construir a Tabela de Contingência segundo o cruzamento das variáveis correspondentes a idade e número de dias de internação, entrada do Sistema Intuicionista *Fuzzy*;
- Aplicar o método Elbow para determinação de um número de *clusters*;
- Aplicar o teste de estratificação para verificar a homogeneidade intra *cluster;*
- Aplicar o Método *Fuzzy* Intuicionista *C-Means* para categorizar agrupamentos de risco;
- Aplicar a modelagem *Fuzzy* Intuicionista segundo similaridade cosseno para classificação em "Classes Padrões", de acordo com sintomas e comorbidades previamente selecionados.

1 **FUNDAMENTAÇÃO TEÓRICA**

1.1 **Novo coronavírus**

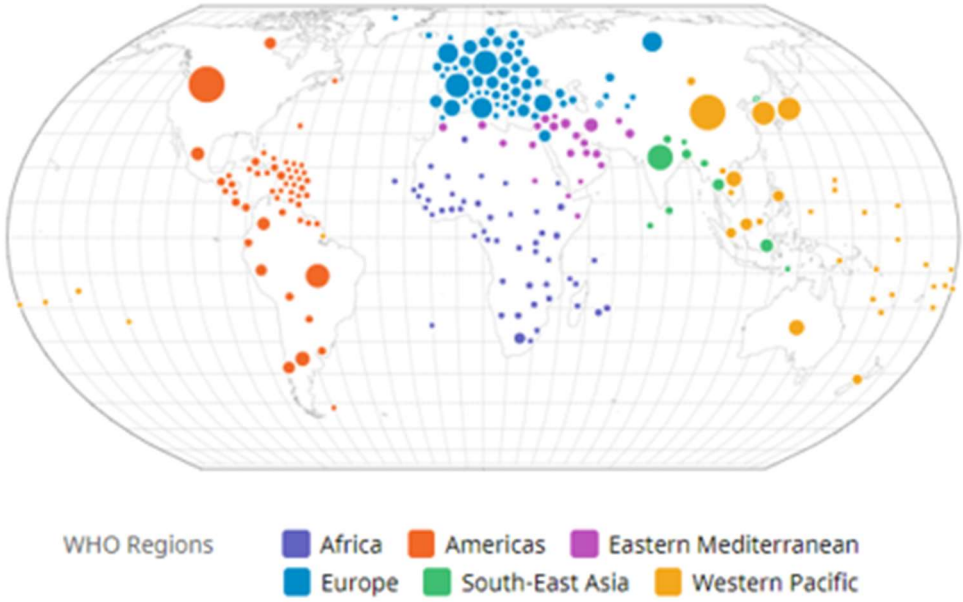
Em 31 de dezembro de 2019, mais precisamente, na cidade Wuhan pertencente à província de Hubei, na China, foram emitidos alertas para a Organização Mundial da Saúde (OMS), do inglês *World Health Organization* (WHO) pois foram registradas tromboses pulmonares referentes a pneumonia silenciosa referente a uma nova cepa (tipo) de coronavírus (OPAS, 2020a). Apesar da família Coronavírus já ser conhecida desde 1937, foi a primeira vez que ocorreu a aparição deste subtipo específico (Cossa *et al.*, 2021).

Em 7 de janeiro de 2020, as autoridades chinesas confirmaram que haviam identificado um novo tipo de coronavírus, denominado de *"Severe Acute Respiratory Syndrome Coronavirus-2"* (Meirelles *et al.*, 2020; Pereira *et al.*, 2022). No dia 30 de janeiro de 2020, a OMS declara em Genebra (Suíça) que o surto do novo coronavírus constitui uma Emergência de Saúde Pública de Importância Internacional (ESPII) (OPAS, 2020b) e recomenda que o nome provisório da doença deveria ser "doença respiratória aguda 2019-nCoV.

No Brasil, a detecção de seu primeiro caso ocorreu no dia 26 de fevereiro de 2020, em São Paulo, em um passageiro que voltava de viagem da Itália. No entanto, a cidade do Rio de Janeiro registrou sua primeira detecção 11 dias após, no dia 6 de março. Rapidamente, o Covid-19 começa a se espalhar pelo mundo, fazendo com que a Organização Mundial da Saúde (OMS), declarasse a Pandemia na data de 11 de março de 2020 (Mano *et al.*, 2023; Meirelles *et al*., 2020).

O termo "pandemia" refere-se à distribuição geográfica de uma doença e não à sua gravidade. A designação reconhecia que, no momento, existiam surtos de Covid-19 em vários países e regiões do mundo (OPAS, 2020b). A Figura 1, mostra a disposição dos casos confirmados no mundo, por região, até 30 de dezembro de 2023 (WHO, 2023), a partir das informações do painel interativo da OMS, visualizando-se os casos em todo o mundo.

Figura 1 - Casos de Covid-19 acumulados até 30 de dezembro de 2023



Fonte: WHO, 2023.

Diante ao cenário de incertezas que se instalou, muitos foram os desafios a serem vencidos pelos países, principalmente para garantir um rápido controle pelas políticas públicas de saúde dos países. Por exemplo, o uso de dados coletados anteriormente sobre a sintomatologia e progressão da doença em surtos anteriores (Cossa *et al.*, 2021).

O Coronavírus é a segunda principal causa de resfriado comum, após rinovírus e, até as últimas décadas, raramente causavam doenças mais graves em humanos do que o resfriado comum. A OPAS (2020c) indica os mais conhecidos:

a) HCoV-229E;

b) HCoV-OC43;

c) HCoV-NL63;

d) HCoV-HKU1;

e) *Sars-CoV* (causa Síndrome Respiratória Aguda Grave);

f) MERS-*CoV* (causa Síndrome Respiratória do Oriente Médio);

g) *Sars-CoV-2* (responsável por causar a doença Covid-19).

De acordo com o Instituto Butantan, os sintomas variam a cada nova variante e onda encontrada (Butantan, 2021) e, conforme a OMS, são considerados sintomas mais comuns da Covid-19 (OPAS, 2020c) e Ministério da Saúde (Brasil, 2020):

- Febre;

- Cansaço;
- Tosse seca.

Outros sintomas que podem afetar alguns pacientes são (OPAS, 2020c):

- Perda de paladar ou olfato;
- Congestão nasal;
- Conjuntivite;
- Dor de garganta;
- Dor de cabeça;
- Dores nos músculos ou juntas;
- Diferentes tipos de erupção cutânea;
- Náusea ou vômito;
- Diarreia;
- Calafrios ou tonturas.

Quanto às comorbidades para o Covid-19, a lista completa com suas especificidades e subdivisões encontra-se no Plano Nacional de Operacionalização (PNO) do Ministério da Saúde (Brasil, 2022), cujas principais estão classificadas em:

- Diabetes;
- Doença cardiovascular;
- Doença cerebrovascular;
- Doença pulmonar;
- Doença renal;
- Hipertensão;
- Imunocomprometido/imunossuprimido;
- Obesidade;
- Dentre outros.

Em 2021, para fins de melhor compreensão e acompanhamento por parte da população, a OMS atribuiu nomenclaturas simples, fáceis de pronunciar e lembrar para as principais variantes do *Sars-CoV*-2, o vírus causador da Covid-19, utilizando letras do alfabeto grego (Bio-Manguinhos, 2021). Atribui-se nomenclaturas às variantes como Variantes de Preocupação (VOC) ou Variantes de Interesse (VOI).

Toda vez que uma nova variante surge, a OMS atualiza um quadro do seu site com informações preliminares sobre as mesmas. A VOC induz uma avaliação comparativa, onde estão associadas alterações com certo grau de significância para a saúde pública global (OPAS, 2021):

- Aumento da transmissibilidade ou alteração prejudicial na epidemiologia da Covid-19; ou
- Aumento da virulência ou mudança na apresentação clínica da doença; ou
- Diminuição da eficácia das medidas sociais e de saúde pública ou diagnósticos, vacinas e terapias disponíveis.

A VOI é considerada como uma variante em comparação com a variante original se seu genoma contiver mutações que alteram o fenótipo do vírus (OPAS, 2021) e se:

- Tiver sido identificada como causadora de transmissão comunitária, de múltiplos casos ou de *clusters* (agrupamentos de casos) de Covid-19 ou tiver sido detectada em vários países; ou
- Ser de outra forma avaliada como uma VOI pela OMS em consulta com o grupo de trabalho de evolução do vírus *Sars-CoV*-2.

Todas as informações relacionadas nesta secção são importantes o processo de classificação necessária para o uso do aprendizado de máquina, com o objetivo de reconhecer padrões segundo sintomas e comorbidades em prontuários do HUPE no período de março a setembro de 2019.

1.2 **Mineração de texto**

A mineração de texto (do inglês, *text mining*) é uma área derivada da mineração de dados (do inglês, *data mining*) que surgiu a partir da necessidade do processamento dos textos, de forma que as palavras do texto sejam transformadas em dados numéricos, a partir do processamento do computacional. Desta forma, os dados passam a ser estruturados, o que permite aplicações de diferentes modelagens para análises.

Conforme citado no Capítulo de Introdução, dentre as áreas que pertencem à mineração de texto e que foram usadas nesta tese, encontram-se a inteligência computacional, estatística, Processamento da Linguagem natural (PLN), Lógica *Fuzzy* (LF), Reconhecimento de Padrão e Aprendizado de Máquina (AM).

O AM, conhecido como *machine learning* (ML), é uma abordagem da aprendizagem na qual uma máquina, por meio de um processo indutivo, constrói um classificador, aprendendo a partir de um conjunto de exemplos pré-classificados (Feldman; Sanger, 2007).

Esta técnica tem sido bastante usada em tarefas de pré-processamento de textos, dentre as quais se destacam a classificação automática de textos (Chakrabarti, 2003) e deve ser capaz de realizar três tarefas (Russell; Norvig, 2003): armazenar conhecimento; aplicar o conhecimento armazenado para resolver problemas; adquirir novo conhecimento. As técnicas de AM são classificadas como:

- **Supervisionada:** existe orientação prévia para avaliar a resposta obtida pelo modelo de classificação e há comparação com as classes pré-definidas, pois o objetivo é induzir conceitos a partir de exemplos e estimar taxas de acerto e erro obtidas por um classificador (Lorena; Carvalho, 2003);

- **Por reforço:** refere a problemas que o agente ou sistema deve aprender a selecionar, ações disponíveis que alteram o estado do ambiente, utilizando reforço para definir a qualidade da sequência de ações (Mitchell, 1997);

- **Não-supervisionada**: há incerteza sobre a expectativa da saída a ser gerada, e envolve o uso de diversas técnicas para a identificação de agrupamento, havendo uma maior autonomia do algoritmo para extrair características dos dados fornecidos ao agrupar classes (Lorena; Carvalho, 2003) e reconhecer padrões (Correia; Ferreira, 2007).

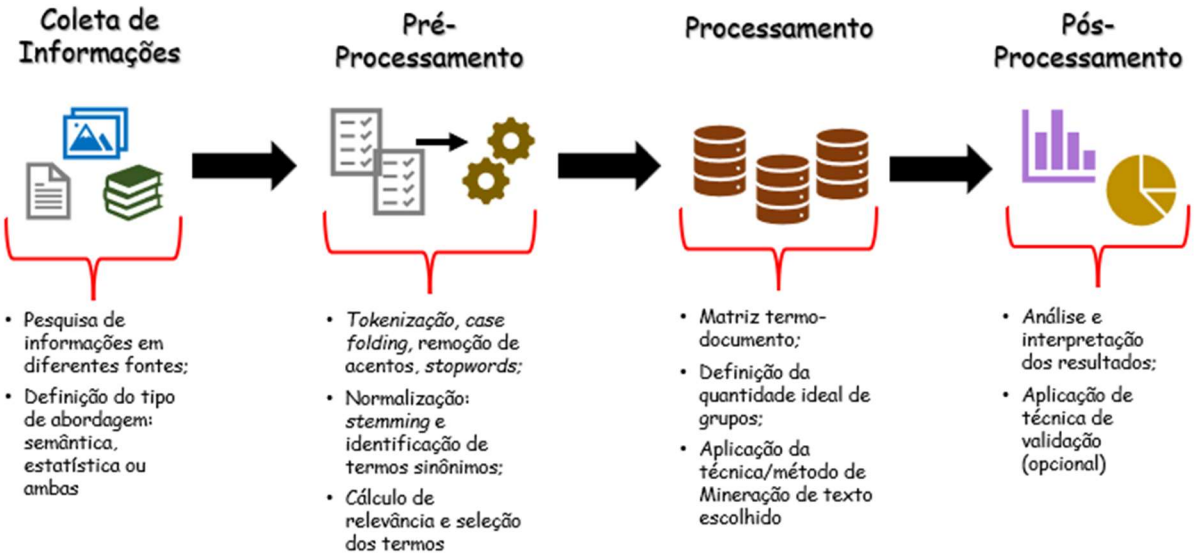
A principal diferença entre o processo de mineração de dados tradicional e o de textos é que, enquanto a primeira trabalha exclusivamente com dados estruturados, a segunda abordagem lida com dados não-estruturados. Deste modo, é importante saber que o tipo de dados a ser tratado, o qual pode se apresentar de formas conforme explicitados na Figura 2, que apresenta um diagrama descrevendo os tipos dados.

Figura 2 - Classificação dos tipos de dados

Fonte: A autora, 2024.

A extração de informações importantes existentes nos dados não-estruturados compreende as técnicas que foram estudadas e desenvolvidas com o objetivo de auxiliarem no processo (Barion; Lago, 2015). Aos dados textuais existe o *Knowledge Discovery in Text* (KDT) que se baseia nas etapas da Mineração de Dados. Dixon (1997), Tan (1999) e Morais e Ambrósio (2007) propuseram etapas do desenvolvimento, ilustradas na Figura 3 e explicitadas nas Subseções 1.2.1, 1.2.2, 1.2.3 e 1.2.4.

Figura 3 - Etapas do *Knowledge Discovery in Text*



Fonte: A autora, 2024.

1.2.1 Coleta das informações

Conhecida na literatura como *corpus* ou *corpora*, é a etapa de busca, coleta e armazenamento dos dados a ser analisados. Nesta etapa define-se o tipo de abordagem a ser aplicada nos textos: semântica ou estatística (Ebecken; Lopes; Costa, 2003).

A abordagem estatística trata os termos pela frequência de aparição, mas sem a contextualização, que leva em conta a inserção do mesmo no parágrafo, a antecedência e a forma de inserção quanto ao posicionamento (antecedência e subsequência), e a funcionalidade da relação dos termos (Morais; Ambrósio, 2007; Limiro; Silva; Cordeiro, 2022).

A Abordagem Semântica tem como foco a funcionalidade do texto em relação ao significado das palavras, isto é, explora aspectos mais ligados à representatividade de um termo em relação a outros, tendo sua base no processamento de linguagem natural (PLN) (Limiro; Silva; Cordeiro, 2022). Nesta tese, ambas as abordagens são utilizadas na etapa de pré-processamento, Subseção 1.2.2 e Seção 2.2, respectivamente.

Na área da saúde, Souza e Felipe (2021) utilizam reconhecimento de entidades nomeadas para extração de termos técnicos de *anamneses*[2] de prontuários eletrônicos do domínio da ginecologia. Já Ridgway *et al*. (2021) usam o PLN em anotações clinicas para detectar doenças mentais e uso de substâncias entre pessoas vivendo com HIV. Sorin *et al*. (2020) utilizam o PLN e *Deep learning* (DL) em radiologia.

O processamento de um texto escrito e desenvolvido por humanos, é necessário seguir uma série de etapas iterativas envolvendo o PLN. Estas etapas incluem o pré-processamento dos dados, a representação de palavras como vetores numéricos, o treinamento de um modelo de aprendizado de máquina (AM). Segundo Bulegon e Moro (2010), descrevem o PLN em quatro etapas, ordenado e exposto na Figura 4:

Figura 4 – Etapas do Processamento da Linguagem Natural

Fonte: A autora, 2024.

[2] Na medicina, a *anamnese* é o diálogo estabelecido entre profissional de saúde e paciente com o objetivo de ajudá-lo a lembrar de situações e fatos que podem estar relacionados a sua doença (Portal Telemedicina, 2022).

- **Análise morfológica:** responsável pela identificação da estrutura das palavras, por definir artigos, substantivos, verbo e adjetivos armazenando-os em um dicionário (*Thesaurus*);
- **Análise sintática:** segue a construção do dicionário e procura mostrar relacionamento entre as características morfológicas e, em segundo momento, as características gramaticais, sujeito, predicado, complementos nominais e verbais, adjuntos e apostos;
- **Análise semântica:** ocorre o encontro de termos ambíguos, sufixos e afixos, ou seja, questões dos significados associados aos morfemas componentes de uma palavra, demonstrando o sentido real da frase ou palavra;
- **Análise pragmática:** faz a junção de todo o mecanismo e mostra visualmente o resultado. Neste caso, existem alguns algoritmos que apresentam os passos em forma de escada até a conclusão, identifica os termos de uma sentença;
- **Análise do discurso:** é o conhecimento de como as sentenças imediatamente precedentes afetam a interpretação da próxima sentença (Morais; Ambrósio, 2007).

Cabe ressaltar que o uso dos métodos de análise de textos e sua utilização, dependerá da pesquisa realizada (Morais; Ambrósio, 2007).

1.2.2 Pré-processamento

Corresponde ao conjunto de ações tomadas sobre os documentos textuais a fim de torná-los manipuláveis para a extração do conhecimento. Considerada como a parte mais importante do processo, pois consiste na preparação dos textos, na definição do tipo de abordagem dos dados, preparação, indexação, normalização, cálculo/relevância e seleção dos termos;

O principal objetivo do pré-processamento é organizar e ajustar os dados de forma que as diversas técnicas possam ser aplicadas e até mesmo combinadas. É considerada a etapa mais complexa do processo, pois organiza e ajusta os dados para possibilitar a construção de uma base de dados textual associado à incidência de Covid -19 nos casos atendidos no HUPE, que poderá ser utilizado em outras pesquisas.

O primeiro passo é chamado de *tokenização* e, segundo Xavier, Silva e Gomes (2015), consiste na separação de um texto em suas unidades mínimas, deve-se separar as palavras e demais características da escrita formal. Estas unidades mínimas são conhecidas como *tokens*.

Existem abordagens que agrupam dois *tokens* para formar um *token* com significado agregado, por exemplo, o nome composto de uma entidade tem seu valor no seu conjunto de palavras, e não em cada palavra separadamente. Cada *token* é separado por aspas.

Pakhomov, Pedersen e Chute (2005) relatam que 30% de *tokens* em texto clínico são considerados ruídos, isto é, escrito de forma não padronizada, pois geralmente os registros dos pacientes são escritos por médicos e enfermeiros que, oram termos específicos de domínio, ora usam termos comuns. A Tabela 1 traz exemplos de frases antes e depois do processo de *tokenização:*

Tabela 1 - Exemplos de frases antes e após o processo de *tokenização*

Tipos	Exemplo 1	Exemplo 2
Frase Original	Ontem fomos ao cinema.	Para maiores informações, acessar: https://www.gov.br
***Tokens* preliminares (observações de espaços em branco e delimitadores)**	“Ontem” “fomos” “ao” “cinema” “.”	“Para” “maiores” “informações” “,” “acessar” “:” “https” “:” “/” “/” “www” “.” “gov” “.” “br”
Abreviações e palavras combinadas	“Ontem” “fomos” “ao” “cinema” “.”	“Para” “maiores” “informações” “,” “acessar” “:” “https:// www.gov.br”
***Tokens* multivocabulares**	“Ontem” “fomos” “ao” “cinema” “.”	“Para” “maiores” “informações” “,” “acessar” “:” “https:// www.gov.br”

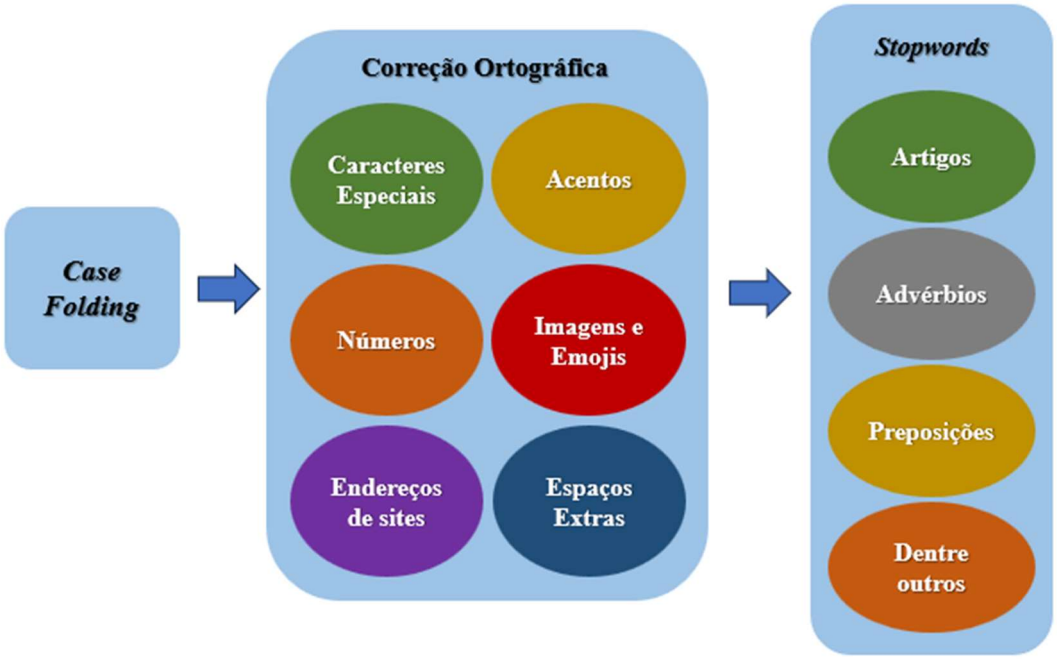
Fonte: A autora, 2024.

O próximo passo é a seleção dos textos, onde a ideia principal é a redução da quantidade a ser processados por intermédio da identificação das similaridades ou significado. Nesta etapa encontram-se os processos de Indexação e Normalização.

A Indexação é a fase responsável pela criação de estruturas de dados, seguindo índices que permitem uma busca sem que seja necessária a análise de toda a base de dados (Manning;

Raghavan; Schütze, 2009). Contempla os processos de *Case Folding*, Correções Ortográficas e *Stopwords*, conforme ilustra a Figura 5:

Figura 5 – Processo de *Case Folding*, Correções Ortográficas e *Stopwords*



Fonte: A autora, 2024.

- ***Case folding*:** trata-se da conversão de todos os termos do documento de maiúsculos para minúsculos ou vice-versa, com o objetivo de padronizar e proporcionar maior agilidade na análise textual (Caputo; Bastos; Ebecken, 2006). Preferencialmente é feita a conversão para o minúsculo. Segaran (2007) indica que o procedimento reduz de forma significativa à quantidade de atributos, entretanto deve-se tomar cuidado para não eliminar informações que possam ser relevantes para o estudo. O autor cita como exemplo a filtragem de mensagens de *spam* que comumente são escritas em letras maiúsculas. Nesse caso, o uso do *case folding* pode levar ao descarte de informações importantes de forma errônea, caracterizadas como um *spam*;

- **Correções ortográficas:** identifica as palavras do documento e exclui símbolos e caracteres do arquivo. Além disso, encontra termos que aparecem no texto de forma composta, pois verifica expressões frequentemente integradas aos documentos correspondentes a uma lista para validação ou se faz uso de um dicionário de expressões. Nesta fase, ocorrem transformações como remoção de números, espaços extras, endereços de *sites*, figuras e *emojis*;

- ***Stopwords*:** são palavras que não agregam informação e não devem ser consideradas na contagem dos termos mais frequentes de um texto por não possuírem relevância significativa. A técnica utilizada para a eliminação das *stopwords* consiste, primeiramente, na criação de uma lista chamada *stoplist,* formada por artigos, preposições, conjunções, pronomes e verbos de ligação. Pode não haver uma lista completa com as *stopwords*. Neste caso, é necessária a complementação por outras fontes, pois um mesmo idioma possui variação de termos. A lista com as *stopwords* usadas nesta tese se encontra no Anexo B. Podem aparecer outras palavras que tenham de serem retiradas, consideradas sem importância. Nesta fase é criada uma lista específica com esses termos, designada de *stoplist* específica.

A Tabela 2 traz exemplos de frases após os processos de *case folding*, correções ortográficas e remoção das *stopwords*.

Tabela 2 - Exemplos de frases antes e após os processos de *Case Folding*, Correções Ortográficas e remoção das *Stopwords*

Antes dos processos	Após os processos
“As” “garotas” “de” “vestidos” “coloridos”	“garotas” “vestidos” “coloridos”
“Emprestei” “um” “livro” “para” “minha” “prima” “,” “que” “adorou”	“emprestei” “livro” “prima” “adorou”
“Você” “gosta” “de” “praticar” “esportes” “?”	“gosta” “praticar” “esportes”
“A” “inscrição” “do” “ENEM” “pode” “ser” “feita” “no” “site” ”.”	“inscricao” “enem” “site”

Fonte: A autora, 2024.

Em seguida, inclui-se a Normalização, que reduz a quantidade de termos diferentes através de um agrupamento de palavras que compartilham de um mesmo padrão. Esta técnica introduz melhoras significativas nos sistemas de mineração de textos, porém pode variar de acordo com o escopo, tamanho da massa textual e a proposta para a saída do sistema (Carrilho Junior, 2007).

Dentre as possíveis técnicas para a realização da normalização das palavras no texto, a lematização, *stemming* (Lovins, 1968) e rotulação (identificação) de termos. O *Stemming* realiza a conversão das palavras presentes no texto para sua forma de base.

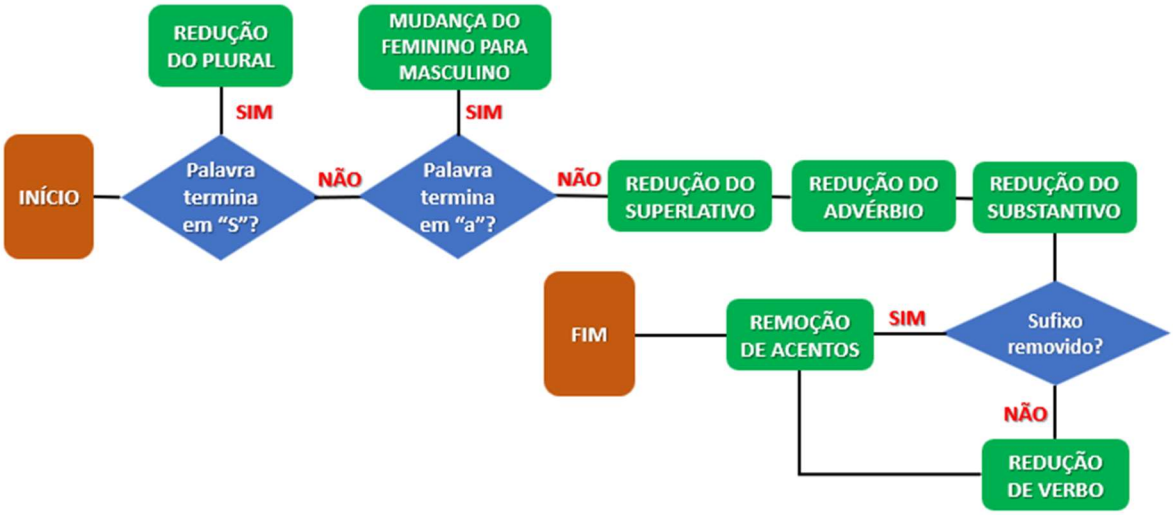
Por exemplo, os verbos em diferentes tempos verbais podem ser transformados para a palavra de origem. Existem diversos métodos para a realização deste procedimento. Destacam-se *Lovins* (Lovins, 1968); *Porter* (Porter, 1980), o mais comum; e *Stemmer S* (Harman, 1991).

No entanto, este processo pode apresentar erros durante a execução do algoritmo, pois depende dos idiomas para os quais foram escritos. Isto ocorre poque se baseiam nas regras linguísticas da formação de palavras para, então, ocorrer a detecção e remoção (Honrado *et al.*, 2000).

A língua portuguesa apresenta variações que causam alterações no radical das palavras, por isso, é necessário utilizar processo de *stemming* específico. Entre os algoritmos propostos para o idioma português, destaca-se o Removedor de Sufixos da Língua Portuguesa (RSLP), publicado por Orengo e Huyck (2001).

Esta é uma versão do algoritmo de 1980, que consiste na identificação e substituição das diversas inflexões e derivações de uma mesma palavra por um mesmo radical. A ideia era agregar importância de um termo pela identificação de suas possíveis variações. A Figura 6 traz a estrutura do algoritmo.

Figura 6 - Processo do Removedor de Sufixos em Língua Portuguesa



Fonte: Adaptado de Morais e Ambrósio, 2007.

A rotulação de termos é o processo de aplicação de um *Parser*, analisador léxico, que consiste na conversão de uma cadeia de caracteres de entrada em uma cadeia de palavras e/ou *tokens* (Fox, 1992). É possível utilizar um dicionário para validação das sequências desses

caracteres e correção de erros ortográficos, denominado *Dictionary Lookup* (Salton; Macgill, 1983).

De acordo com Kochen (1974), utiliza-se um *Thesaurus,* ferramenta atribui a termos distintos, mas relacionados, uma palavra que os represente. Um dos problemas tratados, corresponde aos termos sinônimos. Como exemplo, tem-se as palavras “carro”, “veículo” e “automóvel” que, por serem palavras sinônimas, podem ser associadas à palavra “carro”.

Quanto à rotulação de termos compostos, denominada *Word-Phrase Formation*, o uso de *Thesaurus* é aceitável, mas desde que o termo composto expresse um conceito único pela associação dos termos considerados pois são palavras que, reunidas, apresentam um significado diferente do que separadamente (Lopes, 2004).

Na área da Saúde, os Descritores em Ciência da Saúde (DeCS) foram desenvolvidos pela *Medical Subject Headings* (MeSH) da *U. S. National Library of Medicine* para a indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos e outros tipos de materiais.

Também para ser usado na pesquisa e recuperação de assuntos da literatura científica. Contêm dados referentes aos termos médicos e das áreas de saúde pública, homeopatia, ciência e saúde e vigilância sanitária (Brasil, 2013).

Embora o uso do dicionário tenha ajudado com problemas de palavras sinônimas, o método ainda possui dificuldades com relação às palavras homônimas, pois possuem sentido diferente, porém com mesma pronúncia ou grafia (Lopes, 2004). A Tabela 3, traz os exemplos da Tabela 2 antes e depois do processo de Normalização:

Tabela 3 - Termos antes de depois do processo de Normalização

Frases finais da Tabela 2	**Frases após o processo de Normalização**
“garotas” “vestidos” “coloridos”	“garoto” “vestido” “colorido”
“emprestei” “livro” “prima” “adorou”	“emprestar” “livro” “prima” “adorar”
“gosta” “praticar” “esportes”	“gostar” “praticar” “esporte”
“inscricao” “enem” “site”	“inscricao” “enem” “site”

Fonte: A Autora, 2024.

Depois da etapa de Normalização, o próximo passo é a estruturação dos termos com o intuito de se tornarem processáveis para o computador. O modelo mais utilizado para a

representação dos dados textuais (termos) é o modelo espaço-vetorial, uma representação matemática de termos e documentos de uma coletânea textual.

A coletânea é modelada por meio de uma matriz chamada termo-documento (MTD), estrutura utilizada para os próximos passos da Mineração de Textos. Ao considerar uma coleção composta por *N* documentos e *M* termos, a matriz termo-documento A_{NxM}, é composta por *N* colunas e *M* linhas, onde os documentos são armazenados na coluna e os termos nas linhas.

A Tabela 4, ilustra a forma desta matriz (TP) que representa a matriz designada por *term presence* para explicitar a ocorrência ou não do termo em cada documento onde um significa a presença do termo e zero, caso contrário.

Tabela 4 - Matriz Termo-Documento

Termos	**Documento 1**	**Documento 2**	**...**	**Documento N**
Termo 1	0	0	...	1
Termo 2	1	0	...	0
...	...	⋮	⋱	...
Termo M	1	1	...	1

Fonte: A autora, 2024.

Na etapa seguinte são realizados o cálculo de relevância e a seleção dos termos. Este processo tem o intuito de retirar palavras que não agregam o contexto do estudo. Primeiramente, é necessário analisar a relação entre os termos e o texto, designada por peso (Morais; Ambrósio, 2007).

Cabe ressaltar que a seleção dos termos precisa ser realizada com cautela, pois alguns métodos são influenciados pelos termos de menor relevância como Agrupamento, por exemplo. Cabe ao pesquisador a decisão quanto a relevância ou não para sua pesquisa (Leal, 2003), uma vez que se calcula a relevância de cada termo segundo a ocorrência. Por isso, aplica-se um limiar de corte, denominado *threshold*.

Há inúmeros métodos reportados na literatura, o mais utilizado de forma constante trata da Frequência Absoluta (F_{abs}). É de fácil aplicação, e faz parte da etapa denominada seleção baseada no "peso" do termo (Yan; Pedersen, 1997). O cálculo baseado na frequência absoluta do termo *i* $\left(F_{abs(i)}\right)$, conhecido como *term frequency* (TF), corresponde ao quantitativo de termos *i* que aparece em um documento *j*, descrito pela expressão:

$$F_{abs\,(i)} = \sum_{j=1}^{f_i} f_{ij} \quad (1)$$

Apesar de ser uma medida de peso considerada mais simples, seu uso exclusivo não é aconselhado, visto que não faz distinção entre termos que aparecem em poucos documentos e os que aparecem em muitos. O TF não considera a qualidade dos termos, mas apenas a quantidade.

No entanto, mesmo depois de realizada a seleção, o número de termos resultantes ainda pode ser alto. Dentre outros métodos presentes na literatura que podem ser aplicados em um segundo momento, está o Método de Luhn (Luhn, 1958). Ele se baseia na Lei de Zipf.

Em textos, coloca-se os termos ordenados do mais frequente para o menos frequente dispostos em um histograma. Forma-se a chamada Curva de Zipf. Os termos de maior e menor frequência são julgados irrelevantes, por aparecerem na maioria dos textos ou por considera-los raros, selecionando-se apenas os termos com frequência intermediária. Este valor fica a cargo do pesquisador.

No entanto, nesta tese será utilizada a frequência absoluta sem a aplicação Método de Luhn. Por se tratar de avaliação de prontuários médicos, todos os termos foram considerados importantes.

1.2.3 Processamento

Conhecido como Mineração da Informação, onde os textos são transformados em dados estruturados. Busca-se partes relevantes de um texto em um documento para extrair informações específicas (Machado *et al*., 2010).

Após as etapas realizadas na Subseção 1.2.2, realiza-se o processamento das informações dos termos selecionados, a partir das técnicas e dos métodos necessários e coerentes com o tipo de trabalho a ser desenvolvido. Em Mineração de texto, existem diversas técnicas e, de acordo com Carrilho Junior (2007), as mais usuais são:

- **Sumarização:** localiza palavras ou frases que tenham importância dentro de um texto ou conjunto de textos, criando um resumo ou sumário do conteúdo. Esta técnica é

especialmente útil para textos muito extensos, pois o objetivo é reduzir o seu tamanho, mantendo as partes principais sem perder seu significado geral;

- **Categorização/classificação:** determina a categoria ou classe a qual um documento pertence baseado em seu conteúdo. A implementação desta técnica é baseada em métodos de aprendizagem de máquina supervisionado. O principal uso corresponde a indexação de documentos baseado em assunto, porém encontra-se em outros cenários na etapa de pré-processamento de textos;
- **Agrupamento:** nesta técnica, os documentos são agrupados de acordo com suas semelhanças e co-relacionamentos. Diferentemente do que o ocorre na categorização, *Clustering,* não existe um conhecimento anterior das classes possíveis ou existentes, e a descoberta dos grupos ocorre na execução do algoritmo, baseado unicamente no conteúdo dos documentos. A aplicação da técnica produz um tipo de conhecimento que, em alguns casos, faz-se necessária a avaliação de um especialista para a contextualização.

É possível utilizar as técnicas simultaneamente. Neste trabalho será utilizada a metodologia até agora considerada na avaliação dos prontuários que constavam a contaminação pela Covid-19 no período de dezembro 2019 a meados de 2021, um novo tipo de coronavírus que identificado pelos chineses em 27 de dezembro de 2019.

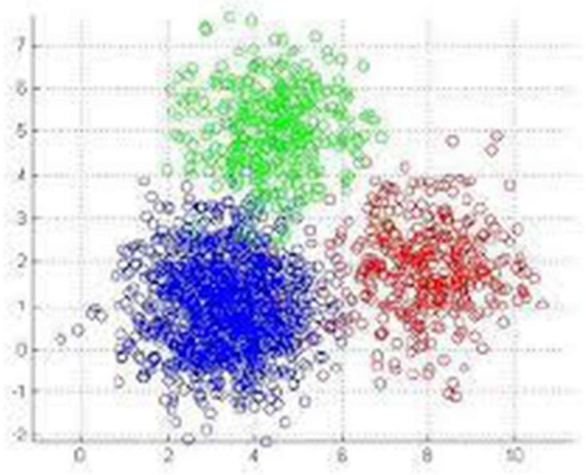
Foi utilizado o Agrupamento, conhecido como Análise de Conglomerados ou *Cluster*. Conforme explica Mingoti (2013), é uma técnica estatística multivariada que objetiva dividir uma amostra (ou população) em grupos, de acordo com as características medidas. Nesta fase aplica-se os métodos de mineração de texto. Na aplicação do c*lustering*, faz-se necessário seguir algumas etapas (Fávero; Belfiore, 2017):

- Seleção da medida de similaridade entre os elementos;
- Seleção do algoritmo de agrupamento, método hierárquico ou não-hierárquico;
- Definição da quantidade de grupos;
- Interpretação e validação dos agrupamentos.

Segundo Bussab, Miazaki e Andrade (1990), critérios diferentes a respeito de medidas de distância e de esquemas de aglomeração podem levar a formações distintas de agrupamentos. A homogeneidade depende fundamentalmente dos objetivos estipulados na pesquisa. O Gráfico

1, ilustra uma pesquisa reportada em Guzzetti (2002), na qual os dados estão agrupados em três *clusters*.

Gráfico 1 - Visualização de agrupamentos



Fonte: Guzzetti, 2002.

A seleção do algoritmo de agrupamento, designado como esquema de aglomeração, é tão importante quanto a definição da medida de distância que afere a semelhança. Essa decisão precisa ser tomada com base no interesse do pesquisador em relação ao objeto da pesquisa (Johnson; Wichern, 2007; Fávero; Belfiore, 2017). Os algoritmos podem ser classificados em:

- **Hierárquicos:**
  - **Aglomerativo:** cada elemento amostral é considerado como um conglomerado que, sucessivamente, sofre uma série de fusões com outros conglomerados até que todos os elementos estejam em um único conglomerado (Johnson; Wichern, 2007);
  - **Divisivo:** todas as observações são consideradas agrupadas e, estágio após estágio, são formados grupos menores pela separação de cada observação, até que essas subdivisões gerem grupos individuais, observações totalmente separadas, que retrata um processo divisivo (Fávero; Belfiore, 2017);
- **Não-Hierárquicos:**
  - **Especificação:** número de agrupamentos;
  - **Novos grupos:** formados pela divisão ou junção de grupos anteriormente combinados;

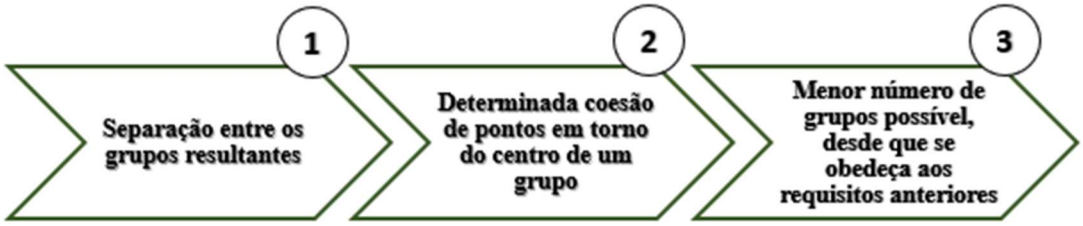
- **Matrizes de distâncias:** armazenar somente a matriz final com as distâncias, não fazendo-se necessário as demais matrizes;
- **Eficiência computacional:** análise de volume de dados expressivos (Johnson; Wichern, 2007).

O critério de agrupamento adotado correspondeu ao *Fuzzy Clustering* que se enquadra no método não-hierárquico, em que os objetos (termos) podem não ser atribuídos a um *cluster* específico, ou seja, podem pertencer a mais de um (Everitt, *et al*., 2011), o que é muito comum em termos (palavras) que possuem mais de um significado. Para esta tese, é escolhido o algoritmo *Intuitionistic Fuzzy C-Means* (IFCM), cuja explicação pode ser vista na Subseção 2.2.

A medida de similaridade ou medida de distância, verifica onde o maior valor foi observado e indica a semelhança dos objetos. A literatura dispõe de diversas medidas de distância, mas, nesta tese, será aplicada a distância euclidiana (DE). Segundo Johnson e Whichern (2007), existe subjetividade na seleção da medida de distância, pois cabe ao pesquisador adotar uma opção em relação aos tipos de variáveis: razão, categórica, ordinal e intervalar.

Após as etapas de seleção do método, do algoritmo de execução e da medida de similaridade para a quantidade de grupos (g), Gath e Geva (1989) descreveram três requisitos que servem como critério para se definir uma partição, isto é, número de grupos considerada aceitável, conforme Figura 7.

Figura 7 – Requisitos para definição do número de grupos



Fonte: A autora, 2024.

No entanto, a determinação de grupos é subjetiva e dependente do valor de similaridade dos parâmetros considerados para agrupamento. Existem alguns métodos/processos/técnicas que auxiliam na escolha dos grupos e, entre os mais utilizados estão: o método Direto Average Silhouette (Rousseeuw, 1987), o método de Teste Estatístico GAP (Tibshirani; Walther; Hastie, 2001) e o método Elbow (Thorndike, 1953).

Este último é utilizado nesta tese, e sua escolha se justifica, pois, seu cálculo é baseado no somatório do erro total dos grupos em cada agrupamento. Desta forma, ele determina o número de grupos (*g*) ideal. Logo, a ideia do método é que o erro total tenda a diminuir até que se aproxime de zero, quando terá um objeto por grupo (Thorndike, 1953). Seu uso explicitado na Seção 2.2.

1.2.4 Pós-processamento

Também conhecida como Análise dos Resultados, tem como principal objetivo apoiar na verificação de até que ponto estes padrões contribuem na solução do problema inicialmente identificado (Milani; Carvalho, 2013).

Diferentemente do contexto no qual os algoritmos que descobrem regras de associação foram originalmente propostos, na área da Saúde fica evidenciado que ou o algoritmo que descobre as regras de associação considera a sequenciação de eventos durante a etapa de descoberta, ou os padrões após descobertos são pós-processados recebendo esta indicação (Milani; Carvalho, 2013).

Os resultados gerados dizem respeito as medidas estatísticas descritivas, aplicação do modelo adotado e uso de representações gráficas. Nesta etapa, podem ser utilizadas técnicas de validação dos resultados a fim de verificar as taxas de erro e de acurácia do método utilizado.

Ferreira (2015) sugere avaliar os resultados do agrupamento obtido pela validação de *clusters* (agrupamentos) de maneira quantitativa e objetiva, a partir do índice de validação, uma opção para caracterizar de forma ampla os diferentes procedimentos avaliativos.

Em outras palavras, avalia a eficiência dos métodos utilizados no que diz respeito à classificação dos grupos e a respectiva taxa de erro. Neste trabalho, optou-se por aplicar algum método específico de validação. Ou seja, o trabalho apresentará os resultados referentes ao modelo *Fuzzy* Intuicionista adaptado de Intarapaiboon (2016) com similaridade cosseno, que será explicitado na Seção 2.2.

1.3 **Modelagem *fuzzy***

A Lógica *Fuzzy* (LF), conhecida como Lógica Difusa, Lógica Nebulosa ou Lógica Multivalorada, objetiva fornecer uma ferramenta capaz de tratar informações de caráter impreciso ou vago (Tanscheit, 2004) em que a Lógica Clássica (LC), também conhecida como Lógica *Booleana* ou Lógica *Crisp*, não consegue contemplar. Sua concepção e divulgação foi consolida na avaliação de textos conceituais por intermédio de diferentes pesquisadores.

A ideia principal da teoria *fuzzy* (TF), considerada simples e natural, é a de que quando não há capacidade de determinar os limites exatos da contingência de um objeto ou texto quando definido de forma ambígua, é necessário buscar uma escala que permita tomar a decisão sobre sua relevância. Isso permite estabelecer valores de pertinência que estão afeitos ao tema de possibilidade (Bellman; Zadeh, 1970).

Esta alternativa tem-se mostrado mais adequada para tratar imperfeições da informação do que a Teoria das Probabilidades (Da Silva *et al.*, 2019). A Lógica *Fuzzy* tornou-se mais reconhecida em meados de 1965 com Zadeh (1965), mas sua origem está ligada à Lógica Matemática.

No século IV a.C., Aristóteles criou a Lógica Aristotélica, conhecida como Lógica Clássica (Kosko, 1994), que estendeu a convicção de Pitágoras ao processo dos indivíduos pensarem e tomarem decisões, aliando a precisão da Matemática com a pesquisa da verdade. Esta lógica, no século X d.C., serviu de base para o pensamento na Europa e no Oriente Médio (De Pontes Saraiva, 2006).

No século XIX, o matemático inglês George Boole (1815-1864), interessado por Matemática e Lógica, aprofundou seus estudos e publicou diversos artigos sobre estas temáticas, intitulados, *"Mathematical Analysis of Logic: Being an Essay Towards a Calculus of Deductive Reasoning"* (Boole, 2011) e *"The Laws of Thought"* (Boole, 2017). Os desdobramentos dos artigos de Boole criaram Álgebra *Booleana,* extremamente importante e necessária para os circuitos dos microprocessadores dos computadores atuais (De Pontes Saraiva, 2006).

Ao longo do tempo surgiram as contestações de que a Lógica *Crisp* (LC) produzia contradições, não passíveis de serem gerenciadas. Estas contestações foram popularizadas no início do século XX pelo filósofo e matemático inglês Bertrand Russell (1872-1970), primeiro a escrever e debater o assunto publicamente. Em 1903, Russell ao trabalhar em seu livro

"Principles of Mathematics" (Russell, 2015), incluiu citações das diversas obras dos matemáticos alemães George Kantor (1845-1918) e Gottlob Fredge (1842-1925).

No entanto, percebeu-se que havia um paradoxo nas teorias de Fredge, quando estava para lançar um livro que trataria sobre fundamentação lógica para a Aritmética, algo que vinha sendo estudado há 20 anos.

Russell acabou enviando uma carta para Fredge a respeito da descoberta do paradoxo (Coury; Vilela, 2019) e o teor desta carta ficou conhecido como Paradoxo de Russell, cuja ideia era mostrar que a Lógica *Crisp* criada por Kantor e as teorias de Fredge poderiam produzir contradições inconclusivas (De Pontes Saraiva, 2006), pois a sua descrição traz a seguinte proposição (Souza; Andrade; Lobo, 2019):

Proposição 1. Considere M o conjunto de todos os conjuntos que não se apresentam como elementos. Se todos os conjuntos estão formando outro conjunto, logo não pode ser considerado um conjunto, assim surge o paradoxo: não há conjunto de todos os conjuntos. Então, têm-se:

$$M = \{x \mid x \notin x\} \tag{2}$$

e desta forma, cita-se vários conjuntos que não pertencem a M. Um exemplo disso é o conjunto:

$$x = \{1\}, \text{ pois } \{1\} \in x \tag{3}$$

Portanto, se M é o conjunto de todos os conjuntos que não são elementos de si mesmos, tem-se dois caminhos:

$$\text{Se } M \in M, \text{ então } M = \{x \mid x \notin x\}, \text{ logo } M \notin M \tag{4}$$
$$\text{Se } M \notin M, \text{ então } M = \{x \mid x \notin x\}, \text{ logo } M \in M \tag{5}$$

A partir de (4) e (5), tem-se que:

$$M \in M \leftrightarrow M \notin M \tag{6}$$

que é uma contradição. Assim, conclui-se que $\nexists \{x \mid x \notin x\}$.

Entre 1920 e 1930, o filósofo e lógico polonês Jan Lukasiewicz (1878-1956) no avançar de suas pesquisas em Lógica, criou a "Lógica da Incerteza", chamado na época de Sistema Trivalente (Salatiel, 2017). O exemplo de Lukasiewicz (Łukasiewicz, 1952) mostra com o exemplo:

"Estarei em Varsóvia ao meio dia de 21 de dezembro do próximo ano. Se essa proposição for verdadeira ou falsa hoje, será necessário ou impossível que eu esteja em Varsóvia ao meio dia de 21 de dezembro do ano que vem, mas isso contraria a suposição de que é apenas possível, não necessário, que eu esteja em Varsóvia ao meio dia em 21 de dezembro do próximo ano. Portanto a proposição considerada é, no momento, nem verdadeira e nem falsa, e deve possuir um terceiro valor, identificado como "possível" ou "indeterminado".

Este fato se deu através do estudo de termos linguísticos como "alto", "velho" e "quente", por exemplo, e propôs que o intervalo de valores de um conjunto pudesse ser 0 (zero) para falso; 1 (um) para verdadeiro e ½ (meio) para possível ou indeterminado, chamado também de terceiro valor nos estudos iniciais (De Pontes Saraiva, 2006; Salatiel, 2017).

Mais tarde, a proposta anterior foi estendida para um conjunto infinito de valores no intervalo [0,1], onde o valor do elemento pertencer ou não ao conjunto é denominado de valor de pertinência e não-pertinência, respectivamente. A Figura 8 exibe as matrizes dos conjuntos de valores de Lukasiewicz.

Figura 8 - Matrizes trivalentes de Lukasiewicz

A	~A
1	0
½	½
0	1

A ∧ B			
A \ B	1	½	0
1	1	½	0
½	½	½	0
0	0	0	0

A ∨ B			
A \ B	1	½	0
1	1	1	1
½	1	½	½
0	1	½	0

A → B			
A \ B	1	½	0
1	1	½	0
½	1	1	½
0	1	1	1

Fonte: Satatiel, 2017.

Em 1937, foi escrito o artigo: "*Vagueness: An Exercise in Logical Analysis"* (Black, 1937), onde se definiu "imprecisão" como algo em que os estados possíveis de proposição não estão claramente definidos e citou a palavra "jovem" ao alegar que para cada indivíduo a definição de jovem poderia variar, levando em consideração outros aspectos. Black mediu a pertinência em graus de utilização e defendeu a teoria geral de "incerteza", no entanto, à época, não deram tanto destaque para seus estudos (Souza, 2003).

Em 1965, Lofti A. Zadeh publicou o artigo *"Fuzzy Sets"* (Zadeh, 1965) que ficou conhecido como a origem da Teoria dos Conjuntos *Fuzzy* (TCF). A motivação para o estudo de Zadeh (1965) surgiu da observação dos recursos tecnológicos disponíveis à época que eram incapazes de automatizar as atividades relacionadas a problemas de natureza industrial,

biológica ou química que compreendessem situações ambíguas, não passíveis de processamento através da lógica computacional fundamentada na Lógica *Booleana.*

O texto de Zadeh teve influência no pensamento sobre a incerteza, pois desafiou não apenas a Teoria da Probabilidade como a única representação da incerteza, mas também a teoria da Lógica Clássica (Ross, 2010). Desta forma, Zadeh, "pai" da Lógica *Fuzzy*, permite indicar as vantagens:

a) Requer poucas regras, valores e decisões;

b) Maior número de variáveis observáveis pode ser valorado e o uso das variáveis linguísticas propicia proximidade do pensamento humano;

c) Simplifica a solução de problemas e proporciona um rápido protótipo dos sistemas.

No entanto, indica-se algumas desvantagens, principalmente ao trabalhar com relações físicas naturais, pois há necessidade maior para verificar a causa e o efeito das variáveis (Da Silva *et al.*, 2019). Agrega-se que os Conjuntos *Fuzzy* Tipo-1 consolidados em imagens geométricas, têm capacidades limitadas quanto ao contexto, sobretudo em situações em que se considera a Inferência Estatística pelo Princípio de Extensão *Fuzzy* ou que sua evolução mude ao longo do tempo.

O princípio de extensão *fuzzy* é utilizado em casos quando se considera uma análise multivariada em Lógica *Fuzzy* ou quando há muitos conjuntos para se trabalhar, o que possibilita uma "explosão" das regras lógicas *"Se... então"*, antecedente/consequente. O princípio será contemplado posteriormente.

Para Braga, Barreto e Machado (1995), apesar da pertinência de um Termo Linguístico variar entre 0 (zero) e 1 (um), não se deve encarar como uma probabilidade, e sim como uma medida de compatibilidade do objeto com o conceito apresentado pelo Conjunto *Fuzzy*.

Em 1975, Zadeh reconheceu certas limitações do Sistema *Fuzzy* do Tipo-1 e propôs o Sistema *Fuzzy* Tipo-2 ou *Fuzzy* Intervalar Tipo-2, segundo o Princípio de Extensão *Fuzzy*, uma vez que as funções de pertinência incluem certa incerteza, de forma a tornar possível a quantificação e modelagem de incertezas tanto numéricas como linguísticas (Mendel, 2017).

Rizol, Mesquita e Saotome (2011) contextualizaram que: seja A um Conjunto *Fuzzy* do Tipo-2 sobre X, caracterizado por uma Função de Pertinência do Tipo-2 igual a $U_A(x,u)$, tem-se que:

$$A = \left\{\left((x,u), U_A(x,u)\right) \middle| \forall\, x \in X \;\; e \;\; \forall \;\; u \;\in\; J_x \;\subseteq [0,1]\right\}, \tag{7}$$

e: $0 \leq U_A(x,u) \leq 1$ (8)

1.3.1 Conjuntos *fuzzy* e princípio de extensão *fuzzy*

A teoria dos conjuntos clássicos, no pertencimento (pertinência) de um elemento no conjunto destaca-se a proposição: Seja *A* um conjunto pertencente ao conjunto universo *X,* onde cada elemento de *X* pode ou não pertencer a *A*, sendo expresso pela função f_A (Tanscheit, 2004):

$$f_A(x) = \begin{cases} 1, \ se\ x \in A \\ 0, \ se\ x \notin A \end{cases} \quad (9)$$

A teoria dos conjuntos *fuzzy* (TCF) apresentada por Zadeh, é uma "extensão" da Teoria dos Conjuntos Clássicos que objetiva fornecer uma ferramenta Matemática para tratar de informações vagas ou imprecisas, pois se afere a possibilidade de um determinado elemento poder pertencer a mais de um conjunto com um respectivo grau de pertinência (Rezende, 2005). Um subconjunto *fuzzy* (*F*) do conjunto Universo (*X*) é definido por uma função de pertinência u, tal que:

$$u_F: X \rightarrow [0,1] \quad (10)$$

onde cada elemento de *X* pode ou não pertencer ao termo linguístico *F*, expressa pela função u_F (Rezende, 2005):

$$u_F(x) = \begin{cases} 1, \ se\ x \in F \\ 0, \ se\ x \notin F \end{cases} \quad (11)$$

e $u_F(x) = 1$ e $u_F(x) = 0$, indicam "o pertencer" e o "não pertencer" de um elemento àquele conjunto, respectivamente.

De forma similar ao Conjunto Clássico, deseja-se saber se o elemento pertence a um ou mais conjuntos e com que o grau de pertinência (u), sendo que quanto mais próximo de zero, mais significa que o elemento se distancia do pertencimento.

Utiliza-se o operador matemático União (∪) para indicar o pertencer a um dos conjuntos e o operador de Intersecção (∩), pertencer simultaneamente. O intuito de facilitar a

interpretação dos Conjuntos *Fuzzy,* categorizado por Variáveis Linguísticas, cuja representação inclui sentenças em uma linguagem especificada construídas por adjetivos.

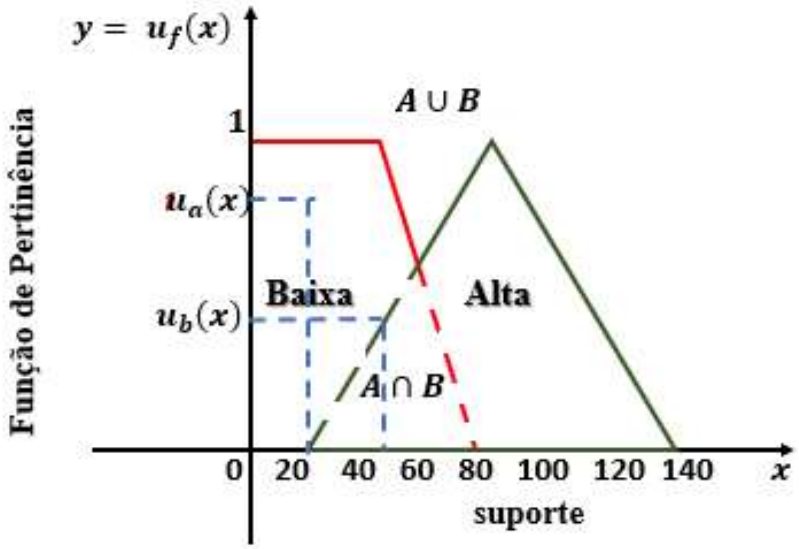
A operacionalidade dessas sentenças é viabilizada pelo uso de conectivos lógicos de negação (NÃO), ligação (E e OU). Tanscheit (2004) relata que a principal função dos termos linguísticos é fornecer uma maneira sistemática para reconhecer padrões referentes a fenômenos complexos ou mal definidos, explicitando:

- ***N:*** nome da variável;
- ***T*(*N*):** conjunto de termos linguísticos de ***N***;
- ***X*:** universo de discurso;
- ***G*:** regra sintática para gerar os valores de ***N*** na composição de termos de ***T*(*N*)**, conectivos lógicos, modificadores e delimitadores;
- ***M:*** regra semântica, associa a cada valor gerado por ***G*** a um Conjunto *Fuzzy* em ***X***.

O Gráfico 2 permite visualizar dois conjuntos *fuzzy*, o Trapezoidal *A* e o Triangular *B* relacionados à diferentes velocidades em km². O eixo das abcissas é designado como "suporte". O envoltório dos conjuntos *fuzzy* é delineado pelo conectivo lógico União (∪) que afere as pertinências $u_A(x)$ e $u_b(x)$, na ordenada gráfica. O conectivo Interseção (∩) desenha a interseção entre os conjuntos $(A \cap B)$, nela considera-se a menor pertinência (Mamdani; Assilian, 1975). Exemplificando:

- ***T(N):*** termos linguísticos "Baixa" e "Alta";
- ***N:*** velocidade;
- ***X:*** valores atribuíveis ao suporte representado pela velocidade segundo a escala de razão (abcissa);
- ***G:*** regra sintática para gerar os valores da velocidade como uma composição de termos linguísticos ***T*(*N*)** segundo conectivos lógicos, modificadores e delimitadores referentes a velocidade;
- ***M*:** regra semântica *"Antecedente... Consequente"*, *"Se...então"* associando a regra sintática ***G*** ao conjunto *fuzzy*, explicitando a respectiva pertinência.

Gráfico 2 - Funções de pertinência dos conjuntos *fuzzy:* velocidade baixa e velocidade alta



Fonte: A autora, 2024.

Nesta tese, o processo de *fuzzyficação* inicia-se pela conversão dos valores padronizados em unidades do desvio padrão em valores de pertinência e respectivos termos linguísticos segundo o uso do Conjuntos *Fuzzy* Tipo-2.

Agrega-se um Conjunto de Regras Lógicas Relacionais *Fuzzy "Se ...e/ou...Então"* (Marques *et al.*, 2006). Também pode-se optar pela *defuzzyficação*, que traduz as pertinências em valores da escala de razão, pois em aplicações práticas geralmente são requeridas saídas precisas (Tanscheit, 2004).

O Princípio de Extensão *Fuzzy,* segundo Tsoukalas e Urig (1997), é uma ferramenta que possibilita utilizar os operadores de matemática, de forma a facilitar a *fuzzyficação*. Sejam X e Y dois universos de conjuntos distintos e f uma função de X em Y (Botelho; Santos; Souza, 2012):

$$f: F(X) \rightarrow F(Y) \tag{12}$$

tal que, para cada $x \in X$:

$$f(x) = y \in Y \tag{13}$$

Seja A um Conjunto *Fuzzy* em X, definido por Rizol, Mesquita e Saotome (2011):

$$A = \sum_{i=1}^{n} \frac{u_A(x_i)}{x_i} = \frac{u_A(x_1)}{x_1} + \frac{u_A(x_2)}{x_2} + \cdots + \frac{u_A(x_n)}{x_n} \quad (14)$$

A imagem de *A* pela função $f(x)$ é um conjunto $B = f(A)$ em *Y*, ou seja, *B* é a saída da função e se apresenta:

$$B = f(A) = f\left(\sum_{i=1}^{n} \frac{u_A(x_i)}{x_i}\right) \quad (15)$$

onde a imagem de x_i sob *f*, $y_i = f(x_i)$, possui um grau de pertinência $u_A(x_i)$.

1.3.2 Algoritmo de agrupamento *Fuzzy* Intuicionista

O algoritmo *fuzzy* é um conjunto ordenado de instruções *fuzzy* que, após execução, produzem uma solução aproximada para um determinado problema (Zadeh, 1973). Os algoritmos *fuzzy* são utilizados em Análise de Agrupamento, que é tratada segundo medida de similaridade ou distância, que verificam o quão distantes podem estar os elementos, ou seja, quanto menor o valor da distância mais similares são os elementos que estão sendo comparados (Mingoti, 2013).

Logo, o Agrupamento objetiva a agregação segundo a similaridade, reconhecendo padrões relativos as características (Mingoti, 2013). Os algoritmos *fuzzy* mais utilizados são: *Fuzzy Analysis Clustering* (FANNY), *Fuzzy C-Means* (FCM) e o *Fuzzy C-Medoids* (FCMdd). Independentemente do algoritmo, todos privilegiam similaridade/dissimilaridade entre os elementos no sentido de agrupar os dados de entrada do Sistema *Fuzzy* (Ferreira, 2015).

Johnson e Wichern (2007) comentam que há subjetividade na escolha de uma medida de distância, isto é, cabe ao pesquisador escolher a melhor. Portanto é importante considerar a escala de medida das variáveis (nominal, ordinal, intervalar e razão). Neste trabalho, no entanto, não será utilizado nenhum dos três mais usuais, mas sim o *Fuzzy Intuicionista C-Means*.

Atanassov (1986) apresentou a teoria dos conjuntos *fuzzy* intuicionista (TCFI) e os fundamentos da lógica *fuzzy* intuicionista (LFI), que inicialmente foi considerada como uma alternativa ao Princípio de Extensão *Fuzzy* (Da Costa, 2012). Esta abordagem permite representar não só a incerteza dos dados, mas a hesitação dos decisores.

A vantagem de utilizar o *fuzzy* intuicionista em vez da teoria dos conjuntos *fuzzy*, é a habilidade para lidar com diferentes tipos de incertezas (Da Costa, 2012). Nesta proposta, os

graus de pertinência e não-pertinência são independentes, pois tanto a função de pertinência quanto a função de não pertinência, são aplicadas simultaneamente, a cada elemento da base de dados (Mert, 2020).

A soma dos graus de pertinência e não-pertinência devem ser menores que as unidade (Mert, 2020). Em teoria seja $A \subset X$, onde X é o conjunto universo e, $x \in X$ um elemento de A que possui pertinência e não pertinência em X. Atanassov (1986) explicitou:

$$A = \{\langle x, u_A(x), 1 - u_A(x)\rangle | \, x \in X\} \tag{16}$$

onde $u(x)$ e $v_A(x)$ são as funções de pertinência e não-pertinência em A, respectivamente, no intervalo [0,1]:

$$u_A(x): X \to [0,1] \text{ e;} \tag{17}$$

$$v_A(x): X \to [0,1] \tag{18}$$

tal que:

$$0 \leq u(x) + v_A(x) \leq 1, \forall \, x \in X \tag{19}$$

O grau de não-pertinência, $v_A(x) = 1 - u_A(x)$, reflete uma possível hesitação do decisor no momento de atribuir a pertinência. Para averiguar tal hesitação, calcula-se o grau de hesitação (grau de indeterminação) de x em A (Atanassov, 1986), introduzido por Bustince e Burillo (1996), com a notação:

$$\pi_A(x) = 1 - u_A(x) - v_A(x) \tag{20}$$

Nesta expressão, π_A corresponde à hesitação de x pertencer ao conjunto *Fuzzy* Intuicionista A, onde $u_A(x)$ e $v_A(x)$ são as funções de pertinência e não-pertinência de A, respectivamente. Este conceito satisfaz a condição:

$$0 \leq \pi_A(x) \leq 1, \forall \, x \in X \tag{21}$$

Caso o grau de pertinência seja o complemento do grau de não-pertinência, não haverá hesitação (Da Costa, 2012) e o conjunto não será *fuzzy* intuicionista.

A Modelagem *Fuzzy,* atribuído a Mamdani (1977), segundo a inferência *fuzzy,* corresponde a um sistema lógico que objetiva reconhecer padrões por regras previamente

definidas, com o intuito de representar experiências da vida real. A Figura 9 mostra um exemplo da operacionalidade de um Sistema *Fuzzy* com o uso do Princípio de Extensão.

Figura 9 – Exemplo de funcionamento de um Sistema *Fuzzy*

Fonte: A autora, 2024.

Nesta proposta o *Fuzzy* Intuicionista *C-Means* é utilizado como método não hierárquico de agrupamento em Mineração de Texto com o intuito de reconhecimento de padrões em uma base de dados da Covid-19, explicação disponibilizada na Seção 2.2.

2 **MATERIAIS E MÉTODOS**

Neste Capítulo são descritos os materiais utilizados na tese, bem como os conceitos, técnicas e métodos considerados necessários para a implementação da modelagem *Fuzzy* sob a ótica do Princípio de Extensão *Fuzzy* para a Mineração de Textos.

Todas as análises são realizadas pelos *softwares* R versão 4.3.0 (R Core Team, 2023), e pelo IDE *RStudio* (Posit Team, 2023), linguagem de programação para gráficos e cálculos estatísticos.

2.1 **Materiais**

A base de dados foi composta de informações provenientes dos 2089 prontuários referentes ao período de dezembro de 2019 até setembro de 2021, do Hospital Universitário Pedro Ernesto (HUPE) das em enfermarias e UTI.

Estas informações foram aprovadas pelo Comitê de Ética, CAAE: 3592920.2.0000.5282, obtidas por meio do projeto em parceria com a Telemedicina da UERJ, edital n° 12/2020, bolsa CAPES-EPIDEMIAS e por meio da assinatura do termo de sigilo e privacidade, além de respeitar as regras da Lei Geral de Proteção de Dados (LGPD).

Os 2089 prontuários perfaziam 88.197 linhas dispostos em colunas para:

- ***CD_PACIENTE*****:** código de identificação do paciente (ID) modificado, para preservar a identidade, porém indicando e associando os registros das outras colunas a um único ID;
- ***DH_REFERENCIA*****:** data em que o prontuário do paciente foi anotado referente ao registro diário do paciente nesta Unidade de Saúde, podendo o paciente ter mais de um registro na mesma data;
- ***LO_VALOR*****:** texto do prontuário do paciente referente as informações pessoais, sociodemográficas e de saúde, sendo atualizado diariamente.

Utilizou-se o *Knowledge Discovery in Text* (Subseção 1.2.2) que inclui o pré-processamento de dados não-estruturados para textos, para transformar as palavras em atributos

comportamentais. Objetivou-se fazer com que o texto pudesse ser interpretado segundo um cenário.

O *RStudio* (Posit Team, 2023), livre e integrado para *R* (R Core Team, 2023), foi escolhido, por ser uma linguagem de programação para gráficos e cálculos estatísticos. É considerado como uma interface "facilitada" para o programador. A Figura 10 mostra os pacotes do *RStudio* necessários para o tratamento dos dados.

Figura 10 - Pacotes do *RStudio* utilizados no pré-processamento e suas descrições

PACOTE	DESCRIÇÃO	AUTOR E ANO
dplyr	Manipulação da base de dados	Wickham (2023)
tidyverse	Manipulação da base de dados	Wickham et al. (2019)
NLP	Processamento da Linguagem Natural. Funções de análise morfológica, sintática, etc.	Hornik (2020)
stringr	Manipulação de caracteres	Wickham (2022)
stringi	Manipulação de caracteres alfanuméricos	Gagolewski (2022)
ludridate	Manipulação de data e hora	Grolemund e Wickham (2011)
tm	Mineração de texto. Funções de conversão de letras de maiúsculo para minúsculo, remoção de números, espaços extras, remoção de pontuação	Feinerer e Hornik (2023)
stopwords	Lista de *stopwords*. Foi necessária complementação, pois não há uma lista completa, em virtude de um mesmo idioma ter variação. Para a tese se utilizou da lista referenciada pela maioria dos sistemas e referências utilizadas, que é a do site Ranks NL Webmaster Tools (1998). A lista com as *stopwords* se encontra no Anexo A	Benoit, Muhr e Watanabe (2021)
rslp	Processo de *stemming* para a língua portuguesa	Falbel (2020)
txt4cs	Remoção de acentos em língua portuguesa	Moreira (2023)

Fonte: A autora, 2024.

A extração das informações levou em consideração as variações dos termos pelo seu sinônimo, singular/plural, além de possíveis erros gramaticais. Seguindo as etapas do *Knowledge Discovery in Text*, Seção 1.2, para a Coleta de Informações (Etapa 1) foram utilizadas as abordagens estatística e semântica.

Na abordagem semântica, observando as etapas do PLN, foram utilizadas todas as análises do processo, vistas na Figura 3. No Pré-processamento (Etapa 2), foram aplicados primeiramente o *case folding* e a Indexação para um tratamento inicial dos termos. Em seguida, foram realizadas:

a) Análises das informações da base de dados, com intuito de verificar variações das palavras;

b) Pesquisas de termos médicos (técnicos) em documentos levantados pelo projeto CAPES-COVID, por haver palavras que possuíam duas formas de escritas;

c) Validação realizada por um especialista da área de saúde.

Estas informações permitiram criar palavras-chave a serem utilizadas que se encontram: no Apêndice B, Quadros B.1, B.2, B.3, B.4 e B.5; Apêndice C e Apêndice D. Todos os processos da etapa de pré-processamento foram realizados na variável inicial *LO_VALOR,* que continha os prontuários dos pacientes.

A criação das variáveis foi feita em dois momentos distintos: a primeira com as *stopwords* e sem o processo de Normalização para que fossem detectadas variações de palavras em feminino e masculino.

Para tal, foi criada uma nova lista de *stopwords*, que pode ser encontrada no Apêndice A, contendo as pertencentes do pacote do *software* R e a de sites especializados como Ranks NL (1998) e no Github do Antônio Lopes (2013), que trazem outros termos, já que a língua portuguesa possui variações de acordo com o país.

Foram mantidas na lista das *stopwords* os termos “bem”, “bom”, “mal”, “certamente”, “certeza”, ”mau”, “não”, “negativamente”, “nunca”, “provavelmente”, “sabe”, “sim”, pois caso fossem retiradas, poderiam prejudicar a coleta de informações ou levar a informações errôneas.

A segunda busca foi realizada sem as *stopwords* e com o processo de Normalização, para uma pesquisa mais refinada. Para ambos casos, ocorreu o processo de seleção dos termos. Os resultados de ambas as buscas foram cruzados e comparados, de forma a melhorar a apuração das informações, já que todos os algoritmos de *machine learning* são passíveis de erro.

Após a aplicação e verificação e exclusão de dados duplicados, e os prontuários que apontaram a não presença de Covid-19 contabilizou 2.0 prontuários discriminando 33 variáveis inseridas no Quadro 1 que discrimina as características sociodemográficas, no Quadro 2 estão os sintomas; e no Quadro 3 ficaram as comorbidades. Algumas variáveis seguem a distribuição de probabilidade Multinominal, pois os relatos escritos indicaram mais de duas opções de classificação, enquanto outras, apenas duas alternativas, “sim” ou “não”, seguida consequentemente, a distribuição de probabilidade Binomial. Ao todo foram 32 variáveis e 2089 pacientes.

Quadro 1 - Variáveis sociodemográficas

Variável	Descrição
ID_PACIENTE	Identificação do paciente com PAC_(número em que foi registrado (exemplo: PAC_001, etc.)
TESTE_COVID19	Positivo; Negativo; Suspeito; Não informado
OBITO_ALTA	Alta; Óbito; Não informado
DATA_ENTRADA	Data em que o paciente entrou
DATA_SAIDA	Data em que o paciente teve alta ou óbito
N_DIAS_INTERNACAO	Número de dias que o paciente ficou internado
GENERO	Masculino; Feminino; Não informado
IDADE	Anos: De 1 a 100; Meses: De 1 a 12; Dias: De 0 a 45
CLASS_ETARIA (anos)	Criança: De 0 a 11; Adolescente: De 12 a 18; Adulto: De 19 a 59; Idoso: De 60 anos ou mais
FUMANTES	Se o paciente fuma: 1-Sim; 0-Não
ETILISTA	Se o paciente bebe/alcoólatra: 1-Sim; 0-Não

Fonte: A autora, 2024.

Quadro 2 – Variáveis de sintomas

Variável	Descrição
CANSACO	1-Sim; 0-Não
CORIZA	1-Sim; 0-Não
DOR_CABECA	1-Sim; 0-Não
DOR_GARGANTA	1-Sim; 0-Não
DOR_CORPO	1-Sim; 0-Não
ERUPCAO_PELE	1-Sim; 0-Não
FALTA_APETITE	1-Sim; 0-Não
FALTA_AR	1-Sim; 0-Não
FALTA_OLFATO	1-Sim; 0-Não
FALTA_PALADAR	1-Sim; 0-Não
FEBRE	1-Sim; 0-Não
TONTURA	1-Sim; 0-Não
TOSSE	1-Sim; 0-Não
VOMITO	1-Sim; 0-Não

Fonte: A autora, 2024.

Quadro 3 - Variáveis de comorbidades

Comorbidades	Descrição
DIABETES	1-Sim; 0-Não
DOENCA_CARDIOVASCULAR	1-Sim; 0-Não
DOENCA_CEREBROCASCULAR	1-Sim; 0-Não
DOENCA_RENAL	1-Sim; 0-Não
DOENCA_PULMONAR	1-Sim; 0-Não
HIPERTENSAO	1-Sim; 0-Não
IMUNOCOMPROMETIDO	1-Sim; 0-Não
OBESIDADE	1-Sim; 0-Não

Fonte: A autora, 2024.

Em seguida, foi aplicado um novo pré-processamento, para a construção de uma nova base, chamada “base ajustada”, com o objetivo de armazenar apenas as informações necessárias para a proposta de aplicação da modelagem *fuzzy* intuicionista, além de eliminar inconsistências e tornar a base mais concisa.

Para este trabalho, era necessário que a base tivesse informações dos prontuários de forma completa, ou seja, que todas as informações estivessem preenchidas. Portanto, criou-se uma nova base com 282 prontuários e 33 variáveis. Desta forma, excluiu-se 1.807 registros das informações iniciais, registros sinalizados por:

a) Na variável *TESTE_COVID19*, registros classificados como “Não informado”, “Suspeito” e “Negativo”, pois para a aplicação da modelagem é de interesse somente os casos positivos;

b) Na variável *OBITO_ALTA*, registros definidos como “Não informado” ou “Transferidos”, pois é de interesse somente os casos com desfecho no HUPE;

c) Na variável *GENERO*, registros definidos como “Não informado”;

d) Na variável *N_DIAS_INTERNACAO*, frequência inferior a 2 (dois) dias;

e) Idades não informadas ou preenchidas com 999 ou não informada.

Esta base foi utilizada para aplicação dos métodos explicitados na Seção a seguir. A Figura 11 ilustra os processos realizados na base até o presente momento.

Figura 11 - Processos realizados antes da aplicação da modelagem *Fuzzy* Intuicionista

Fonte: A autora, 2024.

2.2 **Métodos**

Primeiramente, foi realizada a escolha do método entre os hierárquico ou não-hierárquico, apresentados na Subseção 1.2.3. Foi escolhido o segundo método, pois se trata de dados não-estruturados. De forma geral, os métodos não-hierárquicos seguem a lógica de Huang (1997), que faz uso dos agrupamentos como uma forma de organização, exibida no Quadro 4.

Quadro 4 - Algoritmo do método não-hierárquico

Entrada: k - quantidade de grupos; ε - critério de parada	
Passo1	Seleciona k grupos para os agrupamentos iniciais;
Passo2	Emprega um critério ε de parada, definido pelo usuário ou pela aplicação;
Passo3	Atribua a cada observação ao agrupamento mais próximo;
Passo4	Re-atribua cada observação a um dos k agrupamentos;
Passo5	**SE** não há mais realocação de dados ou a re-atribuição satisfaz um conjunto de critérios determinados pela regra de parada empregada; **FIM** **SENÃO** Retorna ao **Passo2**.
Saída: k agrupamentos	

Fonte: A autora, 2024.

Em seguida, foi utilizada a distância euclidiana como medida de similaridade, função $d: \Re^M \times \Re^M$ que associa cada par ordenado de elementos $x, y \in \Re^M$ a um número real $d(x, y)$, chamado distância entre x e y. Essa medida satisfaz as seguintes condições para quaisquer $x, y, z \in \Re^M$ (Lima, 2014):

a) $d(x, x) = 0$;

b) Se $x \neq y$, então $d(x, y) > 0$;

a) $d(x, y) = d(y, x)$ (simetria) e;

b) $d(x, z) \leq d(x, y) + d(y, z)$ (desigualdade triangular).

Nesta tese, optou-se pela distância euclidiana (DE), que é a distância linear entre dois pontos em um espaço *M*-dimensional. Sejam *X* e *Y* dois vetores (elementos) amostrais, a distância euclidiana entre eles é definida por:

$$d(x, y) = \sqrt{\sum_{k=1}^{M} (x_k - y_k)^2} = \sqrt{(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2} \quad (23)$$

onde:

- x_k, *k*-ésima variável da unidade amostral *x*;
- y_k, *k*-ésima variável da unidade amostral *y*.

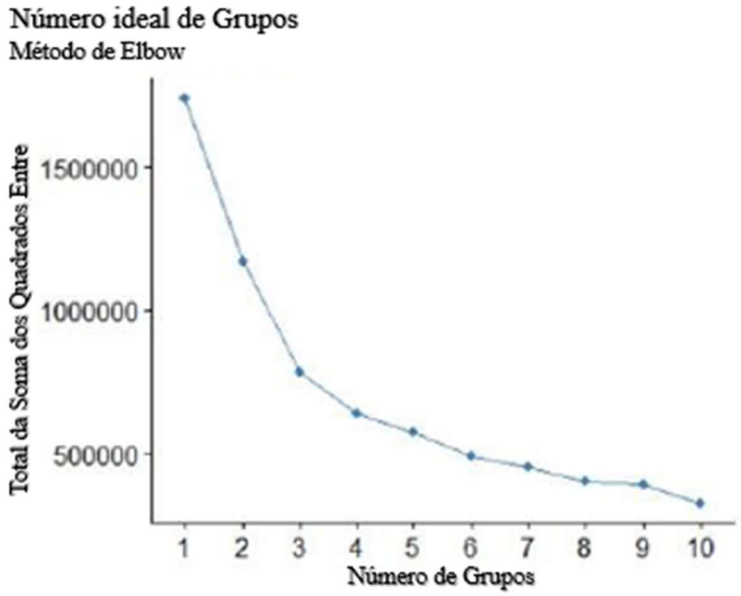
Utilizou-se o método Elbow para auxiliar a decisão da determinação de uma quantidade de grupos (g) a ser considerado ideal. Este número é adotado para o algoritmo do método de agrupamento e é representado pela equação (24) e pelo Gráfico 3 (Renjith; Sreekumar; Jathavedan, 2022), através do pacote *factoextra* (Kassambara; Mundt, 2022):

$$WSS = \sum_{k=1}^{g} \sum_{E_i=1}^{Ck} (\mu_k - E_i)^2 \quad (24)$$

onde:

- g: número de grupos considerando a cada iteração;
- E_i: elemento de cada agrupamento;
- C_k: *cluster*;
- μ_k: centroide.

Gráfico 3 - Gráfico da escolha do número de grupos (g) considerado ”ideal” pelo método Elbow



Fonte: Adaptado de Renjith; Sreekumar; Jathavedan, 2022.

Para o critério de agrupamento, optou-se pelo algoritmo *Intuitionistic Fuzzy C-Means* (IFCM), pois é um método não-hierárquico em que o número de subconjuntos é considerado conhecido e a função de pertinência de cada objeto em cada *cluster* é estimada, de forma interativa, geralmente uma técnica de otimização padrão baseada em uma função objetivo (Everitt, *et al*., 2011). Na implementação do Sistema *Fuzzy* Intuicionista *C-Means* há alguns critérios.

O primeiro diz respeito a geração das pertinências iniciais. Entre as opções presentes na literatura, escolheu-se a de Takagi-Sugeno (1985), que consiste em um sistema de inferência capaz de descrever quaisquer funções de forma aproximada (Ross, 2010).

Entre as vantagens do uso do modelo de Takagi-Sugeno sobre os outros, está o fato dos consequentes serem sistemas dinâmicos, demandando uma menor quantidade de regras *“Se... Então”*, o que facilita substancialmente a tarefa de identificação de modelos (Johansen; Shorten; Murray-Smith, 2000).

O vetor das pertinências iniciais (μ_i) é calculado de forma semelhante ao modelo de Mamdani (1977) com a diferença que, na saída do sistema, a função pode ser linear ou constante. As pertinências de cada grupo são determinadas pela razão entre valores aleatórios gerados no intervalo [0,1], chamado de peso (r), e a soma deste conjunto de valores aleatórios:

$$\mu_i = \frac{r_i}{\Sigma_1^n r_i} \tag{25}$$

No caso *Fuzzy* Intuicionista, os mais utilizados são os métodos de Sugeno e Terano (1977) e Yager (1979). A função de hesitação de Sugeno e Terano (1977), conhecido como Hesitação de Sugeno, é definido por uma constante λ, que pode variar de (-∞ ; ∞). Se a pertinência de um elemento a um conjunto *Fuzzy* Intuicionista é *u*, então:

$$\left(\frac{(1-\mu_i)}{(1+\lambda\mu_i)}\right) \tag{26}$$

A Função de Hesitação ($\boldsymbol{\pi}$) é descrita como:

$$\pi_i = (1-\mu_i) - [(1-\mu_i)/(1+\lambda\mu_i)] \tag{27}$$

Cabe ressaltar que, quando $\lambda = 1$, a função de hesitação de Sugeno torna-se a Função de Hesitação de Zadeh. Na Função de Hesitação de Yager (1979), o operador é definido por uma constante (α) no intervalo [0; +∞). Se a pertinência de um elemento a um conjunto *Fuzzy* Intuicionista é u, então:

$$(1-\mu_i^{\alpha})^{1/\alpha} \tag{28}$$

Logo, a Função de Hesitação ($\boldsymbol{\pi}$) é descrita como:

$$\pi_i = (1-\mu_i) - (1-\mu_i^{\alpha})^{1/\alpha} \tag{29}$$

onde α é um parâmetro que pode ser ajustado. Quando $\alpha = 1$, a Função de Hesitação de Yager se torna a Função de Hesitação de Zadeh.

Em seguida, utiliza-se a Função Objetivo (*J*) para a minimizar. Após *N* iterações do algoritmo, os resultados são obtidos para tornar *J* o menor possível. Quanto menor for o número de iterações necessárias, melhor o desempenho do algoritmo, pois isso indica uma maior eficiência no processo de agrupamento.

Suponha que exista um número g de grupos pré-definidos $C_{a=1,2,\dots g}$, e cada grupo contenha $D_{i=1,2,\dots n}$ documentos e $X_{j=1,2,\dots m}$ termos (Revanasiddappa; Harish, 2018). O IFCM busca minimizar a função objetivo dada por (Chaira, 2011; Chaira, 2019):

$$J(V,X) = \sum_{i=1}^{n}\sum_{j=1}^{g} u_{ij}^{m} d^2(x_i, c_j) \quad (30)$$

onde:

- *n*, quantidade de elementos (termos);
- *g,* número de grupos (*cluster*s) pré-determinado. Nesta tese foi determinado pelo método Elbow;
- u_{kj}, pertinências *Fuzzy* Intuicionista, geradas pela expressão:

$$u_{ik} = \frac{1}{\sum_{j=1}^{g}\left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{\frac{1}{m-1}}} \; ; \quad (31)$$

- *m,* parâmetro de *fuzzificação*, cujo valor sugerido por Pal e Bezdek (1995), provavelmente no intervalo [1,25; 2,5], sendo o ponto médio *m* = 2, escolha mais usual na literatura;
- c_j, centro do agrupamento *fuzzy* para cada agrupamento obtido segundo a média ponderada em função dos valores padronizados em unidades do desvio padrão (*z*):

$$z_{kj} = \frac{x_{kj} - c_j}{\sigma_j} \quad (32)$$

correspondentes aos termos, cujas pertinências correspondem aos ponderadores:

$$c_j = \frac{\sum_{j=1}^{p} u_{kj}^{m} X_{kj}}{\sum_{j=1}^{p} u_{kj}^{m}} \quad (33)$$

- (d_j, c_j), distância entre o elemento amostral (termo), x_i, e o centro do agrupamento, c_j.

Há um critério de parada para a Função Objetivo (*J*), chamado de erro estimado (ε), quando $|J \leq \varepsilon|$. Não há um valor predefinido para ε, porém Egrioglu, Bas e Kizilaslan (2023) sugerem que deveria ser menor que o erro estipulado $\varepsilon = 0{,}03$.

Para o modelo *Fuzzy* Intuicionista *C-Means*, é necessário fornecer a matriz de entrada *X*, que deve conter duas variáveis numéricas. O Quadro 5, descreve as etapas do algoritmo IFCM. Na aplicação realizada no *software* R, os pacotes utilizados foram *inaparc* (Cebeci; Cebeci, 2022) e *rinfis* (Egriolu; Bas; Kizilaslan, 2023):

Quadro 5 – Algoritmo do Método *Fuzzy* Intuicionista *C-Means*

Algoritmo do Método *Fuzzy* Intuicionista *C-Means*	
Entrada	*X = Matriz com os escores ranqueados padronizados em unidades do desvio-padrão correspondentes a* n_{kj} *termos;* *m = Parâmetro de fuzzificação (m = 2);* *k = Número de agrupamentos obtidos pelo método Elbow;* $n = \sum_{k=1}^{n_k}\sum_{j=1}^{p} n_{kj}$ *, total de termos da matriz X;* $\alpha =$ *graus de hesitação (*$\alpha = 0.85$*);* $\lambda = constante\ (\lambda = 2)$; ε *= Erro estimado, utilizado como critério de parada da função objetivo*
Passo1	Gerar matriz valores aleatórios *r* no intervalo [0,1], para a criação das pertinências iniciais u_i.
Passo2	Gerar as pertinências iniciais u, utilizando os valores gerados no **Passo1**, adaptado a ao odelo de Takagi-Sugeno**:** $u_i = \frac{r_i}{\sum_{i=1}^{n} r_i}$
Passo3	Escolha do método para a primeira geração do grau de hesitação π. **SE** o método escolhido for **YAGER:** $\boldsymbol{\pi = 1 - u - (1 - u^{\alpha})^{\frac{1}{\alpha}}}$ **SENÃO** **SE** o método escolhido for **SUGENO:** $\boldsymbol{\pi = 1 - u - \left(\frac{(1-u)}{(1+\lambda u)}\right)}$
Passo4	Obtenha o centroide c_j: $\frac{\sum_{j=1}^{p} u_{kj}^{m} X_{kj}}{\sum_{j=1}^{p} u_{kj}^{m}}$
Passo5	Calcule a distância euclidiana entre os elementos e o centro do agrupamento $d(z,x) = \sqrt{\sum_{k=1}^{M}(z_k - x_k)^2}$
Passo6	Obtenha as pertinências atualizadas u_{ik}: $\frac{1}{\sum_{j=1}^{g}\left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{\frac{1}{m-1}}}$
Passo7	Calcule a função objetivo: $J(Z,C) = \sum_{i=1}^{n}\sum_{j=1}^{g} u_{ij}^{m} d^2(z_i, c_j)$
Passo8	**SE** a função objetivo não estiver minimizada, ou seja, $\lvert J > \varepsilon \rvert$: volte para o **Passo2**; **SENÃO** **SE** a função objetivo estiver minimizada: **FIM** Calcule as não-pertinências (complemento): $\boldsymbol{v = 1 - u - \pi}$
Saída	*Agrupamentos e os valores finais de pertinência, não pertinência e grau de hesitação e função objetivo.*

Fonte: A autora, 2024.

De posse dos agrupamentos, desenhou-se o gráfico dos mesmos, através do pacote *cluster* (Maechler *et al.*, 2023), prosseguiu-se a modelagem *Fuzzy* atribuído a Intarapaiboon (2016) que utiliza a distância cosseno adaptada para ao *Fuzzy* Intuicionista na classificação de dados textuais. A Tabela de Contingência, Tabela 5, agrega as frequências (a_{ij}) segundo a discriminação Documentos, Termos e Classes.

Tabela 5 - Frequência dos documentos segundo termos e respectivas Classes

Documentos ***(i=1,2,3,4,5)***	**Termos** ***(j=1,2,3)***			**Classe** ***(c = 1,2)***
	T_1	T_2	T_3	
D_1	a_{11}	a_{12}	a_{13}	C_1
D_2	a_{21}	a_{22}	a_{23}	C_1
D_3	a_{31}	a_{32}	a_{33}	C_1
D_4	a_{41}	a_{42}	a_{43}	C_2
D_5	a_{51}	a_{52}	a_{53}	C_2

Fonte: Adaptado das informações de Intarapaiboon, 2016.

A classificação em Grupos Padrões foi consubstanciada no índice de estratificação (Câmara, 1966), em que a razão entre a variância dentro e a variância total indica o coeficiente de estratificação.

A variância dentro dos estratos é obtida segundo a média ponderada das variâncias dos estratos, tendo como ponderador o número de observações. Este índice deve apresentar decréscimo em função da opção adotada ao número de estratos: quanto menor a variância, mais homogêneos são os valores pertencentes ao estrato.

Variância dentro:

$$\sigma^2_{Dentro} = \left(\frac{1}{N}\right)\left((N_1)\left(\frac{\sum_{i=1}^{N_1}(X_1-\bar{X}_1)^2}{N_1}\right) + (N_2)\left(\frac{\sum_{i=1}^{N_1}(X_2-\bar{X}_2)^2}{N_2}\right) + \cdots + (N_n)\left(\frac{\sum_{i=1}^{n}(X_n-\bar{X}_n)^2}{N_n}\right)\right) \quad (34)$$

onde: $N = \Sigma(N_i)$

Variância total:

$$\sigma^2_{Total} = \frac{\sum_{i=1}^{N}\left(X_i-\bar{X}\right)^2}{N} \quad (35)$$

onde:

- N_i: amostragem do estrato *i*;
- $\bar{X}_{1i}$: médias da variável *X* em cada estrato *i*;
- $\bar{X}$: média global dos estratos;
- X_i: variável *X* no estrato *i*.

Índice de estratificação:

$$IE_{stratificação} = \frac{\sigma^2_{Dentro}}{\sigma^2_{Total}} \tag{36}$$

É recomendado expressar os valores em unidades do desvio padrão (*z*) para obter os valores de pertinência (μ) e não-pertinência (ν). No entanto, uma vez que os valores a serem utilizados na tese são binários, os valores utilizados ocorreram segundo a escala categórica dicotômica. Assim, têm-se as pertinências e não-pertinências (Intarapaiboon, 2016):

$$\mu_{ij} = \frac{r_j}{1+e^{-z_{ij}}}, \tag{37}$$

$$\nu_{ij} = \frac{r_{j^*}}{1+e^{z_{ij}}} \tag{38}$$

onde:

- r_j: valor da constante igual a 0,9, em similitude a função de ativação sigmoide das redes neurais com propagação positiva;
- z_{ij}: valores binários zero ou um da linha *i*, coluna *j*, dispostos na Tabela 6.

Tabela 6 - Pertinências e não-pertinências dos termos por documentos e classes

Documentos	**Termos *(j)***						**Classes**
(i=1,2,3,4)	$\mu \boldsymbol{T}_1$	$\vartheta \boldsymbol{T}_1$	$\mu \boldsymbol{T}_2$	$\vartheta \boldsymbol{T}_2$	$\mu \boldsymbol{T}_3$	$\vartheta \boldsymbol{T}_3$	**($C_{i=1,2}$)**
$\boldsymbol{D_1}$	μ_{11}	ϑ_{11}	μ_{12}	ϑ_{12}	μ_{13}	ϑ_{13}	C_1
$\boldsymbol{D_2}$	μ_{21}	ϑ_{21}	μ_{22}	ϑ_{22}	μ_{23}	ϑ_{23}	C_1
$\boldsymbol{D_3}$	μ_{31}	ϑ_{31}	μ_{32}	ϑ_{32}	μ_{33}	ϑ_{33}	C_1
$\boldsymbol{D_4}$	μ_{41}	ϑ_{41}	μ_{42}	ϑ_{42}	μ_{43}	ϑ_{43}	C_2
$\boldsymbol{D_5}$	μ_{51}	ϑ_{51}	μ_{52}	ϑ_{52}	μ_{53}	ϑ_{53}	C_2

Fonte: Adaptado das informações de Intarapaiboon, 2016.

onde:

- μ_{ij}: Pertinência do termo do documento D_i da classe C_j;
- ϑ_{ij}: Não-pertinência do termo do documento D_i da classe C_j.

No prosseguimento ao método, obtêm-se a média das pertinências por classe e o respectivo o desvio-padrão (DP) inseridos na Tabela 7.

Tabela 7 - Médias e desvios-padrões das pertinências e não-pertinências por classe

Classe *(j=1,2)*	**Documentos** *(i=1,2,3) – linha1* *(i=1,2) – linha2*	**Termos** *(j=1,2,3) – linha1* *(j=1,2) – linha2*					
		Média μ_{Ti}	**DP** σ_{Ti}	**Média** μ_{Ti}	**DP** σ_{Ti}	**Média** μ_{Ti}	**DP** σ_{Ti}
$\boldsymbol{C_1}$	$D_1 + D_2 + D_3$	μ_{11}	σ_{11}	μ_{12}	σ_{12}	μ_{13}	σ_{13}
$\boldsymbol{C_2}$	$D_4 + D_5$	μ_{21}	σ_{21}	μ_{22}	σ_{22}	μ_{23}	σ_{23}

Fonte: Adaptado das informações de Intarapaiboon, 2016.

Para Intarapaiboon (2016), deseja-se saber como o documento (D_6) seria classificado, em relação as frequências observadas para os termos T_1, T_2 e T_3. Na Tabela 8 visualiza-se a indagação a respeito da classificação do documento D_6 em uma das classes, C_1 ou C_2.

Tabela 8 - Frequência dos documentos segundo termos e respectivas classes

Documentos *(i=1,2,3,4,5)*	**Termos** *(j=1,2,3)*			**Classe C** *(j=1,2)*
	$\boldsymbol{T_1}$	$\boldsymbol{T_2}$	$\boldsymbol{T_3}$	
$\boldsymbol{D_1}$	a_{11}	a_{12}	a_{13}	C_1
$\boldsymbol{D_2}$	a_{21}	a_{22}	a_{23}	C_1
$\boldsymbol{D_3}$	a_{31}	a_{32}	a_{33}	C_1
$\boldsymbol{D_4}$	a_{41}	a_{42}	a_{43}	C_2
$\boldsymbol{D_5}$	a_{51}	a_{52}	a_{53}	C_2
$\boldsymbol{D_6}$	$\boldsymbol{a_{61}}$	$\boldsymbol{a_{62}}$	$\boldsymbol{a_{63}}$	**?**

Fonte: Adaptado das informações de Intarapaiboon, 2016.

De forma similar aos procedimentos utilizados nas Tabelas 6, 7 e 8, são calculados as médias, os desvios-padrão e as pertinências e não-pertinências para o novo conjunto de dados. Com base nessas informações, aplica-se à similaridade cosseno adaptada para o modelo *Fuzzy* proposta por Ye (2011). A expressão é descrita segundo equação (39):

$$S_c(A,B) = \frac{1}{h}\sum_{i=1}^{h} \frac{\left(\mu_a(x_i)\mu_b(x_i)\right) + \left(v_a(x_i)v_b(x_i)\right)}{\sqrt{\mu_a^2(x_i) + v_a^2(x_i)}\,\sqrt{\mu_b^2(x_i) + v_b^2(x_i)}} \tag{39}$$

onde:

- h: quantidade de classes;
- μ_a: pertinências sem o novo documento;
- μ_b: pertinências com o novo documento;
- v_a: não-pertinências sem o novo documento;
- v_b: não-pertinências com o novo documento;
- x_i: frequência da classe.

Segundo Ye (2011), tem-se que $h = 2$ pois há duas classes C_1 e C_2, logo, a distância será calculada em relação a cada classe. A aplicação da equação (39) permite avaliar a similaridade S_{c1} e S_{c2}, de ao coma proposta de Intarapaiboon (2016) e Ye (2011), sendo o "D6" classificado em C_1, quando $S_{c1} > S_{c2}$.

Por se tratar de similaridade cosseno, realizou-se a transformação angular a fim de interpretar o resultado a partir de uma representação gráfica. Isso permite analisar que quanto mais próxima de um, S_{c1} e S_{c2} descrevem uma angulação de 90° com maior chance de o documento pertencer a uma determinada classe.

Nesta tese, de acordo com o exposto, o reconhecimento de padrão em registros de prontuário médico com o uso da modelagem *Fuzzy* ocorre da seguinte forma:

- Na linha, constam os pacientes;
- Na coluna, as variáveis dicotômicas referentes a sintomas e comorbidades previamente selecionadas, preenchidas de forma dicotômica zero e um;
- As classes *C* são adotadas segundo os agrupamentos obtidos pelo *Fuzzy* Intuicionista *C-Means* e nomeados de "Classe Padrão";

- Modelagem *fuzzy* com similaridade cosseno, como modelo de classificação de um novo paciente em uma das "Classes Padrões" já existentes.

A Tabela 9 exibe uma representação da base de dados dos pacientes para os sintomas.

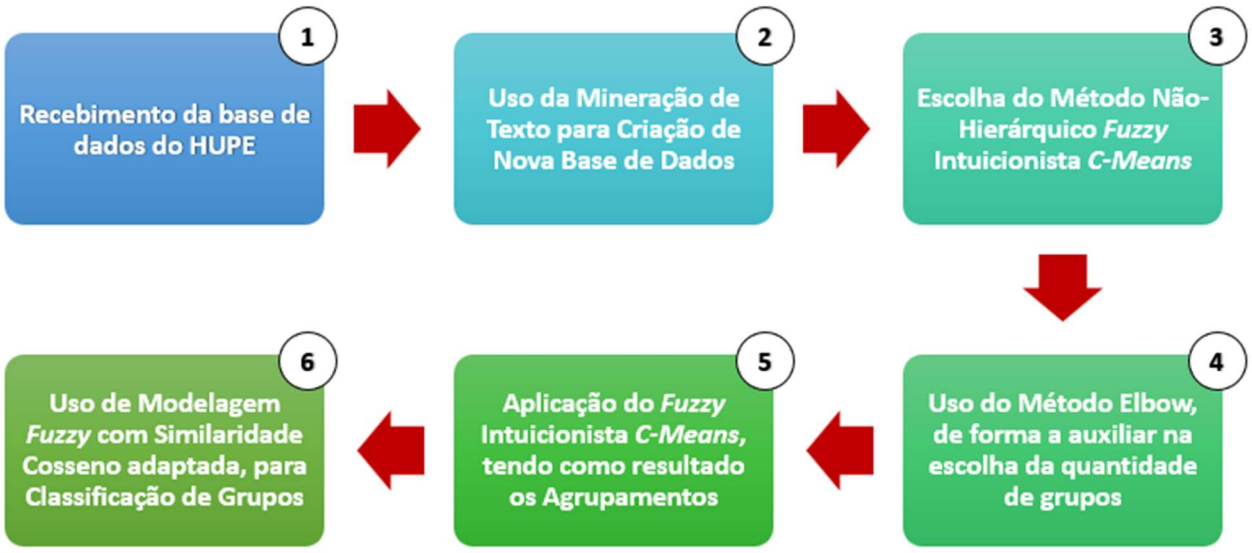
Tabela 9 - Representação da base de dados dos pacientes para sintomas

Pacientes (PAC) ***(i=1,2, ...,N)***	**Sintomas** ***(S=1,2...n)***					**Classe** ***(C)***
	S_1	S_2	S_3	...	S_n	
PAC001	a_{11}	a_{12}	a_{13}	...	a_{1n}	C
PAC002	a_{21}	a_{22}	a_{23}	...	a_{2n}	C
PAC003	a_{31}	a_{32}	a_{33}	⋱	a_{3n}	C
⋮	...	...	...	...	...	C
PAC00N	a_{m1}	a_{m2}	a_{m3}	...	a_{mn}	C

Fonte: A autora, 2024.

A Figura 12 exibe os passos explicitados no processo, sendo os resultados incluídos no Capítulo 3.

Figura 12 - Descrição do processo das técnicas aplicadas



Fonte: A autora, 2024.

3 **RESULTADOS**

3.1 **Análise descritiva**

Dos 282 registros da base ajustada, constam os pacientes curados ou os óbitos designados pela variável *OBITO_ALTA*, conforme apresentado na Tabela 10. Os óbitos correspondem à 31,2%, enquanto as altas obtiveram 68,1%.

Tabela 10 – Número de óbitos e altas na base de dados ajustada

Resposta	**Quantidade**	**Percentual**
Alta	194	68,8
Óbito	88	31,2
Total	**282**	**100,0**

Fonte: A autora, 2024.

Na Tabela 11 agregam-se as variáveis *GENERO* e *CLASS_ETARIA* (“Criança”, “Adolescente”, “Adulto”, “Idoso”), permitindo verificar que, dos 282 pacientes, 154 são idosos (54,6%) predominando o gênero masculino (60,39%).

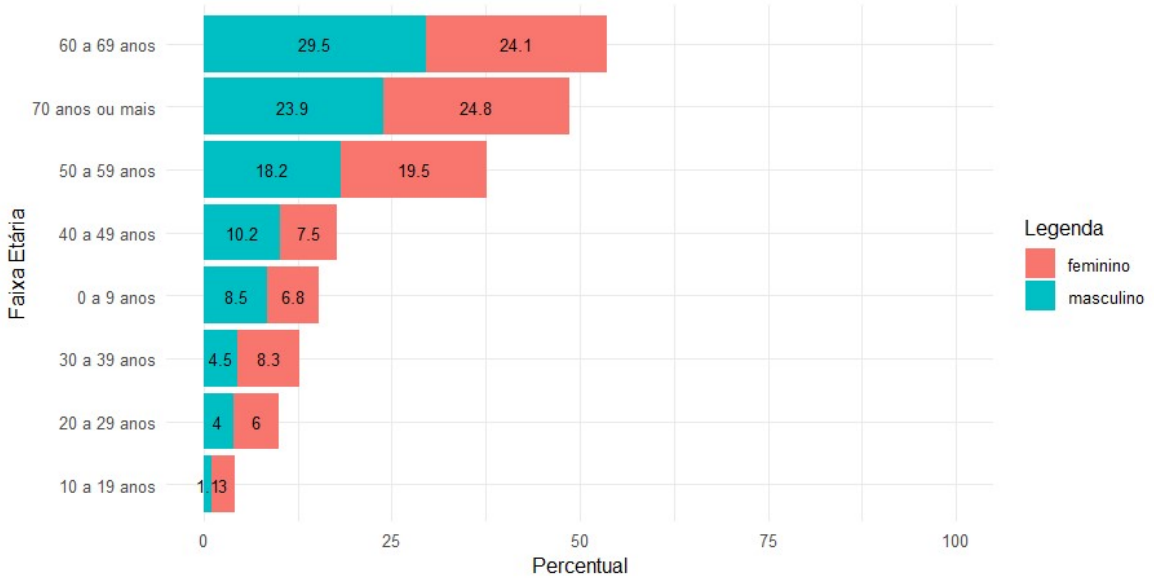
Tabela 11 - Pacientes por Grupo Etário e Gênero

Grupo etário	**Gênero**		**Total**
	Feminino	**Masculino**	
Adolescente	1	2	**3**
Adulto	46	56	**102**
Criança	8	15	**23**
Idoso	61	93	**154**
Total	**116**	**166**	**282**

Fonte: A autora, 2024.

O Gráfico 4, ilustra que as variáveis *GENERO* e *FAIXA_ETARIA,* mostraram que a maioria dos pacientes possuem 60 anos ou mais, tendo prevalência de 56,0% e 52,6% para os gêneros masculino e feminino, respectivamente.

Gráfico 4 - Distribuição relativa de pacientes por gênero e faixa etária



Fonte: A autora, 2024.

A Tabela 12 discrimina os 282 pacientes segundo sintomas. Os mais expressivos foram febre (95,1%), falta de ar (73,1%), cansaço (36,0%), tosse (29,8%), vômito ou náusea (28,2%) e diarreia (22,7%). Apesar do sintoma erupção na pele ter sido indicado pelo Ministério da Saúde, nos prontuários do HUPE não foi contemplado.

Tabela 12 - Quantitativo e percentual de pacientes segundo sintomas (continua)

Sintomas	Pacientes	% em relação ao total (n = 282)
Febre	268	95,0
Falta de ar	210	74,5
Cansaço	103	36,5
Tosse	89	31,6
Vômito ou náusea	74	26,2
Diarreia	65	23,0
Dor no corpo/muscular	44	15,6
Dor de cabeça	39	13,8
Perda, alteração de olfato ou paladar	34	12,1
Coriza	23	8,2
Falta de apetite	15	5,3
Dor de garganta	13	4,6

Tabela 12 - Quantitativo e percentual de pacientes segundo sintomas (conclusão)

Sintomas	Pacientes	% em relação ao total (n = 282)
Tontura ou vertigem	11	3,9
Calafrio	10	3,5
Espirro	4	1,4
Erupção na pele	0	0,0

Fonte: A autora, 2024.

Na Tabela 13 foram discriminadas, em ordem decrescente, as prevalências das comorbidades encontradas nos prontuários dos pacientes. Hipertensão (66,7%) e Doença cerebrovascular (58,9%) superam a metade dos pacientes, seguidas da Doença cardiovascular (41,5%), Imunocomprometido (16,7%), Obesidade (15,6%), Doença Pulmonar (14,5%), Diabetes (13,8%) e Doença Renal (1,1%).

Tabela 13 - Quantitativo e percentual de pacientes segundo comorbidades

Comorbidades	Pacientes	% em relação ao total (n = 282)
Hipertensão	188	66,7
Doença cerebrovascular	166	58,9
Doença cardiovascular	117	41,5
Imunocomprometido	47	16,7
Obesidade	44	15,6
Doença pulmonar	41	14,5
Diabetes	39	13,8
Doença renal	3	1,1

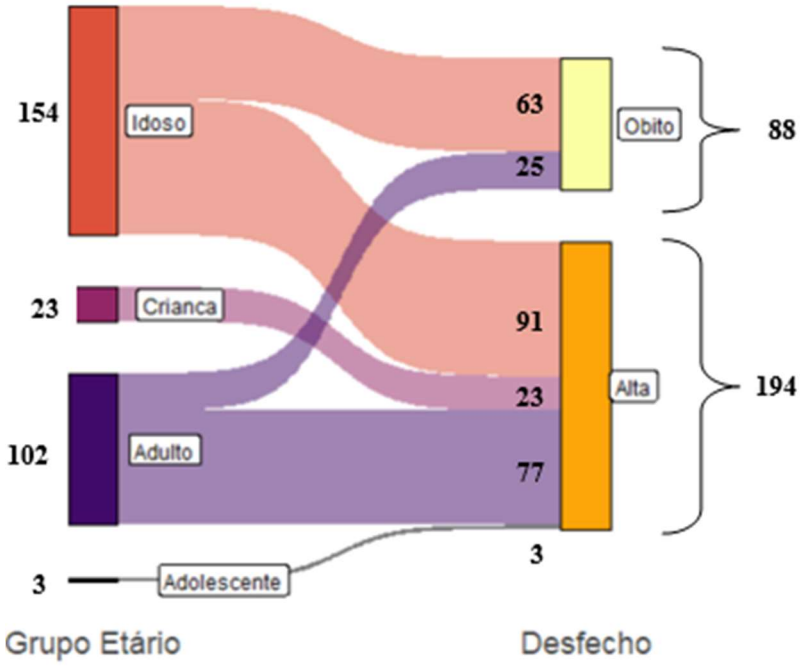
Fonte: A autora, 2024.

O Diagrama de Sankey, ilustrado Gráfico 5, foram utilizadas as variáveis *GRUPO_ETARIO* e *OBITO_ALTA*. Ele traz informações da variável que trata dos grupos etários e seus respectivos desfechos em óbito ou alta.

Dos 154 idosos e 102 adultos, 40,9% e 24,5%%, respectivamente, foram a óbito. Nenhuma criança e adolescente tiveram este desfecho. Em função deste resultado, optou-se por

retirar as crianças e adolescentes, logo, a nova base a ser trabalhada nas Seções 3.2 e 3.3 contará com 256 pacientes.

Gráfico 5 - Diagrama de Sankey de óbitos e altas segundo grupos etários



Fonte: A autora, 2024.

3.2 **Algoritmo de agrupamento *Fuzzy* Intuicionista *C-Means***

A aplicação do *Fuzzy* Intuicionista *C-Means* necessita que se estabeleça a quantidade de grupos a ser considerada ideal. Para tal, utilizou-se os estabelecidos nos artigos lidos para a composição desta tese: parâmetro de *fuzzyficação m* = 2; grau de Hesitação $\alpha = 0{,}85$; constante $\lambda = 2$; e erro[3] para critério de parada $\varepsilon = 0{,}01$.

Foi aplicado primeiramente o método Elbow para avaliar o número de *cluster* para dar início ao processo intuicionista *C-Means*, onde a abscissa corresponde ao número de *clusters* e a ordenada, à soma quadrática da diferença entre o elemento de cada grupo em relação ao "centro".

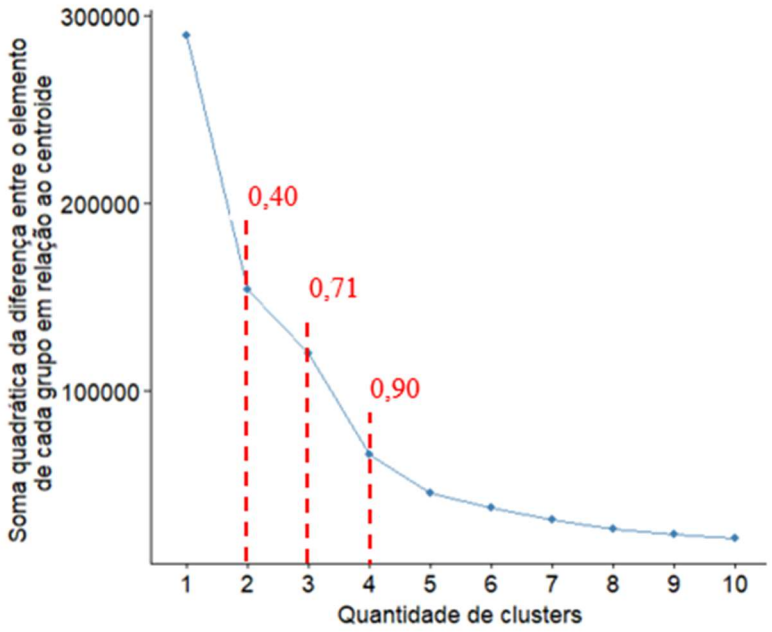
A entrada do sistema lógico *fuzzy* inicia-se conforme a matriz que agregou as variáveis *"Número de dias de internação"* e *"Idade"*. O Gráfico 6, consubstanciado no método de

[3] Retirado do artigo: "*Text classification using similarity measures on Intuitionistic Fuzzy*" de 2016.

Elbow, indica uma mudança de comportamento nos pontos entre dois e quatro, que corresponde à avaliação na minimização do número da grupos.

Esta repartição foi tratada segundo o coeficiente de aferição da estratificação pela razão entre a variância dentro dos estratos e a variância total para a variável *"Idade"*, pois esta informação é uma obrigatoriedade do prontuário hospitalar[4]. Os coeficientes de estratificação assumiram valores 0,90, 0,71 e 0,40 para as opções dois, três e quatro, optando-se pelo menor valor.

Gráfico 6 – Número de *clusters* considerada ideal, segundo método Elbow



Fonte: A autora, 2024

O método *Fuzzy* Intuicionista *C-Means* foi utilizado segundo a proposta de Sugeno e os valores da Função Objetivo segundo as interações do método *Fuzzy* Intuicionista estão na Tabela 14. Na matriz de entrada foram utilizadas as variáveis *"Número de dias de internação"* e *"Idade"*, pois atendiam o requisito de variáveis numéricas.

[4] Conselho Regional de Medicina do Distrito Federal, 2006.

Tabela 14 - Iterações e valores da função Objetivo

Iterações	Função Objetivo
Iteração 1	5,08
Iteração 2	0,62
Iteração 3	0,82
Iteração 4	0,97
Iteração 5	1,01
Iteração 6	0,90
Iteração 7	0,67
Iteração 8	0,42
Iteração 9	0,24
Iteração 10	0,13
Iteração 11	0,07
Iteração 12	0,04
Iteração 13	**0,02 (limiar)**

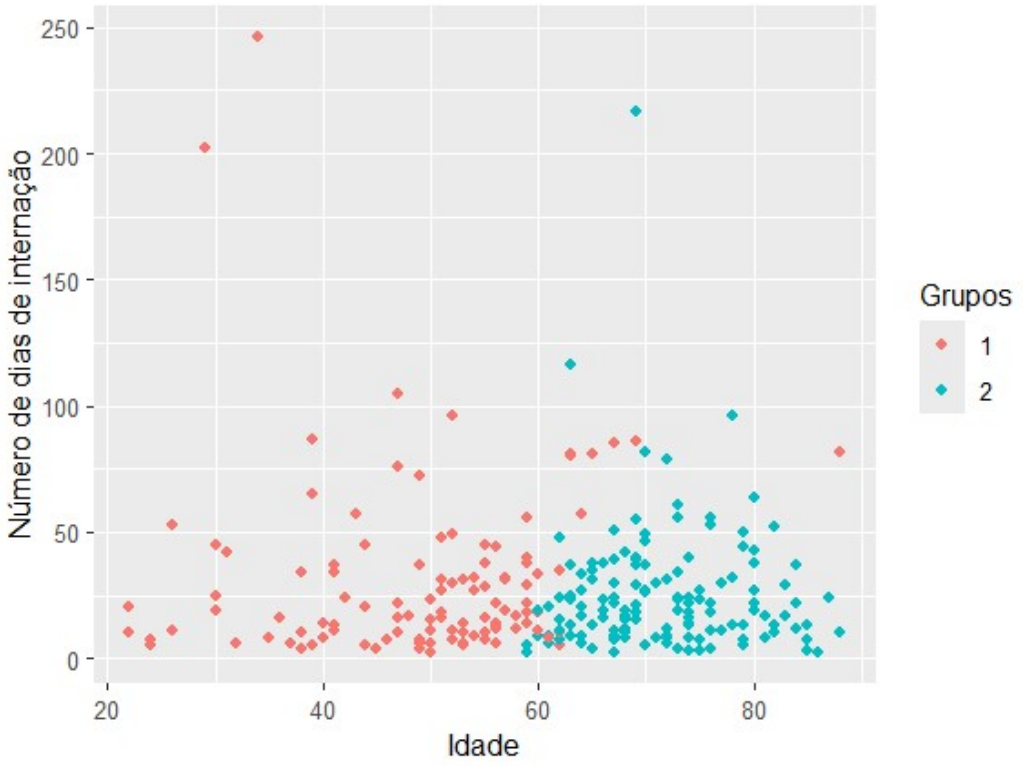
Fonte: A autora, 2024

Prosseguiu-se, então, estabelecendo-se as pertinências máximas e as respectivas hesitações referentes a cada um dos dois *clusters* (grupos). Cabe ressaltar que neste primeiro momento, foi observado o grupo integralmente, e não os pacientes do grupo isoladamente.

Os *clusters* 1 e 2 apresentam graus de hesitação 0,61 e as pertinências máximas de 0,99. Nos Apêndices E e F, constam as tabelas com os pacientes, os respectivos grupos que pertencem e os valores de pertinência, não-pertinência e hesitação.

Os resultados indicados pelo agrupamento *Fuzzy* Intuicionista *C-Means* permitiram a construção do Gráfico 7, por meio do qual pode ser visualizado o montante de pacientes classificados em dois grupos e aqueles que se destacam dos demais.

Gráfico 7 - *Clustering results*



Fonte: A autora, 2024.

Na Tabela 15 estão as variáveis sociodemográficas. O Grupo 1 apresenta mais pacientes que no Grupo 2, 145 e 111, respectivamente. No Grupo 1, a idade mediana corresponde a 70 anos, enquanto no Grupo 2, 51 anos.

Em relação ao *"Gênero"*, o sexo masculino predomina no Grupo 1, 149 pacientes (58,20%) contra 107 pacientes (41,79%) do sexo feminino. No que tange ao *"Tabagismo"*, o Grupo 1 apresentou a prevalência em maior destaque 21,87%.

Tabela 15 - Pacientes segundo idades, gênero e tabagismo por grupos

Grupos	**Pacientes**	**Idade**			**Gênero**		**Tabagismo**	
		Mín.	**Mediana**	**Máx.**	**M**	**F**	**Sim**	**Não**
1	145	34	70	88	87	58	44	101
2	111	22	51	88	62	49	12	99

Fonte: A autora, 2024.

As Tabelas 16 e 17 associam os pacientes segundo sintomas e comorbidades em relação aos Grupos, respectivamente, discriminados segundo ao *Fuzzy* Intuicionista *C-Means*. Em

seguida, partiu-se para a aplicação da modelagem *Fuzzy* Intuicionista com similaridade cosseno, Seção 3.3.

Tabela 16 - Pacientes segundo sintomas por grupos

Sintomas	Grupo 1	Grupo 2
Calafrio	6	4
Cansaço	**45**	**48**
Coriza	8	7
Diarreia	38	18
Dor de Cabeça	14	24
Dor de Garganta	6	7
Dor no Corpo/Muscular	18	25
Espirro	2	1
Erupção na Pele	0	0
Falta de Apetite	8	6
Falta de Ar	**107**	**87**
Febre	**147**	**106**
Perda/Alteração do Olfato ou Paladar	11	23
Tontura	6	5
Tosse	43	38
Vômito ou Náusea	42	24

Fonte: A Autora, 2024.

Tabela 17 - Pacientes segundo comorbidade por grupos

Comorbidade	Grupo 1	Grupo 2
Diabetes	22	15
Doença cardiovascular	**70**	**37**
Doença cerebrovascular	**97**	**55**
Doença pulmonar	19	14
Doença renal	1	2
Imunocomprometido	15	22
Hipertensão	**110**	**68**
Obesidade	20	23

Fonte: A autora, 2024.

3.3 **Aplicação da modelagem *Fuzzy* Intuicionista com similaridade cosseno**

A modelagem *Fuzzy* Intuicionista *C-Means* utilizou as variáveis *"Número de dias de internação"* e *"Idade"* e convergiu para dois clusters, ora designados por Classe Padrão 1 e Classe Padrão 2, ratificadas a homogeneidade segundo estas variáveis pelo Índice de estratificação segundo a razão entre a Variância Dentro e a Variância Total (Seção 2.2), referente a sintomas e comorbidades. Os resultados obtidos estão dispostos nas Subseções 3.3.1 e 3.3.2, a seguir.

3.3.1 Caso 1: sintomas

Na aplicação optou-se pelos três sintomas que apresentaram maior expressividade na Tabela 16: *"Febre"*, *"Falta de ar"* e *"Cansaço"*, para a classificação dos pacientes nas classes Classe Padrão 1 e Classe Padrão 2. Em seguida, foram calculadas as pertinências (P) e não-pertinências (NP) para cada paciente. A Tabela 18 apresenta as médias e os desvios-padrão (DP) para cada Classe Padrão.

Tabela 18 - Médias e desvios-padrão das pertinências e não-pertinências segundo sintomas e Classe Padrão

Média e DP	**Sintomas**						**Classe Padrão**
	Febre		**Falta de ar**		**Cansaço**		
	P	**NP**	**P**	**NP**	**P**	**NP**	
Média	0,49	0,41	0,47	0,44	0,42	0,48	C_1
DP	0,00	0,00	0,19	0,19	0,20	0,20	
Média	0,49	0,42	0,48	0,42	0,47	0,43	C_2
DP	0,00	0,00	0,18	0,18	0,21	0,21	

Fonte: A autora, 2024.

O cálculo da Similaridade Cosseno segundo Seção 2.2, objetiva alocar um paciente (PC) selecionado aleatoriamente em uma das classes nas características: *"Febre"*, *"Falta de ar"* e

"Cansaço", registradas segundo a escala categórica. Os valores foram padronizados e, em seguida, foram calculados os valores de pertinência e não-pertinência, conforme mostra a Tabela 19.

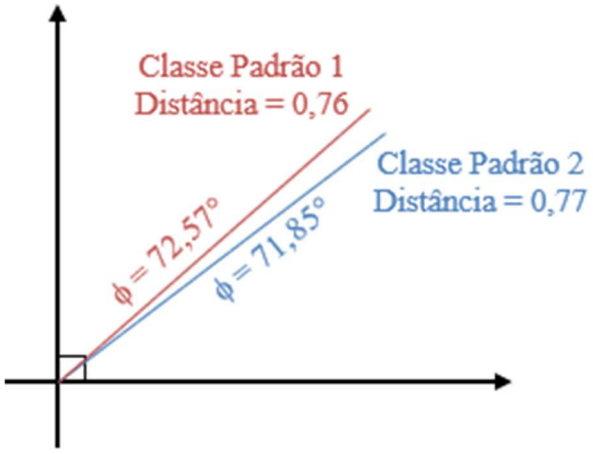
Tabela 19 - Pertinências e não-pertinências do novo paciente segundo sintomas

Paciente	Febre	Falta de ar	Cansaço
Pertinência	0,01	0,13	0,71
Não-pertinência	0,89	0,77	0,19

Fonte: A autora, 2024.

O cálculo da Similaridade Cosseno para o paciente apresentou valores próximos de 0,76 e ângulo $\phi = 72{,}57°$ para a Classe Padrão 1, e de 0,77 e ângulo $\phi = 71{,}85°$ para a Classe Padrão 2. Isso significa que o trio de sintomas conduz o indivíduo para a Classe Padrão 1, uma vez que a distância aferida pela similaridade cosseno tem menor valor. O Gráfico 8 ilustra o resultado.

Gráfico 8 - Distância aferida pela similaridade cosseno para sintomas segundo Classes Padrões



Fonte: A autora, 2024.

3.3.2 Caso 2: comorbidades

"Hipertensão", *"Doença cerebrovascular"* e *"Doença cardiovascular"* foram as comorbidades observadas na Tabela 17 para os pacientes que apresentaram maiores incidências segundo as classes padrões.

De forma análoga à Subseção 3.3.1, foram calculadas as pertinências (P) e não-pertinências (NP), sendo os resultados colocados no Apêndice D, que permitiram obter as médias e desvio-padrão (DP) para pertinências e não-pertinência por Classe Padrão (CP), mostrados na Tabela 20.

Tabela 20 - Médias e desvios-padrão das pertinências e não-pertinências do novo paciente segundo comorbidades e Classe Padrão

Média e DP	**Comorbidades**						**Classe Padrão**
	Hipertensão		**Cerebrovascular**		**Cardiovascular**		
	P	**NP**	**P**	**NP**	**P**	**NP**	
Média	0,00	0,41	0,13	0,47	0,71	0,41	C_1
DP	0,00	0,00	0,21	0,21	0,21	0,21	
Média	0,89	0,41	0,77	0,55	0,19	0,47	C_2
DP	0,00	0,00	0,22	0,22	0,20	0,20	

Fonte: A autora, 2024.

O novo paciente foi novamente solicitado para as características: *"Hipertensão"*, *"Doença cerebrovascular"* e *"Doença cardiovascular"*, registradas segundo a escala categórica. Os valores foram padronizados e, em seguida, foram calculados os valores de pertinência e não-pertinência, conforme mostra a Tabela 21.

Tabela 21 - Pertinências e não-pertinências do novo paciente segundo comorbidades

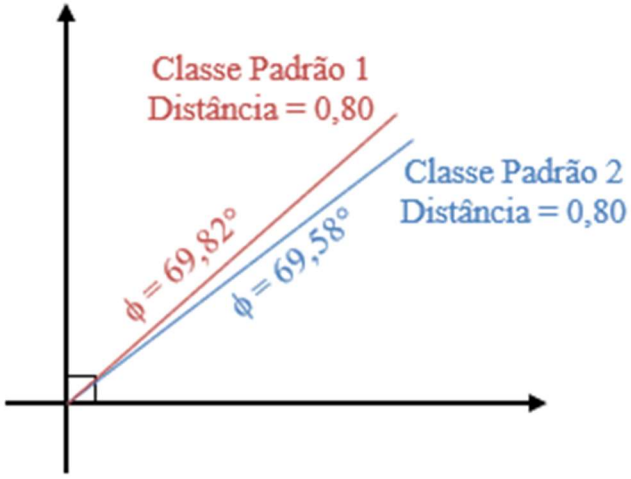
Paciente	**Hipertensão**	**Doença cerebrovascular**	**Doença cardiovascular**
Pertinência	0,00	0,13	0,71
Não-Pertinência	0,90	0,77	0,19

Fonte: A autora, 2024.

O cálculo da similaridade cosseno referente as comorbidades apresentou valores próximos de 0,80 e ângulo ϕ = 69,82° para a Classe Padrão 1 e 0,80 e ângulo ϕ = 69,58° para a Classe Padrão 2.

Isso indica que o trio de comorbidades o indivíduo deve ser classificado na Classe Padrão 1, uma vez que a projeção do seu fator referente tem menor variabilidade. O Gráfico 9, exibe o resultado.

Gráfico 9 - Distância Aferida pela similaridade cosseno para comorbidades segundo Classes Padrões



Fonte: A autora, 2024.

CONCLUSÃO

A análise do Estado da Arte mostrou que o método *Fuzzy* Intuicionista ainda não está sendo utilizado em estudos que tratam do reconhecimento de padrões com o uso de agrupamentos nas diversas áreas do conhecimento. Esta proposta oportuniza a divulgação de uma alternativa que pode ser utilizada por pesquisadores das áreas da Computação e da Matemática Aplicada com a prefixação de um índice que afere a hesitação na modelagem *Fuzzy*.

A proposta da tese foi aplicar a modelagem Intuicionista *Fuzzy* segundo a similaridade cosseno em dados clínicos, mas especificamente, em dados dos pacientes internados com Covid-19 no HUPE no período de dezembro de 2019 até setembro de 2021. O objetivo é propor a modelagem como classificação em "Classes Padrão", de acordo com características previamente estabelecidas.

Com isso, este estudo tem como contribuição a análise o possível reconhecimento de padrões que possam contribuir para entender o comportamento dos grupos, de acordo com características não só estabelecidas dos dados já existentes, mas também de futuros dados a serem incluídos. A denominação das classes ficaria a critério de especialistas, em estudos futuros.

A análise ocorre a partir de dados primários, oriundos a partir dos registros de prontuários dos pacientes. Essa característica torna este estudo uma contribuição inédita e com certo grau de complexidade de reprodução, pelo acesso limitado aos dados, que ocorreu via aprovação pelo Comitê de Ética, CAAE: 3592920.2.0000.5282, obtidas por meio do projeto em parceria com a Telemedicina da UERJ, edital n° 12/2020, bolsa CAPES-EPIDEMIAS, assinatura do termo de sigilo e privacidade, além de respeitar as regras da Lei Geral de Proteção de Dados (LGPD).

Os resultados mostraram-se relevantes, pois classificou o "novo paciente" na mesma "Classe Padrão" em ambos os casos propostos: sintomas e comorbidades. Ou seja, houve convergência do resultado, o que mostra uma possível capacidade da modelagem apresentada unir informações semelhantes em um mesmo grupo de acordo com características apresentadas, de forma a facilitar novas análises.

Além disso, confirmam que um grupo de pacientes pode ser enquadrado em determinada classe considerada padrão mediante a estimativa das pertinências e não-pertinências prefixando

um valor para a hesitação no sistema lógico Intuicionista *Fuzzy,* mesmo levando em consideração a relevância da similaridade. Apesar da opção ter sido aplicada a duas Classe Padrões, o uso para mais classes ainda se torna válido.

Inicialmente, não foram detectadas limitações na modelagem. No entanto, o pesquisador, pode recorrer a alguma técnica de validação se assim desejar, de forma a reafirmar os resultados da alocação do novo paciente. Cabe ressaltar que a aplicação da validação não é obrigatória.

A modelagem foi aplicada na área Biomédica tendo sido usado na identificação de sintomas e comorbidades da Covid-19 em uma Unidade de Saúde Pública, mas poderia ser aplicada às demais áreas do conhecimento, visto seu caráter multifacetado.

REFERÊNCIAS

ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. *RESI – Revista Eletrônica de Sistemas de Informação*, v. 5, n. 2, 2006. Disponível em: https://www.periodicosibepes.org.br/index.php/reinfo/article/viewFile/171/66. Acesso em: 10 nov. 2024.

ATANASSOV, K. T. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, v. 20, n. 1, p. 87-96, 1986. Disponível em: https://www.sciencedirect.com/science/article/pii/S0165011486800343. Acesso em: 04 ago. 2023. DOI: https://doi.org/10.1016/S0165-0114(86)80034-3.

BARION, E. C. N.; LAGO, D. Mineração de textos. *Revista de Ciências Exatas e Tecnologia,* v. 3, n. 3, p. 123-140, 2015. Disponível em: https://exatastecnologias.pgsskroton.com.br/article/view/2372. Acesso em: 10 jan. 2024. DOI: https://doi.org/10.17921/1890-1793.2008v3n3p123-140.

BAUD, R. H.; *et al*. Reconciliation of ontology and terminology to cope with linguistics. *Stud Health Technol Inform*, n. 129, pt. 1, p. 796-801, 2007. Disponível em: https://www.researchgate.net/publication/5932858_Reconciliation_of_ontology_and_terminology_to_cope_with_linguistics. Acesso em: 03 mar. 2024.

BELLMAN, R. E.; ZADEH, L. A. Decision-Making in a fuzzy environment. *Management Science*, v. 17, n. 4, p. 141-164, dez. 1970. Disponível em: http://www.dca.fee.unicamp.br/~gomide/courses/CT820/artigos/DecisionMakingFuzzyEnvironmentBellmanZadeh1970.pdf. Acesso em: 05 out. 2023. DOI: https://doi.org/10.1287/mnsc.17.4.B141.

BENOIT, K.; MUHR, D.; WATANABE, K. *Stopwords:* multilingual stopword lists. R package version 2. 3, 2021. Disponível em: https://cran.r-project.org/web/packages/stopwords/stopwords.pdf. Acesso em: 21 fev. 2024.

BLACK, M. Vagueness: an exercise in logical analysis. *Philosophy of Science*, v. 4, n. 4, p. 427-455, 1937. Disponível em: http://www.jstor.org/stable/184414. Acesso em: 11 out. 2023. DOI: https://doi.org/10.1086/286476.

BOOLE, G. *The mathematical analysis of logic:* being an essay towards a calculus of deductive reasoning. London: Gutemberg Project, 2011. Disponível em: https://www.gutenberg.org/files/36884/36884-pdf.pdf. Acesso em: 25 out. 2023.

BOOLE, G. *The laws of thought*. London: Gutemberg Project, 2017. Disponível em: https://www.gutenberg.org/files/15114/15114-pdf. Acesso em: 25 out. 2023.

BOTELHO, T. G.; SANTOS, O. R.; SOUZA, S. M. Implementação computacional rigorosa do princípio de extensão de Zadeh. *In*: II Congresso Brasileiro de Sistemas *Fuzzy*. *Anais eletrônicos[...]*, v. 1, p. 490-503, 2012. Disponível em: https://dimap.ufrn.br/~cbsf/pub/anais/2012/10000490.pdf. Acesso em: 03 nov. 2023.

BRAGA, M. J. F.; BARRETO, J. M.; MACHADO, M. A. S. *Conceitos da matemática nebulosa na análise de risco*. Rio de Janeiro: Editora Artes & Rabiscus, 1995.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Política do acervo de recursos educacionais em saúde*. Brasília: UNA-SUS, 2ª ed., 2013. Disponível em: https://ares.unasus.gov.br/acervo/bitstream/ARES/3597/1/POL%C3%8DTICA_DO_ARES_18-07-16.pdf. Acesso em: 10 nov. 2024.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Coronavírus Brasil*. Brasília: Ministério da Saúde [Brasil, 2020]. Disponível em: https://covid.saude.gov.br/. Acesso em: 10 mar. 2024.

BRASIL. Ministério da Saúde. *Plano Nacional de Operacionalização contra a Covid-19*, Lista de Comorbidades, p. 27, 2022. Disponível em: https://www.gov.br/saude/pt-br/assuntos/covid-19/publicacoes-tecnicas/guias-e-planos/plano-nacional-de-operacionalizacao-da-vacinacao-contra-covid-19.pdf. Acesso em: 06 ago. 2024.

BULEGON, H.; MORO, C. M. C. Mineração de texto e o processamento de linguagem natural em sumários de alta hospitalar. *Journal of Health Informatics*, v. 2, n. 2, 2010. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/5. Acesso em: 10 jan. 2024.

BUNEMAN, P. Semistructured data. *In:* Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 1997. *Proceedings* [...], p. 117-121, mai. 1995. Disponível em: https://dl.acm.org/citation.cfm?id=263675. Acesso em: 02 jan. 2024.

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. F. Introdução à análise de agrupamentos. *ln: IX Simpósio Brasileiro de Probabilidade e Estatística*, IME-USP, São Paulo, 105 p. 1990.

BUSTINCE, H.; BURILLO, P. Vague sets are intuitionistic sets. *Fuzzy Sets and Systems*, v. 79, n. 3, p. 403-405, 1996. Acesso em: 25 out. 2023. DOI: https://doi.org/10.1016/0165-0114(95)00154-9.

CÂMARA, L. A respeito da construção de estratos. *Revista Brasileira de Estatística*, ano 27, n. 107, p. 152-172, jul./set. 1966.

CAPUTO, G. M.; BASTOS, V. M.; EBECKEN, N. F. F. Using text mining to understand the call center customers' claims. *In:* Data Mining VII: Data, Text and Web Mining and their Business Applications. *Proceedings* [...], v. 37, p. 177-184, 2006. Disponível em: https://www.witpress.com/elibrary/wit-transactions-on-information-and-communication-technologies/37/16715. Acesso em: 11 jan. 2024. DOI: https://www.doi.org/10.2495/DATA060181.

CARRILHO JUNIOR, J. R. *Desenvolvimento de uma Metodologia para Mineração de Textos.* 2007. 96 f. Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, 2007. Disponível em: https://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=11675@1. Acesso em: 10 jan. 2024. DOI: https://doi.org/10.17771/PUCRio.acad.11675.

CEBECI, Z.; CEBECI, C. *Inaparc:* initialization algorithms for partitioning cluster analysis. R package version 3.3.0., 2022. Disponível em: https://cran.r-project.org/web/packages/inaparc/inaparc.pdf. Acesso em: 02 fev. 2024.

CHAIRA, T. A novel intuitionistic fuzzy c-means clustering algorithm and its application to medical images. *Applied Soft Computing*, v. 11, n. 2, p. 1711-1717, 2011. Acesso em: 23 jan. 2024. DOI: https://doi.org/10.1016/j.asoc.2010.05.005.

CHAIRA, T. *Fuzzy set and its extension:* the intuitionistic fuzzy set. USA: John Wiley & Sons, Inc., 1ª ed., 2019.

CHAKRABARTI, S. *Mining the web:* discovering knowledge from hypertext data. California: Elsevier Sciences, 2003.

CHEN, H. *Knowledge management systems:* a text mining perspective. Arizona: Knowledge Computing Corporation, 2001. Disponível em: http://hdl.handle.net/10150/106481. Acesso em: 10 set. 2023.

CONSELHO REGIONAL DE MEDICINA DO DISTRITO FEDERAL. *Prontuário médico do paciente* - guia do uso prático, 2006. Disponível em: https://www.saudedireta.com.br/docsupload/1370271458PEP.pdf. Acesso em: 13 set. 2024.

CORREIA, P. R. M.; FERREIRA, M. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. *Química Nova*, v. 30, p. 481-487, 2007. Disponível em:

https://www.scielo.br/j/qn/a/RnMHKDXFtzCGS56kkTGyFKQ/. Acesso em: 11 nov. 2024. DOI: https://doi.org/10.1590/S0100-40422007000200042.

COSSA, G. S.; *et al.* Medidas de enfrentamento à pandemia da Covid-19 e influência dos sistemas de saúde: uma análise comparativa entre Brasil, Itália e EUA. *Revista O Mundo da Saúde*, v. 45, p. 379-389, 2021. Disponível em: https://bvsms.saude.gov.br/bvs/periodicos/mundo_saude_artigos/medidas_pandemiacovid19_brasilitaliaeua.pdf. Acesso em: 02 jan. 2024. DOI: https://doi.org/10.15343/0104-7809.202145379389.

COURY, A. G. F.; VILELA, D. S. Estudo histórico do paradoxo de Russell: a fecundidade de uma matemática falível. *Acta Latino-americana de Matemática Educativa*, v. 32, n. 2, p. 544-553, Cidade do México, 2019. Disponível em: http://funes.uniandes.edu.co/14073/1/Fonseca2019Estudo.pdf. Acesso em: 10 out. 2023.

DA COSTA, C. G. *Probabilidades Imprecisas:* Intervalar, *Fuzzy* e *Fuzzy* Intuicionista. 2012. 159 f. Tese (Doutorado em Engenharia Elétrica e de Computação) – Universidade Federal do Rio Grande do Norte, Natal, 2012. Disponível em: https://repositorio.ufrn.br/handle/123456789/15202. Acesso em: 06 out. 2023.

DA SILVA, L. M.; *et al*. Estado da arte dos fundamentos e ideias da lógica *fuzzy* aplicada as ciências e tecnologia. *Revista Brasileira de Geomática*, v. 7, n. 3, p. 149-169, jul./set. 2019. Disponível em: https://periodicos.utfpr.edu.br/rbgeo/article/view/9365. Acesso em 07 out. 2023. DOI: https://doi.org/10.3895/rbgeo.v7n3.9365.

DA SILVA, T. D.; DOMINGUES, P. M. Inteligência Artificial e Privacidade: Os desafios do Privacy by Design. *Advances of Knowlegde Representation*, v. 1, n. 2, p. 160-194, 2024. Disponível em: https://periodicos.ufmg.br/index.php/fronteiras-rc/article/download/52813/44775/199635. Acesso em: 10 nov. 2024. DOI: https://www.doi.org/10.5281/zenodo.13152354.

DAGLIATI, A.; *et al*. Health informatics and EHR to support clinical research in the Covid-19 pandemic: an overview. *Briefings in Bioinformatics*, v. 22, n. 2, p. 812-822, 2021. Disponível em: https://academic.oup.com/bib/article/22/2/812/6103007. Acesso em: 05 ago. 2024. DOI: https://doi.org/10.1093/bib/bbaa418.

DE FRANÇA, G. V. A quem pertence ao prontuário? *GENJURÍDICO*, 2017. Disponível em: Acesso em: https://blog.grupogen.com.br/juridico/postagens/artigos/quem-pertence-o-prontuario/. 05 ago. 2024.

DE PONTES SARAIVA, G. J. Lógica *fuzzy*. *Revista da Escola Superior de Guerra*, v. 21,

n. 45, Rio de Janeiro, 2006. Disponível em: https://revista.esg.br/index.php/revistadaesg/article/view/335. Acesso em: 09 out. 2023. DOI: https://doi.org/10.47240/revistadaesg.v21i45.335.

DIXON, M. *An overview of document mining technology*, 1997. Disponível em: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=581fc9ddf8457f1e6cb9ccb76c8ad79d8e9c22d0. Acesso em: 13 jan. 2024.

EBECKEN, N; LOPES, M; COSTA, M. Mineração de Textos, chapter 13, p. 337–370. *In: Sistemas inteligentes:* fundamentos e aplicações. Barueri: Manole, 2005.

EGRIOGLU, E.; BAS, E.; KIZILASLAN, B. *rinfis:* Intuitionistic fuzzy inference systems. R package version 0.0.0.9000, 2023. Disponível em: https://github.com/busenurk/rinfis/blob/main/README.md. Acesso em: 10 fev. 2024.

EVERITT, B. S.; *et al*. *Cluster analysis*. UK: John Wiley & Sons Ltd., 5 ed., 2011.

FACELI, K.; CARVALHO, A. C. P. L. F. de; SOUTO, M. C. P. de. Algoritmos de Agrupamento de Dados. *Relatório Técnico*. Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, 2005. Disponível em: https://web.icmc.usp.br/SCATUSU/RT/Relatorios_Tecnicos/RT_249.pdf. Acesso em: 11 nov. 2024.

FALBEL, D. *rslp:* A stemming algorithm for the Portuguese Language. R package version 0.2.0, 2020. Disponível em: https://cran.r-project.org/web/packages/rslp/rslp.pdf. Acesso em: 21 fev. 2024. DOI: https://doi.org/10.1109/SPIRE.2001.10024.

FALEIROS JÚNIOR, J. L. M.; NOGAROLI, R.; CAVET, C. A. Telemedicina e proteção de dados: reflexões sobre a pandemia da Covid-19 e os impactos jurídicos da tecnologia aplicada à saúde. *Revista dos Tribunais*, n. 1016, jun. 2020. Disponível em: https://dspace.almg.gov.br/handle/11037/37577. Acesso em: 05 ago. 2024.

FANTINELLI, A. A. *Sistema de classificação de pacientes*. Blog LEVi. Disponível em: https://www.ufrgs.br/levi/classificacao-de-pacientes/. Acesso em: 06 nov. 2024.

FÁVERO, L. P.; BELFIORE, P. *Manual de análise de dados:* estatística e modelagem multivariada em Excel, SPSS e Stata. Rio de Janeiro: Elsevier, 1 ed., 2017.

FAYYAD, F.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowlegde discovery. *AI Mazagine*, v. 17, n. 3, p. 37-54, 1996. Disponível em:

https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230. Acesso em: 11 nov. 2024. DOI: https://doi.org/10.1609/aimag.v17i3.1230.

FEINERER I.; HORNIK, K. *tm:* text mining package. R package version 0.7-11, 2023. Disponível em: https://cran.r-project.org/web/packages/tm/tm.pdf. Acesso em: 21 fev. 2024.

FELDMAN, R.; SANGER, J. *The text mining handbook* – advanced approaches in analyzing unstructured data. Cambridge University Press, 2007. DOI: https://doi.org/10.1017/CBO9780511546914.

FERNANDES, G. S.; *et al*. Avaliação da qualidade de prontuários médicos de uma Unidade Básica de Saúde: Desafio para caracterização do perfil epidemiológico dos usuários atendidos. *Revista Médica de Minas Gerais*, v. 29, n. e-2032, 2019. Disponível em: https://www.rmmg.org/artigo/detalhes/2551. Acesso em: 16 nov. 2024. DOI: http://dx.doi.org/10.5935/2238-3182.20190023.

FERREIRA, R. N. Análise *fuzzy cluster* em ambiente R: uma aplicação dos algoritmos FANNY, *C-means* e *C-medoids* na aplicação dos municípios brasileiros segundo indicadores de bem-estar social. *Cadernos do Leste*, v. 15, n. 15, 2015. Disponível em: https://periodicos.ufmg.br/index.php/caderleste/article/view/13061. Acesso em: 08 nov. 2023. DOI: https://doi.org/10.29327/249218.15.15-4.

FERREIRA, D. G. *Análise de dados em linguagem R*, 2020. Disponível em: http://repositorio.enap.gov.br/handle/1/7872. Acesso em: 05 ago. 2024.

FOX, C. J. Lexical analysis and stoplists. *In:* FRAKES, W. B.; BAEZA-YATES, R. A. (Eds.). *Information Retrieval:* Data Structures & Algorithms. New Jersey: Prentice Hall PTR, p. 102-130, 1992.

FRANÇOIS, R.; *et al. dplyr:* A grammar of data manipulation. R package version 1.1.4, 2023. Disponível em: https://cran.r-project.org/web/packages/dplyr/dplyr.pdf. Acesso em: 22 fev. 2024.

FUNDAÇÃO OSWALDO CRUZ. *OMS anuncia novas nomenclaturas para variantes do Sars-CoV-2*. Fundação Oswaldo Cruz [Brasil: Bio-Manguinhos, 2021]. Disponível em: https://www.bio.fiocruz.br/index.php/br/noticias/2421-oms-anuncia-novas-nomenclaturas-para-variantes-do-sars-cov-2. Acesso em: 05 jan. 2024.

FUNDAÇÃO OSWALDO CRUZ. *Pesquisadores desenvolvem técnica de mineração de texto para análise da Covid longa*. Fundação Oswaldo Cruz [Brasil: Fiocruz, outubro de 2024].

Disponível em: https://agencia.fiocruz.br/pesquisadores-desenvolvem-tecnica-de-mineracao-de-texto-para-analise-da-covid-longa. Acesso em: 10 nov. 2024.

FUGULIN, F. M. T.; GAIDZINSKI, R. R.; KURCGANT, P. Sistema de classificação de pacientes: identificação do perfil assistencial dos pacientes das unidades de internação do HU-USP. *Revista Latino-Americana de Enfermagem*, v. 13, n. 1, p. 72-78, 2005. Disponível em: Acesso em: https://www.scielo.br/j/rlae/a/C5p9kcnnFkxV3Cm3JJfVQjx/?format=pdf&lang=pt. 06 nov. 2024. DOI: https://doi.org/10.1590/S0104-11692005000100012.

GAGOLEWSKI, M. stringi: Fast and Portable Character String Processing in R. *Journal of Statistical Software*, v. 103, n. 2, p. 1-59, 2022. Disponível em: https://www.jstatsoft.org/article/view/v103i02. Acesso em: 22 fev. 2024. DOI: https://doi.org/10.18637/jss.v103.i02.

GATH, I.; GEVA, A. B. Unsupervised Optimal Fuzzy Clustering. *IEEE Transactions Pattern Analysis and Machine Intelligence,* v. 11, n. 7, p. 773-780, jul. 1989. Acesso em: 26 jan. 2024. DOI: https://doi.org/10.1109/34.192473.

GROLEMUND, G.; WICKHAM, H. Dates and times made easy with Lubridate. *Journal of Statistical Software*, v. 40, n. 3, p. 1-25, 2011. Disponível em: https://www.jstatsoft.org/v40/i03/. Acesso em: 22 fev. 2024. DOI: https://doi.org/10.18637/jss.v040.i03.

GUZZETTI, B. J. (Ed.). *Literacy in america:* an encyclopedia of history, theory and practice [2 volumes]. USA: Bloomsbury Publishing USA, 2002.

HAND, D.; MANNILA, H.; SMYTH, P. *Principles of data mining*. EUA: MIT Press, 2001.

HARMAN, D. How effective is suffixing? *Journal of the American Society for Information Science*, v. 42, n. 1, p. 7-15, 1991. Acesso em: 15 jan. 2024. DOI: https://doi.org/10.1002/(SICI)1097-4571(199101)42:13.0.CO;2-P.

HONRADO, A.; *et al*. A word stemming algorithm for the Spanish Language. *In:* Seventh International Symposium on String Processing and Information Retrieval SPIRE 2000, Spain. *Proceedings* [...]**,** p. 139-145, set. 2000. Acesso em: 17 jan. 2023. DOI: http://doi.org/10.1109/SPIRE.2000.878189.

HORNIK, K. *NLP:* natural language processing infrastructure. R package version 0.2-1, 2020. Disponível em: https://cran.r-project.org/web/packages/NLP/NLP.pdf. Acesso em: 21 fev. 2024.

HUANG, Z. Clustering large data sets with mixed numeric and categorical values. *In:* In The First Pacific-Asia Conference on Knowledge Discovery and Data Mining, 1997. *Proceedings* […], p. 21-34, 1997. Disponível em: https://grid.cs.gsu.edu/~wkim/index_files/papers/kprototype.pdf. Acesso em: 20 jan. 2024.

INTARAPAIBOON, P. Text classification using similarity measures on Intuitionistic Fuzzy Sets. *ScienceAsia*, n. 42, p. 52-60, 2016. Disponível em: https://www.scienceasia.org/2016.42.n1/scias42_52.pdf. DOI: https://doi.org/10.2306/scienceasia1513-1874.2016.42.052.

INSTITUTO BUTANTAN. *Conheça os sintomas mais comuns da ômicron e de outras variantes da Covid-19* [São Paulo: Butantan, 2021]. Disponível em: https://butantan.gov.br/noticias/conheca-os-sintomas-mais-comuns-da-omicron-e-de-outras-variantes-da-covid-19. Acesso em: 05 jan. 2024.

JOHANSEN, T. A.; SHORTEN, R.; MURRAY-SMITH, R. On the interpretation and identification of dynamic Takagi-Sugeno models. *IEEE Transactions on Fuzzy Systems*, v. 8, n. 3, p. 297-313, 2000. Disponível em: https://ieeexplore.ieee.org/document/855918. Acesso em: 11 nov. 2024. DOI: https://doi.org/10.1109/91.855918.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Practice Hall, 6ª ed., 2007.

KASSAMBARA, A.; MUNDT, F. *Factoextra:* extract and visualize the results of multivariate data analyses. R package version 3.1.2, 2022. Disponível em: https://cran.r-project.org/web/packages/factoextra/factoextra.pdf. Acesso em: 04 fev. 2024.

KOCHEN, M. *Principles of Information Retrieval*. New York: John Wiley & Sons, 1974.

LEAL, L. M. N. *Text Mining*, 2003. Relatório Técnico. Disponível em: https://www.dei.isep.ipp.pt/~paf/proj/Set2003/Text%20Mining.pdf. Acesso em: 10 jan. 2024.

LI, G.; *et al.* EASE: An Effective 3-in-1 Keyword Search Method for Unstructured, Semi-Structured and Structured Data. *In:* 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, Canadá. *Proceedings* […], p. 903-914, jun. 2008. Disponível em: https://dl.acm.org/citation.cfm?id=1376616.1376706. Acesso em: 10 out. 2023.

LIMA, E. L. *Espaços Métricos.* Rio de Janeiro: IMPA, Projeto Euclides, 1ª ed., 2014.

LIMIRO, R. M.; SILVA, N. R.; CORDEIRO, D. F. Mineração de Textos para Agrupamento de Teses e Dissertações por meio de Análise de Similaridade. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 18, p. 1-20, 2022. Disponível em: https://rbbd.febab.org.br/rbbd/article/view/1736. Acesso em: 10 jan. 2024.

LOPES, M. C. S. *Mineração de Dados Textuais Utilizando Técnicas de Clustering para o Idioma Português.* 2004. 191 f. Tese (Doutorado Ciências em Engenharia Civil) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004. Disponível em: www.coc.ufrj.br/pt/teses-de-doutorado/148-2004/980-maria-celia-santos-lopes. Acesso em: 14 jan. 2024.

LOPES, A. *Portuguese Stop Words*. Página do repositório do Github, 2013. Disponível em: https://gist.github.com/alopes/5358189. Acesso em: 03 nov. 2023.

LORENA, A. C.; CARVALHO, A. C. P. L. F. *Introdução às Máquinas de Vetores Suporte,* São Carlos, n. 192, abr. 2003. Relatório Técnico. Disponível em: http://www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_192.pdf. Acesso em: 02 out. 2023.

LOVINS, J. B. Development of a Stemming Algorithm. *Mechanical Translation and Computational Linguistics*, v. 11, n. 2, p. 22-31, mar./jun. 1968. Disponível em: https://aclanthology.org/www.mt-archive.info/MT-1968-Lovins.pdf. Acesso em: 11 jan. 2024.

LUHN, H. P. The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, v. 2, n. 2, p. 159-165, abr. 1958. Acesso em: 16 jan. 2024. DOI: https://doi.org/ 10.1147/rd.22.0159.

LUKASIEWICZ, J. *Aristotle's Syllogistic:* from The Standpoint of Modern Formal Logic, 1957. DOI: https://doi.org/10.2307/628276.

MAECHLER, M.; *et al*. *cluster:* Finding Groups in Data - Cluster Analysis Extended Rousseeuw *et al*. R package version 2.1.6, 2023. Disponível em: https://cran.r-project.org/web/packages/cluster/cluster.pdf. Acesso em: 01 fev. 2024.

MAMDANI, E. H.; ASSILIAN, S. An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *International Journal of Man-Machine Studies*, v. 7, p. 1-13, 1975. Disponível em: https://www.sciencedirect.com/science/article/pii/S0020737375800022. Acesso em: 24 set. 2023. DOI: https://doi.org/10.1016/S0020-7373(75)80002-2.

MAMDANI, E. H. Applications of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Transations on Computers*, v. 126, n. 12, p. 1182-1191, 1977. Disponível em: https://ieeexplore.ieee.org/document/1674779. Acesso em: 10 nov. 2023. DOI: https://doi.org/10.1109/TC.1977.1674779.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *An introduction to information retrieval.* Cambridge: Cambridge University Press, Online Edition, 2009. Disponível em: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf. Acesso em: 08 out. 2023.

MANO, L.Y.; *et al.* Machine Learning Applied to Covid-19: A Review of the Initial Pandemic Period. *International Journal of Computational Intelligence Systems*, v. 16, n. 73, p. 1-24, mai. 2023. Disponível em: https://link.springer.com/article/10.1007/s44196-023-00236-3. Acesso em: 07 jan. 2024. DOI: https://doi.org/10.1007/s44196-023-00236-3.

MARQUES, J. A. L.; *et al.* Sistema de Inferência *Fuzzy* para Interpretação da Frequência Cardíaca Fetal em Exames Cardiotocográficos. *In:* CBIS 2006, Florianópolis. *Anais [...]*, p. 1-7, 2006. Disponível em: https://rbejournal.org/doi/10.4322/rbeb.2012.051. Acesso em: 12 out. 2023. DOI: http://dx.doi.org/10.4322/rbeb.2012.051.

MEIRELLES, A. F. V.; *et al. Covid-19 e Saúde da Criança e do Adolescente*. Instituto Nacional da Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira (IFF/Fiocruz), 2020. Disponível em: https://www.arca.fiocruz.br/bitstream/handle/icict/43274/?sequence=2. Acesso em: 08 out. 2023.

MENDEL, J. M. *Uncertain Rule-Based Fuzzy Systems:* Introduction and New Directions. Springer, 2ª ed., 2017.

MERT, A. On the WABL Defuzzification Method for Intuitionistic Fuzzy Numbers. *In:* KAHRAMAN, C.; *et al.* (Eds.). *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making* INFUS 2019. Advances in Intelligent Systems and Computing. Springer, v. 1029, 2020. DOI: https://doi.org/10.1007/978-3-030-23756-1_7.

MILANI, C. S.; CARVALHO, D. R. Pós-Processamento em KDD. *Revista de Engenharia e Tecnologia*, v. 5, n. 1, p. 151-162, abr. 2013. Disponível em: https://revistas.uepg.br/index.php/ret/article/view/11473. Acesso em: 11 nov. 2024.

MINGOTI, S. A. *Análise de Dados Através de Métodos de Estatística Multivariada:* Uma Abordagem Aplicada. Belo Horizonte: UFMG, 2ª ed., 2013.

MITCHELL, T. M. *Machine Learning.* USA: McGraw-Hill, 1997. Disponível em: https://archive.org/details/machine-learning-tom-mitchell_202402. Acesso em: 19 jul. 2023.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. *Mineração de Textos.* Goiânia: Universidade Federal de Goiás, 2007, 30 p. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf. Acesso em: 10 jan. 2024.

MOREIRA, D. *txt4cs:* Text as Data para Ciências Sociais. R package version 0.1.0., 2023. Disponível em: https://github.com/davi-moreira/txt4cs. Acesso em 02 fev. 2024.

NEUMAM, C. *Entenda a Diferença dos Sintomas da Gripe e da Covid-19 e da Importância de Continuar se Vacinando Anualmente*. São Paulo: Instituto Butantan, 2023. Disponível em: https://butantan.gov.br/noticias/entenda-a-diferenca-dos-sintomas-da-gripe-e-da-covid-19-e-a-importancia-de-continuar-se-vacinando-anualmente. Acesso em: 31 dez. 2023.

ORENGO, V. M.; HUYCK, C. R. A Stemming Algorithm for The Portuguese Language. *In: 8th International Symposium on String Processing and Information Retrieval (SPIRE)*. Laguna de San Raphael, Chile, 2001, p. 183-193. Disponível em: https://ieeexplore.ieee.org/document/989755. Acesso em: 27 dez. 2023. DOI: https://doi.org/10.1109/SPIRE.2001.10024.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. *Histórico da Pandemia de Covid-19* [Brasil: OPAS, 2020a]. Disponível em: https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19. Acesso em: 04 jan. 2024.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. *OMS Declara Emergência de Saúde Pública de Importância Internacional por Surto de Novo Coronavírus*. [Brasil: OPAS, 2020b]. Disponível em: https://www.paho.org/pt/news/30-1-2020-who-declares-public-health-emergency-novel-coronavirus. Acesso em: 10 dez. 2023.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. *Folha Informativa sobre Covid-19:* Histórico da Pandemia de Covid-19 [Brasil: OPAS, 2020c]. Disponível em: https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19. Acesso em: 10 dez. 2023.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. *OMS Anuncia Nomenclaturas Simples e Fáceis de Pronunciar para Variantes de Interesse e de Preocupação do Sars-CoV-2* [Brasil: OPAS, 2021]. Disponível em: https://www.paho.org/pt/noticias/1-6-2021-oms-anuncia-nomenclaturas-simples-e-faceis-pronunciar-para-variantes-interesse-e. Acesso em: 05 jan. 2024.

PAL, N. R.; BEZDEK, J. C. On Cluster Validity for the Fuzzy C-Means Model. *IEEE Transactions on Fuzzy Systems*, v. 3, p. 370-379, ago. 1995. Disponível em: https://ieeexplore.ieee.org/document/413225. Acesso em: 24 abr. 2024. DOI: https://doi.org/10.1109/91.413225

PAKHOMOV, S.; PEDERSEN, T.; CHUTE, C. G. Abbreviation and Acronym Disambiguation in Clinical Discourse. *In:* American Medical Informatics Association. *AMIA Annual Symposium Proceedings*, p. 589-593, 2005. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560669/. Acesso em 01 abr. 2024. PMID: 16779108.

PEREIRA, M. F. I.; *et al.* Estudo Descritivo da Mortalidade por Covid-19 segundo Sexo, Escolaridade, Faixa Etária, Região de Saúde e Série Histórica: Estado do Rio de Janeiro, Janeiro de 2020 a Agosto de 2021. *In: SciELO Preprints*, 2022. Disponível em: https://preprints.scielo.org/index.php/scielo/preprint/view/3614. Acesso em: 15 nov. 2023. DOI: https://doi.org/10.1590/SciELOPreprints.3614.

PORTAL TELEMEDICINA. O Que é *Anamnese* e Como é Feita? [Portal Telemedicina, 2022], 2022. Disponível em: https://portaltelemedicina.com.br/o-que-e-anamnese-e-como-e-feita. Acesso em: 02 ago. 2024.

PORTER, M. F. An Algorithm for Suffix Stripping. *Program Electronic Library and Information Systems,* v. 14, n. 3, p. 130-137, 1980. Disponível em: https://www.emeraldinsight.com/doi/pdfplus/10.1108/eb046814. Acesso em: 11 jan. 2024. DOI: https://doi.org/10.1108/eb046814.

POSIT TEAM. *RStudio*: Integrated Development Environment for R. Posit Software, PBC, Boston, MA, 2023. Disponível em: http://www.posit.co/. Acesso em 17 fev. 2024.

R CORE TEAM. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2023. Disponível em: https://www.R-project.org/. Acesso em: 17 fev. 2024.

RANKS NL. Portuguese Stopwords [Ranks NL, 1998], 1998. Disponível em: https://www.ranks.nl/stopwords/portuguese. Acesso em: 03 out. 2023.

RECTOR, A. L. Clinical Terminology: Why Is it so Hard? *Methods of Information in Medicine*, v. 38, n. 04/05, p. 239-252, 1999. DOI: https://doi.org/10.1055/s-0038-1634418.

RENJITH, S.; SREEKUMAR, A.; JATHAVEDAN, M. An empirical Research and Comparative Analysis of Clustering Performance for Processing Categorical and Numerical Data Extracts from Social Media. *Acta Scientiarum: Technology*, v. 44, n. 1, 2022. Disponível em: https://periodicos.uem.br/ojs/index.php/ActaSciTechnol/article/view/58653. Acesso em: 25 jan. 2024. DOI: https://doi.org/10.4025/actascitechnol.v44i1.58653.

REVANASIDDAPPA, M. B.; HARISH, B. S. A new feature selection method based on intuitionistic fuzzy entropy to categorize text documents. *International Journal of Interactive Multimedia and Artificial Intelligence*, v. 5, n. 3, p. 106-117, dez. 2018. Disponível em: https://www.ijimai.org/journal/sites/default/files/files/2018/04/ijimai_5_3_12_pdf_16348.pdf. Acesso em: 06 mar. 2024. DOI: https://doi.org/10.9781/ijimai.2018.04.002.

REZENDE, S. O. *et al* (Org.). *Sistemas inteligentes*: fundamentos e aplicações. Barueri: Manole, 2005.

RIDGWAY, J. P.; *et al*. Natural language processing of clinical notes to identify mental illness and substance use among people living with hiv: retrospective cohort study. *JMIR Medical Informatics*, v. 9, n. 3, p. e23456, 2021. Acesso em: 10 nov. 2024. DOI: https://doi.org/10.2196/23456.

RIZOL, P. M. S. R.; MESQUITA, L.; SAOTOME, O. Lógica *fuzzy* tipo-2. *Revista SODEBRAS*, v. 6, n. 70, p. 27-46, 2011. Disponível em: http://www.sodebras.com.br/edicoes/N70.pdf. Acesso em: 03 nov. 2023.

ROSS, T. J. *Fuzzy Logic with Engineering Applications*. USA: John Wiley & Sons, 3ª ed., 2010. DOI: https//doi.org/10.1002/9781119994374.

ROUSSEEUW, P. J. Silhouettes: A Graphical aid To The Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53-65, 1987. Disponível em: https://www.sciencedirect.com/science/article/pii/0377042787901257. Acesso em: 05 mar. 2024. DOI: https://doi.org/10.1016/0377-0427(87)90125-7.

RUSSELL, S. J.; NORVIG, P. *Inteligência Artificial:* Un Enfoque Moderno, Madrid: Pearson Prentice Hall, 2ª ed., 2003.

RUSSELL, B. *Principles of Mathematics*. London: Routledge, 1ª ed., 2015.

SALATIEL, J. R. Tempo, Modalidade e Lógica Trivalente em Peirce e Lukasiewicz. *Kínesis-Revista de Estudos dos Pós-Graduandos em Filosofia*, v. 9, n. 20, p. 151-173, 2017. Disponível em: https://revistas.marilia.unesp.br/index.php/kinesis/article/view/7731. Acesso em: 23 out. 2023. DOI: https://doi.org/10.36311/1984-8900.2017.v9n20.10.p151.

SALTON, G.; MACGILL, M. J. *Introduction to Modern Information Retrieval*. New York: McGRAW-Hill, 1983.

SANTOS, R. E. S.; *et al*. Técnicas de Processamento de Linguagem Natural Aplicadas ao Processo de Mineração de Textos: resultados preliminares de um Mapeamento Sistemático. *Revista de Sistemas e Computação*, v. 4, n. 2, p. 116-125, jul./dez. 2014. Disponível em: https://revistas.unifacs.br/index.php/rsc/article/view/3030. Acesso em: 10 nov. 2024.

SEGARAN, T. *Programming Collective Intelligence.* California: O'Reilly Media, 1ª ed., 2007.

SORIN, V.; *et al*. Deep Learning for Natural Language Processing in Radiology – Fundamentals and a Systematic Review. *Journal of the American College of Radiology*, v. 17, n. 5, p. 639-648, 2020. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S154614402030003X. Acesso em: 10 nov. 2024. DOI: https://doi.org/10.1016/j.jacr.2019.12.026.

SOUZA, S. A. O. Alguns Comentários sobre a Teoria *Fuzzy*. *Exacta*, v. 1, p. 139-148, jan./dez 2003. Disponível em: https://periodicos.uninove.br/exacta/article/view/527. Acesso em: 09 out. 2023. DOI: https://doi.org/10.5585/exacta.v1i0.527.

SOUZA, D. S.; ANDRADE, E. C. S.; LOBO, M. P. O Axioma da Seleção resolve o Paradoxo de Russell. *In:* XVI Semana Acadêmica, VII Encontro Regional de Educação Matemática e III Encontro de Pós-Graduação Lato Sensu em Educação e Educação Matemática: A Matemática, suas Áreas e suas Aplicações, Jaguariúna, Tocantins, 2019. *Anais Eletrônicos[...]*, p. 1-6, 2019. Disponível em: http://www.uft.edu.br/matematicaaraguaina/includes/eventos/2019/po/PO1.pdf. Acesso em: 10 out. 2023.

SOUZA, A. D.; DE ALMEIDA, M. B. Análise de Dados Clínicos Textuais de Prontuários Eletrônicos do Paciente para Integração com Terminologias Médicas Padronizadas. *In:* 12º Seminário em Pesquisa de Ontologia no Brasil, ONTOBRAS, Porto Alegre, 2019. *Anais[...]*, p. 1-6, 2019. Disponível em: https://ceur-ws.org/Vol-2519/doctorate3.pdf. Acesso em: 10 jun. 2024.

SOUZA, A. D. de; FELIPE, E. R. Processamento de Linguagem Natural Aplicado a *Anamneses* do Domínio da Ginecologia. *Frontiers of Knowlegde Representation*, v. 1, n. 2, p. 51-69, 2021. Disponível em: https://periodicos.ufmg.br/index.php/fronteiras-rc/article/view/37308. Acesso em: 10 nov. 2024.

SUGENO, M.; TERANO, T. A Model of Learning Based on Fuzzy Information. *Kybernetes*, v. 6, n. 3, p. 157-166, 1977. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/eb005448/full/html. Acesso em 15 fev. 2024. DOI: https://doi.org/10.1108/eb005448.

TAKAGI, T.; SUGENO. M. Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Transactions on Systems, Man and Cybernetics*, v. smc 15, n. 1, p. 116-132, jan./fev. 1985. Disponível em: https://ieeexplore.ieee.org/document/6313399. Acesso em: 13 fev. 2024. DOI: https://doi.org/10.1109/TSMC.1985.6313399.

TAN, A-H. Text Mining: The State of The Art and The Challenges. *In: Proceedings of The PAKDD Workshop on Knowledge Discovery from Advanced Databases*, 1999. Pequim: Kent Ridge Digital Labs, p. 65-70, 1999. Disponível em: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9a80ec16880ae43dc20c792ea3734862d85ba4d7. Acesso em: 13 jan. 2024.

TANSCHEIT, R. *Sistemas Fuzzy*. Relatório Técnico. Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004. Disponível em: https://www.inf.ufsc.br/~mauro.roisenberg/ine5377/Cursos-ICA/LN-Sistemas%20Fuzzy.pdf. Acesso em: 27 jul. 2023.

TESSAROLO, P. H.; MAGALHÃES, W. B. *A Era do Big Data no Conteúdo Digital:* os Dados Estruturados e Não Estruturados. Universidade Paranaense, Paranavaí, 2015. Disponível em: http://web.unipar.br/~seinpar/2015/_include/artigos/Pedro_Henrique_Tessarolo.pdf. Acesso em: 12 mar. 2024.

THORNDIKE, R. Who Belongs in the Family? *Psychometrika,* v. 18, n. 4, p. 267-276, 1953. Acesso em: 01 fev. 2024. DOI: http://doi.org/10.1007/bf02289263.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating The Number of Clusters in a Data Set Via The Gap Statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* v. 63, n. 2, p. 411-423, jul. 2001. Disponível em: https://academic.oup.com/jrsssb/article/63/2/411/7083348. Acesso em: 15 fev. 2024. DOI: http://doi.org/10.1111/1467-9868.00293.

TSOUKALAS, L. H.; UHRIG, R. E. *Fuzzy and Neural Approaches in Engineering*. New York: Wiley, 1997.

WANG, Z.; *et al*. Extracting Diagnoses and Investigation Results from Unstructured Text in Electronic Health Records by Semi-Supervised Machine Learning. *PLoS One*, v. 7, n. 1, jan. 2012. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0030412. Acesso em: 03 mar. 2024. DOI: https://doi.org/10.1371/journal.pone.0030412.

WICKHAM, H.; *et al*. Welcome to the Tidyverse. *Journal of Open Source Software*, v. 4, n. 43, 2019. Disponível em: https://joss.theoj.org/papers/10.21105/joss.01686. Acesso em: 05 mar. 2024. DOI: https://doi.org/10.21105/joss.01686.

WICKHAM, H. *stringr:* Simple, Consistent Wrappers for Common String Operations. R package version 1.5.0, 2022. Disponível em: https://cran.r-project.org/web/packages/stringr/stringr.pdf. Acesso em: 22 fev. 2024.

WORLD HEALTH ORGANIZATION. *WHO Coronavirus Disease (Covid-19) Dashboard*. Geneva: World Health Organization [WHO, 2020]. Disponível em: https://covid19.who.int/. Acesso em: 17 set. 2023.

XAVIER, B. M.; SILVA, A. D.; GOMES, G. Uma Arquitetura Hibrida para a Indexação de Documentos do Diário Oficial do Município de Cachoeiro de Itapemirim. *Scielo*, p. 83-95, 2015. Disponível em: http://www.scielo.br/pdf/tinf/v27n1/0103-3786-tinf-27-01-00083.pdf. Acesso em: 10 nov. 2024.

YAGER, R. R. On The Measure of Fuzziness and Negation Part I: Membership in the Unit Interval. *International Journal of General Systems*, v. 5, n. 4, p. 221-229, 1979. Acesso em 15 fev. 2024. DOI: https://doi.org/10.1080/03081077908547452.

YANG, Y.; PEDERSEN, J. O. A comparative study on features selection in text categorization. *In:* FISHER, D. H. ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning. *Proceedings* [...]*,* p. 412-420, jul. 1997. Acesso em: 11 nov. 2004.

YE, J. Cosine Similarity Measures for Intuitionistic Fuzzy Sets and Their Applications. *Mathematical and Computer Modeling,* v. 53, n. 1, p. 91-97, 2011. Disponível em: https://www.sciencedirect.com/science/article/pii/S0895717710003651. Acesso em: 02 dez. 2023. DOI: https://doi.org/10.1016/j.mcm.2010.07.022.

ZADEH, L. A. Fuzzy Sets. *Information and Control*, v. 8, n. 3, p. 338-353, 1965. Disponível em: https://www.sciencedirect.com/science/article/pii/S001999586590241X. Acesso em: 01 out. 2023. DOI: https://doi.org/10.1016/S0019-9958(65)90241-X.

ZADEH, L. A. Outline of a New Approach to The Analysis of Complex Systems and Decision Processes. *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-3, n. 1, p. 28-44, 1973. Disponível em: https://ieeexplore.ieee.org/document/5408575. Acesso em: 07 nov. 2023. DOI: https://doi.org/10.1109/TSMC.1973.5408575.

ZADEH, L. A. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning-I. *Information Sciences*, v. 8, n. 3, p. 199-249, 1975. Disponível em: https://www.sciencedirect.com/science/article/pii/0020025575900365. Acesso em: 12 dez. 2023. DOI: https://doi.org/10.1016/0020-0255(75)90036-5.

APÊNDICE A – Lista das *Stopwords* usadas na etapa de pré-processamento

a, à, abaixo, acaso, acerca, acima, acolá, acordo, afora, agora, aí, ainda, além, algo, alguém, algum, alguma, algumas, alguns, ali, amanhã, ambas, ambos, ampla, amplas, amplo, amplos, ante, antes, ao, aos, apenas, apesar, após, aquela, àquela, àquelas, aquelas, aquele, àquele, aqueles, àqueles, aqui, aquilo, àquilo, aquém, as, às, assim, até, atrás, através, baixo, basta, bastante, cá, cada, cada, caso, causa, cedo, chega, cima, coisa, coisas, com, comigo, companhia, conforme, conosco, conseguinte, consigo, consoante, contigo, contra, contras, contudo, convosco, cuja, cujas, cujo, cujos, da, daquela, daquelas, daquele, daqueles, daquilo, das, de, debaixo, dela, delas, dele, deles, dentre, dentro, depois, depressa, desde, dessa, dessas, desse, desses, desta, destas, deste, destes, diante, disso, disto, do, donde, dos, dum, duma, duns, dumas, durante, e, eis, ela, elas, ele, eles, em, embaixo, embargo, embora, enquanto, entanto, então, entre, entretanto, essa, essas, esse, esses, esta, este, estes, eu, exceto, fim, fora, frente, grande, grandes, há, hão, hei, hoje, isso, isto, já, jamais, junto, lá, la, las, lo, logo, longe, los, lhe, lhes, maior, mais, maneira, mas, me, mediante, medida, meia, meio, meios, menor, menos, mesma, mesmas, mesmo, mesmos, meu, meus, mim, minha, minhas, modo, muita, muitas, muito, muitos, na, nada, naquela, naquelas, naquele, naqueles, naquilo, nas, nela, nelas, nele, neles, nem, nenhum, nenhuma, nessa, nessas, nesse, nesses, nesta, nestas, neste, nestes, ninguém, nisso, nisto, no, nos, nós, nossa, nossas, nosso, nossos, num, numa, numas, nuns, nunca, o, obstante, onde, ontem, ora, os, ou, outra, outras, outro, outros, para, partir, pela, pelas, pelo, pelos, pequena, pequenas, pequeno, pequenos, per, perante, perto, pois, por, porém, porquanto, porque, porquê, portanto, possível, posso, pouca, poucas, pouco, poucos, pra, praquele, praquilo, pras, primeira, primeiras, primeiro, primeiros, pro, própria, próprias, próprio, próprios, pros, prós, prum, pruma, prumas, pruns, pude, quais, quaisquer, qual, qualquer, quando, quanta, quantas, quanto, quantos, quão, que, quê, quem, quer, razão, respeito, salvo, se, segundo, seja, sem, sempre, ser, seu, seus, si, só, sob, sobre, sois, somente, somos, sou, sua, suas, tais, tal, talvez, também, tampouco, tanta, tantas, tanto, tantos, tão, tarde, te, tem, teu, teus, toda, todas, todavia, todo, todos, trás, tu, tua, tuas, tudo, última, últimas, último, últimos, um, uma, umas, uns, vez, vezes, visto, você, vocês, vos, vós, vossa, vossas, vosso, vossos

APÊNDICE B – Palavras-chave Utilizadas na Busca para Construção das Variáveis Sociodemográficas

Quadro 6 - Palavras-chave utilizadas na busca para construção da variável TESTE_COVID19 (Continua)

Variável	Classificação	Palavras-chave
TESTE_COVID19	Suspeito	"covid suspeito", "suspeito covid", "suspeita covid", "aguarda resultado teste rapido", "aguarda pcr", "aguarda swab", "aguarda resultado swab", "aguarda teste covid", "suspeita covid", "covid suspeita", "complementares teste rapido covid19"
	Positivo	"covid positivo", "positivo covid", "positivo para covid", "covid 19 positivo", "positivo covid 19", "positivo para covid 19", "covid19 positivo", "positivo covid19", "positivo para covid19", "coronavirus positivo", "positivo coronavirus", "positivo para coronavirus", "pcr positivo", "positivo pcr", "pcr rt positivo", "positivo pcr rt", "rt pcr positivo", "positivo rt pcr", "pcrrt positivo", "positivo pcrrt", "rtpcr positivo", "positivo rtpcr", "positivo sarv cov", "sarv cov positivo", "positivo sarv cov 2", "sarv cov 2 positivo", "positivo sarv cov2", "sarv cov2 positivo", "positivo sarvcov 2", "sarvcov 2 positivo", "positivo sarvcov2", "sarvcov2 positivo", "positivo sars cov", "sars cov positivo", "positivo sars cov 2", "sars cov 2 positivo", "positivo sars cov2", "sars cov2 positivo", "positivo sarscov 2", "sarscov 2 positivo", "positivo sarscov2", "sarscov2 positivo", "swab positivo", "positivo swab",

Quadro 6 - Palavras-chave utilizadas na busca para construção da variável TESTE_COVID19 (Continuação)

Variável	Classificação	Palavras-chave
TESTE_COVID19	Positivo	“swabs positivo”, “positivo swabs”, “teste rapido swab positivo”, “positivo teste rapido swab”, “positivo teste covid”, “teste covid positivo”, “coronavirus covid positivo”, “swab covid positivo”, “proteina reativa positivo”, “positivo proteina reativa”, “covid 19 detectado”, “teste rapido positivo”, “positivo teste covid19”, “confirmado covid”, “covid confirmado”, “covid9 positivo”, “positivo covid9”, “srag por covid 19”, “srag covid”, “positivo adenovirus”, “adenovirus positivo”, “positivo para adenovirus”
	Negativo	“negativo covid”, “negativo para covid”, “covid 19 negativo”, “negativo covid 19”, “negativo para covid 19”, “covid19 negativo”, “negativo covid19”, “negativo para covid19”, “coronavirus negativo”, “negativo coronavirus”, “negativo para coronavirus”, “pcr negativo”, “negativo pcr”, “pcr rt negativo”, “negativo pcr rt”, “rt pcr negativo”, “negativo rt pcr”, “pcrrt negativo”, “negativo pcrrt”, “rtpcr negativo”, “negativo rtpcr”, “negativo sarv cov”, “sarv cov negativo”, “negativo sarv cov 2”, “sarv cov 2 negativo”, “negativo sarv cov2”, “sarv cov2 negativo”, “negativo sarvcov 2”, “sarvcov 2 negativo”, “negativo sarvcov2”, “sarvcov2 negativo”, “negativo sars cov”, “sars cov negativo”, “negativo sars cov 2”, “sars cov 2 negativo”, “negativo sars cov2”, “sars cov2 negativo”, “negativo sarscov 2”, “sarscov 2 negativo”, “negativo sarscov2”, “sarscov2 negativo”, “swab negativo”, “negativo swab”, “swabs negativo”,

Quadro 6 - Palavras-chave utilizadas na busca para construção da variável TESTE_COVID19 (Conclusão)

Variável	Classificação	Palavras-chave
TESTE_COVID19	Negativo	“negativo swabs”, “teste rapido swab negativo”, “negativo teste rapido swab”, “negativo teste covid”, “teste covid negativo”, “coronavirus covid negativo”, “swab covid negativo”, “proteina reativa negativo”, “negativo proteina reativa”, “negativo adenovirus”, “adenovirus negativo”, “negativo para adenovirus”

Quadro 7 - Palavras-chave utilizadas na busca para construção da variável OBITO_ALTA (Continua)

Variável	Classificação	Palavras-chave
OBITO_ALTA	Transferido	“transferido para”, “transferida para”, “transferencia para”, “trasnferencia para”, “trasnferido para”
	Óbito	“declaro obito”, “obito declaro”, “obito declarado”, “declarado obito”, “obito as”, “declaro obito as”, “atesto obito”, “obito atesto”, “atestado obito”, “obito atestado”, “confirmo obito”, “obito confirmo”, “confirmado obito”, “obito confirmado”, “obito constato”, “constato obito”, “obito constatado”, “constatado obito”, “veio a obito”, “veio obito”, “evolucao obito”, “evolucao para obito”, “evoluiu obito”, “evoluiu para obito”, “evoluiu obito”, “evidencio obito”, “obito evidenciado”, “obito registrado”, “registrado obito”, “registro obito”, “obito registro”, “obito”, “paciente veio a falecer”, “falecimento do paciente”, “declaro morte do paciente”, “declaro morte da paciente”, “foi a obito”, “obito as”

Quadro 7 - Palavras-chave utilizadas na busca para construção da variável OBITO_ALTA (Conclusão)

Variável	Classificação	Palavras-chave
OBITO_ALTA	Alta	“alta paciente”, “alta do paciente”, “alta de paciente”, “alta da paciente”, “alta paciente”, “paciente alta”, “paciente recebe alta”, “paciente recebeu alta”, “alta para o dia”, “alta dia”, “previsao alta”, “alta declarada”, “declaro alta”, “declaro alta do paciente”, “alta declarada”, “declaro alta da paciente”, “declaro alta de paciente”, “declaro alta paciente”, “prescrevo alta”, “previsao de alta”, “prescrevo alta paciente”, “prescrevo alta de paciente”, “prescrevo alta do paciente”, “prescrevo alta da paciente”, “liberado alta”, “paciente liberado”, “confirmo alta”, “atesto alta”, “atesto alta paciente”, “atesto alta do paciente”, “atesto alta da paciente”, “alta confirmada”, “liberacao paciente”, “alta paciente confirmada”, “liberacao de paciente”, “liberacao do paciente”, “liberacao da paciente”, “paciente liberado”, “recebeu alta”, “confirmo alta paciente”, “confirmo alta do paciente”, “confirmo alta da paciente”, “confirmo alta de paciente”, “avaliar alta hospitalar”, “recebe alta hospitar”, “avaliar abordagem x alta”, “apta por alta hospitalar”, “apto por alta hospitalar”, “programo alta”, “alta hospitalar”, “alta hoje”, “realizo alta”, “alta com”, “alta programada”

Quadro 8 - Palavras-chave utilizadas na busca para construção das variável GENERO

Variável	Classificação	Palavras-chave
GENERO	Masculino	“masculino”, “masculina”, “masc”, “sexo masculino”, “sexo masc”, “sexo m”, “sexo h”, “sexo homem”, “idoso”, “senhor”, “homem”, “o jovem”, “rapaz”, “moco”, “menino”, “garoto”, “solteiro”, “noivo”, “casado”, “divorciado”, “separado”, “viúvo”, “o paciente”, “marido de”, “esposo de”, “acompanhado”, “sozinho”, “encaminhado”, “transferido”, “lucido”, “sedado”, “sonolento”, “agitado”, “internado”, “ativo”, “calmo”, “nervoso”, “ansioso”, “deitado”, “choroso”, “cooperativo”, “corado”, “reanimado”, “hidratado”, “sentado”, “desnutrido”, “hipertenso”, “nutrido”, “cardíaco”, “diabético”, “intubado”, “extubado”, “colaborativo”, “entubado”, “eupneico”, “paciente masculino”
	Feminino	“feminino”, “feminina”, “fem”, “sexo feminino”, “sexo fem”, “sexo f”, “sexo mulher”, “idosa”, “senhora”, “garota”, “moca”, “a jovem”, “menina”, “solteira”, “noiva”, “casada”, “separada”, “divorciada”, “viúva”, “a paciente”, “acompanhada”, “sozinha”, “encaminhada”, “transferida”, “lucida”, “sedada”, “sonolenta”, “agitada”, “deitada”, “internada”, “ativa”, “calma”, “nervosa”, “ansiosa”, “deitada”, “chorosa”, “cooperativa”, “corada”, “reanimada”, “hidratada”, “desnutrida”, “hipertensa”, “nutrida”, “cardíaca”, “diabética”, “intubada”, “extubada”, “entubada”, “colaborativa”, “eupneica”, “sentada”, “paciente feminino”

Quadro 9 - Palavras-chave utilizadas na busca para construção das variáveis TABAGISTA e ETILISTA

Variável	Classificação	Palavras-chave
TABAGISTA	Tabagista	"tabagismo sim", "sim tabagismo", "positivo tabagismo", "tabagismo positivo", "tabagista sim", "sim tabagista", "tabagista positivo", "positivo tabagista", "tabagista ativo", "ativo tabagista", "fuma sim", "sim fuma", "fuma positivo", "positivo fuma", "confirma tabagismo", "tabagismo confirmado", "afirmado tabgismo", "tabagismo afirmado", "e tabagista", "ex tabagista", "historico tabagismo", "historico de tabagista", "historico tabagista", "foi tabagista", "tabagista ativo", "ativo tabagismo", "ativo tabagista", "ex fumante", "foi fumante", "e fumante", "fumante sim", "sim fumante", "fumante positivo", "positivo fumante", "fumante ha", "tabagista ha", "fumante ativo", "ativo fumante", "tabagista ha anos", "fumante ha anos", "grande fumante", "grande tabagista", foi tabagista", "foi fumante", "confirma fumante", "fumante confirmado", "fumante atualmente", "atualmente fumante", "foi tabagista ha", "foi fumante ha"
ETILISTA	Alcoólatra	"alcoolatra", "etilista"

Quadro 10 - Palavras-chave utilizadas na busca para construção da variável IDADE
(Continua)

Variável	Classificação	Palavras-chave
IDADE	Anos (xx varia de 1 até 110)	"paciente xx anos", "paciente de xx anos", "paciente com xx anos", "idade xx anos", "id xx anos", "paciente xx anos de idade", "paciente de xx anos de idade", "paciente xx ano", "paciente de xx ano", "paciente com xx ano", "idade xx ano", "id xx ano", "paciente xx ano de idade", "paciente de xx ano de idade", "pacinte xx anos", "pacinte de xx anos", "pacinte com xx anos", "pacinte xx anos de idade", "pacinte de xx anos de idade", "pacinte xx ano", "pacinte de xx ano", "pacinte com xx ano", "pacinte xx ano de idade", "pacinte de xx ano de idade"
	Meses (xx varia de 1 até 12)	"paciente xx meses", "paciente de xx meses", "paciente com xx meses", "idade xx meses", "id xx meses", "paciente xx meses de idade", "paciente de xx meses de idade", "paciente xx mes", "paciente de xx mes", "paciente com xx mes", "idade xx mes", "id xx mes", "paciente xx mes de idade", "paciente de xx mes de idade", "pacinte xx meses", "pacinte de xx meses", "pacinte com xx meses", "pacinte xx meses de idade", "pacinte de xx meses de idade", "pacinte xx mes", "pacinte de xx mes", "pacinte com xx mes", "pacinte xx mes de idade", "pacinte de xx mes de idade", "paciente xx mes e", "paciente de xx mes e", "paciente xx mes de vida"
	Dias (xx varia de 1 até 72)	"paciente de xx dias", "paciente xx dias", "pacinte xx dias", "pacinte de xx dias", "paciente de xx dias de vida", "paciente xx dias de vida", "pacinte xx dias de vida",

Quadro 10 - Palavras-chave utilizadas na busca para construção da variável IDADE
(Conclusão)

Variável	Classificação	Palavras-chave
IDADE	Dias (xx varia de 1 até 72)	"pacinte de xx dias de vida", "paciente de xx dia", "paciente xx dia", "pacinte xx dia", "pacinte de xx dia", "paciente de xx dia de vida", "paciente xx dia de vida", "pacinte xx dia de vida", "pacinte de xx dia de vida", "lactente de xx dias", "lactente xx dias", "lactente de xx dia", "lactente xx dia", "lactente de xx dias de vida", "lactente xx dias de vida", "lactente de xx dia de vida", "lactente xx dia de vida", "recem nascido xx dia de vida", "recem nascido de xx dia de vida", "recem nascido xx dias de vida", "recem nascido de xx dias de vida", "recem nascido xx hora de vida", "recem nascido de xx hora de vida", "recem nascido xx horas de vida", "recem nascido de xx horas de vida"

APÊNDICE C - Palavras-chave utilizadas na busca para construção das variáveis de sintomas

Quadro 11 - Palavras-chave utilizadas na busca para construção das variáveis de sintomas (Continua)

Variável	Classificação	Palavras-chave
CALAFRIO	1 – sim; 0 - não	“calafrio”, “calafrios” , “arrepio”, “arrepios”
CANSAÇO	1 – sim; 0 - não	“cansaco”, “cansaco persistente”, “cansada”, “cansado”, “moleza”, “frequeza”, “fadiga”, “fadiga persistente”, “prostracao”, “prostrada”, “prostrado”, “astenia”
CORIZA	1 – sim; 0 – não	“coriza”, “nariz escorrendo”, “nariz entupido”, “rinorreia”
DIARREIA	1 – sim; 0 - não	“diarreia”
DOR_CABECA	1 – sim; 0 - não	“dor de cabeca”, “enxaqueca”, “cefaleia”
DOR_GARGANTA	1 – sim; 0 - não	“dor de garganta”, “dor na garganta”, “garganta inflamada”
DOR_CORPO	1 – sim; 0 - não	“dor no corpo”, “dores no corpo”, “dor muscular”, “dores musculares”, “dor no musculo”, “dores nos musculos”, “mialgia”
ERUPCAO_PELE	1 – sim; 0 - não	“erupcao na pele”, “erupcao cutanea”
ESPIRRO	1 – sim; 0 – não	“espirro”, “espirros”, “esternutacao”
FALTA_APETITE	1 – sim; 0 - não	“falta de apetite”, “perda de apetite”, “perda do apetite”, “sem apetite”, “inapetencia”

Quadro 11 - Palavras-chave utilizadas na busca para construção das variáveis de sintomas (Conclusão)

Variável	Classificação	Palavras-chave
FALTA_AR	1 – sim; 0 - não	“falta de ar”, “dificuldade para respirar”, “dificuldade de respirar”, “dispneia”
FEBRE	1 – sim; 0 - não	“com febre”, “febril”, “apresenta febre”, “paciente com febre”, “paciente febril”, “paciente apresenta febre”, “paciente se apresenta febril”, “febre alta”, “pirexia”, “piretico”, “piretica”, “paciente com pirexia”, “paciente com hipertermia”
PERDA_OLFATO_PALADAR	1 – sim; 0 - não	“perda do olfato”, “perda do paladar”, “olfato alterado”, “paladar alterado”, “nao sente cheiro”, “nao sente gosto”, “hiposmia”, “anosmia”, “digeusia”, “ageusia”
TONTURA	1 – sim; 0 - não	“tonteira”, “tontura”, “paciente apresentou tontura”, “vertigem”
TOSSE	1 – sim; 0 - não	“tosse seca”, “tosse persistente”, “tosse continua”, “tosse produtiva”
VOMITO	1 – sim; 0 - não	“vomito”, “vomitos”, “nausea”, “nauseas”, “emese”

APÊNDICE D - Palavras-chave utilizadas na busca para construção das variáveis de comorbidades

Quadro 12 - Palavras-chave utilizadas na busca para construção das variáveis de Comorbidades (Continua)

Variável	Classificação	Palavras-chave
DOENCA_CARDIOVASCULAR	1 – sim; 0 – não	"doenca cardiovascular", "doenca coracao", "doenca miocardio", "doenca cardiaca", "doenca coronariana", "doenca arterial coronariana", "insuficiencia cardiaca", "insuficiencia coracao", "insuficiencia miocardio", "isquemia cardiaca", "isquemica cordiaca", "isquemica miocardio", "arritmia cardiaca", "infarto", "infarte", "infarto agudo do miocardio", "iam", "angina", "angina pictoris", "coronariopata", "cardiopata", "cardiopatia congenita", "cardiopatia hipertensiva", "sindrome coronariana", "valvopatia", "valvopatias", "miocardiopatias", "pericardiopatias", "doenca da aorta", "fistulas arteriovenosas"
DOENCA_CEREBROVASCULAR	1 – sim; 0 – não	"doenca cerebrovascular", "avc", "avc isquemico", "avc hemorragico", "acidente vascular cerebral", "demencia", "acidente vascular cerebral isquemico", "acidente vascular isquemico",

Quadro 12 - Palavras-chave utilizadas na busca para construção das variáveis de Comorbidades (Continuação)

Variável	Classificação	Palavras-chave
DOENCA_CEREBROVASCULAR	1 – sim; 0 – não	"avi", "acidente vascular hemorragico", "ait", "ataque isquemico transitorio", "demencia vascular", "aneurisma", "isquemia"
DOENCA_PULMONAR	1 – sim; 0 – não	"doenca pulmonar", "doenca respiratoria", "asma", "bronquite", "bronquite asmatica", "fibrose pulmonar", "asmatico", "asmatica", "doenca pulmonar obstrutiva cronica", "pneumoconioses", "displasia broncopulmonar", "asma grave"
DOENCA_RENAL	1 – sim; 0 – não	"doenca rim", "doenca rins", "doenca renal", "doenca renal cronica", "sindrome nefrotica"
DIABETES	1 – sim; 0 – não	"diabetes", "diabetes mellitus", "diabetes tipo", "diabetico", "diabetica"
HIPERTENSAO	1 – sim; 0 – não	"hipertensao", "hipertensao arterial", "har", "hipertensao arterial lesao orgao", "hipertenso", "hipertensa", "pressao alta"
IMUNOCOMPROMETIDO	1 – sim; 0 – não	"imunocomprometido", "imunossuprimido", "hiv", "aids", "transplantado", "transplantada", "oncologico", "oncologica", "tratamento oncologico", "lupus", "artrite", "artrite reumatoide", "neoplasia hematologica",

Quadro 12 - Palavras-chave utilizadas na busca para construção das variáveis de Comorbidades (Conclusão)

Variável	Classificação	Palavras-chave
IMUNOCOMPROMETIDO	1 – sim; 0 – não	"imunodeficiencia primaria", "imunidade baixa", "doenca crohn", "esclerose sistemica", "anemia falciforme", "doenca falciforme", "talassemia maior", "doenca reumatica"
OBESIDADE	1 – sim; 0 – não	“obesidade”, “obesa”, ”obeso”

APÊNDICE E - Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means*

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Continua)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC001	0,6257	0,1882	0,1861	**1**
PAC002	0,9891	0,0067	0,0043	**2**
PAC003	0,6854	0,1494	0,1652	**2**
PAC004	0,6072	0,1596	0,2332	**1**
PAC005	0,6258	0,1256	0,2486	**2**
PAC006	0,9471	0,0270	0,0258	**1**
PAC007	0,7520	0,1140	0,1340	**2**
PAC008	0,8039	0,0805	0,1156	**1**
PAC009	0,8752	0,0120	0,1128	**2**
PAC010	0,7759	0,1020	0,1221	**2**
PAC011	0,7647	0,1150	0,1203	**1**
PAC012	0,9666	0,0246	0,0088	**2**
PAC013	0,7569	0,0954	0,1477	**1**
PAC014	0,9976	0,0009	0,0015	**2**
PAC015	0,7658	0,0253	0,2089	**1**
PAC016	0,9915	0,0009	0,0076	**1**
PAC017	0,9333	0,0092	0,0574	**2**
PAC018	0,7657	0,0076	0,2268	**1**
PAC019	0,8319	0,1546	0,0135	**2**
PAC020	0,9276	0,0002	0,0722	**1**
PAC021	0,9411	0,0259	0,0330	**1**
PAC022	0,8305	0,0673	0,1022	**2**
PAC023	0,8757	0,0467	0,0776	**2**
PAC024	0,8228	0,0783	0,0990	**1**
PAC025	0,7950	0,0981	0,1070	**2**
PAC026	0,9461	0,0159	0,0380	**1**
PAC027	0,9168	0,0365	0,0467	**1**
PAC028	0,7438	0,0995	0,1566	**2**
PAC029	0,7115	0,0599	0,2286	**2**
PAC030	0,9634	0,0182	0,0183	**1**
PAC031	0,5606	0,3706	0,0688	**2**
PAC032	0,6639	0,1658	0,1703	**1**
PAC033	0,7228	0,1056	0,1717	**2**
PAC034	0,7365	0,1249	0,1385	**1**
PAC035	0,8595	0,0119	0,1286	**1**
PAC036	0,9898	0,0051	0,0051	**2**

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Continuação)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC037	0,7257	0,1131	0,1612	**2**
PAC038	0,7135	0,0621	0,2244	**2**
PAC039	0,9621	0,0056	0,0323	**1**
PAC040	0,7016	0,1572	0,1412	**2**
PAC041	0,7642	0,2154	0,0204	**2**
PAC042	0,8230	0,1360	0,0410	**1**
PAC043	0,7098	0,1134	0,1769	**2**
PAC044	0,8409	0,0656	0,0935	**1**
PAC045	0,6144	0,2549	0,1307	**1**
PAC046	0,6385	0,1502	0,2112	**2**
PAC047	0,8033	0,0869	0,1098	**2**
PAC048	0,7733	0,1119	0,1147	**2**
PAC049	0,7556	0,0966	0,1478	**2**
PAC050	0,7676	0,0879	0,1445	**2**
PAC051	0,9155	0,0213	0,0632	**1**
PAC052	0,9223	0,0378	0,0399	**2**
PAC053	0,7879	0,0741	0,1380	**2**
PAC054	0,7118	0,1270	0,1612	**1**
PAC055	0,7634	0,0905	0,1461	**2**
PAC056	0,9035	0,0636	0,0329	**1**
PAC057	0,8833	0,0120	0,1047	**2**
PAC058	0,9928	0,0003	0,0069	**1**
PAC059	0,8688	0,1052	0,0261	**2**
PAC060	0,6544	0,1265	0,2191	**2**
PAC061	0,7356	0,1161	0,1483	**2**
PAC062	0,9839	0,0072	0,0089	**1**
PAC063	0,9281	0,0517	0,0202	**2**
PAC064	0,9441	0,0045	0,0514	**2**
PAC065	0,9598	0,0020	0,0382	**1**
PAC066	0,8753	0,0550	0,0696	**2**
PAC067	0,7924	0,0762	0,1314	**2**
PAC068	0,9749	0,0059	0,0192	**1**
PAC069	0,8760	0,0403	0,0836	**1**
PAC070	0,9376	0,0046	0,0578	**1**
PAC071	0,8299	0,1178	0,0523	**2**
PAC072	0,9459	0,0016	0,0525	**2**
PAC073	0,6307	0,1457	0,2236	**2**
PAC074	0,9692	0,0112	0,0196	**2**
PAC075	0,9143	0,0015	0,0841	**1**
PAC076	0,6050	0,3575	0,0375	**2**
PAC077	0,8009	0,0782	0,1209	**2**
PAC078	0,6773	0,1411	0,1816	**2**

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Continuação)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC079	0,7204	0,1183	0,1613	**1**
PAC080	0,8980	0,0494	0,0526	**1**
PAC081	0,8892	0,0117	0,0991	**2**
PAC082	0,8786	0,0130	0,1084	**2**
PAC083	0,7374	0,1145	0,1481	**2**
PAC084	0,9760	0,0137	0,0103	**1**
PAC085	0,7560	0,0881	0,1559	**2**
PAC086	0,7975	0,0997	0,1028	**2**
PAC087	0,6102	0,1477	0,2421	**1**
PAC088	0,9157	0,0205	0,0637	**1**
PAC089	0,8740	0,0287	0,0973	**1**
PAC090	0,9005	0,0531	0,0464	**1**
PAC091	0,6677	0,0183	0,3140	**2**
PAC092	0,8706	0,1079	0,0215	**2**
PAC093	0,6026	0,2541	0,1433	**1**
PAC094	0,8601	0,0554	0,0845	**1**
PAC095	0,8581	0,0218	0,1201	**2**
PAC096	0,6525	0,0589	0,2886	**2**
PAC097	0,8551	0,0064	0,1385	**1**
PAC098	0,6240	0,1731	0,2030	**2**
PAC099	0,8990	0,0639	0,0371	**1**
PAC100	0,7819	0,0845	0,1336	**1**
PAC101	0,6803	0,0489	0,2709	**2**
PAC102	0,9849	0,0051	0,0100	**1**
PAC103	0,8975	0,0270	0,0755	**1**
PAC104	0,7983	0,0881	0,1136	**1**
PAC105	0,9297	0,0276	0,0427	**1**
PAC106	0,6981	0,1025	0,1994	**2**
PAC107	0,9816	0,0066	0,0118	**2**
PAC108	0,8576	0,0689	0,0735	**2**
PAC109	0,9035	0,0327	0,0637	**1**
PAC110	0,8223	0,0794	0,0983	**1**
PAC111	0,9552	0,0227	0,0221	**1**
PAC112	0,7956	0,0794	0,1249	**1**
PAC113	0,6811	0,0733	0,2456	**1**
PAC114	0,7209	0,1541	0,1250	**1**
PAC115	0,9923	0,0032	0,0045	**1**
PAC116	0,8449	0,1289	0,0262	**2**
PAC117	0,7526	0,1745	0,0730	**2**
PAC118	0,9240	0,0333	0,0426	**1**
PAC119	0,6272	0,1291	0,2437	**1**
PAC120	0,9731	0,0095	0,0174	**1**

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Continuação)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC121	0,7865	0,0896	0,1239	**2**
PAC122	0,7837	0,0034	0,2129	**2**
PAC123	0,8168	0,0680	0,1152	**1**
PAC124	0,6032	0,2660	0,1308	**2**
PAC125	0,9655	0,0155	0,0190	**1**
PAC126	0,9022	0,0846	0,0132	**1**
PAC127	0,6414	0,2456	0,1130	**2**
PAC128	0,8229	0,0779	0,0992	**1**
PAC129	0,8558	0,1073	0,0369	**1**
PAC130	0,9659	0,0014	0,0327	**1**
PAC131	0,8801	0,0637	0,0562	**1**
PAC132	0,7948	0,0787	0,1266	**1**
PAC133	0,8752	0,0608	0,0640	**2**
PAC134	0,5171	0,1634	0,3196	**1**
PAC135	0,8868	0,1016	0,0115	**1**
PAC136	0,7541	0,0053	0,2406	**1**
PAC137	0,7301	0,1094	0,1605	**1**
PAC138	0,9213	0,0379	0,0408	**1**
PAC139	0,6954	0,0383	0,2663	**1**
PAC140	0,8041	0,1673	0,0286	**1**
PAC141	0,8077	0,0323	0,1600	**1**
PAC142	0,7832	0,0556	0,1612	**2**
PAC143	0,9040	0,0440	0,0521	**1**
PAC144	0,9032	0,0035	0,0933	**1**
PAC145	0,7788	0,0858	0,1354	**1**
PAC146	0,7785	0,0824	0,1390	**2**
PAC147	0,7810	0,0924	0,1266	**1**
PAC148	0,8003	0,0952	0,1045	**2**
PAC149	0,9232	0,0748	0,0020	**1**
PAC150	0,9340	0,0220	0,0440	**1**
PAC151	0,7602	0,1531	0,0867	**1**
PAC152	0,5943	0,2607	0,1451	**1**
PAC153	0,8143	0,0745	0,1112	**1**
PAC154	0,6467	0,2225	0,1309	**1**
PAC155	0,9390	0,0244	0,0366	**1**
PAC156	0,8696	0,0384	0,0920	**2**
PAC157	0,8073	0,0738	0,1189	**2**
PAC158	0,8702	0,0158	0,1140	**1**
PAC159	0,7731	0,1127	0,1142	**1**
PAC160	0,9424	0,0102	0,0474	**1**
PAC161	0,8287	0,1307	0,0406	**2**
PAC162	0,9905	0,0019	0,0076	**2**

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Continuação)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC163	0,9196	0,0684	0,0120	**2**
PAC164	0,6785	0,1447	0,1768	**2**
PAC165	0,7453	0,0664	0,1883	**1**
PAC166	0,7723	0,1714	0,0564	**2**
PAC167	0,9546	0,0076	0,0379	**1**
PAC168	0,9689	0,0112	0,0199	**2**
PAC169	0,8347	0,0113	0,1540	**2**
PAC170	0,8145	0,1274	0,0580	**1**
PAC171	0,8883	0,0866	0,0251	**2**
PAC172	0,5693	0,1518	0,2788	**1**
PAC173	0,7619	0,0795	0,1586	**1**
PAC174	0,8566	0,0944	0,0490	**1**
PAC175	0,8068	0,0935	0,0997	**1**
PAC176	0,6631	0,1515	0,1854	**1**
PAC177	0,8587	0,0710	0,0703	**1**
PAC178	0,7562	0,0821	0,1617	**1**
PAC179	0,9257	0,0058	0,0686	**2**
PAC180	0,7778	0,1042	0,1180	**1**
PAC181	0,8588	0,0106	0,1306	**1**
PAC182	0,8280	0,0390	0,1329	**1**
PAC183	0,8541	0,0077	0,1382	**2**
PAC184	0,7711	0,0810	0,1479	**1**
PAC185	0,9165	0,0782	0,0053	**1**
PAC186	0,8087	0,0927	0,0986	**1**
PAC187	0,6283	0,2881	0,0836	**2**
PAC188	0,9211	0,0216	0,0573	**1**
PAC189	0,7789	0,0277	0,1934	**1**
PAC190	0,6952	0,1273	0,1774	**2**
PAC191	0,8763	0,0562	0,0675	**1**
PAC192	0,9052	0,0001	0,0948	**2**
PAC193	0,8369	0,0822	0,0809	**2**
PAC194	0,8649	0,0621	0,0730	**2**
PAC195	0,9010	0,0435	0,0555	**2**
PAC196	0,7750	0,0981	0,1269	**1**
PAC197	0,6938	0,1056	0,2006	**1**
PAC198	0,7749	0,0936	0,1315	**1**
PAC199	0,7831	0,1002	0,1167	**1**
PAC200	0,8126	0,0785	0,1089	**1**
PAC201	0,7919	0,0007	0,2075	**1**
PAC202	0,9344	0,0026	0,0630	**1**
PAC203	0,7877	0,1951	0,0171	**1**
PAC204	0,7672	0,1062	0,1265	**1**

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Continuação)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC205	0,6173	0,1588	0,2239	**2**
PAC206	0,6810	0,0082	0,3108	**1**
PAC207	0,8044	0,1319	0,0637	**1**
PAC208	0,9467	0,0237	0,0296	**1**
PAC209	0,8572	0,0960	0,0468	**1**
PAC210	0,7803	0,0780	0,1417	**2**
PAC211	0,8457	0,0385	0,1158	**1**
PAC212	0,8723	0,0475	0,0801	**1**
PAC213	0,6749	0,1268	0,1983	**1**
PAC214	0,8885	0,0256	0,0859	**1**
PAC215	0,8765	0,0842	0,0393	**1**
PAC216	0,8614	0,0339	0,1047	**2**
PAC217	0,8281	0,0173	0,1546	**1**
PAC218	0,8693	0,0457	0,0850	**2**
PAC219	0,9578	0,0011	0,0411	**1**
PAC220	0,9883	0,0052	0,0066	**2**
PAC221	0,5763	0,1540	0,2697	**2**
PAC222	0,9355	0,0106	0,0538	**1**
PAC223	0,8587	0,0237	0,1175	**1**
PAC224	0,8334	0,1030	0,0636	**1**
PAC225	0,8541	0,0593	0,0866	**1**
PAC226	0,7588	0,0090	0,2322	**1**
PAC227	0,9116	0,0061	0,0823	**2**
PAC228	0,8208	0,0840	0,0952	**2**
PAC229	0,6909	0,0337	0,2754	**2**
PAC230	0,7365	0,0727	0,1908	**2**
PAC231	0,8697	0,0103	0,1200	**1**
PAC232	0,8307	0,0347	0,1346	**1**
PAC233	0,8148	0,0460	0,1392	**2**
PAC234	0,8670	0,0292	0,1038	**2**
PAC235	0,7427	0,0936	0,1637	**2**
PAC236	0,8385	0,0695	0,0920	**2**
PAC237	0,9202	0,0387	0,0410	**1**
PAC238	0,6825	0,0887	0,2287	**2**
PAC239	0,9033	0,0048	0,0920	**1**
PAC240	0,8833	0,0015	0,1153	**1**
PAC241	0,9014	0,0572	0,0414	**1**
PAC242	0,8577	0,0046	0,1377	**2**
PAC243	0,8788	0,0126	0,1086	**1**
PAC244	0,8943	0,0369	0,0688	**1**
PAC245	0,8433	0,0665	0,0902	**1**
PAC246	0,8106	0,0925	0,0969	**1**

Tabela 22 – Pertinências, não-pertinências, hesitação e grupos dos pacientes, segundo *Fuzzy* Intuicionista *C-Means* (Conclusão)

Id_paciente	Pertinência	Não-pertinência	Hesitação	*Cluster*
PAC247	0,9101	0,0440	0,0458	**1**
PAC248	0,6935	0,1423	0,1641	**2**
PAC249	0,8631	0,0515	0,0854	**1**
PAC250	0,8332	0,0412	0,1256	**2**
PAC251	0,8182	0,0274	0,1544	**2**
PAC252	0,8401	0,0013	0,1586	**2**
PAC253	0,7923	0,0839	0,1237	**2**
PAC254	0,7677	0,0803	0,1520	**1**
PAC255	0,9299	0,0257	0,0444	**2**
PAC256	0,7385	0,1232	0,1382	**2**

APÊNDICE F - Informações dos sintomas e comorbidades selecionadas e seus respectivos *clusters*

Legenda:

- **ID:** Identificação do paciente;
- **FALTA AR:** Falta de ar;
- **CARDIO:** Doença cardiovascular;
- **CEREBRO:** Doença Cerebrovascular;
- **C:** *Cluster* ao qual o paciente foi alocado;
- **SIM:** 1; **NÃO:** 0.

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Continua)

Id	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC001	Não	Nao	Sim	Sim	Sim	Sim	**1**
PAC002	Sim	Nao	Sim	Não	Não	Sim	**2**
PAC003	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC004	Não	Sim	Sim	Sim	Sim	Sim	**1**
PAC005	Não	Sim	Sim	Sim	Não	Sim	**2**
PAC006	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC007	Não	Sim	Sim	Não	Sim	Sim	**2**
PAC008	Sim	Sim	Não	Sim	Sim	Não	**1**
PAC009	Não	Sim	Sim	Não	Não	Não	**2**
PAC010	Não	Sim	Sim	Não	Não	Não	**2**
PAC011	Sim	Sim	Sim	Não	Sim	Sim	**1**
PAC012	Não	Sim	Sim	Sim	Sim	Sim	**2**
PAC013	Não	Sim	Sim	Não	Não	Sim	**1**
PAC014	Sim	Sim	Sim	Sim	Sim	Não	**2**
PAC015	Não	Sim	Sim	Sim	Não	Não	**1**
PAC016	Não	Nao	Sim	Não	Não	Sim	**1**
PAC017	Não	Sim	Sim	Não	Sim	Sim	**2**
PAC018	Sim	Sim	Sim	Sim	Nao	Sim	**1**
PAC019	Não	Sim	Sim	Não	Nao	Sim	**2**
PAC020	Não	Nao	Sim	Não	Sim	Sim	**1**
PAC021	Não	Sim	Sim	Sim	Nao	Sim	**1**
PAC022	Sim	Sim	Sim	Não	Não	Sim	**2**
PAC023	Não	Sim	Sim	Não	Não	Sim	**2**

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Continuação)

Id	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC024	Sim	Sim	Sim	Não	Sim	Sim	**1**
PAC025	Não	Sim	Sim	Não	Não	Sim	**2**
PAC026	Não	Sim	Sim	Sim	Sim	Sim	**1**
PAC027	Não	Sim	Sim	Não	Sim	Sim	**1**
PAC028	Não	Sim	Sim	Não	Não	Sim	**2**
PAC029	Sim	Sim	Sim	Não	Não	Sim	**2**
PAC030	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC031	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC032	Não	Sim	Sim	Sim	Sim	Não	**1**
PAC033	Não	Sim	Sim	Sim	Sim	Não	**2**
PAC034	Não	Sim	Sim	Nao	Não	Não	**1**
PAC035	Não	Sim	Sim	Sim	Sim	Sim	**1**
PAC036	Não	Sim	Sim	Nao	Sim	Não	**2**
PAC037	Não	Sim	Sim	Nao	Sim	Sim	**2**
PAC038	Não	Sim	Sim	Sim	Sim	Sim	**2**
PAC039	Não	Sim	Sim	Sim	Sim	Não	**1**
PAC040	Sim	Sim	Sim	Nao	Sim	Sim	**2**
PAC041	Não	Sim	Sim	Nao	Não	Não	**2**
PAC042	Sim	Nao	Sim	Nao	Não	Sim	**1**
PAC043	Não	Nao	Sim	Nao	Sim	Sim	**2**
PAC044	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC045	Sim	Sim	Sim	Sim	Sim	Não	**1**
PAC046	Sim	Sim	Sim	Sim	Não	Sim	**2**
PAC047	Sim	Sim	Sim	Nao	Não	Não	**2**
PAC048	Sim	Sim	Sim	Nao	Sim	Sim	**2**
PAC049	Sim	Nao	Nao	Nao	Sim	Não	**2**
PAC050	Sim	Sim	Sim	Nao	Não	Sim	**2**
PAC051	Não	Sim	Sim	Sim	Sim	Sim	**1**
PAC052	Sim	Sim	Sim	Nao	Não	Sim	**2**
PAC053	Sim	Sim	Sim	Nao	Não	Não	**2**
PAC054	Sim	Sim	Sim	Nao	Não	Sim	**1**
PAC055	Sim	Sim	Sim	Sim	Sim	Não	**2**
PAC056	Não	Sim	Sim	Nao	Sim	Sim	**1**
PAC057	Sim	Sim	Nao	Nao	Não	Não	**2**
PAC058	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC059	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC060	Sim	Sim	Sim	Nao	Não	Sim	**2**
PAC061	Sim	Sim	Sim	Nao	Não	Não	**2**
PAC062	Sim	Sim	Sim	Nao	Não	Não	**1**
PAC063	Sim	Sim	Sim	Sim	Sim	Não	**2**
PAC064	Não	Sim	Sim	Nao	Sim	Não	**2**
PAC065	Não	Nao	Sim	Nao	Sim	Sim	**1**

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Continuação)

ID	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC066	Sim	Sim	Sim	Sim	Não	Sim	**2**
PAC067	Não	Sim	Sim	Nao	Não	Não	**2**
PAC068	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC069	Não	Sim	Sim	Nao	Não	Não	**1**
PAC070	Não	Sim	Sim	Sim	Não	Não	**1**
PAC071	Sim	Sim	Sim	Nao	Não	Não	**2**
PAC072	Não	Nao	Sim	Nao	Sim	Sim	**2**
PAC073	Sim	Sim	Sim	Nao	Não	Sim	**2**
PAC074	Sim	Sim	Sim	Nao	Sim	Sim	**2**
PAC075	Não	Nao	Sim	Sim	Sim	Sim	**1**
PAC076	Sim	Nao	Sim	Nao	Sim	Sim	**2**
PAC077	Não	Sim	Sim	Nao	Não	Não	**2**
PAC078	Não	Nao	Sim	Nao	Não	Não	**2**
PAC079	Sim	Sim	Nao	Nao	Sim	Não	**1**
PAC080	Não	Sim	Sim	Nao	Sim	Sim	**1**
PAC081	Não	Sim	Sim	Nao	Sim	Sim	**2**
PAC082	Não	Sim	Sim	Nao	Não	Sim	**2**
PAC083	Não	Sim	Sim	Nao	Não	Não	**2**
PAC084	Sim	Sim	Sim	Nao	Não	Sim	**1**
PAC085	Sim	Sim	Sim	Nao	Sim	Sim	**2**
PAC086	Sim	Sim	Nao	Nao	Sim	Não	**2**
PAC087	Não	Sim	Sim	Nao	Nao	Sim	**1**
PAC088	Não	Sim	Sim	Nao	Sim	Sim	**1**
PAC089	Sim	Sim	Sim	Sim	Nao	Sim	**1**
PAC090	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC091	Não	Sim	Sim	Sim	Sim	Nao	**2**
PAC092	Não	Sim	Sim	Nao	Nao	Nao	**2**
PAC093	Não	Nao	Sim	Nao	Sim	Sim	**1**
PAC094	Não	Sim	Sim	Nao	Sim	Nao	**1**
PAC095	Não	Sim	Sim	Sim	Nao	Sim	**2**
PAC096	Não	Sim	Sim	Sim	Nao	Sim	**2**
PAC097	Não	Sim	Sim	Sim	Nao	Sim	**1**
PAC098	Não	Sim	Sim	Nao	Nao	Sim	**2**
PAC099	Não	Sim	Sim	Sim	Sim	Sim	**1**
PAC100	Não	Sim	Sim	Sim	Sim	Nao	**1**
PAC101	Não	Sim	Sim	Sim	Sim	Sim	**2**
PAC102	Não	Nao	Sim	Nao	Sim	Nao	**1**
PAC103	Não	Sim	Sim	Sim	Nao	Nao	**1**
PAC104	Sim	Sim	Sim	Sim	Nao	Nao	**1**
PAC105	Sim	Nao	Sim	Não	Nao	Sim	**1**
PAC106	Sim	Sim	Sim	Não	Nao	Nao	**2**
PAC107	Sim	Sim	Sim	Não	Sim	Nao	**2**
PAC108	Sim	Sim	Sim	Não	Sim	Sim	**2**

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Continuação)

ID	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC109	Nao	Sim	Sim	Não	Sim	Sim	**1**
PAC110	Sim	Sim	Sim	Não	Não	Sim	**1**
PAC111	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC112	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC113	Nao	Sim	Sim	Nao	Nao	Nao	**1**
PAC114	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC115	Sim	Sim	Sim	Nao	Nao	Sim	**1**
PAC116	Nao	Sim	Sim	Nao	Nao	Nao	**2**
PAC117	Nao	Nao	Sim	Sim	Sim	Sim	**2**
PAC118	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC119	Sim	Sim	Sim	Nao	Nao	Sim	**1**
PAC120	Sim	Sim	Sim	Nao	Sim	Nao	**1**
PAC121	Nao	Sim	Sim	Nao	Nao	Sim	**2**
PAC122	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC123	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC124	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC125	Nao	Nao	Sim	Nao	Sim	Não	**1**
PAC126	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC127	Nao	Sim	Sim	Sim	Sim	Não	**2**
PAC128	Nao	Sim	Sim	Nao	Nao	Não	**1**
PAC129	Nao	Sim	Sim	Nao	Nao	Não	**1**
PAC130	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC131	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC132	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC133	Nao	Sim	Sim	Sim	Sim	Sim	**2**
PAC134	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC135	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC136	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC137	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC138	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC139	Sim	Sim	Sim	Nao	Sim	Não	**1**
PAC140	Nao	Sim	Sim	Sim	Nao	Não	**1**
PAC141	Nao	Nao	Sim	Sim	Sim	Sim	**1**
PAC142	Sim	Sim	Sim	Nao	Sim	Sim	**2**
PAC143	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC144	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC145	Nao	Sim	Sim	Sim	Sim	Não	**1**
PAC146	Nao	Nao	Sim	Nao	Sim	Sim	**2**
PAC147	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC148	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC149	Nao	Nao	Sim	Sim	Sim	Sim	**1**
PAC150	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC151	Nao	Sim	Sim	Sim	Nao	Sim	**1**

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Continuação)

ID	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC152	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC153	Nao	Nao	Sim	Sim	Nao	Não	**1**
PAC154	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC155	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC156	Sim	Sim	Sim	Sim	Não	Não	**2**
PAC157	Nao	Sim	Nao	Sim	Não	Nao	**2**
PAC158	Nao	Nao	Sim	Sim	Sim	Sim	**1**
PAC159	Sim	Sim	Sim	Nao	Não	Sim	**1**
PAC160	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC161	Nao	Nao	Sim	Sim	Sim	Sim	**2**
PAC162	Nao	Sim	Sim	Sim	Sim	Sim	**2**
PAC163	Sim	Sim	Sim	Nao	Nao	Sim	**2**
PAC164	Sim	Sim	Sim	Nao	Nao	Nao	**2**
PAC165	Nao	Nao	Sim	Sim	Sim	Sim	**1**
PAC166	Nao	Nao	Sim	Nao	Nao	Nao	**2**
PAC167	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC168	Nao	Sim	Sim	Nao	Sim	Nao	**2**
PAC169	Nao	Nao	Sim	Sim	Nao	Nao	**2**
PAC170	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC171	Nao	Nao	Sim	Sim	Nao	Sim	**2**
PAC172	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC173	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC174	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC175	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC176	Nao	Nao	Sim	Sim	Sim	Nao	**1**
PAC177	Sim	Nao	Sim	Sim	Sim	Sim	**1**
PAC178	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC179	Nao	Sim	Sim	Sim	Nao	Sim	**2**
PAC180	Nao	Sim	Sim	Sim	Nao	Sim	**1**
PAC181	Nao	Sim	Sim	Sim	Sim	Nao	**1**
PAC182	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC183	Sim	Sim	Sim	Nao	Nao	Nao	**2**
PAC184	Sim	Nao	Sim	Nao	Nao	Sim	**1**
PAC185	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC186	Nao	Sim	Sim	Sim	Nao	Nao	**1**
PAC187	Nao	Sim	Sim	Nao	Sim	Sim	**2**
PAC188	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC189	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC190	Nao	Sim	Sim	Nao	Sim	Sim	**2**
PAC191	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC192	Nao	Sim	Sim	Nao	Sim	Nao	**2**
PAC193	Sim	Nao	Sim	Nao	Nao	Nao	**2**
PAC194	Nao	Nao	Sim	Nao	Sim	Nao	**2**

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Continuação)

ID	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC195	Nao	Sim	Sim	Nao	Sim	Sim	**2**
PAC196	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC197	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC198	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC199	Nao	Nao	Sim	Nao	Nao	Sim	**1**
PAC200	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC201	Nao	Nao	Sim	Nao	Nao	Nao	**1**
PAC202	Sim	Nao	Sim	Sim	Sim	Sim	**1**
PAC203	Nao	Nao	Nao	Sim	Sim	Nao	**1**
PAC204	Nao	Sim	Sim	Sim	Nao	Sim	**1**
PAC205	Nao	Sim	Sim	Sim	Nao	Sim	**2**
PAC206	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC207	Nao	Nao	Sim	Nao	Nao	Sim	**1**
PAC208	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC209	Nao	Nao	Sim	Nao	Nao	Nao	**1**
PAC210	Sim	Sim	Sim	Sim	Sim	Nao	**2**
PAC211	Nao	Nao	Sim	Sim	Nao	Nao	**1**
PAC212	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC213	Nao	Sim	Sim	Nao	Nao	Nao	**1**
PAC214	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC215	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC216	Nao	Sim	Sim	Nao	Nao	Sim	**2**
PAC217	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC218	Nao	Nao	Sim	Sim	Nao	Sim	**2**
PAC219	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC220	Nao	Nao	Sim	Sim	Sim	Sim	**2**
PAC221	Nao	Nao	Nao	Nao	Sim	Nao	**2**
PAC222	Nao	Sim	Sim	Nao	Nao	Sim	**1**
PAC223	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC224	Nao	Nao	Sim	Nao	Nao	Sim	**1**
PAC225	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC226	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC227	Nao	Nao	Sim	Nao	Nao	Sim	**2**
PAC228	Nao	Nao	Sim	Nao	Sim	Sim	**2**
PAC229	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC230	Sim	Sim	Sim	Nao	Nao	Sim	**2**
PAC231	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC232	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC233	Sim	Nao	Sim	Nao	Nao	Sim	**2**
PAC234	Nao	Nao	Sim	Nao	Nao	Sim	**2**
PAC235	Sim	Sim	Sim	Nao	Nao	Nao	**2**
PAC236	Nao	Sim	Sim	Sim	Sim	Nao	**2**
PAC237	Nao	Sim	Sim	Nao	Sim	Sim	**1**

Tabela 23 – Informações dos pacientes segundo sintomas e comorbidades sslecionadas e seus respectivos *clusters* (Conclusão)

ID	Cansaço	Falta ar	Febre	Cardio	Cerebro	Hipertensão	C
PAC238	Nao	Sim	Sim	Sim	Sim	Sim	**2**
PAC239	Nao	Nao	Sim	Nao	Sim	Sim	**1**
PAC240	Nao	Nao	Nao	Sim	Nao	Nao	**1**
PAC241	Sim	Sim	Sim	Sim	Sim	Sim	**1**
PAC242	Sim	Sim	Sim	Sim	Sim	Sim	**2**
PAC243	Nao	Sim	Sim	Nao	Nao	Nao	**1**
PAC244	Nao	Sim	Sim	Sim	Sim	Sim	**1**
PAC245	Nao	Nao	Sim	Sim	Sim	Sim	**1**
PAC246	Nao	Sim	Sim	Sim	Sim	Nao	**1**
PAC247	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC248	Sim	Sim	Sim	Nao	Nao	Nao	**2**
PAC249	Nao	Sim	Sim	Nao	Sim	Sim	**1**
PAC250	Nao	Sim	Sim	Nao	Sim	Sim	**2**
PAC251	Sim	Nao	Sim	Nao	Sim	Sim	**2**
PAC252	Sim	Nao	Sim	Nao	Nao	Nao	**2**
PAC253	Nao	Sim	Sim	Nao	Nao	Nao	**2**
PAC254	Sim	Sim	Sim	Nao	Sim	Sim	**1**
PAC255	Nao	Sim	Sim	Nao	Sim	Sim	**2**
PAC256	Nao	Nao	Sim	Nao	Nao	Sim	**2**

APÊNDICE G - Pertinências e não-pertinências da modelagem *fuzzy* intuicionista segundo similaridade cosseno dos sintomas selecionados, por Classe Padrão

Tabela 24 - Pertinências e não-pertinências de febre, falta de ar e cansaço, por Classe Padrão (Continua)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	padrão
PAC001	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC002	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC003	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC004	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC005	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC006	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC007	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC008	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC009	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC010	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC011	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC012	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC013	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC014	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC015	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC016	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC017	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC018	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC019	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC020	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC021	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC022	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC023	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC024	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC025	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC026	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC027	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC028	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC029	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC030	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC031	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC032	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC033	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC034	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC035	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**

Tabela 24 - Pertinências e não-pertinências de febre, falta de ar e cansaço, por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe Padrão
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	
PAC036	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC037	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC038	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC039	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC040	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC041	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC042	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC043	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC044	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC045	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC046	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC047	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC048	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC049	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC050	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC051	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC052	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC053	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC054	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC055	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC056	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC057	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC058	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC059	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC060	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC061	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC062	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC063	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC064	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC065	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC066	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC067	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC068	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC069	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC070	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC071	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC072	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC073	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC074	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC085	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC086	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC087	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**

Tabela 24 - Pertinências e não-pertinências de febre, falta de ar e cansaço, por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC088	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC089	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC090	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC091	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC092	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC093	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC094	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC095	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC096	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC097	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC098	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC099	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC100	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC101	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC102	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC103	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC104	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC105	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC106	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC107	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC108	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC109	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC110	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC111	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC112	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC113	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC114	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC115	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC116	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC117	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC118	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC119	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC120	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC121	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC122	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC123	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC124	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC125	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC126	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC127	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC128	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC129	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**

Tabela 24 - Pertinências e não-pertinências de febre, falta de ar e cansaço, por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe Padrão
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	
PAC130	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC131	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC132	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC133	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC134	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC135	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC136	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC137	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC138	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC139	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC140	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC141	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC142	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC143	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC144	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC145	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC146	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC147	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC148	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC149	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC150	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC151	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC152	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC153	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC154	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC155	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC156	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC157	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC158	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC159	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC160	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC161	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC162	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC163	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC164	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC165	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC166	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC167	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC168	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC169	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC170	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC171	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**

Tabela 24 - Pertinências e não-pertinências de febre, falta de ar e cansaço, por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC172	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC173	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC174	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC175	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC176	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC177	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC178	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC179	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC180	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC181	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC182	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC183	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC184	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC185	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC186	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC187	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC188	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC189	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC190	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC191	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC192	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC193	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC194	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC195	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC196	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC197	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC198	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC199	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC200	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC201	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC202	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC203	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC204	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC205	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC206	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC207	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC208	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC209	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC210	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC211	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC212	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC213	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**

Tabela 24 - Pertinências e não-pertinências de febre, falta de ar e cansaço, por Classe Padrão (Conclusão)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC214	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC215	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC216	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC217	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC218	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC219	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC220	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC221	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC222	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC223	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC224	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC225	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC226	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC227	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC228	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC229	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC230	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC231	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC232	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC233	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC234	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC235	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC236	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC237	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC238	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC239	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC240	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC241	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC242	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC243	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC244	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC245	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC246	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC247	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC248	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC249	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC250	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC251	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC252	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC253	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC254	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC255	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC256	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**

APÊNDICE H - Pertinências e não-pertinências da modelagem *fuzzy* segundo similaridade cosseno dos sintomas selecionados, por Classe Padrão

Tabela 25 - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Continua)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC001	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC002	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC003	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC004	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC005	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC006	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC007	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC008	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC009	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC010	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC011	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC012	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC013	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC014	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC015	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC016	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC017	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC018	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC019	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC020	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC021	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC022	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC023	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC024	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC025	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC026	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC027	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC028	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC029	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC030	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC031	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC032	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC033	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC034	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC035	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC036	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**

Tabela 25 - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC037	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC038	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC039	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC040	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC041	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC042	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC043	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC044	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC045	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC046	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC047	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC048	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC049	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC050	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC051	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC052	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC053	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC054	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC055	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC056	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC057	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC058	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC059	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC060	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC061	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC062	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC063	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC064	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC065	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC066	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC067	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC068	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC069	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC070	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC071	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC072	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC073	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC074	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC075	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC076	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC077	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC078	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**

Tabela 25 - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC079	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC080	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC081	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC082	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC083	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC084	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC085	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC086	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC087	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC088	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC089	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC090	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC091	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC092	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC093	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC094	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC095	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC096	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC097	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC098	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC099	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC100	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC101	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC102	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC103	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC104	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC105	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC106	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC107	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC108	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC109	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC110	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC111	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC112	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC113	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC114	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC115	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC116	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC117	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC118	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC119	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC120	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**

Tabela 25 - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC121	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC122	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC123	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC124	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC125	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC126	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC127	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC128	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC129	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC130	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC131	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC132	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC133	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC134	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC135	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC136	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC137	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC138	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC139	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC140	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC141	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC142	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC143	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC144	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC145	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC146	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC147	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC148	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC149	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC150	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC151	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC152	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC153	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC154	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC155	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC156	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC157	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC158	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC159	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC160	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC161	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC162	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**

Tabela 25 - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC163	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC164	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC165	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC166	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC167	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC169	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC170	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC171	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC172	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC173	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC174	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC175	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC176	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC177	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC178	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC179	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC180	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC181	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC182	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC183	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC184	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC185	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC186	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC187	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC188	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC189	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC190	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC191	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC192	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC193	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC194	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC195	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC196	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC197	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC198	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC199	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC200	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC201	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC202	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC203	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC204	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC205	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**

Tabela H - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Continuação)

Id	Febre		Falta ar		Cansaço		Classe Padrão
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	
PAC206	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC207	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC208	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC209	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC210	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC211	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC212	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC213	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC214	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC215	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC216	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC217	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC218	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**2**
PAC219	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC220	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC221	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC222	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC223	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC224	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC225	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC226	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC227	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC228	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC229	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC230	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC231	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC232	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC233	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC234	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC235	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC236	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC237	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC238	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC239	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC240	0,4927	0,4073	0,1315	0,7685	0,7105	0,1895	**1**
PAC241	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC242	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**2**
PAC243	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**1**
PAC244	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC245	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC246	0,4927	0,4073	0,5737	0,3263	0,7105	0,1895	**1**
PAC247	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**

Tabela H - Pertinências e não-pertinências das comorbidades hipertensão, doença cerebrovascular e doença cardiovascular por Classe Padrão (Conclusão)

Id	Febre		Falta ar		Cansaço		Classe
	Pert	Não-pert	Pert	Não-pert	Pert	Não-pert	Padrão
PAC248	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC249	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC250	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC251	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC252	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC253	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**
PAC254	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**1**
PAC255	0,4927	0,4073	0,5737	0,3263	0,2880	0,6120	**2**
PAC256	0,4927	0,4073	0,1315	0,7685	0,2880	0,6120	**2**

APÊNDICE I - Trecho do código em linguagem R, referente à Modelagem *Fuzzy* Intuicionista *C-Means* e com similaridade cosseno

```
#############################################################################
#                                Programação da tese                        #
#                                                                           #
#          Trecho da programação referente ao Fuzzy Intuicionista C-Means   #
#                                       e                                   #
#                  modelagem fuzzy com similaridade cosseno                 #
#                                                                           #
#                          Autora: Carla Cristina Passos Cruz               #
#                          Orientadora: Regina Lanzillotti                  #
#                                Ano: 2024                                  #
#############################################################################



#############################################################################
#
# Leitura da base de dados geral
#
#############################################################################

# Leitura dos pacotes
library(readxl)
library(writexl)
library(dplyr)
library(tidyverse)
library(ggplot2)
library(cluster)
library(factoextra)
library(inaparc)
library(rinfis)


# Leitura da base de dados
base_ifcm   <- readxl::read_excel("Base_final.xlsx")


# Verificação da quantidade de crianças e adolescentes na base
crianca     <- which(base_ifcm$CLASS_ETARIA=="Crianca")
adolescente <- which(base_ifcm$CLASS_ETARIA=="Adolescente")


# Retirando as crianças e adolescentes na base
base_ifcm   <- base_ifcm[-c(crianca,adolescente),]
```

```
# Criação da base com as variáveis numéricas para o ifcm
base_modelo <- base_ifcm[,c(4,6)]
base_modelo <- as.data.frame(base_modelo)


# Método Elbow: quantidade de grupos considerado ideal
options(scipen=999)
fviz_nbclust(base_modelo,kmeans,method = "wss")+
  labs(title="")+
  xlab("Quantidade de clusters")+
  ylab("Soma quadrática da diferença entre o elemento \n de cada grupo em relação ao centroi
de")
#############################################################################



#############################################################################
#
# PARTE 1: APLICAÇÃO DO MODELO FUZZY INTUICIONISTA C-MEANS
#
#############################################################################

#Calculando o fuzzy c-means intuicionista
# Funcao: ifcm()
# x       = base de dados;
# c       = numero de agrupamentos;
# m       = numero mais que 1 (grau de fuzzyficacao)
# epsilon = valor do erro (0.03)
# fgen    = tipo de gerador fuzzy: "yager" ou "sugeno" (padrao do R: "yager")
# alpha   = grau de hesitacao
# lambda  = constante (padrao do R: lambda = 2)

# Com g = 2
ifs2  <- ifcm(x=base_modelo, c=2, m=2, alpha=0.1, fgen="Sugeno", verbose=TRUE)


# Grupos gerados pelo fuzzy intucionista c-means, por paciente
grupos <- ifs2$hardcluster


# Salvando uma base com as informações dos grupos
writexl::write_xlsx(as.data.frame(grupos),"ifs_per_grupos_2.xlsx")
#-----------------------------------------------------------------------------------------------------------


# Leitura e união dos dados dos grupos
nova_base_ifcm2 <- base_ifcm
```

```
nova_base_ifcm2$GRUPOS <- as.numeric(unlist(grupos))
writexl::write_xlsx(as.data.frame(nova_base_ifcm2),"Base_final_grupos2.xlsx")

nova_base_ifcm2 <- as.data.frame(nova_base_ifcm2)
g1 = nova_base_ifcm2[which(nova_base_ifcm2$GRUPOS==1),]
g2 = nova_base_ifcm2[which(nova_base_ifcm2$GRUPOS==2),]
#-----------------------------------------------------------------------------------------------------------


# Informações descritivas por Classe Padrão

#----- Classe Padrão 1
nrow(g1) #Quantidade de pacientes
mean(g1$IDADE) #média das idades
median(g1$IDADE) #mediana das idades
min(g1$IDADE) #idade mínima
max(g1$IDADE) #idade máxima
table(g1$GENERO) #Verificação do gênero
table(g1$FUMANTE) #Verificaçaõ dos fumantes

#----- Classe Padrão 2
nrow(g2) #Quantidade de pacientes
mean(g2$IDADE) #média das idades
median(g2$IDADE) #mediana das idades
min(g2$IDADE) #idade mínima
max(g2$IDADE) #idade máxima
table(g2$GENERO) #Verificação do gênero
table(g2$FUMANTE) #Verificaçaõ dos fumantes
#-----------------------------------------------------------------------------------------------------------


# Verificação das informações por Classe Padrão

#----- Grupo padrao 1
vgp1 <- var(g1$IDADE) #198.3284
n1   <- nrow(g1) #n1 = 172

#----- Grupo padrao 2
n2   <- nrow(g2)
vgp2 <- var(g2$IDADE)

#----- Calculo das variancias dentro e total
var_dentro <- ((vgp1 * n1) + (vgp2 * n2))/(n1+n2)
var_total  <- var(nova_base_ifcm2$IDADE)
indice_estratificacao <- var_dentro/var_total
#-----------------------------------------------------------------------------------------------------------


# Informações obtidas do cluster fuzzy intuicionista c-means
#----- Pertinência
```

```
pertinencia <- ifs2$membership
pertinencia <- as.data.frame(pertinencia)
writexl::write_xlsx(pertinencia,"pertinencias.xlsx")

#----- Não-Pertinência
naopertinencia <- ifs2$nonmembership
naopertinencia <- as.data.frame(naopertinencia)
writexl::write_xlsx(naopertinencia,"naopertinencia.xlsx")

#----- Hesitação
hesitacao <- ifs2$hesitation
writexl::write_xlsx(as.data.frame(hesitacao),"hesitacao.xlsx")
#-------------------------------------------------------------------------------------------------------------


# Verificação das pertinências e não-pertinências por Classe Padrão

#----- Classe Padrão 1
pertg1 <- pertinencia$V1
max(pertg1) #Pertinência máxima
min(pertg1) #Pertinência mínina

#----- Classe Padrão 2
pertg2 <- pertinencia$V2
max(pertg2) #Pertinência máxima
min(pertg2) #Pertinência mínina
#-------------------------------------------------------------------------------------------------------------


# Sintomas e comorbidades por grupos

#----- Classe Padrão 1 - Sintomas
table(g1$CALAFRIO)
table(g1$CANSACO)
table(g1$CORIZA)
table(g1$DIARREIA)
table(g1$DOR_DE_CABECA)
table(g1$DOR_GARGANTA)
table(g1$DOR_CORPO)
table(g1$ESPIRRO)
table(g1$ERUPCAO_CUTANEA)
table(g1$FALTA_APETITE)
table(g1$FALTA_AR)
table(g1$FEBRE)
table(g1$OLFATO_PALADAR)
table(g1$TONTURA)
table(g1$TOSSE)
table(g1$VOMITO)
```

```
#----- Classe Padrão 1 - Comorbidades
table(g1$DIABETES)
table(g1$DOENCA_CARDIOVASCULAR)
table(g1$DOENCA_CEREBROVASCULAR)
table(g1$DOENCA_PULMONAR)
table(g1$DOENCA_RENAL)
table(g1$IMUNOCOMPROMETIDO)
table(g1$DOENCA_obesidade)
#-------------------------------------------------------------------------------------------------------


#----- Classe Padrão 2 - Sintomas
table(g2$CALAFRIO)
table(g2$CANSACO)
table(g2$CORIZA)
table(g2$DIARREIA)
table(g2$DOR_DE_CABECA)
table(g2$DOR_GARGANTA)
table(g2$DOR_CORPO)
table(g2$ESPIRRO)
table(g2$ERUPCAO_CUTANEA)
table(g2$FALTA_APETITE)
table(g2$FALTA_AR)
table(g2$FEBRE)
table(g2$OLFATO_PALADAR)
table(g2$TONTURA)
table(g2$TOSSE)
table(g2$VOMITO)

#----- Classe Pasrão 2 - Comorbidades
table(g2$DIABETES)
table(g2$DOENCA_CARDIOVASCULAR)
table(g2$DOENCA_CEREBROVASCULAR)
table(g2$DOENCA_PULMONAR)
table(g2$DOENCA_RENAL)
table(g2$IMUNOCOMPROMETIDO)
table(g2$DOENCA_obesidade)
#############################################################################



#############################################################################
#
# PARTE 2: APLICAÇÃO DO MODELO FUZZY COM SIMILARIDADE COSSENO
#
#############################################################################


#=================================================
```

```
# CASO 1: SINTOMAS

#=====================================================

# Leitura da base de dados
base_mfi <- read_excel("Base_final_grupos2.xlsx")
base_mfi <- as.data.frame(base_mfi)
#-------------------------------------------------------------------------------------------------------

#----- MODELAGEM FUZZY INTUICIONISTA

# Nova base ajustada com as informações necessárias de sintomas
nova_base_mfi <- as.data.frame(base_mfi)
nova_base_mfi %>% select(ID_PACIENTE,FEBRE,FALTA_AR,CANSACO,GRUPOS) %
>%
  arrange(GRUPOS)

nova_base_mfi$FEBRE <- ifelse(nova_base_mfi$FEBRE=="SIM",1,0)
nova_base_mfi$FALTA_AR <- ifelse(nova_base_mfi$FALTA_AR=="SIM",1,0)
nova_base_mfi$CANSACO <- ifelse(nova_base_mfi$CANSACO=="SIM",1,0)

# Médias das colunas da nova base
xbarra1 <- mean(nova_base_mfi$FEBRE);round(xbarra1,2)
xbarra2 <- mean(nova_base_mfi$FALTA_AR);round(xbarra2,2)
xbarra3 <- mean(nova_base_mfi$CANSACO);round(xbarra3,3)

# Desvios-padrão das colunas da nova base
sd1 = sd(nova_base_mfi$FEBRE);round(sd1,2)
sd2 = sd(nova_base_mfi$FALTA_AR);round(sd2,2)
sd3 = sd(nova_base_mfi$CANSACO);round(sd3,2)

# Padronização da Classe Padrão 1
p1 <- base_mfi[which(base_mfi$GRUPOS==1),]
p1 <- as.data.frame(p1)
p1 <- p1 %>% select(ID_PACIENTE,FEBRE,FALTA_AR,CANSACO,GRUPOS)
p1$FEBRE <- ifelse(p1$FEBRE=="SIM",1,0)
p1$FALTA_AR <- ifelse(p1$FALTA_AR=="SIM",1,0)
p1$CANSACO <- ifelse(p1$CANSACO=="SIM",1,0)
z1 <- matrix(NA, ncol=3, nrow=nrow(p1))
for (i in 1:nrow(p1)){
  z1[i,1] <- (p1[1,2] – xbarra1)/sd1
  z1[i,2] <- (p1[i,3] – xbarra2)/sd2
```

```
  z1[i,3] <- (p1[i,4] – xbarra3)/sd3
}


# Padronização da Classe Padrão 2
p2 <- base_mfi[which(base_mfi$GRUPOS==2),]
p2 <- as.data.frame(p2)
p2 <- p2 %>% select(ID_PACIENTE,FEBRE,FALTA_AR,CANSACO,GRUPOS)
p2$FEBRE <- ifelse(p2$FEBRE==”SIM”,1,0)
p2$FALTA_AR <- ifelse(p2$FALTA_AR==”SIM”,1,0)
p2$CANSACO <- ifelse(p2$CANSACO==”SIM”,1,0)
z2 <- matrix(NA, ncol=3, nrow=nrow(p2))
for (i in 1:nrow(p2)){
  z2[i,1] <- (p2[1,2] – xbarra1)/sd1
  z2[i,2] <- (p2[i,3] – xbarra2)/sd2
  z2[i,3] <- (p2[i,4] – xbarra3)/sd3
}


# Peso (neste caso, o valor foi chutado)
rj = 0.9


# Pertinências Classe Padrão 1
mp1 <- NULL
mp1 <- rj/(1+exp(-z1));round(mp1,2)
# Não-Pertinências Classe Padrão 1
vp1 <- NULL
vp1 <- rj/(1+exp(z1));round(vp1,2)


# Pertinências Classe Padrão 2
mp2 <- NULL
mp2 <- rj/(1+exp(-z2));round(mp2,2)


# Não-Pertinências Classe Padrão 2
vp2 <- NULL
vp2 <- rj/(1+exp(z2));round(vp2,2)
#-------------------------------------------------------------------------------------------------------------


# Matriz com as pertinências e não-pertinências com as classes padrão

#----- Classe Padrão 1

# Pertinências
mcp1 <- matrix(NA, ncol=5,nrow=nrow(mp1))
mcp1 <- as.data.frame(mcp1)
mcp1[,1] <- paste0("PAC",seq(1,nrow(mp1)))
```

```
mcp1[,2] <- mp1[,1]
mcp1[,3] <- mp1[,2]
mcp1[,4] <- mp1[,3]
mcp1[,5] <- "1"
colnames(mcp1) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp1,"ifs_pertinencias_cp1.xlsx")


# Não-Pertinências
mcp1v <- matrix(NA, ncol=5,nrow=nrow(vp1))
mcp1v <- as.data.frame(mcp1v)
mcp1v[,1] <- p1$ID_PACIENTE
mcp1v[,2] <- vp1[,1]
mcp1v[,3] <- vp1[,2]
mcp1v[,4] <- vp1[,3]
mcp1v[,5] <- "1"
colnames(mcp1v) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp1v,"ifs_naopertinencias_cp1.xlsx")
#-------------------------------------------------------------------------------------------------------------


#----- Classe Padrão 2

# Pertinências
mcp2 <- matrix(NA, ncol=5,nrow=nrow(mp2))
mcp2 <- as.data.frame(mcp2)
mcp2[,1] <- paste0("PAC",seq(1,nrow(mp2)))
mcp2[,2] <- mp2[,1]
mcp2[,3] <- mp2[,2]
mcp2[,4] <- mp2[,3]
mcp2[,5] <- "2"
colnames(mcp2) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp2,"ifs_pertinencias_cp2.xlsx")


# Não-Pertinências
mcp2v <- matrix(NA, ncol=5,nrow=nrow(vp2))
mcp2v <- as.data.frame(mcp2v)
mcp2v[,1] <- p2$ID_PACIENTE
mcp2v[,2] <- vp2[,1]
mcp2v[,3] <- vp2[,2]
mcp2v[,4] <- vp2[,3]
mcp2v[,5] <- "2"
colnames(mcp2v) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp2v,"ifs_naopertinencias_cp2.xlsx")
#-------------------------------------------------------------------------------------------------------------


# Média das Pertinências por classe Padrão
mbarrap11 <- mean(mp1[,1]);round(mbarrap11,2)
```

```
mbarrap12 <- mean(mp1[,2]);round(mbarrap12,2)
mbarrap13 <- mean(mp1[,3]);round(mbarrap13,2)
mbarrap21 <- mean(mp2[,1]);round(mbarrap21,2)
mbarrap22 <- mean(mp2[,2]);round(mbarrap22,2)
mbarrap23 <- mean(mp2[,3]);round(mbarrap23,2)


# Desvio-Padrão das Pertinências por classe Padrão
sdbarrap11 <- sd(mp1[,1]);round(sdbarrap11,2)
sdbarrap12 <- sd(mp1[,2]);round(sdbarrap12,2)
sdbarrap13 <- sd(mp1[,3]);round(sdbarrap13,2)
sdbarrap21 <- sd(mp2[,1]);round(sdbarrap21,2)
sdbarrap22 <- sd(mp2[,2]);round(sdbarrap22,2)
sdbarrap23 <- sd(mp2[,3]);round(sdbarrap23,2)


# Média das Não-Pertinências por classe Padrão
vbarrap11 <- mean(vp1[,1]);round(vbarrap11,2)
vbarrap12 <- mean(vp1[,2]);round(vbarrap12,2)
vbarrap13 <- mean(vp1[,3]);round(vbarrap13,2)
vbarrap21 <- mean(vp2[,1]);round(vbarrap21,2)
vbarrap22 <- mean(vp2[,2]);round(vbarrap22,2)
vbarrap23 <- mean(vp2[,3]);round(vbarrap23,42)


# Desvio-Padrão das Não-Pertinências por classe Padrão
sdvbarrap11 <- sd(vp1[,1]);round(sdvbarrap11,2)
sdvbarrap12 <- sd(vp1[,2]);round(sdvbarrap12,2)
sdvbarrap13 <- sd(vp1[,3]);round(sdvbarrap13,2)
sdvbarrap21 <- sd(vp2[,1]);round(sdvbarrap21,2)
sdvbarrap22 <- sd(vp2[,2]);round(sdvbarrap22,2)
sdvbarrap23 <- sd(vp2[,3]);round(sdvbarrap23,2)
#-------------------------------------------------------------------------------------------------------------


# Novo paciente
pc <- c(0,0,1)


# Padronização dos valores do novo paciente
zp1 <- (pc[1] - xbarra1)/sd1;round(zp1,2)
zp2 <- (pc[2] - xbarra2)/sd2;round(zp2,2)
zp3 <- (pc[3] - xbarra3)/sd3;round(zp3,2)


# Pertinências do novo paciente
pp1 = rj/(1+exp(-zp1));round(pp1,2)
pp2 = rj/(1+exp(-zp2));round(pp2,2)
pp3 = rj/(1+exp(-zp3));round(pp3,2)
```

```
# Não-pertinências do novo paciente
vp1 = rj/(1+exp(zp1));round(vp1,2)
vp2 = rj/(1+exp(zp2));round(vp2,2)
vp3 = rj/(1+exp(zp3));round(vp3,2)
#=========================================================


#----- DISTÂNCIA COSSENO

#h = 3 -> quantidade de sintomas
# Cálculo da distância cosseno - classe padrão 1
parte1 = ((mbarrap11*pp1)+(vbarrap11*vp1)) / (sqrt((mbarrap11^2)+(vbarrap11^2))*sqrt((
pp1^2)+(vp1^2)))
parte2 = ((mbarrap12*pp2)+(vbarrap12*vp2)) / (sqrt((mbarrap12^2)+(vbarrap12^2))*sqrt((
pp2^2)+(vp2^2)))
parte3 = ((mbarrap13*pp3)+(vbarrap13*vp3)) / (sqrt((mbarrap13^2)+(vbarrap13^2))*sqrt((
pp3^2)+(vp3^2)))
round((parte1 + parte2 + parte3)/3,2)


# Cálculo do ângulo - classe padrão 1
dist1 <- (parte1 + parte2 + parte3)/3
round(cos(dist1),2) * 100
#-------------------------------------------------------------------------------------------------------

# Cálculo da distância cosseno - classe padrão 2
parte4 = ((mbarrap21*pp1)+(vbarrap21*vp1)) / (sqrt((mbarrap21^2)+(vbarrap21^2))*sqrt((
pp1^2)+(vp1^2)))
parte5 = ((mbarrap22*pp2)+(vbarrap22*vp2)) / (sqrt((mbarrap22^2)+(vbarrap22^2))*sqrt((
pp2^2)+(vp2^2)))
parte6 = ((mbarrap23*pp3)+(vbarrap23*vp3)) / (sqrt((mbarrap23^2)+(vbarrap23^2))*sqrt((
pp3^2)+(vp3^2)))
round((parte4 + parte5 + parte6)/3,2)


# Cálculo do ângulo - classe padrão 2
dist2 <- (parte4 + parte5 + parte6)/3
round(cos(dist2),2) * 100
#=========================================================


#=========================================================

# CASO 2: COMORBIDADES

#=========================================================


# Base de dados com as informações relacionadas as comorbidades selecioandas
```

```
#----- Padronização Classe Padrão 1
p1 <- base_mfi[which(base_mfi$GRUPOS==1),]
p1 <- as.data.frame(p1)
p1c <- p1 %>% select(ID_PACIENTE,DOENCA_hipertensao,DOENCA_CEREBROVASC
ULAR,DOENCA_CARDIOVASCULAR,GRUPOS)
p1c$DOENCA_hipertensao <- ifelse(p1c$DOENCA_hipertensao=="SIM",1,0)
p1c$DOENCA_CEREBROVASCULAR <- ifelse(p1c$DOENCA_CEREBROVASCULAR=
="SIM",1,0)
p1c$DOENCA_CARDIOVASCULAR <- ifelse(p1c$DOENCA_CARDIOVASCULAR=="S
IM",1,0)
z1c <- matrix(NA, ncol=3, nrow=nrow(p1c))
for (i in 1:nrow(p1c)){
  z1c[i,1] <- (p1c[1,2] - xbarra1)/sd1
  z1c[i,2] <- (p1c[i,3] - xbarra2)/sd2
  z1c[i,3] <- (p1c[i,4] - xbarra3)/sd3
}



# Padronização Classe Padrão 2
p2 <- base_mfi[which(base_mfi$GRUPOS==2),]
p2 <- as.data.frame(p2)
p2c <- p2 %>% select(ID_PACIENTE,DOENCA_hipertensao,DOENCA_CEREBROVASC
ULAR,DOENCA_CARDIOVASCULAR,GRUPOS)
p2c$DOENCA_hipertensao <- ifelse(p2c$DOENCA_hipertensao=="SIM",1,0)
p2c$DOENCA_CEREBROVASCULAR <- ifelse(p2c$DOENCA_CEREBROVASCULAR=
="SIM",1,0)
p2c$DOENCA_CARDIOVASCULAR <- ifelse(p2c$DOENCA_CARDIOVASCULAR=="S
IM",1,0)
z2c <- matrix(NA, ncol=3, nrow=nrow(p2c))
for (i in 1:nrow(p2c)){
  z2c[i,1] <- (p2c[i,2] - xbarra1)/sd1
  z2c[i,2] <- (p2c[i,3] - xbarra2)/sd2
  z2c[i,3] <- (p2c[i,4] - xbarra3)/sd3
}


# Peso (neste caso, o valor foi chutado)
rj = 0.9

# Pertinências - Classe Padrão 1
mp1c <- NULL
mp1c <- rj/(1+exp(-z1c));round(mp1c,2)


# Não-Pertinências - Classe Padrão 1
vp1c <- NULL
vp1c <- rj/(1+exp(z1c));round(vp1c,2)
```

```
# Pertinências - Classe Padrão 2
mp2c <- NULL
mp2c <- rj/(1+exp(-z2c));round(mp2c,2)

# Não-Pertinências Classe Padrão 2
vp2c <- NULL
vp2c <- rj/(1+exp(z2c));round(vp2c,2)
#-------------------------------------------------------------------------------------------------------------


#----- Classe Padrão 1

# Matriz de Pertinências
mcp1c <- matrix(NA, ncol=5,nrow=nrow(mp1c))
mcp1c <- as.data.frame(mp1c)
mcp1c[,1] <- p1$ID_PACIENTE
mcp1c[,2] <- mp1c[,1]
mcp1c[,3] <- mp1c[,2]
mcp1c[,4] <- mp1c[,3]
mcp1c[,5] <- "1"
colnames(mcp1c) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp1c,"ifs_pertinencias_comor_cp11.xlsx")

# Matriz de Não-pertinências
mcp1vc <- matrix(NA, ncol=5,nrow=nrow(vp1c))
mcp1vc <- as.data.frame(vp1c)
mcp1vc[,1] <- p1$ID_PACIENTE
mcp1vc[,2] <- vp1c[,1]
mcp1vc[,3] <- vp1c[,2]
mcp1vc[,4] <- vp1c[,3]
mcp1vc[,5] <- "1"
colnames(mcp1vc) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp1vc,"ifs_naopertinencias_comor_cp1.xlsx")
#-------------------------------------------------------------------------------------------------------------


#----- Classe Padrão 2
# Matriz de Pertinências
mcp2 <- matrix(NA, ncol=5,nrow=nrow(mp2c))
mcp2 <- as.data.frame(mcp2)
mcp2[,1] <- p2$ID_PACIENTE
mcp2[,2] <- mp2c[,1]
mcp2[,3] <- mp2c[,2]
mcp2[,4] <- mp2c[,3]
mcp2[,5] <- "2"
colnames(mcp2) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp2,"ifs_pertinencias_comor_cp2.xlsx")

# Matriz de Não-Pertinências
mcp2v <- matrix(NA, ncol=5,nrow=nrow(vp2c))
```

```
mcp2v <- as.data.frame(mcp2v)
mcp2v[,1] <- p2$ID_PACIENTE
mcp2v[,2] <- vp2c[,1]
mcp2v[,3] <- vp2c[,2]
mcp2v[,4] <- vp2c[,3]
mcp2v[,5] <- "2"
colnames(mcp2v) <- c("ID_PACIENTE","FEBRE","FALTA_AR","CANSACO","CP")
writexl::write_xlsx(mcp2v,"ifs_naopertinencias_comor_cp2.xlsx")

# Média das Pertinências por classe Padrão
mbarrap11c <- mean(mp1c[,1]);round(mbarrap11c,2)
mbarrap12c <- mean(mp1c[,2]);round(mbarrap12c,2)
mbarrap13c <- mean(mp1c[,3]);round(mbarrap13c,2)
mbarrap21c <- mean(mp2c[,1]);round(mbarrap21c,2)
mbarrap22c <- mean(mp2c[,2]);round(mbarrap22c,2)
mbarrap23c <- mean(mp2c[,3]);round(mbarrap23c,2)

# Desvio-Padrão das Pertinências por classe Padrão
sdabrrap11c <- sd(mp1c[,1]);round(sdabrrap11c,2)
sdabrrap12c <- sd(mp1c[,2]);round(sdabrrap12c,2)
sdabrrap13c <- sd(mp1c[,3]);round(sdabrrap13c,2)
sdabrrap21c <- sd(mp2c[,1]);round(sdabrrap21c,2)
sdabrrap22c <- sd(mp2c[,2]);round(sdabrrap22c,2)
sdabrrap23c <- sd(mp2c[,3]);round(sdabrrap23c,2)

# Média Não-Pertinências por classe Padrão
vbarrap11c <- mean(vp1c[,1]);round(vbarrap11c,2)
vbarrap12c <- mean(vp1c[,2]);round(vbarrap12c,2)
vbarrap13c <- mean(vp1c[,3]);round(vbarrap13c,2)
vbarrap21c <- mean(vp2c[,1]);round(vbarrap21c,2)
vbarrap22c <- mean(vp2c[,2]);round(vbarrap22c,2)
vbarrap23c <- mean(vp2c[,3]);round(vbarrap23c,2)

# Desvio-Padrão das Não-Pertinências por classe-Padrão
sdvbarrap11c <- sd(mp1c[,1]);round(sdvbarrap11c,2)
sdvbarrap12c <- sd(mp1c[,2]);round(sdvbarrap12c,2)
sdvbarrap13c <- sd(mp1c[,3]);round(sdvbarrap13c,2)
sdvbarrap21c <- sd(mp2c[,1]);round(sdvbarrap21c,2)
sdvbarrap22c <- sd(mp2c[,2]);round(sdvbarrap22c,2)
sdvbarrap23c <- sd(mp2c[,3]);round(sdvbarrap23c,2)
#-------------------------------------------------------------------------------------------------------------


# Novo paciente
pc <- c(0,0,1)

# Padronização dos valores do novo paciente
zp1c <- (pc[1] - xbarra1)/sd1;round(zp1c,2)
zp2c <- (pc[2] - xbarra2)/sd2;round(zp2c,2)
zp3c <- (pc[3] - xbarra3)/sd3;round(zp3c,2)
```

```
# Pertinências do novo paciente
pp1c <- rj/(1+exp(-zp1c));round(pp1c,2)
pp2c <- rj/(1+exp(-zp2c));round(pp2c,2)
pp3c <- rj/(1+exp(-zp3c));round(pp3c,2)

# Não-Pertinências do novo paciente
vp1c <- rj/(1+exp(zp1c));round(vp1c,2)
vp2c <- rj/(1+exp(zp2c));round(vp2c,2)
vp3c <- rj/(1+exp(zp3c));round(vp3c,2)
#-------------------------------------------------------------------------------------------------------------


#----- DISTÂNCIA COSSENO
#h = 3 -> quantidade de Comorbidades

#--- Cálculo da distância cosseno - classe padrão 1
parte1c = ((mbarrap11c*pp1c)+(vbarrap11c*vp1c)) / (sqrt((mbarrap11c^2)+(vbarrap11c^2))
*sqrt((pp1c^2)+(vp1c^2)))
parte2c = ((mbarrap12c*pp2c)+(vbarrap12c*vp2c)) / (sqrt((mbarrap12c^2)+(vbarrap12c^2))
*sqrt((pp2c^2)+(vp2c^2)))
parte3c = ((mbarrap13c*pp3c)+(vbarrap13c*vp3c)) / (sqrt((mbarrap13c^2)+(vbarrap13c^2))
*sqrt((pp3c^2)+(vp3c^2)))
round((parte1c + parte2c + parte3c)/3,2)

#--- Cálculo do ângulo - classe padrão 1
dist1c <- (parte1c + parte2c + parte3c)/3
round(cos(dist1c),2) * 100

#--- Cálculo da distância cosseno - classe padrão 2
parte4c = ((mbarrap21c*pp1c)+(vbarrap21c*vp1c)) / (sqrt((mbarrap21c^2)+(vbarrap21c^2))
*sqrt((pp1c^2)+(vp1c^2)))
parte5c = ((mbarrap22c*pp2c)+(vbarrap22c*vp2c)) / (sqrt((mbarrap22c^2)+(vbarrap22c^2))
*sqrt((pp2c^2)+(vp2c^2)))
parte6c = ((mbarrap23c*pp3c)+(vbarrap23c*vp3c)) / (sqrt((mbarrap23c^2)+(vbarrap23c^2))
*sqrt((pp3c^2)+(vp3c^2)))
round((parte4c + parte5c + parte6c)/3,2)

#--- Cálculo do ângulo - classe padrão 2
dist2c <- (parte4c + parte5c + parte6c)/3
round(cos(dist2c),2) * 100
```