



ANAIS DO II SIMPÓSIO DO CI-IA SAÚDE DA UFMG



**Cristiane dos Santos Dias
Elisa Tuler de Albergaria
Zilma Silveira Nogueira Reis**

Realização



Apoio



Parceria



Financiadores



**Belo Horizonte, MG
2024**



ANAIS DO II SIMPÓSIO DO CI-IA SAÚDE DA UFMG

Cristiane dos Santos Dias
Elisa Tuler de Albergaria
Zilma Silveira Nogueira Reis



**Belo Horizonte, MG
2024**



Esta obra é disponibilizada nos termos da Licença Creative Commons – Atribuição – Não Comercial – Compartilhamento pela mesma licença 4.0 Internacional. É permitida a reprodução parcial ou total desta obra, desde que citada a fonte.

S612 Simpósio CI-IA Saúde da UFMG (2. : 2024 : Belo Horizonte, MG).

Anais do II Simpósio CI-IA Saúde da UFMG [recurso eletrônico] : inteligência artificial responsável no ensino, pesquisa e práticas em saúde. / organizado por Cristiane dos Santos Dias, Elisa Tuler Albergaria, Zilma Silveira Nogueira Reis. - Belo Horizonte : Faculdade de Medicina da UFMG, 2024.

177f.

Requisitos do sistema: Adobe Reader

1. Inteligência Artificial. 2. Saúde. 3. Assistência à saúde. I. Dias, Cristiane dos Santos. II. Albergaria, Elisa Tuler. III. Reis, Zilma Silveira Dias. IV. Anais do II Simpósio do Centro de Inovação em Inteligência Artificial para a Saúde da UFMG.

NLM: W 26.5

Bibliotecário responsável: Marina Nogueira Ferraz. CRB-6/2194



Universidade Federal de Minas Gerais

Reitora

Sandra Regina Goulart Almeida

Vice-Reitor

Alessandro Fernandes Moreira



Coordenador e Pesquisador Responsável do CI-IA Saúde da UFMG e do Comitê Executivo: Virgílio Augusto Fernandes Almeida

Vice-coordenador do CI-IA Saúde e Representante da UNIMED no Comitê Executivo: Antônio Luiz Pinho Ribeiro

Diretor do CI-IA Saúde e Coordenador do CT Saúde: Wagner Meira Júnior

Coordenador de Transferência de Tecnologia do Comitê Executivo UFMG: Gilberto Medeiros Ribeiro

Coordenadora de Educação e Difusão do Conhecimento do Comitê Executivo - UFMG: Zilma Silveira Nogueira Reis

Representante da UFMG no Conselho Técnico-Científico: Mauro Martins Teixeira

Representante no Conselho Técnico-Científico - UNIMED BH: Silvana Márcia Bruschi Kelles

Representante no Conselho Técnico-Científico - UNIMED BH: Leonardo Pereira Florêncio

Representante no Conselho Técnico-Científico - UNIMED BH: Marcos Borin

Gerência de projetos: Fabiana Costa Pereira Peixoto, Letícia Santos Neto

Comissão Organizadora

Cristiane dos Santos Dias - Universidade Federal de Minas Gerais - UFMG

Zilma Silveira Nogueira Reis - Universidade Federal de Minas Gerais - UFMG

Elisa Tuler de Albergaria - Universidade Federal de São João Del Rei - UFSJ

André Felipe de Andrade Santos - Universidade Federal de Minas Gerais - UFMG

André Soares da Silva - Universidade Federal de Minas Gerais - UFMG

Enzo Furtado Goulart - Universidade Federal de Minas Gerais - UFMG

Giovanna Martins Vilas Boas - Universidade Federal de Minas Gerais - UFMG

Lívia Felberg - Universidade Federal de Minas Gerais - UFMG

Mizael Andrade Rodrigues - Universidade Federal de Minas Gerais - UFMG

Comissão Científica

Presidente da Comissão Científica

Cristiane dos Santos Dias - Universidade Federal de Minas Gerais - UFMG

Adriano Lages dos Santos

Instituto Federal de Educação, Ciência e
Tecnologia de Minas Gerais - IFMG

Alexandre Serpa

Universidade Presbiteriana Mackenzie

Altigran Soares da Silva

Universidade Federal do Amazonas - UFAM

André Felipe Leal Bernardes

Fundação Ezequiel Dias - FUNED

Antônio Luiz Pinho Ribeiro

Universidade Federal de Minas Gerais - UFMG

Benigna Maria de Oliveira

Universidade Federal de Minas Gerais - UFMG

Cristiano Xavier Lima

Universidade Federal de Minas Gerais - UFMG

Eduardo Araújo de Oliveira

Universidade Federal de Minas Gerais - UFMG

Elisa Tuler de Albergaria

Universidade Federal de São João Del Rei - UFSJ

Eura Martins Lage

Universidade Federal de Minas Gerais - UFMG

Fabiano Conde Araujo

Universidade Federal de Minas Gerais - UFMG

Fábio Gagliardi Cozman

Universidade de São Paulo - USP

Fernando Martín Biscione

UNIMED-BH

Flávia Ribeiro de Oliveira

Universidade Federal de Minas Gerais - UFMG

Flávio Guimarães da Fonseca

Universidade Federal de Minas Gerais - UFMG

Giovanna Martins Vilas Boas

Universidade Federal de Minas Gerais - UFMG

Gisele Lobo Papa

Universidade Federal de Minas Gerais - UFMG

João Paulo Pereira de Almeida

Universidade Federal de Minas Gerais - UFMG

Juliano de Souza Gaspar

Universidade Federal de Minas Gerais - UFMG

Júlio Guerra Domingues

Universidade Federal de Minas Gerais - UFMG

Leonardo Chaves Dutra da Rocha

Universidade Federal de São João Del Rei - UFSJ

Márcia Gomes Penido Machado

Universidade Federal de Minas Gerais - UFMG

Marcos André Gonçalves

Universidade Federal de Minas Gerais - UFMG

Mariangela Leal Cherchiglia

Universidade Federal de Minas Gerais - UFMG

Mauro Martins Teixeira

Universidade Federal de Minas Gerais - UFMG

Priscila Menezes Ferri Liu

Universidade Federal de Minas Gerais - UFMG

Ricardo João Cruz Correa

Faculdade de Medicina da Universidade do
Porto - FMUP

Silvana Márcia Bruschi Kelles - Pontifícia

Universidade Católica e UNIMED-BH

Virgílio Augusto Fernandes Almeida -

Universidade Federal de Minas Gerais - UFMG

Wagner Meira Júnior

Universidade Federal de Minas Gerais - UFMG

Zilma Silveira Nogueira Reis

Universidade Federal de Minas Gerais - UFMG

Editorial

II Simpósio CI-IA Saúde da UFMG

O Centro de Inovação em Inteligência Artificial para a Saúde da Universidade Federal de Minas Gerais (CI-IA Saúde UFMG) é uma iniciativa estratégica e interdisciplinar dedicada à pesquisa e desenvolvimento de tecnologias avançadas de inteligência artificial (IA) aplicadas ao setor de saúde. Fundado em 2021 com o apoio do edital FAPESP e MCTIC (Fundação de Amparo à Pesquisa do Estado de São Paulo e Ministério da Ciência, Tecnologia, Inovações e Comunicações), o CI-IA Saúde busca estimular avanços científicos e promover o intercâmbio de conhecimento entre pesquisadores, estudantes e profissionais da área.

Anualmente, o simpósio do CI-IA Saúde consolida-se como um espaço de referência para discussões sobre a aplicação ética e responsável da IA na saúde. Em 2024, o evento, realizado em parceria com o Congresso Brasileiro de Informática em Saúde (CBIS'24) na Associação Médica de Minas Gerais, teve como foco central fomentar o debate sobre o uso consciente da IA na superação de desafios críticos no setor. Esse propósito refletiu-se na ampla adesão de participantes, com 378 inscritos e 166 credenciados no dia do evento.

A abertura do simpósio foi marcada pela inspiradora fala do diretor executivo do CI-IA Saúde, Prof. Virgílio Almeida, que ressaltou a importância da colaboração efetiva entre academia, governo e setor privado para que os avanços em IA sejam aplicados de maneira inclusiva e responsável. O professor reforçou a visão multidisciplinar e multi-institucional do CI-IA Saúde e destacou o potencial impacto de soluções de IA para enfrentar desafios nacionais em saúde.

Dentre os destaques da programação, a mesa-redonda “Desafios da IA no ensino, pesquisa e práticas em saúde” reuniu profissionais de referência para discutir a criação de diretrizes para o uso de IA nas universidades e a preparação dos alunos para um ambiente de evolução tecnológica acelerada. Sob a moderação do Prof. Antônio Ribeiro e com as contribuições dos professores Virgílio Almeida e Silvana Kelles, o painel abordou tópicos críticos sobre o papel da IA no ensino e nas práticas de saúde, enfatizando a necessidade de preparar os estudantes para atuarem em um contexto de rápidas inovações tecnológicas.

As sessões dedicadas à pesquisa e inovação abordaram temas sensíveis e de alta relevância, como predição de comportamento suicida em crianças e adolescentes, uso de IA para diagnóstico de dengue e desinformação médica em redes sociais. A participação entusiástica da audiência evidenciou a importância das discussões e o interesse crescente pelo impacto da IA na saúde pública.

O eBook com os Anais do simpósio disponibiliza os 55 resumos científicos aprovados pela revisão por pares e apresentados no evento. Entre estes, foram premiados os três melhores trabalhos e um produto inovador, enquanto os demais foram apresentados em formato de pôster. Os trabalhos premiados abordaram temas de alta complexidade e inovação, como avaliação de prematuridade, diagnóstico de fibrilação atrial, remoção de ruídos em sinais biomédicos e uso de modelos de linguagem para apoio a prescrições médicas. Os comentários dos professores Benigna Oliveira e Ricardo Cruz Correia enriqueceram as discussões, trazendo reflexões sobre os impactos da IA no SUS e a importância da formação ética dos futuros profissionais.

Na parte cultural, o evento proporcionou momentos de celebração e integração para o público presente. Após a mesa de abertura, houve a apresentação do pianista Bruno Cruz e do Coral Vozes da Saúde. Para encerrar o dia, a charanga Psicose, formada por estudantes de Medicina da UFMG, realizou uma animada performance que finalizou o simpósio com entusiasmo.

Agradecemos aos palestrantes, autores, revisores e organizadores pelo empenho e dedicação que fizeram deste evento um sucesso. Que o II Simpósio CI-IA Saúde inspire novas parcerias e avanços na inteligência artificial para a saúde, promovendo um futuro onde tecnologia e ética caminhem juntas em prol de um sistema de saúde mais eficiente e responsável.

Cristiane dos Santos Dias

Presidente do 2º Simpósio do CI-IA Saúde UFMG:

Inteligência Artificial responsável no ensino, pesquisa e práticas em saúde

Membro da Diretoria de Educação e Difusão do Conhecimento do CI-IA Saúde UFMG

Sumário

Agrupamento Hierárquico de Setores Censitários para Estimativas mais Precisas de Riscos à Saúde em Pequenas Áreas	13
AI-Driven Solutions for Low Back Pain: A Pilot Study on Diagnosis and Treatment Planning	16
Análise crítica das métricas de avaliação usando escores preditivos de aprendizado de máquina para desfechos relacionados à covid-19	19
Análise da conformidade aos princípios FAIR em dados de pesquisa em saúde: Adaptação	22
Análise da validade da classificação de exames de retinografia por um modelo de IA na Atenção Primária à Saúde	25
Análise de variações de custos de uma grande operadora de saúde 1 suplementar no pós-covid-19	28
Análise temporal do impacto de variáveis individuais e hospitalares na mortalidade em pacientes com covid-19	31
Aplicação de inteligência artificial no planejamento, avaliação e monitoramento de políticas e programas do Sistema Único de Saúde	34
Aprendizado de máquina para a saúde baseada em valor: predição de quedas em pessoas idosas	37
Aprendizado de máquina para priorização de acesso a procedimentos eletivos na rede especializada do SUS	40
Avaliação do modelo YOLOv8 para detecção de oocistos de Plasmodium vivax em mosquitos do gênero Anopheles sp.	43
Avaliação do uso de IA em estratégia de recuperação de cursistas não concluintes em MOOC: Educa e-SUS APS	46
Avaliando as limitações de redes adversariais gerativas na construção 1 de modelos preditivos relacionados à saúde a partir de dados sintéticos	49
Bioética Aplicada ao Uso da Inteligência Artificial na Assistência à Saúde: Análises e Perspectivas	52

Breast Cancer Subtypes Multi-Layer Classification for Worst Prognosis	55
Characterizing and Understanding Temporal Shifts in Healthcare Data Using the DIS Methodology	58
Combinação de Dados de Inquéritos Populacionais para Aprimorar a Predição dos Indicadores de Saúde por meio de IA	61
Resumo Expandido - Como a Inteligência Artificial está transformando o ambiente hospitalar: Breve relato da experiência da implantação de IA em serviço de farmácia clínica em hospital de alta complexidade	64
Comparison between DCGAN and PGGAN in Synthesizing Magnetic Resonance Images of Frontotemporal Dementia	67
Concepção de dataset de epítomos lineares de células B de organismos do gênero Trypanosoma para treinamento de algoritmos baseados em aprendizado de máquina	70
Convolutional Neural Network for Clinical Diagnosis Support of Spondylolisthesis Through Imaging: Expanded Abstract	73
Convolutional Neural Networks in the Diagnosis of Attention Deficit Hyperactivity Disorder in Children Using Electroencephalogram Signals: A Systematic Review	76
Data Mining associated with electrocardiogram for atrial fibrillation diagnosis: an integrative review	79
Desenvolvimento de tecnologia para o monitoramento da fluência de leitura	82
Educação em Saúde: O uso de Inteligência Artificial e Ambientes Imersivos para apoiar a Aprendizagem	85
Elucidativa: usando modelos de linguagem para explicar resultados de exames complementares	88
Estratégia para Padronização de Dados em Imagens Mamográficas para Análises por Radiômica	91
HOMOGENiSE: Um método para organizar o conhecimento de dados em análise quali-quantitativa	94
Identificação do risco neonatal com algoritmos de inteligência artificial embarcados em um dispositivo: Premie-teste - Inovação	97

In silico screening of p-coumaric acid derivatives with antileishmanial activity using methods coupled to Artificial Intelligence	100
Innovative Solutions for Deaf Patient Care: a Review of Sign Language Recognition Technologies	103
Inteligência Artificial Generativa para Personalização do Cuidado de Saúde Integral: Processo de caracterização dos dados	106
Inteligência Artificial na Fisioterapia Gerontológica	109
Inteligência Artificial para a Saúde e a divulgação do conhecimento de forma responsável: Um Relato de Experiência	112
Learning Bayesian Networks for Coronary Artery Disease Diagnoses	115
LLM assessment of LLM generated patient instructions on how to take medications	118
Mapeando ontologias oncopediátricas com foco em tumores do Sistema Nervoso Central: Uma Revisão de Escopo	121
Melhorando qualidade de modelos preditivos de custo hospitalar por enriquecimento descritivo e geração de dados sintéticos	124
Método Automatizado de Registro de Métricas do Instagram: resumo expandido - Iniciação Científica	127
Métodos de Aprendizagem de Máquina para Previsão da Quantidade Semanal de Casos de Dengue*.....	130
Métodos de Estimação de Expressões Gênicas de Câncer de Mama Baseados em Correlação*	133
Modelagem de ontologia de domínio com foco em tumores do Sistema Nervoso Central em crianças e adolescentes	136
Predição de desfechos em saúde em pequenas áreas	139
Predição do uso de serviços de Urgência e Emergência em Pelotas-RS, 2021-2023	142
Prescrições Médicas Apoiadas por LLMs: Um Estudo Comparativo entre ChatGPT 4-o e LLaMa 3 8B	145

Protótipo de Ferramenta para o Auxílio de Exercícios Fisioterapêuticos	148
Título: RAGECAN: Framework de seleção automática de genes para classificação de amostras com câncer	151
Random Forest para predição de risco cardiovascular em internos no processo de abstinência por uso de álcool e drogas	154
Segmentação automática da próstata em imagens médicas com deep learning por meio da técnica de federated learning	157
SEREIA – Document Store Exploration through Keywords	160
Subset of PAM50 Genes for Breast Cancer Subtypes Classification¹	163
Um Estudo Experimental de Técnicas de Aumentação de Dados para Reconhecimento de Entidades Nomeadas em Domínios Low-Resource	166
Unsupervised Learning for Noise Removal in Biomedical Signals: A Case Study with Autoencoders	169
Uso de Inteligência Artificial para a Prevenção de Acidentes Vasculares Encefálicos	171
Utilização de IA Generativa na Construção do Plano Diretor Estratégico (PDE) em um Hospital Universitário: Uma Experiência Exitosa	174

Agrupamento Hierárquico de Setores Censitários para Estimativas mais Precisas de Riscos à Saúde em Pequenas Áreas

Samuel N. Alves, Beatriz Barbosa, Thania M. Faria, Marisa A. Vasconcelos, Marcos A. Gonçalves, Jussara M. Almeida, Regina Bernal, Deborah Malta, Crizian S. Gomes, Juliana B. de Souza, Fabiano Célio, Laís S. de M. Cardoso, Larissa F. Araújo, Airton T. Stein, Jorge G. V. Mendes

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

Contacting author: samuelnorbertoalves@ufmg.br

Resumo: Este estudo propõe uma estratégia de agrupamento hierárquico de setores censitários para estimar desfechos de saúde em pequenas áreas, por meio da utilização de dados do Censo Demográfico de 2010, realizado pelo IBGE (Instituto Brasileiro de Geografia e Estatística) e do Vigitel - Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (entre 2006 e 2018) em Belo Horizonte, Brasil. O objetivo é reduzir a variabilidade entre setores agrupados conforme o Índice de Vulnerabilidade da Saúde (IVS), utilizado atualmente pela Secretaria Municipal de Saúde, bem como das estimativas de desfechos para estes grupos. Neste resumo, a estratégia foi validada na análise da prevalência de fumantes em dois períodos. Mas esta é geral e pode ser aplicada para diferentes desfechos de saúde, permitindo assim a análise dos dados considerando as particularidades dos vários setores e apoiando tomadas de decisão da gestão local em relação a políticas mais eficazes de saúde pública.

Abstract. This study proposes a hierarchical clustering strategy of census tracts to estimate health outcomes in small areas, through the use of data from the 2010 Demographic Census, carried out by the IBGE (Brazilian Institute of Geography and Statistics), and Vigitel - Surveillance of Risk and Protective Factors for Chronic Diseases by Telephone Survey (from 2006 to 2018) in Belo Horizonte, Brazil. The aim is to reduce the variability of tracts grouped according to the Health Vulnerability Index (IVS), currently in use by the Municipal Health Secretariat, as well as in the estimates of health outcomes for such groups. In this work, the strategy was validated in the analysis of prevalence of smokers in two time periods. Yet, it is general and can be used to study various health outcomes, allowing thus data analysis taking the specificities of different sectors into account, and supporting the decision making of local authorities in effective public health policies.

Palavras-chave: Técnicas de agrupamento; análise de pequenas áreas; estimativas de desfechos de saúde

Nome do projeto: Uso de modelos preditivos de Inteligência Artificial na prevenção de doenças crônicas não transmissíveis (Deborah Malta)

1. CONTEXTO E MOTIVAÇÃO

Acesso a informações atualizadas e precisas em pequenas áreas é fundamental para que autoridades possam tomar suas decisões de gestão em saúde pública direcionando a alocação de recursos de forma adequada à necessidade da população (1). No Brasil, os dados disponíveis para dar suporte a essas decisões são principalmente obtidos do Censo Demográfico e de sistemas de informação e inquéritos de saúde, como o Vigitel - Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico.

A cidade de Belo Horizonte (Minas Gerais, Brasil) apresenta uma particularidade significativa por utilizar o Índice de Vulnerabilidade da Saúde (IVS) como métrica de apoio à descrição das desigualdades sociais e econômicas no território. O IVS é uma medida síntese calculada com base em variáveis do Censo Demográfico de 2010, que captura fatores determinantes em saúde e classifica os setores censitários em quatro níveis de risco à saúde segundo sua vulnerabilidade socioeconômica: baixo, médio, elevado e muito elevado risco (2). Por outro lado, várias coleções

de dados municipais de saúde, como o Vigitel, já passaram por pré-processamento visando a localização espacial, isto é, a associação a setores censitários e, portanto, ao IVS correspondente. Isso possibilita a análise dos dados de saúde por localidade, revelando semelhanças e diferenças no território (3). Porém, devido a diversidades socioeconômicas e à disponibilidade de dados em diferentes regiões da cidade, existe uma alta heterogeneidade dos setores agrupados em um mesmo nível de IVS, especialmente aqueles com maior vulnerabilidade. Isto resulta em grande variabilidade das estimativas de prevalência de desfechos agrupadas por nível de IVS, o que motiva uma reclassificação dos setores em grupos mais homogêneos, visando reduzir a variabilidade das estimativas e torná-las mais confiáveis para fins de gestão de saúde pública.

Nesse contexto, este estudo propõe uma estratégia de análise baseada em técnicas de aprendizado de máquina não-supervisionadas (agrupamento - clustering) para reclassificação dos setores em níveis de IVS mais homogêneos, permitindo assim uma análise mais detalhada dos desfechos em saúde. Espera-se, com isto, prover subsídios para apoiar a tomada de decisão em pequenas áreas no município de Belo Horizonte, Brasil, de modo a permitir a obtenção de um conhecimento mais detalhado acerca da prevalência de desfechos de saúde nos setores censitários da cidade.

1.1 Objetivo

Explorar estratégias de agrupamento de setores censitários utilizando variáveis que compõem o IVS, visando obter grupos de setores mais homogêneos quando comparados às quatro faixas originais de IVS, reduzindo, assim, a variabilidade interna dos resultados destes grupos.

2. ATIVIDADES PRINCIPAIS

Avaliação de diferentes estratégias de agrupamento (clustering), incluindo estratégias hierárquicas, de setores censitários, utilizando os componentes principais do IVS como variáveis de entrada, tendo como alvo de estudo a estimativa da prevalência de fumantes por setor.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais sob parecer nº 6.538.883.

A Secretaria Municipal de Saúde de Belo Horizonte desenvolveu o IVS a partir da análise de componentes principais de dois grupos de variáveis do Censo Demográfico de 2010, um referente às condições socioeconômicas e outro referente às características sanitárias dos setores censitários. A cada um dos setores censitários da cidade foi associado um nível de IVS, considerando os 4 níveis definidos em seção prévia. Neste estudo, os valores dos componentes do IVS foram utilizados para testar diferentes estratégias de (re-)agrupamento destes setores, visando reduzir a variabilidade interna de cada grupo. A melhor estratégia testada consiste em uma abordagem hierárquica na qual o nível mais alto da hierarquia corresponde aos agrupamentos derivados dos 4 níveis originais do IVS. Em seguida, o algoritmo K-Means (4) foi executado para os setores em um mesmo nível de IVS, visando subdividi-lo em grupos mais homogêneos de setores. Este processo foi repetido para cada um dos 4 níveis originais. O índice de Silhouette (5) foi utilizado como métrica de qualidade tanto na escolha da melhor estratégia de agrupamento quanto para determinação do número ideal de grupos para cada nível.

A estratégia proposta resultou em nove grupos (clusters) finais de setores, com variabilidade do IVS significativamente reduzida quando comparada às faixas originais. Para medir a variabilidade, foi utilizado o Coeficiente de Variação (CV), que corresponde à razão entre o desvio-padrão e a média dos dados e mede, nesse caso, a dispersão dos valores de IVS dentro de um agrupamento. Para fins de comparação, ao se considerar as 4 faixas originais, o CV dos valores de IVS associados a setores de uma mesma faixa chegou a 20.5. Para os 9 grupos identificados, o CV máximo reduziu

para 16.0, sugerindo maior homogeneidade entre os setores de um mesmo grupo e, assim, menor variabilidade nas estimativas de risco à saúde associadas a cada grupo.

Para ilustrar este ponto, foi testada a medida de prevalência de fumantes em dois períodos, 2009-2014 e 2015-2018, por meio do uso dos pesos de pós-estratificação da amostra do Vigitel nas quatro faixas IVS e nos nove grupos encontrados (Tabela 1). Com base nos resultados, pode-se realizar uma análise dos grupos que apresentaram maior prevalência. Mais ainda, a análise focada em grupos mais homogêneos permite identificar diferenças relevantes que antes não se destacavam pelo uso de resultados agregados: destaque para as diferenças de prevalência nos três grupos associados aos IVS elevado (EL-0, EL-1, EL-2), anteriormente mascaradas pelo valor médio agregado (EL). O mesmo pode ser dito para as diferenças entre os grupos associados ao IVS muito elevado (ME-0 e ME-1). Com estes resultados, um gestor conseguirá, ao identificar estas diferenças, propor estratégias específicas para cada grupo.

Tabela 1 – Prevalência e número de adultos fumantes nas 4 faixas IVS e 9 grupos de setores para os períodos 2009-2014 e 2015-2018 e o Coeficiente de Variação (CV%) dentro de cada faixa

Faixa de IVS	BA (baixo)		MD (médio)		EL (elevado)			ME (muito elevado)	
Prevalência % (n) - (2009-2014)	11,5 (403)		13,8 (528)		15,9 (227)			16,3 (73)	
Prevalência % (n) - (2015-2018)	9,8 (210)		10,1 (254)		8,5 (104)			12,7 (36)	
CV (%)	20.5		10.4		8.1			13.4	
Cluster	BA-0	BA-1	MD-0	MD-1	EL-0	EL-1	EL-2	ME-0	ME-1
Prevalência % (n) - (2009-2014)	11,1 (220)	12,1 (183)	13,2 (230)	14,3 (298)	17,0 (71)	17,5 (35)	14,9 (121)	19,8 (29)	14,9 (44)
Prevalência % (n) - (2015-2018)	10,1 (143)	9,2 (67)	10,8 (134)	9,5 (120)	9,3 (34)	4,4 (11)	9,1 (59)	15,6 (12)	11,5 (24)
CV (%)	9.1	16.0	6.6	5.1	4.3	7.6	4.4	13.6	7.7

4. DESAFIOS E APRENDIZADOS

O principal desafio foi estabelecer a melhor estratégia de agrupamento para chegar em um resultado que reduzisse a variabilidade, mas pudesse ser utilizada na prática por gestores. Foram testadas outras abordagens, inclusive algumas que desconsideraram os 4 níveis originais de IVS. A abordagem proposta se mostrou superior, estabelecendo grupos com maior homogeneidade. Ela é genérica e pode ser aplicada a outros desfechos na área da saúde (trabalho em andamento). A expectativa é que esta estratégia possa auxiliar gestores públicos a melhor caracterizar as especificidades de diferentes sub-regiões dentro de um mesmo território, direcionando a proposição de ações para saúde pública mais particularizadas e eficazes. Este estudo resultou em um resumo expandido que será apresentado no CBIS'24.

5. REFERÊNCIAS

As referências foram condensadas no link a seguir: https://docs.google.com/document/d/1FRSR41VDVgdzhMvrHA4dZaEZPQyvtv9UwepX_qyMZm9Y/edit?usp=sharing

AI-Driven Solutions for Low Back Pain: A Pilot Study on Diagnosis and Treatment Planning

Neto A. G.¹, Moura R. A.², Cielo G. M., Pegoraro F.¹, D'Alessandro W. B.¹, Herrera S. D.¹

¹Universidade de Gurupi - UNIRG, Gurupi, TO, Brasil.

²Faculdade da Santa Casa de Belo Horizonte, Belo Horizonte, MG, Brasil.

agrinoziogeraldo@gmail.com, rodrigoamoura@gmail.com.br,
fabiopegoraro@unirg.edu.br, saviaherrera@unirg.edu.br

Abstract. Low back pain mainly affects the working-age population, and few specific causes can be identified, making diagnosis difficult and nonspecific. Artificial Intelligence can be a great ally for prognosis, diagnosis, and treatment plans in healthcare. This study describes the development of software aimed at providing prognoses, diagnoses, and treatment suggestions for low back pain with AI support, as well as reporting the functionality and initial limitations through a pilot study. Using data mining, including information described by patients and post-processing of discovered anamnesis patterns (rules), it was possible to develop software for evaluation and intervention in this patient group. Subsequently, a pilot study was initiated with 34 patients residing in the city of Gurupi- TO to test the application's functionality. The software enabled more accurate treatments, diagnoses, and prognoses during the pilot study, directing the patient towards physiotherapeutic intervention based on the presented condition.

Keywords: Low back pain, Artificial Intelligence, Diagnosis, Treatment, Pilot study.

Project name: PROGNOSIS, DIAGNOSIS AND BEST TREATMENT SUGGESTIONS FOR LOW BACK PAIN: A PROPOSAL BASED ON ARTIFICIAL INTELLIGENCE - approved by the Research Ethics Committee of the University of Gurupi UnirG - CAAE:45167321.5.0000.5518.

Links: www.medme.care

1. CONTEXT AND MOTIVATION

Low back pain (LBP) is one of the most common health problems and the number one cause of global disability (1). It is considered a public health issue and is one of the leading causes of physical disability in active populations of industrialized countries, necessitating appropriate measures to control pain and improve the quality of life for these patients. The prevalence of LBP in the adult population is 18.5%, affecting approximately 80% of the world's population at some point in their lives. This project aims to describe the creation of an application to propose diagnoses, prognoses, and treatment suggestions for low back pain with Artificial Intelligence (AI) support, as well as to report the functionality and initial limitations through a pilot study.

2. MAIN ACTIVITIES

The study employed a descriptive developmental research approach conducted through an application created by MEDME.CARE®, utilizing a preliminary database of patients with low back pain, along with information provided by physiotherapists through anamnesis. To demonstrate the application of data mining tasks, a dataset recording the follow-up of 50 patients with a clinical diagnosis of low back pain was utilized. Three algorithms were employed to work with this data: Naive Bayes, Apriori, and J48.

3. METHODOLOGY AND RESULTS

This study employed a descriptive developmental research approach, conducted through an application created by MEDME.CARE®, utilizing a preliminary database of patients presenting with low back pain, along with information provided by physiotherapists through anamnesis conducted from August 2021 to August 2022. This study was approved by the Research Ethics Committee under opinion 45167321.5.0000.5518.

To demonstrate the application of each of the three data mining tasks, a dataset recording the follow-up of 50 patients at the physiotherapy school clinic with a clinical diagnosis of low back pain was utilized.

For the experiment, the Weka Explorer program was adopted. Machine learning can be defined as an area that studies how computers can learn or improve their performance automatically (4).

a. Naive Bayes Algorithm

The algorithm achieved an accuracy of 91.1765%, correctly predicting the class from the provided data. The errors were related to the variables “Pain Duration” and “Fractures,” “Leg Weakness,” and “Physical/Neurological Impairment.”

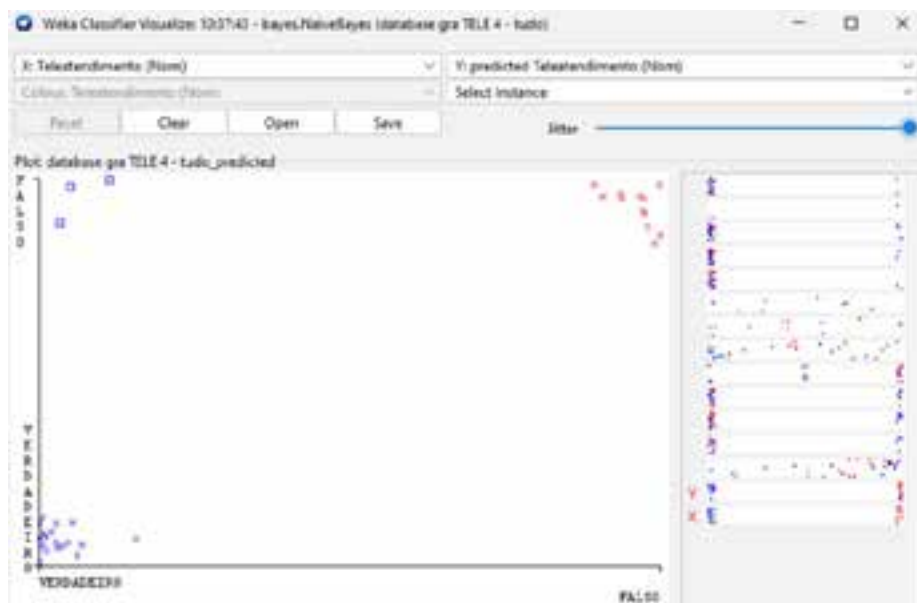
b. Apriori Algorithm

The objective was to group the best attributes for the physiotherapist’s preliminary intervention, with parameters set at 20% Support and 90% Confidence. Two attributes were not correctly associated: Exercise Practice/Leg Numbness and Physical/Neurological Deficiency/Pain Duration/ Leg Numbness.

c. J48 Algorithm

The algorithm correctly classified 88.2353% of the data, failing in 11.7647%. A platform based on evaluative and interventional determinants was created and tested in a pilot study with 34 patients, identifying the functionality and initial limitations of the solution.

Figure 1 - Weka Classifier Visualize



After this analysis, a platform was created, based on evaluative and interventional determinants for this group of patients. In the initial test, a pilot study was conducted with the same number of patients, residents of the city of Gurupi-TO, to identify the functionality and initial limitations of the solution.

The functionality of the application allows the patient to register in advance through a QR code provided by the physiotherapist to start the anamnesis: name, date of birth, email, password, CPF, cell phone, gender, telephone, height, weight, as well as the signature of the terms of use of MEDME. CARE® (Figure 2A).

Figure 2A - Initial description of the anamnesis in the application. **Figure 2B** - Information about the patient's clinical condition. **Figure 2C** - Previously fractured area or correlated pains.



Figures 2A, 2B, 2C source: Authors' own elaboration

Based on the responses, the software can guide the best approach for the patient, which may include scheduling a preliminary assessment and physiotherapeutic follow-up via teleconsultation, or prescribing exercises that the patient can perform at home. A pilot study was conducted with 34 patients from the city of Gurupi-TO, with 22 females and 12 males, average age

≈ 25 years, all with complaints of LBP correlated with work activity, who responded to the application questionnaire.

4. CHALLENGES AND LEARNINGS

During the project's development, the main challenges included the need for algorithm adjustments to improve diagnosis/prognosis, patient understanding of evaluative activities, and technology adaptation for better assessment accuracy.

5. REFERENCES

1. Global Burden of Disease 2021: Findings from the GBD 2021 Study | Institute for Health Metrics and Evaluation [Internet]. [citado 14 de julho de 2024]. Disponível em: <https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study>
2. PERFIL EPIDEMIOLÓGICO DOS PACIENTES COM DOR LOMBAR QUE PROCURAM O SERVIÇO DE TRAUMATOLOGIA E ORTOPEDIA DO HSL-PUCRS - FisoWeb WGate - Referência em Fisioterapia na Internet. [Internet]. [citado 14 de julho de 2024]. Disponível em: https://www.wgate.com.br/conteudo/medicinaesaude/fisioterapia/traumato/dor_lombar.htm
3. WHO guideline for non-surgical management of chronic primary low back pain in adults in primary and community care settings [Internet]. Geneva: World Health Organization; 2023 [citado 14 de julho de 2024]. (WHO Guidelines Approved by the Guidelines Review Committee). Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK599212/>
4. Data Mining: Concepts and Techniques [Internet]. 2011 [citado 14 de julho de 2024]. Disponível em: <https://shop.elsevier.com/books/data-mining-concepts-and-techniques/han/978-0-12-381479-1>

Análise crítica das métricas de avaliação usando escores preditivos de aprendizado de máquina para desfechos relacionados à covid-19

Vanessa G.J. Ventura (doutoranda), Claudio M.V Andrade (doutorando), Jussara M. Almeida, Leonardo Rocha, Marcos A. Gonçalves, Milena S. Marcolino (orientadora)

Universidade Federal de Minas Gerais, Belo Horizonte, MG

nessached@yahoo.com.br, claudiovaliense@gmail.com, jussara@dcc.ufmg.br, lrocha@ufs.br, mgoncalv@dcc.ufmg.br, milenamarc@gmail.com

Resumo: A seleção de métricas de avaliação adequadas ao problema é essencial para avaliar o desempenho de modelos preditivos. Este estudo objetivou comparar métricas tradicionalmente utilizadas na avaliação de modelos preditivos com métricas por classe, e demonstrar como uma avaliação considerando um conjunto de métricas complementares pode produzir avaliações mais robustas, usando modelos preditivos para terapia renal substitutiva (TRS) e mortalidade durante a internação de pacientes com covid-19 como estudo de caso. O modelo foi desenvolvido com base em dados retrospectivos de 17.337 pacientes de 41 hospitais, utilizando Random Forest. Enquanto os valores de acurácia e área sob a curva característica do operador receptor (AUROC) foram 0,909 e 0,930 para TRS, e 0,898 e 0,942 para mortalidade, respectivamente, indicando um excelente desempenho do modelo, analisando outras métricas, para a classe minoritária (necessitou TRS) foram observados precisão 0,662, revocação 0,205 e F1 0,312 para TRS, e para mortalidade (foi a óbito), precisão 0,732, revocação 0,691 e F1 0,711. Tais resultados demonstram que o modelo claramente precisa de melhorias. Dessa forma, usar métricas complementares é essencial para entender adequadamente o desempenho de modelos, especialmente quando há grande desbalanceamento entre as classes.

Abstract. The selection of appropriate evaluation metrics is essential for assessing the performance of predictive models. This study aimed to compare traditionally used metrics in the evaluation of predictive models with metrics for class and demonstrate how an assessment considering a set of complementary metrics can provide more robust evaluations, using a predictive model for kidney replacement therapy (KRT) and mortality during the hospitalization of COVID-19 patients as a case study. The model was developed based on retrospective data from 17,337 patients from 41 hospitals, using Random Forest. While the accuracy and area under the receiving operator characteristic curve (AUROC) values were 0.909 and 0.930 for KRT, and 0.898 and 0.942 for mortality, respectively, indicating excellent model performance, analysis of other metrics for the minority class (KRT requirement) showed precision of 0.662, recall of 0.205, and F1 of 0.312 for KRT, and for mortality (deceased), precision of 0.732, recall of 0.691, and F1 of 0.711. These results indicate that the model clearly requires improvements. Therefore, using complementary metrics is essential for adequately understanding the performance of models, especially when there is a high imbalance between classes.

Palavras-chave: modelos de predição; inteligência artificial; curva ROC; métricas de avaliação

Nome do projeto: Predição de Desfechos Clínicos e Econômicos por Meio de Representações Semânticas MultiModais de Pacientes Resilientes a Drifts Temporais (Marcos André Gonçalves).

1. CONTEXTO E MOTIVAÇÃO

Modelos de predição clínica, construídos a partir de uma ampla gama de dados de pacientes, podem contribuir significativamente para o suporte à decisão na Medicina moderna (1). Entretanto, o desenvolvimento desses modelos requer uma abordagem científica rigorosa, incluindo um projeto experimental minucioso, com validação interna e externa, para garantir resultados confiáveis (1,2). Esses modelos tradicionalmente são avaliados por meio de métricas tais como acurácia e área sob a curva característica do operador receptor (AUROC). Entretanto, a maioria dos estudos na área

de saúde trabalha com problemas onde há um desbalanceamento de classes, o que representa desafio para o aprendizado de modelos supervisionados, que podem apresentar vieses para as classes majoritárias, afetando a efetividade das previsões (3,4). Nesse contexto, avaliar métricas adicionais pode ser interessante para evidenciar possíveis vieses no modelo bem como perda de desempenho em uma das classes, principalmente a classe minoritária, que em problemas de Saúde normalmente é a classe de interesse (5).

1.1 Objetivo

O objetivo deste estudo é comparar métricas tradicionalmente utilizadas na avaliação de modelos preditivos com métricas que captam diferentes perspectivas da performance do modelo e demonstrar como uma avaliação considerando um conjunto de métricas complementares pode produzir uma avaliação mais robusta, usando modelo preditivo para terapia renal substitutiva (TRS) e mortalidade durante a internação de pacientes com covid-19 como estudo de caso.

2. ATIVIDADES PRINCIPAIS

Trabalho colaborativo na seleção das referências bibliográficas, análise e interpretação dos dados e redação do texto científico, por meio de reuniões síncronas e discussões assíncronas.

3. DESENVOLVIMENTO DO TRABALHO

Para o desenvolvimento do modelo, foi usado repositório de dados de coorte retrospectiva, que incluiu pacientes adultos consecutivos (≥ 18 anos), com covid-19 confirmada laboratorialmente, internados em 41 hospitais, de março de 2020 a agosto de 2022. O estudo foi aprovado pela Comissão Nacional de Ética em Pesquisa (CAAE 30350820.5.0000.0008).

Gestantes, pacientes transferidos para outro hospital que não fazia parte do estudo, com história de TRS prévia e em tratamento paliativo foram excluídos. Os desfechos foram a necessidade de TRS e a mortalidade durante a internação. Para esta análise, o método de aprendizado de máquina utilizado foi o Random Forest juntamente com um procedimento experimental 5-fold stratified cross-validation.

As métricas de avaliação foram comparadas considerando as métricas no conjunto de classes, tradicionalmente usadas de forma isolada para avaliar performance de modelos preditivos em saúde, e métricas por classe, que capturam uma perspectiva diferente da efetividade do modelo.

Foram calculadas no conjunto de teste:

- Precisão: número de acertos em relação ao número total de vezes que o algoritmo reconheceu a classe específica;
- Revocação: o quanto o algoritmo conseguiu identificar corretamente a classe alvo;
- Macro F1: média aritmética das pontuações F1 por classe;
- AUROC: representa a probabilidade de um classificador categorizar uma instância positiva escolhida aleatoriamente maior que uma negativa escolhida aleatoriamente.
- Score de Brier: avalia a qualidade da calibração do modelo com pontuação entre 0 e 1, sendo valores próximos a 0 relacionados à proximidade entre as previsões e a realidade. Foram incluídos 17.337 pacientes (idade mediana de 62 anos [intervalo interquartilico 48-71], 45,3% eram mulheres). Foi observado significativo desbalanceamento de dados: 10,1% necessitaram de TRS e 18,1% morreram.

Os resultados das diferentes métricas de avaliação são apresentados na Tabela 1. A acurácia e a AUROC mostram resultados agregados de 0,909 e 0,930 para TRS, e 0,898 e 0,942 para

mortalidade, respectivamente, indicando um excelente desempenho do modelo. De forma similar o Brier está mais próximo de zero, evidenciando boa calibração do modelo. Entretanto, as métricas por classes evidenciam uma diferença na performance para cada classe, com viés para a classe majoritária. Para a classe minoritária (classe TRS = sim) foram observados precisão 0,662, revocação 0,205 e F1 0,312 para TRS, e para mortalidade (classe óbito = sim), precisão 0,732, revocação 0,691 e F1 0,711. Um desempenho bastante aquém, principalmente para as classes minoritárias que são de maior interesse nesse contexto.

Tabela 1. Desempenho do modelo de predição de terapia renal substitutiva e mortalidade intra-hospitalar em pacientes hospitalizados com covid-19.

		Métricas						
Modelo	Classes	Acurácia	AUROC	Macro F1	Brier	Precisão	Revocação	F1
Terapia renal substitutiva	Sim					0,662 (+/- 0,017)	0,205 (+/- 0,016)	0,312 (+/- 0,018)
	Não	0,909 (+/- 0,001)	0,930 (+/- 0,002)	0,632 (+/- 0,009)	0,061 (+/- 0,000)			
Mortalidade	Sim					0,732 (+/- 0,009)	0,691 (+/- 0,014)	0,711 (+/- 0,006)
	Não	0,898 (+/- 0,002)	0,942 (+/- 0,002)	0,825 (+/- 0,004)	0,071 (+/- 0,001)			
						0,933 (+/- 0,003)	0,944 (+/- 0,003)	0,938 (+/- 0,001)

4. DESAFIOS E APRENDIZADOS

Em tarefas preditivas com dados desbalanceados, é importante analisar as métricas de cada classe separadamente para identificar possíveis vieses porque mesmo que o desempenho geral do modelo seja muito bom (medido por AUROC e acurácia), a eficácia na classe minoritária, que em muitos casos é a classe de interesse, pode ser muito baixa. Portanto, ao avaliar a eficácia de um modelo preditivo em cenários com desbalanceamento de classes, é necessário mais cuidado porque podem estar ocultas falhas que não são evidentes na análise considerando métricas tradicionais que possuem um enviesamento para a classe majoritária. Nesse contexto, a avaliação de um conjunto de métricas complementares, incluindo aquelas que avaliam o desempenho por classe individualmente, é fundamental para uma compreensão robusta do desempenho do modelo preditivo. No setor de saúde, esse conhecimento é essencial, pois um erro de classificação na classe de interesse, geralmente a classe minoritária, pode ter um impacto negativo significativo nos resultados dos pacientes, como a falha em realizar uma intervenção necessária para a melhora da qualidade de vida ou da sobrevivência do paciente.

5. REFERÊNCIAS

- Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ*. 2024 Jan 8;384:e074819.
- Riley RD, Pate A, Dhiman P, et al. Clinical prediction models and the multiverse of madness. *BMC Med*. 2023 Dec 18;21(1):502.
- Paiva BBM, Pereira PD, de Andrade CMV, et al. Potential and limitations of machine meta-learning (ensemble) methods for predicting COVID-19 mortality in a large in-hospital Brazilian dataset. *Sci Rep*. 2023 Mar 1;13(1):3463.
- Liu S, Roemer F, Ge Y, et al. Comparison of evaluation metrics of deep learning for imbalanced imaging data in osteoarthritis studies. *Osteoarthritis Cartilage*. 2023 Sep;31(9):1242-1248.
- Adhikari S, Normand SL, Bloom J, et al. Revisiting performance metrics for prediction with rare outcomes. *Stat Methods Med Res*. 2021 Oct;30(10):2352-2366.

ANÁLISE DA CONFORMIDADE AOS PRINCÍPIOS FAIR EM DADOS DE PESQUISA EM SAÚDE: ADAPTAÇÃO DE UMA FERRAMENTA FAIRNESS

Andressa Fiorenzano Nunes (aluna)¹, Cassia Trojahn (coorientador)²,
Silvio César Cazella (orientador)¹

¹Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil.

²Universidade de Toulouse² - Jean Jaurès, Toulouse, Occitânia, França.

andressafn@ufcspa.edu.br, cassia.trojahn@irit.fr, silvioc@ufcspa.edu.br

Resumo: Os princípios FAIR (Findable, Accessible, Interoperable, Reusable) visam garantir a qualidade e reutilização dos dados através de diretrizes claras para o comportamento dos conjuntos de dados. No entanto, sua adoção enfrenta desafios significativos, especialmente em campos como a saúde, onde questões de regulamentação, privacidade e infraestrutura tecnológica são cruciais. Este trabalho tem como objetivo adaptar uma ferramenta para analisar o alinhamento dos dados de pesquisa em saúde aos princípios FAIR. Por meio de um mapeamento sistemático em bases de dados e repositórios, pretende-se identificar ferramentas utilizadas em diversas áreas do conhecimento que servirão como base para a criação de uma nova ferramenta de avaliação FAIRness. Esta nova ferramenta incorporará métricas e funcionalidades adicionais específicas para atender às necessidades do contexto da saúde. Com a adaptação da ferramenta resultante, espera-se contribuir para a promoção das práticas da Ciência Aberta, facilitando a avaliação e o alinhamento dos dados de saúde aos princípios FAIR. Isso promoverá maior transparência, acessibilidade e reutilização dos dados, impactando positivamente a pesquisa em saúde e fomentando novas descobertas e avanços científicos.

Palavras-chave: Princípios FAIR. Ciência Aberta. Dados. FAIRness.

1. CONTEXTO E MOTIVAÇÃO

A Ciência Aberta busca disponibilizar o conhecimento científico de forma aberta, tornando-o acessível e reutilizável (1,2). Dessa forma, planos de gerenciamento de dados de pesquisa têm sido exigidos para assegurar a produção de publicações digitais de alta qualidade e reutilização dos dados de pesquisa. Nesse contexto, os repositórios de dados desempenham um papel crucial, facilitando o armazenamento e a publicação dos conjuntos de dados (3,4).

Para nortear as boas práticas em gestão de dados e melhorar a qualidade dos conjuntos de dados, os princípios FAIR (Findable, Accessible, Interoperable e Reusable) se apresentam como uma recomendação internacionalmente aceita para atingir o objetivo de reutilização de dados por humanos e máquinas, descrevendo um conjunto de atribuições definidas para proporcionar qualidade aos dados e sua capacidade de reuso (5–7).

A adoção e alinhamento aos princípios FAIR são práticas relacionadas à Ciência Aberta prioritárias em política de dados em países que consideram o investimento em pesquisa uma forma direta de desenvolvimento e crescimento do país (8). Mas os princípios ainda apresentam desafios em certas áreas de pesquisa, como as preocupações com confidencialidade, privacidade, interesses comerciais e sensibilidades gerais. Neste contexto, a 'FAIRificação' dos dados tende a proporcionar um ecossistema de compartilhamento confiável e transparente (9).

1.1. Objetivo

Desenvolver e adaptar uma ferramenta específica para analisar o alinhamento dos dados de pesquisa em saúde aos princípios FAIR no contexto brasileiro, utilizando como base ferramentas previamente identificadas.

2. DESENVOLVIMENTO DO TRABALHO

2.1. METODOLOGIA

Através de um mapeamento sistemático em bases de dados e repositórios, serão identificadas as ferramentas especializadas na avaliação do nível de alinhamento aos princípios FAIR de conjuntos de dados. Esse mapeamento considerará diversas características das ferramentas, incluindo suas funcionalidades, métricas de avaliação, metodologias empregadas e limitações, assim como a aplicação em diferentes áreas do conhecimento.

Com base nessa análise, será realizada a adaptação de uma ferramenta FAIRness específica para atender às necessidades dos dados de saúde no Brasil. Considerando as necessidades e desafios específicos dos dados de saúde no contexto brasileiro, incluindo questões relacionadas à regulamentação, privacidade, e infraestrutura tecnológica. A ferramenta irá incorporar características que a tornem adequada para a avaliação de conjuntos de dados de saúde no Brasil. Isso pode incluir a adição de novas métricas, a modificação de metodologias existentes ou a inclusão de funcionalidades adicionais que atendam a requisitos locais.

2.1. RESULTADOS ESPERADOS

Espera-se encontrar ferramentas que possam ser utilizadas como base para a adaptação. Entre as ferramentas automatizadas e manuais já mapeadas estão a ferramenta automatizada F-UJI Automated FAIR Data Assessment Tool, e as ferramentas manuais RDA FAIR Maturity Model e FairDataBR. Com isso será possível facilitar a avaliação e o alinhamento dos dados de saúde aos princípios FAIR, promover transparência, acessibilidade e reutilização dos dados de saúde e impactar positivamente nas pesquisas em saúde, facilitando novas descobertas e avanços científicos.

3. DESAFIOS E APRENDIZADOS

Atualmente, diversas ferramentas têm sido desenvolvidas para avaliar o grau de alinhamento de conjuntos de dados aos princípios FAIR em diversos contextos e áreas do conhecimento. No entanto, há uma lacuna na literatura com relação às ferramentas específicas para dados de pesquisa em saúde. Também é notável a ausência de profissionais com experiência na área FAIR. O conhecimento acerca dos princípios para tornar os dados mais encontráveis, acessíveis, interoperáveis e reutilizáveis, é fundamental para avançar na pesquisa científica e melhorar a colaboração entre diferentes comunidades de pesquisa, a fim de promover uma cultura de compartilhamento de dados e impulsionar o avanço do conhecimento científico.

Através dessa pesquisa será possível disseminar o conhecimento sobre dados FAIR e sobre ferramentas que possibilitem alinhar as plataformas de pesquisa em saúde brasileiras com os princípios, colaborando com a Ciência Aberta.

4. REFERÊNCIAS

1. FIOCRUZ. Fiocruz. [citado 28 de janeiro de 2024]. Ciência Aberta na Fiocruz. Disponível em: <https://portal.fiocruz.br/ciencia-aberta-na-fiocruz>
2. UNESCO. Recomendação da UNESCO sobre Ciência Aberta - UNESCO Digital Library [Internet]. 2021 [citado 15 de novembro de 2023]. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por
3. SAYÃO, L. F., SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre | Revista Eletrônica de Comunicação, Informação & Inovação em Saúde. 30 de junho de 2014 [citado 23 de novembro de 2023]; Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>

4. Henning PC, Ribeiro CJS, Santos LOB da S, Santos PX dos. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. Em Quest. 26 de abril de 2019;389–412.
5. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 15 de março de 2016;3(1):160018.
6. Henning PC, Ribeiro CJS, Sales LF, Moreira LR, Santos LOBDS. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. Pesqui Bras Em Ciênc Informação E Bibl [Internet]. 25 de setembro de 2019 [citado 26 de novembro de 2023]; Disponível em: <https://periodicos.ufpb.br/index.php/pscib/article/view/46969>
7. Ibict N de CS do. Instituto Brasileiro de Informação em Ciência e Tecnologia. [citado 11 de julho de 2024]. Princípios FAIR. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2022/marco-2022/principios-go-fair>
8. UNESCO. Open science outlook 1: status and trends around the world - UNESCO Digital Library [Internet]. 2023 [citado 29 de janeiro de 2024]. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000387324>
9. Almada M, Midão L, Portela D, Dias I, Núñez-Benjumea F, Parra-Calderón CL, et al. A New paradigm in health research: FAIR Data (Findable, Accessible, Interoperable, Reusable). Um Novo Paradigma em Investigação em Saúde: Dados FAIR (Localizáveis, Acessíveis, Interoperáveis, Reutilizáveis) [Internet]. dezembro de 2020 [citado 30 de janeiro de 2024]; Disponível em: <https://digital.csic.es/handle/10261/237>

Análise da validade da classificação de exames de retinografia por um modelo de IA na Atenção Primária à Saúde

Alan Cristian Marinho Ferreira¹, Johan Alejandro Vega Castañeda¹, Rosângela Durso Perillo¹, Mariana Abreu Caporali de Freitas¹, Luiz Carlos Molinari Gomes¹, Alaneir de Fátima Santos¹

¹Faculdade de Medicina da Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

acristianff@gmail.com, alejandrovcast@gmail.com, rosangeladurso.perillo@gmail.com, caporalimariana@gmail.com, luizcarlosmolinari@gmail.com, laines@uol.com.br

Resumo: A Atenção Primária à Saúde (APS) enfrenta desafios devido à alta demanda e subfinanciamento, resultando em longas filas para atendimentos especializados. A inteligência artificial (IA) surge como uma solução promissora, já utilizada no rastreamento de várias doenças. Este estudo avalia a validade da classificação de exames de retinografia por IA, realizada em parceria com municípios de Minas Gerais, comparada à de especialista humano. Desde dezembro de 2023, um modelo de IA foi incorporado para analisar imagens de retina, capturadas por técnicos em unidades básicas de saúde. Foram realizados 920 exames, com 748 incluídos na análise após a exclusão dos inconclusivos. A IA apresentou uma sensibilidade de 67,9% e uma especificidade de 96,9%, indicando alta precisão na identificação de exames normais, mas necessitando de melhorias para identificar exames alterados. A integração da IA mostrou desafios, como a necessidade de aprimorar a sensibilidade do modelo e a importância do treinamento contínuo dos técnicos. A experiência destaca a potencialidade da IA em reduzir a sobrecarga dos especialistas e a importância de avaliações contínuas para maximizar seus benefícios na saúde pública.

Palavras-chave: Saúde Ocular; Inteligência Artificial; Atenção Primária à Saúde, Validade.

Nome do projeto: Incorporação de teleconsultorias no fluxo assistencial do atendimento especializado em Minas Gerais.

1. CONTEXTO E MOTIVAÇÃO

A Atenção Primária à Saúde (APS) tem como atributos o acesso e primeiro contato, integralidade, longitudinalidade e coordenação do cuidado (1). Esses atributos resultam em uma alta carga de atividades e demandas, e, em um contexto de subfinanciamento, representam um grande desafio (2). Na Rede de Atenção à Saúde (RAS), a APS deve referenciar usuários a outros níveis de atenção, identificando suas necessidades e agendando consultas com especialistas (3). Atualmente, a APS enfrenta longas filas para atendimentos especializados com esperas de meses ou até anos, cenário que ocorre, em parte pelo aumento da demanda, em parte pela falta de um sistema de triagem eficaz que poderia priorizar pacientes mais graves e atender outros na própria APS (4). A inteligência artificial (IA) surge como uma ferramenta promissora para resolver esse problema, sendo já utilizada no rastreamento e diagnóstico de várias doenças, como cânceres de pele (5). Modelos de aprendizagem de máquina e processamento de linguagem natural têm ajudado a reorganizar tarefas, automatizar processos e diminuir a sobrecarga dos profissionais de saúde (5). Na rede de atenção oftalmológica, um modelo de IA pode auxiliar na classificação de exames de retinografia realizados por retinógrafos portáteis, categorizando-os em baixa, média e alta probabilidade de alteração (6). Exames de baixa probabilidade de alteração, em teoria, não precisariam entrar na fila para consulta com especialista, podendo ser acompanhados na APS por médicos generalistas (6). Este estudo avalia o impacto da incorporação desse modelo na APS de municípios de Minas Gerais e analisa se sua classificação pode ser usada no processo de triagem dos usuários no acompanhamento da saúde ocular.

1.1 Objetivo

Avaliar a validade da classificação de retinografias realizadas pela IA comparada a realizada por profissional humano.

2. ATIVIDADES PRINCIPAIS

O núcleo de Telessaúde da Faculdade de Medicina da UFMG firmou parceria com alguns municípios de Minas Gerais para reduzir a fila de espera para retinografias. Os exames são realizados em unidades básicas de saúde por técnicos treinados que capturam imagens da retina com retinógrafos portáteis, sem necessidade de dilatação das pupilas. As imagens são enviadas para análise de um especialista, que fornece um relatório indicando se há ou não alterações e quais são elas. Desde dezembro de 2023, um modelo de inteligência artificial baseado em machine learning, o Eyer Maps, foi incorporado para analisar as imagens e classificar a probabilidade de alterações. Mais informações técnicas sobre a IA foram solicitadas à Phelcom, empresa desenvolvedora, no entanto sem retorno até o momento de escrita deste trabalho. Até junho de 2024, foram realizados 920 exames nos municípios parceiros (Contagem, Nova Lima, Santa Luzia e Santana do Riacho). Análises de concordância já foram feitas para verificar a confiabilidade do modelo para reproduzir o diagnóstico do especialista na identificação de exames alterados e normais. Este trabalho pretende estudar a validade do modelo para classificar os pacientes com e sem alterações da retina.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Para a análise de validade (7) necessita-se de um método de avaliação do exame considerado como referência, ou “padrão ouro”. Nesse caso, o padrão de referência foi a classificação feita pelo médico especialista. Excluiu-se exames inconclusivos, resultando em 748 exames analisados. Para comparação, somaram-se os exames com média e alta probabilidade de alteração segundo a IA, sendo esses comparados aos exames alterados de acordo com o profissional. Foi construída uma tabela de contingência 2x2 com os exames alterados e normais conforme o padrão de referência e a análise da IA (Tabela 1). Em seguida calculou-se a sensibilidade e a especificidade da categorização feita pelo modelo de IA. O resumo apresentado utiliza dados do projeto de teleconsultorias realizado pelo Centro de Tecnologia em Saúde da Faculdade de Medicina da Universidade Federal de Minas Gerais, projeto já aprovado pelo comitê de ética em pesquisa. A sensibilidade foi de 67,9%, indicando que a capacidade do modelo de IA para identificar pacientes com alterações da retina foi de quase 7 em cada 10 pacientes. Por outro lado, a especificidade foi de 96,9%, apontando que o modelo identifica adequadamente 96,9% das pessoas sem alterações da retina. Os resultados indicam que o modelo tem alta capacidade de classificar corretamente os indivíduos sem alterações da retina, mas precisa de aprimoramento para identificar com precisão os indivíduos com alterações da retina. Este resultado sugere que o modelo de IA poderia auxiliar na diminuição das filas de oftalmologia, dado que os pacientes com resultado negativo na classificação pela IA tem uma probabilidade de ter uma retina saudável. Desse modo, os pacientes classificados pela IA como alterado, precisariam de avaliação adicional para confirmar se possuem ou não algum tipo de alteração retiniana.

Tabela 1 – Classificação dos exames pelo padrão de referência e o modelo de IA

Classificação pela IA	Classificação pelo padrão de referência		TOTAL
	Normal	Alterado	
Normal	317	135	452
Alterado	10	286	296
TOTAL	327	421	748

Fonte: Elaboração Própria.

4. DESAFIOS E APRENDIZADOS

A incorporação da IA na triagem de exames de retinografia apresentou desafios, como a necessidade de aprimorar a sensibilidade do modelo para identificar com maior precisão os pacientes com alterações da retina. Aprendemos que a integração de novas tecnologias requer treinamento contínuo dos técnicos de saúde e ajustes no modelo para atender às especificidades do contexto da APS. Além disso, o modelo mostrou-se capaz de identificar corretamente quase a totalidade de pacientes sem alterações da retina, destacando sua potencialidade em reduzir a sobrecarga dos especialistas, permitindo um foco maior na atenção dos pacientes que foram classificados como alterados. A experiência reforça a importância de avaliações contínuas e melhorias adaptativas para maximizar os benefícios da IA na saúde pública.

5. REFERÊNCIAS

1. Lima JG, Ligia G, Fausto MCR, Bousquat A, Silva EV. Atributos essenciais da Atenção Primária à Saúde: resultados nacionais do PMAQ-AB. *Saúde em Debate*, v. 42, n. spe1, p. 52–66, set. 2018.
2. Xavier PB, Silva JB, Oliveira LM, Almeida RY, Souza DF. Impactos da COVID-19 no trabalho colaborativo na Atenção Primária à Saúde. *Boletim de Conjuntura (BOCA)*. 2023 Aug 21;15(44):166-81.
3. da Silva Nina LN, Santos MB, Oliveira TL, Ribeiro GF, Carvalho AP. Atenção primária à saúde e redes de atenção à saúde: uma reflexão perante a pandemia. *Saúde Coletiva (Barueri)*. 2022 May 20;12(76):10454-65.
4. Stainkey LA, Seidl IA, Johnson AJ, Tulloch GE, Pain T. The challenge of long waiting lists: how we implemented a GP referral system for non-urgent specialist' appointments at an Australian public hospital. *BMC Health Serv Res* 2010; 10:303.
5. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Rashidi P, et al. Ideal algorithms in healthcare: explainable, dynamic, precise, autonomous, fair, and reproducible. *PLOS Digit Health*. 2022 Jan 18;1(1).
6. Oliveira LES, Silva RM, Santos JG, Almeida FC, Costa LB, Pinto MT, et al. Diagnóstico da retinopatia diabética por inteligência artificial por meio de smartphone. *Rev Bras Oftalmol*. 2024 Feb 9;83.
7. Diagnósticos DA. Avaliação. [place unknown: publisher unknown]; [cited 2024 Jul 25]. Available from: <https://files.cercomp.ufg.br/weby/up/59/o/Modulo2-Avaliacaodetestesdiagnosticos.pdf>.

Análise de variações de custos de uma grande operadora de saúde suplementar no pós-covid-19

Guilherme Dal Bianco^{1,5}, Bruno B. Paiva¹, Guilherme Fonseca², Ricardo Cardoso³, Fernando M. Biscione⁴, Silvana M. B. Kelles⁴, Milena S. Marcolino¹, Jussara M. Almeida¹, Leonardo Rocha², Marcos A. Gonçalves¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, ²Universidade Federal de São João del-Rei, São João del-Rei, MG, ³Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, ⁴Unimed-BH, Belo Horizonte, MG, ⁵Universidade Federal da Fronteira Sul, Chapecó, SC.

guilherme.dalbianco@uffs.edu.br

Resumo: Durante a pandemia de covid-19, adaptações na prestação de serviços médicos foram implementadas, com objetivo principal de reduzir riscos de contágio e de complicações do desenvolvimento da doença. Tais procedimentos, aliados a outros fatores, como a resiliência de pacientes na busca por tratamento médico, geraram alteração nos custos das operadoras de saúde privadas, impactando seus modelos de negócio devido à falta de previsibilidade. Neste trabalho, investigamos a evolução dos custos, especificamente relativos a exames médicos solicitados em regime ambulatorial, durante e após a pandemia de covid-19, de uma grande operadora privada de planos de saúde do Brasil. Analisou-se se houve alterações na demanda por exames ao longo do tempo, considerando o agrupamento dos médicos pelo seu tempo de experiência. As análises apontam que, após a pandemia, a demanda por exames vem sofrendo alterações no seu volume e tipificação, impactando significativamente nos custos da operadora, especialmente devido às mudanças de padrões observadas em grupos de médicos com menor tempo de experiência.

Abstract. During the COVID-19 pandemic, adaptations in the provision of medical services were implemented with the primary objective of reducing contagion risks. Such procedures, combined with other factors, may have generated changes in the costs of private health operators, which could impact their business models due to the lack of predictability. In this study, the main objective was to investigate the evolution of costs, specifically related to medical exams, during and after the COVID-19 pandemic of one of the leading health insurances in Brazil. Our analysis demonstrated that, after the pandemic, the demand for exams has been changing, especially among groups of doctors with less experience.

Palavras-chave: covid-19; análise de dados; análise de custo em saúde

Nome do projeto: Predição de Desfechos Clínicos e Econômicos por Meio de Representações Semânticas Multi-modais de Pacientes Resilientes a Drifts Temporais (Marcos André Gonçalves)

1. CONTEXTO E MOTIVAÇÃO

Apesar do Brasil disponibilizar um sistema público e universal de saúde (SUS), a rede privada representa um mercado de mais de 50 milhões de clientes, com lucro estimado de 3 bilhões de reais somente em 2023 (1). Com tal volume de clientes, a previsibilidade de custos é uma demanda primordial para o planejamento das seguradoras e operadoras de saúde (2). Durante a pandemia de covid-19, a dificuldade de acesso a rede hospitalar devido ao risco de contaminação e a adaptação de serviços (como telemedicina) estimularam alterações no comportamento de usuários e prestadores de serviço. Por exemplo, exames e consultas eletivas foram postergadas para evitar possíveis contaminações. Essas alterações podem ter resultado em mudanças de padrão, de médicos e de pacientes, nas demandas por serviços médicos. Diante desse contexto, o presente estudo visa analisar os custos, que representam um dos maiores montante de gastos de uma grande operadora de planos de saúde. A operadora analisada contém atualmente cerca de 5.000 médicos filiados e com mais de 1 milhão de clientes e é composta por duas categorias de médicos: credenciados e não-credenciados. Os médicos credenciados são prestadores diretos de

serviços médicos e sua contratação é feita a partir de processos seletivos. Já os não-credenciados, são médicos não associados à operadora, mas que têm permissão, conforme Agência Nacional de Saúde (1), para solicitar exames médicos.

1.1 Objetivo

Este estudo visa caracterizar desfechos econômicos resultantes de alteração de padrões de custos de 2019 a 2023 de umas das principais operadoras privadas de planos de saúde do Brasil.

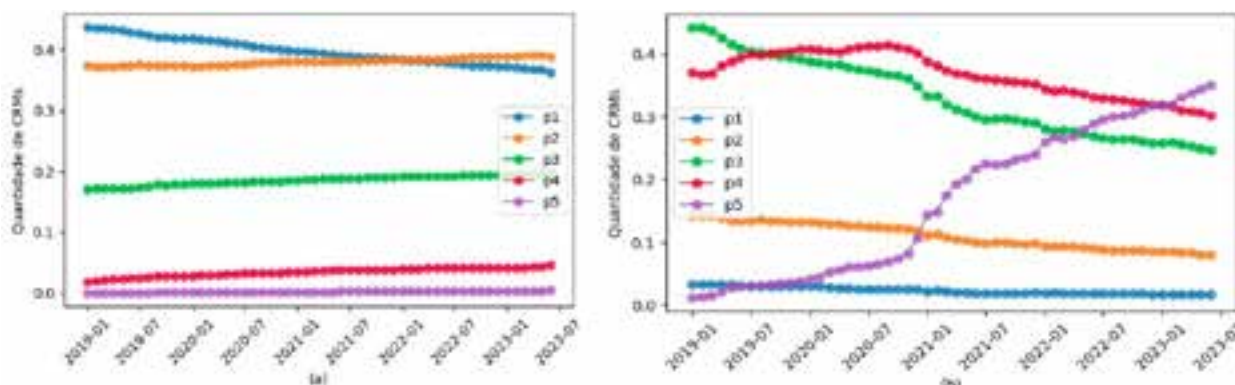
2. ATIVIDADES PRINCIPAIS

As principais atividades deste projeto envolvem a limpeza, análise, e extração de informações de grandes conjuntos de dados envolvendo uma operadora de planos de saúde.

3. DESENVOLVIMENTO DO TRABALHO

Para o desenvolvimento deste estudo, foi utilizada uma coleção de dados contendo informações referentes a médicos, exames solicitados pelos médicos no ambiente do consultório e seus custos, de janeiro de 2019 a julho de 2023. Como os dados são sensíveis e contêm informações sigilosas, foi aplicado um processo de anonimização e serão reportados somente quantitativos normalizados entre 0 (valor mínimo) e 1 (valor máximo). Este estudo foi aprovado no Comitê de Ética e Pesquisa da Universidade Federal de Minas Gerais (CAAE 70801523.7.1001.5149).

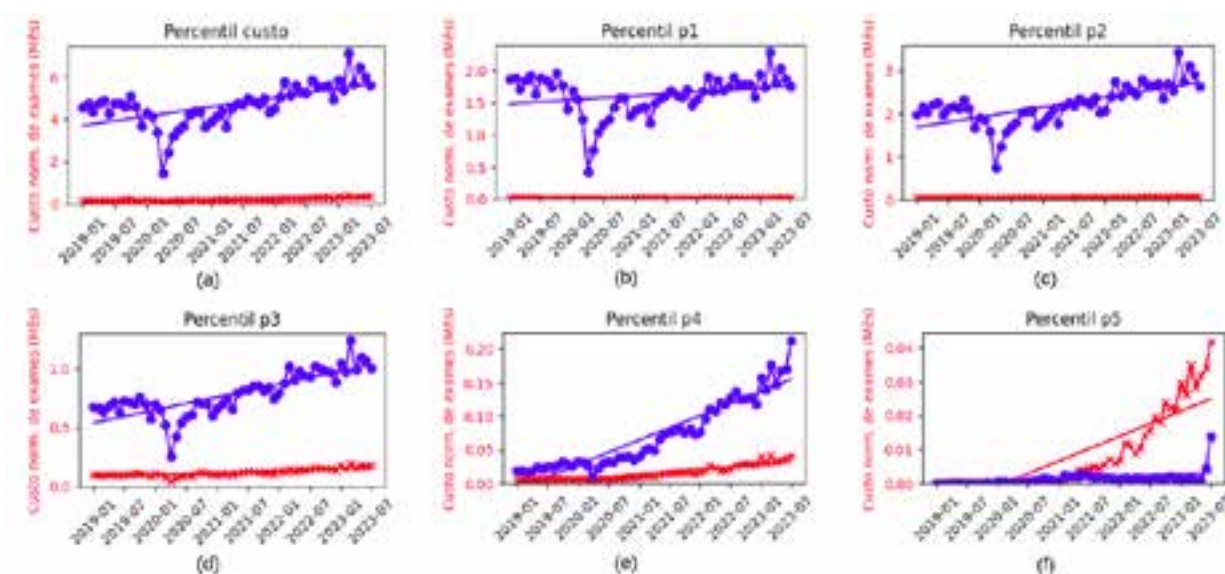
Figura 1 –Número normalizado mensal de CRMs de credenciados (a) e não-credenciados (b).



Inicialmente, foi delineada a evolução quantitativa de médicos credenciados e não-credenciados durante o período analisado, e foi introduzida uma segmentação em grupos de acordo com o tempo de atividade. O número do Conselho Regional de Medicina (CRM) é um inteiro incremental, em que valores menores de CRM são de médicos com maior tempo decorrido desde da graduação, considerado nesta análise como um proxy para experiência. O número do CRM foi adotado como critério para a divisão em 5 percentis com 20% dos CRMs em cada. Por exemplo, o percentil 'p1' agrupa os médicos que estão há mais tempo em atividade. A hipótese é que, a partir da pandemia, médicos com diferentes tempos de experiência têm apresentado diferentes padrões de demanda quanto à solicitação de exames. A Figura 1 apresenta a distribuição de médicos credenciados e não-credenciados de 2019 a 2023 com solicitações por exames. Como pode ser observado, houve um crescimento substancial no número de médicos não-credenciados distintos comparando 2019 e 2023 (cerca de 73% de aumento). Por outro lado, os médicos credenciados apresentaram uma redução de 5% no mesmo período. Assim, essa informação indica que o aumento substancial no número de médicos não-credenciados está impactando diretamente nos custos. No entanto, a Figura 2(a), que apresenta o custo total normalizado com exames dos médicos credenciados (em azul) e não-credenciados (em vermelho) a cada mês, ilustra uma projeção consistente

de crescimento dos credenciados. Para um melhor entendimento, as Figuras 2(b),(c),(d) e (e) reportam os valores de custos de cada grupo. O grupo p2, que contém um número de médicos credenciados aproximadamente constante conforme Figura 1(a), teve uma projeção acentuada de aumento de custos. Outro ponto relevante é o aumento de custos com exames dos grupos p4 e p5 (médicos com menos tempo de atuação). Apesar do número constante de CRMs do grupo p4, o custo com exames aumentou em cerca de 50% se comparado o início e o final da série. O grupo P5 apresentou projeções de aumento substancial de custos, especialmente para o grupo dos não-credenciados. Uma hipótese para a variação de custos em 2020 é a redução na demanda gerada durante a pandemia. Iremos averiguar tais hipóteses na continuação do trabalho.

Figura 2 – Evolução dos custos normalizado geral (a) e em cada percentil(b,c,d,e,f) a cada mês.



4. DESAFIOS E APRENDIZADOS

Este trabalho apresentou um estudo preliminar para entender a alteração de comportamento de custos nos períodos durante e pós pandemia de covid-19, envolvendo umas das principais operadoras nacionais de planos de saúde privados. Para tal, foi necessário processar, analisar grandes volumes de dados e contornar desafios envolvendo a limpeza e remoção de ruídos. As análises demonstraram um aumento acentuado de custos, produzido principalmente por médicos com formação mais recente. Além disso, a demanda por exames de médicos não-credenciados apresentou uma alta projeção de crescimento no último percentil analisado, impactando diretamente nas projeções futuras de custos. Esta informação é relevante para o planejamento das operadoras, uma vez que consolidando a tendência de crescimento deste grupo, poderá gerar impactos substanciais no modelo de negócio. Esperamos a seguir identificar pontualmente o leque de exames com maiores variações e entender as razões para tal aumento.

5. REFERÊNCIAS

1. Agência Nacional de Saúde Suplementar [Internet]. Agência Nacional de Saúde Suplementar. Disponível em: <https://www.gov.br/ans/pt-br>. Acessado em 17/06/2024.
2. Cruz JA, da Cunha MA, de Moraes TP, Marques S, Tuon FF, Gomide AL, de Paula Linhares G. Brazilian private health system: history, scenarios, and trends. BMC Health Services Research. 2022 Dec;22:1-1.

Análise temporal do impacto de variáveis individuais e hospitalares na mortalidade em pacientes com covid-19

Fernanda C. B. Lana¹, Bruno B. M. Paiva¹, Guilherme F. do Nascimento², Leonardo C. D. Rocha²,
Marcos André Gonçalves¹, Carolina C. Marinho¹, Milena S. Marcolino¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil,

²Universidade Federal de São João del-Rei, São João del-Rei, MG, Brasil

{fernandacblana, brunobarbosa.mpaiva, magoncalv, carolinacmarinho, milenamarc} @gmail.com;
guilhermefonseca8426@aluno.ufsj.edu.br; lcrocha@ufsj.edu.br

Resumo: A pandemia por covid-19 representou um desafio para os sistemas de saúde em todo o mundo. A maioria dos estudos realizados carecem de uma análise detalhada do impacto das mudanças temporais sobre a força de associação entre diferentes preditores e importantes desfechos clínicos. Nesse contexto, o presente estudo investiga o impacto de características relacionadas aos pacientes e características hospitalares na predição de mortalidade hospitalar entre pacientes com covid-19 ao longo de três ondas da pandemia. Trata-se de uma coorte retrospectiva multicêntrica de 39 hospitais, que incluiu 16.958 pacientes adultos, internados durante as três ondas da pandemia. A relação entre as variáveis e os desfechos foi avaliada pela correlação de Spearman, combinada com valores de Shapley Additive ExPlanation (SHAP). Entre as características avaliadas, observa-se que a variável idade tornou-se consistentemente menos importante para o desfecho morte ao longo das ondas. Por outro lado, a relação entre saturação de oxigênio e fração inspirada de oxigênio (SatO₂/FiO₂) ganhou importância ao longo das ondas, tonando-se a principal variável para o modelo na onda 3.

Abstract. The COVID-19 pandemic represented a challenge for healthcare systems around the world. Most of the studies lack a detailed analysis of the impact of temporal changes on the strength of association between different predictors and important clinical outcomes. In this context, the present study investigates the impact of patient-related characteristics and hospital characteristics in predicting hospital mortality and transfer to intensive care unit (ICU) among patients with COVID-19 over three waves of the pandemic. This is a multicenter retrospective cohort of 39 hospitals, which included 16,958 adult patients admitted during the three waves of the pandemic. The relationship between variables and outcomes was assessed using Spearman's correlation, combined with Shapley Additive ExPlanation (SHAP) values. Among the evaluated characteristics, it is observed that the variable age became consistently less important for the outcome over the waves. Conversely, oxygen saturation to inspired oxygen fraction (SatO₂/FiO₂) ascended in importance over the waves, becoming the main variable for the model in wave 3.

Palavras-chave: covid-19, fatores de risco, aprendizagem de máquina, mortalidade

Nome do projeto: Predição de Desfechos Clínicos e Econômicos por Meio de Representações Semânticas Multi-modais de Pacientes Resilientes a Drifts Temporais (Marcos André Gonçalves)

1. CONTEXTO E MOTIVAÇÃO

A pandemia de covid-19 constituiu um grande desafio para os sistemas de saúde em todo o mundo (1). Vários fatores, incluindo o surgimento de novas variantes do SARS-CoV-2, flutuações em intervenções de saúde pública e mudanças no comportamento da população, desempenharam um papel crucial nas taxas de infectividade, morbidade e mortalidade, o que possibilitou dividir a pandemia em três ondas (2). Fatores como idade, comorbidades, marcadores inflamatórios, experiência profissional da equipe de saúde em unidades de terapia intensiva (UTI) e o produto interno bruto (PIB) per capita da cidade onde o hospital está localizado, foram encontrados como tendo impacto na mortalidade (2,3). No entanto, a maioria desses estudos não levou em conta como a associação entre diferentes variáveis e desfechos adversos variou ao longo das ondas. Esse estudo procura preencher essa lacuna da literatura.

1.1 Objetivo

Analisar a força da associação entre diferentes variáveis de entrada usadas como preditores (modelos de inteligência artificial - IA) para o desfecho morte, durante três ondas da pandemia por covid-19.

2. ATIVIDADES PRINCIPAIS

Coleta de dados socioeconômicos e hospitalares dos centros participantes do estudo; atualização bibliográfica sobre os últimos estudos envolvendo métricas estatísticas avançadas e doenças infecciosas; análise dos dados e comparação dos resultados entre as ondas da pandemia.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

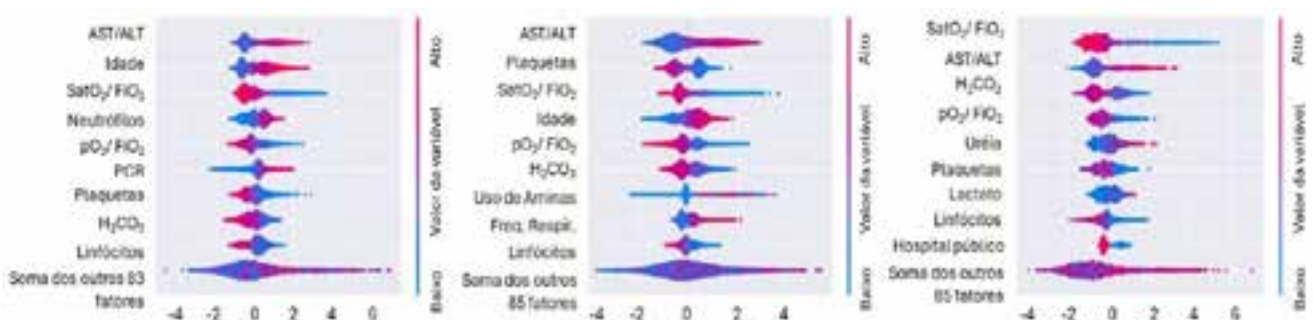
Este é um subestudo do Registro hospitalar multicêntrico nacional de pacientes com doença causada pelo SARS-COV-2 (covid-19), uma coorte retrospectiva com dados de pacientes internados em 39 hospitais brasileiros (4). Foram incluídos pacientes adultos (≥ 18 anos), com confirmação laboratorial da doença, internados entre março de 2020 e agosto de 2022.

O estudo foi aprovado pela Comissão Nacional de Ética em Pesquisa, sob o número CAAE: 30350820.5.0000.0008, com dispensa de aplicação de termo de consentimento livre e esclarecido. Para fins desta análise, a pandemia de covid-19 no Brasil foi dividida em três ondas: a primeira de 10/03/2020 a 14/11/2020, a segunda de 15/11/2020 a 25/12/2021 e a terceira de 26/12/2021 a 03/08/2022 (2). Foram analisadas características próprias do paciente, como comorbidades, sexo e idade; características laboratoriais e clínicas no momento da admissão hospitalar e características hospitalares, como acreditação, fonte de renda hospitalar e produto interno bruto e índice de desenvolvimento humano da cidade de instalação do hospital. Esses potenciais fatores de risco tiveram seu impacto avaliado para o desfecho morte durante cada onda da pandemia por meio da correlação de Spearman, combinados com valores de Shapley Additive ExPlanation (SHAP) obtidos a partir de um modelo lightGBM.

Foram incluídos 16.958 pacientes, sendo 6.246 da primeira onda, 9.385 da segunda e 1.327 da terceira. Na primeira onda ocorreram 1.035 (16,6%) óbitos; na segunda, 1.096 (11,7%) e na terceira, 188 (14,2%) óbitos.

Com relação ao impacto das variáveis no desfecho morte, observamos que características laboratoriais e parâmetros ventilatórios foram os fatores de maior impacto no modelo durante as três ondas. A relação entre aspartato aminotransferase e alanina aminotransferase (AST/ALT) permaneceu com significativa importância durante os três momentos da pandemia. Idade consistentemente se tornou menos relevante para o modelo ao longo das três ondas. Por outro lado, a relação entre saturação de oxigênio e fração inspirada de oxigênio (SatO₂/FiO₂) ganhou importância ao longo das ondas, tonando-se a principal variável para o modelo na onda 3. Na onda 1, ressalta-se também o impacto dos neutrófilos e da proteína C reativa sérica (PCR) como importantes variáveis de risco, o que não persiste nas ondas subsequentes.

Figura 1 Impacto de variáveis no modelo preditivo representadas pelo SHAP para o desfecho morte.



Da esquerda para direita estão representadas as ondas 1, 2 e 3. O eixo Y representa as principais variáveis para o modelo em ordem de importância. O eixo X representa os valores de SHAP. Valores positivos contribuem para o desfecho e valores negativos se associam a redução do desfecho. Tons de vermelho indicam altos valores das variáveis e os tons azuis representam baixos valores.

4. DESAFIOS E APRENDIZADOS

Este estudo envolveu colaboração entre equipes com formações muito distintas (equipe de saúde e equipe da computação). Estabelecer uma comunicação clara que atenda às demandas de ambas as partes é desafiador.

Esse estudo possibilitou um maior conhecimento acerca das métricas de aprendizado de máquina e o impacto que fatores externos podem ter sobre as variáveis preditivas de um modelo de inteligência artificial para diferentes tipos de desfechos de interesse em Saúde. No futuro iremos trabalhar em como identificar automaticamente essas mudanças de importância e adaptar os modelos preditivos (semi-)automaticamente a elas.

5. REFERÊNCIAS

1. World Health Organization. World Health Statistics 2022: monitoring health for the SDGs, sustainable development goals. [Internet]. 2022. Disponível: <http://apps.who.int/bookorders>.
2. Moura EC, Cortez-Escalante J, Cavalcante FV, Barreto IC de HC, Sanchez MN, Santos LMP. Covid-19: temporal evolution and immunization in the three epidemiological waves, Brazil, 2020-2022. Rev Saude Publica. 2022;56:105.
3. Suri A, Singh NK, Perumal V. Association of inflammatory biomarker abnormalities with mortality in COVID-19: a meta-analysis. Bull Natl Res Cent [Internet]. 2022 Dec [cited 2024 Jul 15];46(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/35261542/>
4. Marcolino MS, Ziegelmann PK, Souza-Silva MVR, Nascimento IJB, Oliveira LM, Monteiro LS, et al. Clinical characteristics and outcomes of patients hospitalized with COVID-19 in Brazil: Results from the Brazilian COVID-19 registry. International Journal of Infectious Diseases. 2021 Jun 1;107:300–10.

Aplicação de inteligência artificial no planejamento, avaliação e monitoramento de políticas e programas do Sistema Único de Saúde

Jéssica Ferreira^{1,2}, Denise Valadão¹, Raphael Augusto Teixeira de Aguiar¹

¹Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil.

²Empresa Brasileira de Serviços Hospitalares (EBSERH)/HC-UFMG, Belo Horizonte, MG, Brasil.

jessicaferreira2603@gmail.com, denisevss.enfermagem@gmail.com, raphael.aguiar@gmail.com

Resumo: A inteligência artificial (IA) tem se mostrado promissora na área da saúde, especialmente na gestão de políticas públicas. Assim, este trabalho objetiva identificar e analisar estudos sobre o uso de IA no planejamento, execução, monitoramento e avaliação de políticas e programas de saúde no Sistema Único de Saúde (SUS). A metodologia consistiu em uma revisão narrativa da literatura em cinco bases de dados (MedLine, BVS, Scielo, Embase e Web of Science). Após a aplicação dos critérios de inclusão e exclusão foram identificados quatro estudos relevantes que demonstram o potencial da IA em diversas aplicações, como identificação de risco, predição de indicadores epidemiológicos, apoio à tomada de decisão e avaliação de impacto de políticas. Os resultados apontam para a importância da IA como ferramenta de gestão em saúde, auxiliando na formulação de políticas mais eficazes e personalizadas. No entanto, desafios como a pouca quantidade de estudos sobre o tema, a necessidade de dados de qualidade e a interpretação dos modelos de IA precisam ser superados para sua ampla aplicação.

Palavras-chave: Inteligência Artificial; Políticas de Saúde; Gestão em Saúde; Sistema Único de Saúde.

Nome do projeto: Trabalho produto da disciplina de Tópicos Especiais em Saúde Coletiva: Introdução à Inteligência Artificial em Saúde, ministrada no Programa de Pós-graduação em Saúde Pública da Faculdade de Medicina da UFMG.

1. CONTEXTO E MOTIVAÇÃO

Inteligência Artificial (IA) se trata de um sistema baseado em máquina que infere a partir da entrada que recebe, como gerar saídas como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais (1). Devido ao seu grande potencial de otimização de processos, agilidade de processamento de dados e tomada de decisões, potencial para a melhoria na precisão de diagnósticos e tratamento de doenças, dentre outras aplicações, a IA tem sido amplamente explorada na área da saúde (2).

Na gestão em saúde pública, a IA tem se mostrado promissora ao auxiliar na análise de grandes volumes de dados, identificação de padrões e tendências, e na formulação de políticas mais eficazes e personalizadas (3). No contexto brasileiro, o uso de IA em estratégias governamentais de saúde tem crescido (2). Diante disso, este trabalho busca identificar e analisar estudos sobre iniciativas nacionais que utilizam IA em políticas públicas de saúde.

1.1 Objetivo

O objetivo deste resumo é, por meio de revisão de literatura, identificar e analisar os estudos sobre o uso da IA no planejamento, execução, monitoramento e/ou avaliação de políticas e programas de saúde, com o intuito de aprimorar sua efetividade e impacto no contexto do SUS.

2. ATIVIDADES PRINCIPAIS

As atividades desenvolvidas neste estudo incluíram:

- Participação em aulas expositivas dialogadas, debates e atividades da disciplina;
- Leituras e estudos sobre os conceitos fundamentais, aplicações e desafios da IA em saúde, com foco em políticas públicas;
- Elaboração e apresentação desta revisão de literatura como forma de consolidar os conhecimentos adquiridos e contribuir para a produção científica sobre o tema.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Foi realizada uma revisão narrativa da literatura em cinco bases de dados relevantes: MedLine (PubMed), Biblioteca Virtual em Saúde (BVS), Scientific Electronic Library Online (SciELO), Embase, e Web of Science. Utilizou-se a ferramenta PRISMA para guiar o processo, fluxograma que orienta a eliminação de estudos duplicados e não relacionados ao tema. Os termos em português (e seus correspondentes em inglês) utilizados para a busca foram: (inteligência artificial) OR (aprendizado de máquina) OR AND (planejamento) OR (avaliação) OR (monitoramento) OR (gestão) AND (Sistema Único de Saúde) OR (SUS) OR (saúde pública) OR (políticas públicas de saúde) AND (Brasil).

Buscou-se delimitar o material a ser estudado de modo a responder à seguinte pergunta: Como a IA pode ser utilizada para auxiliar no monitoramento, avaliação e planejamento de políticas de saúde? Assim, os critérios de inclusão adotados foram: estudos que abordam o uso da IA no planejamento, execução, monitoramento e/ou avaliação de políticas de saúde no Brasil. Critérios de exclusão adotados: estudos sem texto completo disponível e artigos de revisão. Inicialmente foram selecionados 413 artigos, dos quais 92 eram duplicados. Após a leitura do título e resumo dos 337 artigos restantes e aplicação dos critérios de inclusão e exclusão, foram selecionados apenas 4 artigos para a leitura completa que foram todos incluídos no estudo. Os principais achados conforme os itens encontrados nos artigos são apresentados na Tabela 1:

Tabela 1 – Principais informações dos artigos selecionados para revisão, 2024.

Autores	Objetivo	Tipo de IA utilizada	Aplicações nas políticas públicas de saúde
Costa Filho et al (2021)	Descrever a trajetória de evolução do GISSA®, ferramenta tecnológica que apoia a tomada de decisão na gestão da Estratégia Saúde da Família no nordeste do Brasil.	<i>Machine learning, Gaussian Naive Bayes (GNB), Random Forest (RF), Decision Tree (DT), e Redes Neurais Artificiais (RNA).</i>	Utiliza dados do SINASC e SIM para empregar classificadores de risco de óbito materno e infantil, auxiliando os gestores de saúde na tomada de decisão. Aplica técnicas de IA para análises preditivas de vigilância epidemiológica utilizando dados do IBGE, INMET e SINAN para recomendar medidas de prevenção e controle da dengue.
Couto et al (2022)	Definir grupos prioritários para a vacinação contra covid-19, com base no risco de morte intra-hospitalar.	<i>Extreme Gradient Boosting e Machine Learning (ML).</i>	Definição de grupos populacionais prioritários para vacinação, com base em diferentes níveis de risco de morte intra-hospitalar por covid-19, auxiliando na formulação de políticas de vacinação.
Rezende et al (2022)	Apresentar a plataforma FLINK, seu processo de construção e implantação na Secretaria de Estado da Saúde de Goiás como instrumento de apoio à tomada de decisão, planejamento, avaliação e gestão em saúde.	<i>Sistemas de Recomendação, Processamento de Linguagem Natural (PLN), Mapa de Palavras em Alta (WordCloud), Chatbot e Machine Learning.</i>	Agrega dados de saúde de vários sistemas e os disponibiliza em um formato que promove acesso à informação e transparência na saúde pública; Suporte à tomada de decisão, monitoramento, avaliação de indicadores, e análise preditiva para prevenção e resposta rápida utilizando IA para identificar padrões e tendências nos dados.
Wink	Mensurar o efeito das	Bayesian Additive	Fornecer dados para a tomada de decisão de
Junior et al (2022)	Unidades de Pronto Atendimento (UPAs) nas taxas de internação por doenças do aparelho respiratório.	Regression Trees (BART).	gestores públicos sobre a alocação de recursos para a implementação, expansão ou redução de UPAs, considerando o efeito que cada uma dessas ações pode ter sobre as taxas de internação por doenças respiratórias.

Os estudos analisados nesta revisão demonstram que o uso da IA já é uma realidade em alguns contextos do SUS, e o seu grande potencial como ferramenta de apoio à gestão em saúde, auxiliando na identificação de riscos, predição de indicadores epidemiológicos, apoio à tomada de decisão e avaliação de impacto de políticas. As diferentes abordagens e técnicas de IA utilizadas nos estudos evidenciam a versatilidade dessa ferramenta e sua capacidade de adaptação a diferentes contextos e necessidades.

4. DESAFIOS E APRENDIZADOS

O número reduzido de estudos encontrados na realidade nacional e a sua heterogeneidade metodológica limita as conclusões e generalizações sobre o uso da IA no contexto do SUS. Apesar disso, os trabalhos identificados demonstram o grande potencial da IA como ferramenta de apoio à gestão e a clínica na execução e monitoramento das políticas de saúde. Os desafios como a necessidade de dados de qualidade, a interpretabilidade dos modelos, implicações éticas no uso da IA e a equidade no acesso aos benefícios dessa tecnologia precisam de avanços para sua aplicação efetiva e responsável na gestão em saúde. Estudos futuros devem ser elaborados para o aprofundamento e acompanhamento dessas fragilidades.

5. REFERÊNCIAS

1. OECD - Organisation for Economic Co-operation and Development. Recommendation of the Council on Artificial Intelligence [Internet]. OECD Legal Instruments; 2024 [cited 2024 Jun 25].
2. Martins Lemes M, Nunes Lopes Espiñeira Lemos A. O uso da inteligência artificial na saúde pela Administração Pública brasileira. *Cadernos Ibero-Americanos de Direito Sanitário*. 2020, 29;9(3):166–82.
3. Couto RC, Pedrosa TMG, Seara LM, Couto CS, Couto VS, Giacomini K, et al. Covid-19 vaccination priorities defined on machine learning. *Revista de Saúde Pública*. 2022 Mar 11;56:11
4. Costa Filho RV, Souza JN de, Andrade LOM de, Oliveira AMB de, Denis JL, Ribeiro LLS, et al. LARIISA: soluções digitais inteligentes para apoio à tomada de decisão na gestão da Estratégia de Saúde da Família. *Ciência & Saúde Coletiva*. 2021 May 28 [cited 2021 Sep 11];26:1701–12.
5. Nogueira A de R, Tibiriça CAG, Moura PMRL, Marques W da S. O uso da inteligência artificial como ferramenta de apoio à gestão das ações em saúde na secretaria de estado da saúde de Goiás. *Revista científica da escola estadual de saúde pública de goiás “Cândido Santiago.”* 2022 Feb 1
6. Wink Junior MV, Santos FL dos, Hoffmann MG, Garcia LP. Impact assessment of emergency care units on hospitalizations for respiratory system diseases in Brazil. *Ciência & Saúde Coletiva* [Internet]. 2022 Aug 15 [cited 2024 Jul 19];27:3627–36.

Aprendizado de máquina para a saúde baseada em valor: predição de quedas em pessoas idosas

Giovana Assis da Matta Machado^{1,3}, Jônatas A. Silva^{2,3}, Thiago Cassiano Campos Abreu^{1,3},
Lênio Correa³, Leticia Rosa Santos Fernandes Faria^{1,3}, Pedro Augusto Gomes Ferreira de
Albuquerque^{3,4}, Leonardo Barbosa Ferreira³, Fernando César Menezes Assunção³,
Geraldo Barcellos de Camargo Neto³, Estevão Alves Valle³, Daniela Castelo Azevedo⁵

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil. ²Universidade Federal de
Ouro Preto, João Monlevade, MG, Brasil ³Mais 60 Saúde, Belo Horizonte, MG, Brasil. ⁴Pontifícia
Universidade Católica de Minas ⁵Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil

Email de contato: estevao@mais60saude.com.br; dca@ufmg.br

Resumo: Este trabalho apresenta o desempenho e comparativo de algoritmos de aprendizado de máquina para predição e classificação de pacientes idosos com risco de ir ao Pronto-Socorro por quedas. Foram utilizados dados de uma plataforma de gestão de saúde, de 12217 indivíduos, incluindo informações sociodemográficas, de mobilidade, comorbidades, uso de medicações e frequência de atendimentos em serviço geriátrico. Estratégias para lidar com o tratamento de dados foram introduzidas visando mitigar o desbalanceamento entre as classes. Os resultados permitem a identificação, priorização e agilidade no cuidado desses pacientes, contribuindo para a cadeia de cuidado baseado em valor.

Abstract: This study presents the performance and comparison of machine learning algorithms for predicting and classifying older persons at risk of emergency room visits due to falls. Data from a health management platform were used, involving 12,127 individuals, including sociodemographic information, mobility, comorbidities, medication use, and frequency of visits to geriatric services. Strategies for data treatment were introduced to mitigate class imbalance. Results enable identification, prioritization, and expedited care of these patients, contributing to the value-based care continuum.

Palavras-chave: aprendizado supervisionado de máquina, idosos fragilizados, modelos de predição, acidentes por quedas

1. CONTEXTO E MOTIVAÇÃO

O acelerado envelhecimento populacional traz grandes desafios para os sistemas de saúde. Dentre os principais problemas, temos as quedas, que são frequentes nessa faixa etária, acometendo até 30% das pessoas com mais 65 anos vivendo na comunidade (1), com consequências graves como a fratura de fêmur (2). Modelos assistenciais baseados em valor, isto é, que consideram os resultados que mais importam aos pacientes em relação aos custos de obtenção desses resultados (3), podem incentivar a melhoria dos cuidados para prevenção de quedas e de suas consequências. Um ponto essencial nessa cadeia de valor é a identificação ágil e eficiente dos pacientes com maior risco de quedas a fim de evitar esse agravo (4).

1.1 Objetivo

Identificar em meio a dados de plataforma digital de gestão de saúde de pacientes idosos (LifeCode, Belo Horizonte, Brasil) e utilizando o aprendizado de máquina supervisionado e profundo, os pacientes em risco de ter atendimento em pronto-socorro (PS) secundário a queda.

2. ATIVIDADES PRINCIPAIS

- Discussão e definição dos algoritmos, técnicas e variáveis de interesse no LifeCode;
- Coleta dos dados e análise exploratória: gráficos; mapa de correlação;

- Tratamento dos dados: valores ausentes ou inconsistentes, categóricos para numéricos com one-hot-encoding, normalização e padronização, balanceamento entre as classes com SMOTE (Synthetic Minority Over-sampling Technique);
- Divisão dos dados para treinamento e teste para os algoritmos de classificação de aprendizado de máquina e treinamento, teste e validação para aprendizado profundo;
- Execução dos algoritmos e análise dos resultados: acurácia, sensibilidade e especificidade, matriz de confusão.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Esse trabalho faz parte do projeto “Análise de dados de plataforma digital de gestão de saúde de pacientes idosos”, cujo objetivo é investigar determinantes de saúde de pacientes idosos acompanhados a nível ambulatorial. Esse projeto foi aprovado pelo Comitê de Ética em Pesquisa, sendo dispensado o TCLE (CAAE: 58545422.5.0000.8787).

A plataforma digital de gestão de saúde de pacientes idosos, LifeCode, é composta por prontuário eletrônico, gestor de tarefas, dashboards integrados ao sistema, que ficam armazenados na nuvem (Microsoft SQL Server) e são disponibilizados sem a identificação individual dos pacientes.

Foram selecionados dados sociodemográficos (idade, sexo, estado civil, bairro e cidade de residência, número de responsáveis, se mora sozinho), de mobilidade (dificuldade para caminhar, de visão, de audição, quedas, índice de massa corporal, velocidade de marcha e circunferência da panturrilha), morbidades (número de diagnósticos, autopercepção da saúde, hipotensão postural, dor crônica, doença de Parkinson, osteoartrite, demência, comprometimento cognitivo leve, diabetes, doença cerebrovascular e etilismo), uso de medicamentos (benzodiazepínicos, sedativos, relaxante muscular, opioides, antipsicóticos, anticolinérgicos, antidepressivos e quantidade total de medicamentos). Por último, foram coletadas informações sobre atendimentos ambulatoriais registrados no LifeCode (tempo de permanência, mediana de consultas médicas, de enfermagem, de fisioterapia, de nutricionista e outros profissionais por ano). A pesquisa considerou o período de 01/07/2020 até 10/07/2024.

Após a coleta e o entendimento da base, foi feito um processamento dos dados para tratamento de outliers, colunas numéricas e colunas categóricas utilizando o one-hot-encoding. Após o processamento, foram encontrados 12.127 pacientes, sendo que 2.112 apresentaram pelo menos uma ida ao PS por queda.

Na tarefa de predição, foram utilizadas técnicas de aprendizado de máquina supervisionada (KNN, Random Forest, Decision Tree, SVG e XGBoost) e rede neural simples, com poucas camadas densas, cujos resultados estão apresentados na Tabela 1. Utilizamos a técnica SMOTE para igualar as amostras no conjunto de treinamento, o que permitiu que os modelos aprendessem mais sobre a classe minoritária, a nossa classe de maior interesse.

Considerou-se a maior sensibilidade, tendo em vista a maior detecção de pacientes com risco de queda. Como as intervenções preventivas têm baixo risco, não se esperam efeitos deletérios nos falsos positivos. O algoritmo permite a sinalização automática do risco, auxiliando os profissionais de saúde e assim contribuindo para a linha de cuidados focados em valor.

Tabela 1 - Resultados da classificação para cada modelo

Algoritmo	Acurácia	Precisão	Sensibilidade	Especificidade	F1-Score
<i>K-Neighbors Classifier</i>	0,74	0,27	0,27	0,84	0,27
<i>Random Forest Classifier</i>	0,82	0,49	0,25	0,94	0,33
<i>DecisionTreeClassifier</i>	0,75	0,29	0,32	0,84	0,3
<i>SVC</i>	0,73	0,33	0,52	0,78	0,4
<i>XGBClassifier</i>	0,75	0,36	0,56	0,79	0,44
<i>Rede Neural Densa</i>	0,70	0,33	0,66	0,71	0,44

4. DESAFIOS E APRENDIZADOS

A plataforma LifeCode permite a coleta sistemática de informações de pacientes idosos e facilita a tarefa de predição. A presença de campos preenchidos manualmente, com possível subnotificação de dados, levam a algumas inconsistências e podem prejudicar os algoritmos. O desbalanceamento de classes revelou um grande desafio no nosso trabalho.

A disponibilidade de grande volume de dados e ferramentas de inteligência artificial tendem a beneficiar cada vez mais o manejo de saúde populacional. Permite a identificação, priorização e agilidade no cuidado daqueles que mais precisam, promovem a sustentabilidade de sistemas de saúde e o bem-estar das pessoas.

5. REFERÊNCIAS

1. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev.* 2012 Sep;12:CD007146. doi: 10.1002/14651858.CD007146.pub3.
2. Azevedo DC, Hoff LS, Kowalski SC, Faria RF, Dias RS, Pereira AC, et al. Risk factors for osteoporotic hip fracture among community-dwelling older adults: a real-world evidence study. *Adv Rheumatol.* 2024;64:8. doi: 10.1186/s42358-024-00350-6.
3. Porter PM, Teisberg EO. *Redefining Health Care: Creating Value-Based Competition on Results.* Boston: Harvard Business School Press; 2006.
4. Das A, Dhillon P. Application of machine learning in measurement of ageing and geriatric diseases: a systematic review. *BMC Geriatr.* 2023; 23:841. doi: 10.1186/s12877-023-04477-x.
5. TensorFlow Core. Classificação em dados desbalanceados. Disponível em: https://www.tensorflow.org/tutorials/structured_data/imbalanced_data?hl=pt-br. Acesso em: 19 jul. 2024.

Aprendizado de máquina para priorização de acesso a procedimentos eletivos na rede especializada do SUS

Fernanda GS Cordeiro¹, Kleber S Araújo²

¹Mestranda em Saúde Digital UERJ e Coordenadora da Saúde Digital IMIP, Recife, PE, Brasil

²Diretor de Inovação IMIP, Recife, PE, Brasil.

fernanda.cordeiro@imip.org.br, kleberaraujo@imip.org.br

Resumo: As Secretarias estaduais e municipais de saúde têm como atribuição facilitar o fluxo assistencial dos pacientes do SUS, tendo a atenção primária a coordenação do cuidado da rede de atenção à saúde. Protocolos de acesso à rede especializada auxiliam na tomada de decisão para referenciamento para agendamentos de procedimentos eletivos. Devido à desproporção entre oferta e demanda, frequentemente há filas de espera em diversas especialidades, sendo necessário priorizar usuários com base em sua condição clínica ou determinantes sociais de saúde. Não identificamos uma classificação sistematizada de abrangência nacional para priorização dos agendamentos para consultas eletivas, similar à classificação de risco de Manchester para atendimentos de urgência. Propomos um sistema baseado em aprendizado de máquina para facilitar a avaliação do risco de agravamento da condição clínica do usuário de saúde, durante o período de espera pelo atendimento eletivo, e o agendamento equânime de consultas especializadas. Palavras-chave: Algoritmos, Equidade, Atenção Secundária à Saúde.

Abstract. State and municipal health departments facilitate patient care flow within SUS, with primary care coordinating the health network. Access protocols to specialized care aid in referral and scheduling elective procedures. Due to supply and demand imbalances, waitlists often exist, requiring prioritization based on clinical condition or social determinants of health. There is no nationwide systematic classification for prioritizing elective appointments, like the Manchester Triage for urgent care. We propose a machine learning-based system to assess the risk of clinical condition worsening during the wait for elective care and ensure equitable scheduling of specialized consultations.

Key words: Algorithms, Equity, Secondary Care.

1. CONTEXTO E MOTIVAÇÃO

Considerando a desproporção entre a oferta de consultas e exames da rede especializada do SUS e a demanda por estes serviços, é frequente a atuação de uma equipe de regulação de ofertas assistenciais, para que haja um processo de priorização no acesso aos serviços da rede especializada de atenção à saúde. Até onde chega o nosso conhecimento, não há uma classificação de risco sistematizada para priorização do agendamento dos atendimentos de consultas eletivas de abrangência nacional ou internacional, à semelhança do que acontece com os atendimentos de urgência, como o sistema de classificação de risco de Manchester.

Em nossos levantamentos preliminares, identificamos a existência de protocolos clínico-assistenciais, assim como de protocolos de acesso à rede especializada de abrangência regional e nacional, que têm como proposta apoiar a tomada de decisão do profissional solicitante do agendamento, assim como apoiar as equipes da regulação, quanto a análise de conformidade do encaminhamento (referência) para a rede de atenção especializada¹⁻⁵.

Entretanto, mesmo após selecionar a população com indicação de acesso à rede especializada, filas de espera são observadas e se faz necessário identificar quais usuários que mesmo tendo indicação para consulta em determinada especialidade devem ser priorizados por condição clínica ou determinantes sociais de saúde que os coloque em condição de risco diferenciada em relação aos demais usuários que também aguardam atendimento.

Em nosso levantamento, apenas as secretarias municipais de saúde de Recife e do Rio de Janeiro dispunham de protocolos de encaminhamento aos serviços ambulatoriais, incluindo priorização estratificada de acesso, mas apenas para algumas das especialidades médicas existentes no Instituto de Medicina Integral Prof. Fernando Figueira (IMIP).

1.1 Objetivo

Propor um sistema baseado em aprendizado de máquina para classificação de prioridade do agendamento de consultas especializadas, segundo avaliação do risco de agravamento da condição clínica do usuário, no Instituto de Medicina Integral Prof. Fernando Figueira (IMIP).

2. ATIVIDADES PRINCIPAIS

As atividades estão sendo realizadas seguindo as etapas 1- Análise documental de documentos extraídos de bases de literatura revisada por pares, como Pubmed e Lilacs, e de domínio público nacional (portarias e manuais técnicos do Ministério da Saúde, protocolos estaduais e municipais); 2- Análise do perfil clínico de cada ambulatório especializado para criar critérios de priorização dos agendamentos ambulatoriais; 3- Proposição e validação de variáveis que influenciam a marcação de consultas especializadas; 4- Identificação da origem dos dados para mineração necessária; 5- Elaboração e validação de protocolo institucional de acesso com classificação de prioridade; 6- Implementação de formulário estruturado no prontuário eletrônico; 7- Identificação do método adequado para classificação de prioridade; 8- Comparação com o método atual.

Atualmente, as atividades em andamento se referem à etapa 7.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Os protocolos de acesso com classificação de prioridade foram criados com base nos motivos de encaminhamento para consulta ambulatorial na instituição, foram incluídos aspectos como, hipóteses diagnósticas, patologias e condições clínicas. A priorização foi baseada em prevalência no referenciamento para o ambulatório especializado, referencial documental de secretarias de saúde, na literatura científica e discussões clínicas entre os especialistas de cada área e o núcleo de regulação interna, classificando os casos em A, B, C e D, em ordem decrescente de priorização de acesso à rede especializada. Atualmente, há 23 especialidades médicas em uso do encaminhamento estruturado no prontuário eletrônico - PEP. Para treinamento e teste do algoritmo, selecionamos a especialidade Cardiologia adulto, por ter sido o primeiro formulário estruturado implantado e possuir um volume de dados adequado para treinamento do modelo proposto.

Foram inicialmente elencados onze critérios para compor o escore de estratificação, baseados em variáveis identificadas nos formulários de encaminhamento elaborados pelos especialistas focais para priorização da marcação de consultas: prioridade por lei, perfil estratégico institucional, conformidade ao protocolo institucional de acesso com classificação de prioridade, origem do encaminhamento, tempo até resultados dos exames, tempo até consulta de retorno, ordem de chegada, tempo desde a última consulta, critérios de alta do ambulatório, faltas consecutivas e idade, sendo ordem de chegada e idade critérios com ordenação contínua.

O projeto foi submetido ao comitê de ética. Atualmente, está sendo avaliado o método de classificação ideal, testando sistematicamente três algoritmos populares de machine learning.

4. DESAFIOS E APRENDIZADOS

Inicialmente, planejamos usar apenas dados de rotina do prontuário eletrônico, mas a solicitação de encaminhamento com dados não estruturados dificultava o uso para aprendizado de máquina. A implantação do formulário com dados estruturados foi essencial para registrar adequadamente

os critérios clínicos de priorização no PEP, acelerando a análise de dados. No entanto, sabe-se que essa estruturação e qualificação do referenciamento na rede especializada alteram a rotina dos profissionais, o que pode gerar desconforto. Recomendamos que além das adequações tecnológicas também haja oferta suficiente de treinamentos, destacando os benefícios para pacientes mais críticos, que passam a ter acesso priorizado com o modelo de referenciamento otimizado.

5. REFERÊNCIAS

1. Recife. Prefeitura Municipal. Protocolo de Acesso Ambulatorial. Recife; 2013. Disponível em: https://www2.recife.pe.gov.br/wp-content/uploads/PROTOCOLO_ACESSO_AMBULATORIAL.pdf
2. Brasil. Ministério da Saúde. Protocolos de acesso ambulatorial: consultas especializadas: Hospitais Federais no Rio de Janeiro. Brasília; 2015. Disponível em: https://bvsmms.saude.gov.br/bvs/publicacoes/protocolos_acesso_ambulatorial_consulta_especializada.pdf
3. Espírito Santo. Secretaria de Saúde do Estado. Protocolo de Regulação Para Acesso A Consultas e Exames Especializados. Vitória; 2016. Disponível em: <https://saude.es.gov.br/Media/sesa/Protocolo/Protocolo%20consultas%20e%20exames.pdf>
4. Universidade Federal do Rio Grande do Sul. TelessaúdeRS. Materiais e Protocolos. Porto Alegre; 2016. Disponível em: <https://www.ufrgs.br/telessauders/materiais-protocolos/>
5. Rio Claro. Secretaria Municipal de Saúde. Protocolos de regulação para acesso a consultas especializadas. Rio Claro; 2023. Disponível em: <https://www.saude-rioclaro.org.br/uac/protocolos%20consultas%20especializadas.pdf>

Avaliação do modelo YOLOv8 para detecção de oocistos de Plasmodium vivax em mosquitos do gênero Anopheles sp.

Jady Shayenne Mota Cordeiro^{1,2}, Alexandre Vilhena da Silva Neto^{1,2}, Gabriel dos Santos Mouta^{1,2}, Felipe Nery Saldanha Braga^{1,2}, Patrícia Carvalho da Silva Balieiro^{1,2}, Maria Gabriela de Almeida Rodrigues^{1,2}, Elisângela Aparecida da Silva¹, Sandro Figueira Moreira^{2,3}, Allan Delon da Costa Bruce^{1,3}, Erika Gomez Martinez^{1,2}, Beatriz Fabiane dos Santos Carvalho^{1,2}, Iago Richard Rodrigues⁶, Rosa Amélia Gonçalves Santana^{1,3}, Patricia Takako Endo⁴ & Vanderson de Souza Sampaio^{1,5}

¹Fundação de Medicina Tropical Dr Heitor Vieira Dourado, Manaus, AM, Brasil.

²Universidade do Estado do Amazonas, Manaus, AM, Brasil.

³Instituto Leônidas & Maria Deane, Fiocruz, Manaus, AM, Brasil.

⁴Universidade de Pernambuco, Brasil.

⁵Instituto Todos Pela Saúde, São Paulo, SP, Brasil.

⁶Universidade Católica de Pernambuco, Recife, PE, Brasil.

motajady8@gmail.com, alexandre.neto94@gmail.com, gabrielSMouta@gmail.com, felipenery408@gmail.com, tricia.cks@gmail.com, rodriguesgabriela016@gmail.com, ellizzahbio@gmail.com, sandrofmstm@gmail.com, allandelon.sk8@gmail.com, eri1578@hotmail.com, bfs.carvalho.bmed@gmail.com, iagorrs@gmail.com, santana.rosaamelia@gmail.com, patriciaendo@gmail.com, vandersons@gmail.com

Resumo: A malária é uma doença febril aguda de notificação compulsória, com cerca de 240 milhões de ocorrências por ano em escala global. No ciclo da malária, a detecção de oocistos no intestino médio do mosquito Anopheles é essencial para confirmar infecções em estudos. No entanto, a dissecação e análise dessas amostras são processos demorados e passíveis de interpretações empiricamente subjetivas. Este estudo avaliou a acurácia de um teste diagnóstico automatizado utilizando variantes do modelo YOLOv8 para contagem de oocistos em imagens de intestinos de mosquitos infectados com Plasmodium vivax. As imagens foram rotuladas e aumentadas, resultando em 357 imagens divididas para treinamento, validação e teste. As variantes YOLOv8 nano e small foram comparadas, destacando a matriz de confusão e a precisão dos resultados. Os testes com diferentes variantes mostraram precisão semelhante, porém houve uma discrepância notável na confusão entre oocistos melanizados e não melanizados, sendo a variante small a que teve melhor desempenho. Este estudo demonstra o potencial da visão computacional para automação de processos laboratoriais, melhorando a precisão e eficiência na contagem de oocistos.

Palavras-chave: Malária; Visão computacional; YOLOv8; Automação laboratorial.

Abstract: Malaria is an acute febrile disease of compulsory notification, with approximately 240 million cases annually worldwide. In the malaria cycle, detecting oocysts in the midgut of the Anopheles mosquito is essential for confirming infections in studies. However, the dissection and analysis of these samples are time-consuming processes prone to empirically subjective interpretations. This study evaluated the accuracy of an automated diagnostic test using YOLOv8 model variants for counting oocysts in images of mosquito intestines infected with Plasmodium vivax. The images were labeled and augmented, resulting in 357 images divided for training, validation, and testing. The YOLOv8 nano and small variants were compared, highlighting the confusion matrix and accuracy of the results. Tests with different variants showed similar accuracy, but there was a notable discrepancy in the confusion between melanized and non-melanized oocysts, with the small variant performing better. This study demonstrates the potential of computer vision for automating laboratory processes, improving the accuracy and efficiency of oocyst counting.

Key-words: Malaria; Computer vision; YOLOv8, Laboratory automation.

1. CONTEXTO E MOTIVAÇÃO

A malária é uma doença infecciosa febril aguda de notificação compulsória. Em 2022, foram notificados 249 milhões de casos no mundo, com 131.224 casos registrados no Brasil (1). O ciclo de vida da malária é heteroxênico, envolvendo humanos como hospedeiro intermediário e mosquito como hospedeiro definitivo. O gênero *Anopheles* compreende os mosquitos vetores, amplamente utilizados em pesquisas visando a erradicação da doença (2). Durante o ciclo evolutivo do parasita, a formação de oocistos no intestino médio dos mosquitos é frequentemente explorada em testes para verificar a presença da infecção. Para observar os oocistos, é necessário dissecar os mosquitos sete dias após a alimentação com sangue infectado. Os oocistos podem ser melanizados ou não melanizados. Esse procedimento é demorado, exigindo paciência e experiência do investigador (3). A etapa de leitura das lâminas pode ser exaustiva devido à grande quantidade de lâminas e a presença de numerosos oocistos (4). A falta de recursos automatizados para analisar imagens pode comprometer a precisão da contagem dos oocistos. Técnicas de inteligência artificial, como a microscopia automatizada, foram implementadas em diversos segmentos para aprimorar a precisão dos testes diagnósticos (5,6). Um exemplo é o uso desse método para confirmação remota de diagnósticos feitos a partir de esfregaços sanguíneos, utilizando algoritmos de computação para identificar células infectadas (7). A aplicação de métodos de aprendizado de máquina na quantificação de oocistos tem o potencial de aprimorar significativamente a pesquisa sobre interações patógeno-hospedeiro, fornecendo aplicações práticas e precisas.

1.1 Objetivo

Comparar e avaliar as variantes do modelo YOLOv8 para detecção e contagem de oocistos de *Plasmodium vivax* em mosquitos do gênero *Anopheles* sp.

2. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A pesquisa foca na análise de imagens dos intestinos de mosquitos do gênero *Anopheles* sp. infectados com *Plasmodium vivax*. Este estudo explora o uso de técnicas de visão computacional para automatizar este processo, potencialmente melhorando a eficiência e a precisão da análise. As lâminas preparadas com oocistos foram fotografadas utilizando um microscópio óptico Carl Zeiss com objetiva de 10x, e as imagens foram salvas no formato JPEG. A base de dados original, composta por 149 imagens, foi aumentada para 357 utilizando técnicas de aumento de dados, como rotação, para criar um conjunto mais robusto. As imagens foram divididas em conjunto de treinamento (70%, 312 imagens), validação (20%, 30 imagens) e teste (10%, 15 imagens), e a conversão do formato VoTT para YOLO foi realizada com a ferramenta Roboflow. Esse estudo não carece de aprovação ética, pois não envolve pesquisa com humanos.

O modelo YOLOv8 foi escolhido devido à sua eficiência e versatilidade, com foco nas variantes nano e small. A variante nano, é conhecida por sua velocidade e economia de recursos, foi comparada com a variante small, que oferece uma combinação de robustez e eficiência (8). As métricas avaliadas foram a matriz de confusão, precisão e F1-score.

A análise dessas métricas permitiu uma comparação detalhada entre as variantes, destacando suas capacidades e limitações. Foram testadas duas variantes do modelo YOLOv8 em dois ambientes distintos: Notebook (Sistema operacional: Windows 10; Processador: Intel Core i7-10750 H; Memória RAM: DDR4 16GB) e servidor (Sistema operacional: Ubuntu Server 22.04 LTS 64bits; Processador: 2x Intel Xeon Gold 5120 (14-core, 56 VCPUs); Memória RAM: 220GB; Disco: 10TB em RAID 5).

Os resultados indicaram que a variante nano teve maior taxa de confusão (61%) entre oocistos melanizados e não melanizados em comparação com a variante small (24%). Os erros de detecção de oocistos não melanizados e melanizados foram semelhantes entre as variantes, com a variante

nano apresentando 52% e a small 42%. A variante small também se destacou por ter menos falsos positivos e uma pontuação F1-score superior (53% versus 37% da variante nano) (Tabela 1).

Tabela 1 – Diferença entre as variantes Nano e Small.

	YOLO.v8 Variantes			
	Nano		Small	
	Windows	Servidor	Windows	Servidor
Tempo de processamento	1h 21min	34min	2h 02 min	1h 20min
Acertos (Não melanizados)	82%	78%	76%	80%
Falsos positivos (Não melanizados)	61%	52%	42%	42%
Confusão entre Melanizado -> Não melanizado	61%	22%	24%	9%
Confusão entre Não melanizado -> Melanizado	18%	52%	42%	42%
F1-score (Desempenho geral)	37%	37%	56%	53%

Este estudo demonstra o potencial da visão computacional para automatizar a contagem de oocistos, com alta precisão e eficiência. Embora a variante small exija maior capacidade computacional, a nano é uma alternativa viável para laboratórios com menos recursos, visto que pode ser executado até em computadores mais simples. Além da contagem de oocistos, a técnica pode ser treinada e avaliada para identificar ovos de mosquitos como Aedes, ovos de parasitas intestinais e identificar anomalias ou presença de tumores em imagens de radiografia, ampliando seu uso nas investigações médicas. Essa capacidade pode otimizar processos em diferentes contextos laboratoriais, beneficiando especialmente regiões endêmicas e com infraestrutura limitada. Futuras pesquisas podem explorar essa versatilidade para aumentar o impacto da automação em diagnósticos médicos e vigilância epidemiológica.

3. REFERÊNCIAS

1. Ministério da saúde. Boletim Epidemiológico - Número Especial, nov. 2021 - Malária — Ministério da Saúde [Internet]. [citado 28 de março de 2023]. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/m/malaria/situacao-epidemiologica-da-malaria-1/boletins-epidemiologicos-de-malaria/boletim-epidemiologico-numero-especial-nov-2021-malaria/view>
2. Biamonte MA, Wanner J, Le Roch KG. Recent advances in malaria drug discovery. *Bioorganic & Medicinal Chemistry Letters*. 15 de maio de 2013;23(10):2829–43.
3. Smith RC, Vega-Rodríguez J, Jacobs-Lorena M. The Plasmodium bottleneck: malaria parasite losses in the mosquito vector. *Mem Inst Oswaldo Cruz*. agosto de 2014;109(5):644–61.
4. Martins-Campos KM, Kuehn A, Almeida A, Duarte APM, Sampaio VS, Rodriguez ÍC, et al. Infection of Anopheles aquasalis from symptomatic and asymptomatic Plasmodium vivax infections in Manaus, western Brazilian Amazon. *Parasites & Vectors*. 4 de maio de 2018;11(1):288.
5. Kaewkamnerd S, Uthaipibull C, Intarapanich A, Pannarut M, Chaotheing S, Tongshima S. An automatic device for detection and classification of malaria parasite species in thick blood film. *BMC Bioinformatics*. 13 de dezembro de 2012;13(17):S18.
6. Srivastava B, Anvikar AR, Ghosh SK, Mishra N, Kumar N, Houry-Yafin A, et al. Computer-vision-based technology for fast, accurate and cost effective diagnosis of malaria. *Malaria Journal*. 30 de dezembro de 2015;14(1):526.
7. Das D, Vongpromek R, Assawariyathipat T, Srinamon K, Kennon K, Stepniewska K, et al. Field evaluation of the diagnostic performance of EasyScan GO: a digital malaria microscopy device based on machine-learning. *Malaria Journal*. 12 de abril de 2022;21(1):122.
8. Ultralytics. Prevê [Internet]. [citado 25 de julho de 2024]. Disponível em: <https://docs.ultralytics.com/pt/modes/predict>

Avaliação do uso de IA em estratégia de recuperação de cursistas não concluintes em MOOC: Educa e-SUS APS

Gabriel. Lébeis¹, Isaias. Oliveira¹, Marco. Costa¹, Juliana. Oliveira¹, Zilma. Reis¹

¹Universidade Federal de Minas Gerais, Belo Horizonte (MG), Brasil.

gmlebeis@gmail.com, isaias.ramos@gmail.com, massiscosta@gmail.com,
julara.med@gmail.com, zilma.medicina@gmail.com

Resumo: A modalidade Massive Open Online Courses (MOOC) possui diversas vantagens como ferramenta de ensino à distância. Foi escolhida para oferta nacional de cursos para profissionais e gestores da Atenção Primária, em estratégia de letramento digital alinhada ao Prontuário Eletrônico do Cidadão (PEC e-SUS APS) numa parceria entre o Ministério da Saúde e a UFMG. Dentre os principais desafios de MOOCs estão o abandono dos cursos. O estudo descreve o uso de IA para mitigar a evasão dos cursistas, com mensagens personalizadas desenvolvidas por meio de engenharia de prompt. Adicionalmente, um questionário para levantar os motivos do abandono foi elaborado. Foram enviados 14.363 e-mails, resultando na recuperação de 684 cursistas até 8 de julho de 2024, indicando um efeito positivo das mensagens personalizadas no engajamento.

Palavras-chave: IA em Saúde Pública Digital; Educação à Distância; Abandono em MOOC.

Nome do projeto: Educa e-SUS APS (Zilma Reis)

Link: <https://educaesusaps.medicina.ufmg.br/>

1. CONTEXTO E MOTIVAÇÃO

O projeto Educa e-SUS APS é fruto de uma parceria entre a Secretaria de Atenção Primária à Saúde do Ministério da Saúde (SAPS-MS) e a UFMG, cujo objetivo é oferecer educação permanente para profissionais da Atenção Primária à Saúde (APS) através de uma plataforma MOOC. O Educa e-SUS APS, alinhado à Estratégia de Saúde Digital (ESD28), busca promover a literacia digital e o uso responsável de recursos no cuidado, vigilância e gestão em saúde(5). A elevada taxa de evasão em MOOCs é documentada na literatura acadêmica(2,3,6). Os altos índices de desistência representam um obstáculo para a efetividade do modelo educacional, representando um desafio a ser superado(3,6). O uso de Large Language Models (LLM) tem potencial de personalizar a interação educacional nos MOOCs, buscando reduzir a evasão.

1.1 Objetivo

Este estudo tem o objetivo de avaliar o uso de IA no engajamento de cursistas não concluintes no projeto Educa e-SUS APS.

2. ATIVIDADES PRINCIPAIS

Para o desenvolvimento e avaliação da estratégia proposta neste estudo, foram desenvolvidas as seguintes atividades:

- Definição de critérios de predição do abandono de MOOCs(7);
- Alinhamento de elementos para personalização das mensagens;
- Escolha da LLM e elaboração dos prompts e entrada de dados;
- Escolha da estratégia de envio das mensagens incentivadoras;
- Avaliação dos resultados em um primeiro disparo-piloto de e-mails aos cursistas;
- Avaliação dos resultados obtidos, com análise de respostas e das taxas de recuperação de cursistas.

3. METODOLOGIA E RESULTADOS

Para minimizar as taxas de abandono, o Educa e-SUS APS incorporou o uso de IA em suas estratégias de reengajamento através da (1) identificação dos cursistas que iniciaram um curso (fizeram pelo menos a primeira aula) e não o acessaram por 90 dias; (2) extração individualizada de dados sobre o uso da plataforma; (3) desenvolvimento de uma engenharia de prompt para personalizar mensagens com o nome do cursista, cargo na APS, ponto do percurso onde ele abandonou o curso, indicação dos próximos conteúdos e mensagem de incentivo customizada; (4) convite para responder um questionário curto; e (5) análise dos resultados sobre motivos do abandono e da recuperação de cursistas.

Na tabela 1, são expostas estatísticas gerais de 30/01/2024 a 08/07/2024 considerando-se cursista o usuário que iniciou pelo menos um dos cursos nos quais fez matrícula.

Tabela 1 - Estatísticas gerais do Educa e-SUS APS

Data	Usuários únicos	Cursistas	Matrículas	Certificados
30/01/2024	9016	4232	26342	3720
25/03/2024	12601	5915	37194	6045
27/05/2024	18046	16080	54480	10202
08/07/2024	21194	19849	63071	12903

Os resultados obtidos pelo Educa e-SUS APS até o momento são animadores quando comparados a outras publicações sobre o tema(1,4,6), com índices de conclusão de 20,44%, e taxa de inatividade por 90 dias ou mais inferiores a 22,7% em 08/07/2024. No entanto, a real taxa de abandono só será conhecida após um maior tempo de oferta de cada um dos cursos.

Inicialmente foram enviados 327 e-mails para avaliação, sendo seguido pelo disparo para todos os cursistas que preencheram o critério de 90 dias sem acesso à plataforma. Esse prazo foi definido baseado em adaptações da literatura, que avalia abordagem direta após inatividade de 2 a 4 semanas para um curso(7,4), enquanto o Educa trabalha com cursos simultâneos por perfil profissional. Os resultados são expostos na Tabela 2.

Tabela 2 - Estatísticas de envios de e-mails para reengajamento

Data	Matrículas	E-mails enviados	Feedbacks respondidos	Cursistas recuperados
20/05/2024	52966	327	17	32
03/06/2024	55498	9083	254	253
17/06/2024	58706	13020	347	530
08/07/2024	63071	14363	385	684

Os resultados demonstram recuperação de 684 cursistas (4,8%) com a intervenção proposta. Também foram identificadas as principais razões para abandono dos cursos, em ordem: falta de tempo no local de trabalho (49,4%), dificuldades para acessar o site (14,5%), outro motivo (11,9%), falta de mensagens de incentivo (11,1%), falta de estrutura mínima no local de trabalho (7,0%), curso longo (3,3%) e não enquadramento no perfil profissional (2,9%).

4. APRENDIZADOS E DESAFIOS

As mensagens personalizadas mostraram-se eficazes no reengajamento de uma parte dos cursistas, mas as razões apontadas para a interrupção, consistentes com outros estudos(4,6), indicam que estratégias de suporte institucional, adaptações às condições de trabalho e melhorias no suporte técnico poderiam ser eficazes para aumentar as taxas de conclusão.

O projeto gerou aprendizados sobre registros em saúde, o arcabouço legal da saúde digital e a importância da parceria entre a Universidade e a Saúde Pública. Destaca-se como aprendizado deste percurso a compreensão de que mensagens personalizadas têm grande potencial de efetividade e que o feedback dos cursistas pode proporcionar oportunidade de melhoria do próprio curso.

Grandes desafios também se impuseram: limitações como a impossibilidade de utilizar mensagens por aplicativos devido a restrições orçamentárias, como pelo WhatsApp, que potencialmente traria maior alcance; a necessidade de realizar alinhamentos para a implementação de cada intervenção, uma vez que as propostas demandaram o aval das equipes da UFMG e da SAPS-MS envolvidas. Preocupações com os limites de uso da IA se fazem presentes, sobretudo diante das previsões legais da LGPD. O equilíbrio entre inovação tecnológica e a proteção dos dados pessoais, sobretudo por meio de consentimento explícito, é essencial para a sustentabilidade dessa estratégia em longo prazo.

5. REFERÊNCIAS

1. Burd EL, Smith SP, Reisman S. Exploring business models for MOOCs in higher education. *Innov High Educ.* 2015 Feb;40(1):37-49.
2. Gomez-Zermeno MG, Aleman De La Garza L. Research analysis on MOOC course dropout and retention rates. *Turk Online J Distance Educ [Internet].* 2016 Apr 1.
3. Greene JA, Oswald CA, Pomerantz J. Predictors of retention and achievement in a massive open online course. *Am Educ Res J.* 2015 Oct;52(5):925-55.
4. Jimenez G, Spinazze P, Matchar D, Koh Choon Huat G, van der Kleij RMJJ, Chavannes NH, et al. Digital health competencies for primary healthcare professionals: A scoping review. *Int J Med Inform.* 2020 Nov;143:104260.
5. Halawa S, Greene D, Mitchell J. Attrition and achievement gaps in online learning. In: *Proceedings of the 2014 ACM Conference on Learning at Scale.* New York: Association for Computing Machinery; 2014. p. 57-66.
6. Jordan K. Initial trends in enrolment and completion of massive open online courses. *Int Rev Res Open Distrib Learn.* 2014;15(1):133-60.
7. Xing W, Du D. Dropout prediction in MOOCs: using deep learning for personalized intervention. *J Educ Comput Res.* 2019 Jun;57(3):547-70.

Avaliando as limitações de redes adversariais gerativas na construção de modelos preditivos relacionados à saúde a partir de dados sintéticos

Guilherme Fonseca², Ana Machado², Gabriel Prenassi², Beatriz Reis¹, Ricardo Cardoso³,
Fernanda Lana¹, Claudio Andrade¹, Milena S. Marcolino¹, Jussara M. Almeida¹,
Leonardo Rocha², Marcos A. Gonçalves¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, ²Universidade Federal de São João del-Rei, São João del-Rei, MG, ³Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.

guilhermefonseca8426@aluno.ufsj.edu.br

Resumo: Um dos principais desafios para construção de modelos preditivos efetivos é a quantidade de dados insuficientes para o processo de aprendizado. Um exemplo desse cenário é a área de predição de custos hospitalares de pacientes devido à sensibilidade desses dados. Nesse contexto, estratégias baseadas em Generative Adversarial Networks (GANs) têm crescido para geração de dados sintéticos. Este estudo realizou uma avaliação comparativa entre quatro diferentes propostas de GANs, analisando a efetividade dos modelos de predição de internação na unidade de terapia intensiva (UTI) de pacientes com covid-19, treinados a partir dos dados gerados por elas.

Realizamos uma variação entre a quantidade de dados reais e dados sintéticos gerados no conjunto de treinamento. As estratégias de GANs analisadas resultaram em modelos preditivos efetivos e promissores, mas a efetividade degrada significativamente com a redução de dados reais, o que representa uma oportunidade para novas propostas.

Abstract. In some areas of healthcare, one of the main challenges for building effective prediction models is the insufficient amount of data for the learning process. An example of this scenario is predicting hospital patient costs due to the sensitivity of this type of data. In this context of data scarcity, we have observed a growth in strategies based on Generative Adversarial Networks (GANs) for generating synthetic data. In this study, we carried out a comparative evaluation among four different GAN proposals, analyzing the model's effectiveness in predicting admission to the intensive care unit (ICU) of patients with covid-19, trained from the data generated by them. We perform a variation between the amount of real and synthetically generated data in the training set, evaluating the effectiveness of the resulting model. All analyzed GAN strategies resulted in effective and promising predictive models. We also observed significant effectiveness losses when the amount of real data is reduced, representing an opportunity for new proposals.

Palavras-chave: redes adversariais gerativas; predição de unidade de terapia intensiva; covid-19.

Nome do projeto: Predição de Desfechos Clínicos e Econômicos por Meio de Representações Semânticas Multi-modais de Pacientes Resilientes a Drifts Temporais (Marcos André Gonçalves)

1. CONTEXTO E MOTIVAÇÃO

Na área da saúde, a utilização de técnicas de aprendizagem de máquina (AM) vem se provando extremamente útil para gerar insights valiosos, aprimorar diagnósticos, e prever desfechos.

Entretanto, algumas áreas da saúde enfrentam um problema já bem conhecido de quantidade insuficiente de dados disponíveis para o processo de aprendizado. Este problema é especialmente crítico na predição de custos hospitalares e de desfechos clínicos dos pacientes, devido à sensibilidade desse tipo de dado (1). Uma maneira para minimizar este problema é aumentar o número de amostras da base de dados por meio da geração de dados sintéticos. Nesse cenário, estratégias baseadas em redes generativas adversariais (GANs) têm emergido como uma solução promissora para fazer esta geração (2).

1.1 Objetivo

Realizar uma avaliação comparativa entre algumas das mais recentes e promissoras estratégias de GANs no cenário de modelos de predição de internação na UTI para pacientes com covid-19 (1).

Em nossa comparação, consideramos os modelos gerados pelos dados reais, contrastando-os com modelos gerados a partir de dados sintéticos produzidos pelas GANs.

2. ATIVIDADES PRINCIPAIS

Estudo e a implementação de quatro estratégias de GANs recentemente publicadas e avaliadas em cenários de saúde. A avaliação das GANs foi feita comparando a eficácia de um modelo de predição de UTI de pacientes de covid-19 (1) com dados reais e dados gerados por cada GAN. O estudo foi aprovado pela CAAE 30350820.5.0000.0008.

3. DESENVOLVIMENTO DO TRABALHO

Foi usada uma base de dados de coorte retrospectiva de pacientes com covid-19 internados durante a primeira onda da pandemia, período compreendido entre 01/02/2020 e 14/11/2020, em 36 hospitais públicos e privados brasileiros (6). Nosso objetivo com a base de dados selecionada foi criar modelos para prever se os pacientes admitidos nos hospitais seriam ou não encaminhados para UTI. Dessa forma, consideramos apenas variáveis que obtidas dos pacientes no momento de sua admissão hospitalar.

Na literatura recente, o estado-da-arte na geração de dados sintéticos é constituído pelas GANs (3). Nesse contexto, implementamos quatro estratégias que vêm sendo bastante utilizadas no contexto de geração de dados tabulares, incluindo dados da saúde: a proposta de GAN de (2), a qual chamaremos de GANOriginal, apenas adaptada para lidar com dados estruturados; pGAN descrita em (3); HealthGAN (4); e tGAN (5). O que diferencia essas propostas, além da estrutura da rede neural interna usada por cada uma, é o modo como os dados são pré-processados antes de serem enviados para a geração de dados sintéticos.

Para cada uma das GANs, foi utilizado o método de pré-processamento descrito em suas respectivas propostas (2-4). Para avaliação da efetividade dos modelos e da qualidade dos dados sintéticos gerados, os dados foram separados em treino e teste utilizando validação cruzada com 10 partições e aplicando o classificador XGBoost para prever o desfecho de UTI. A MacroF1 do modelo foi avaliada usando as partições de teste. Executamos nossos experimentos variando a quantidade de dados sintéticos usados para treinar o classificador: selecionamos X% (variando de 0% a 90%, com passo de 10%) da base de dados reais de treinamento para ser substituída por dados sintéticos gerados pelas GANs. Usamos o restante da partição (100 - X%) para treinar o modelo GAN e, após isso, para treinar o classificador junto com os dados gerados.

A Figura 1 apresenta a efetividade (MacroF1) dos modelos resultantes da avaliação acima descrita. Os resultados com eixo X igual a "0%" representam a efetividade de um modelo treinado apenas com os dados reais. Para os demais resultados, o eixo X representa o total de dados sintéticos considerados no processo de treinamento dos modelos. Inicialmente, notamos que todas as estratégias analisadas apresentam resultados promissores, sendo capazes de manter uma eficácia próxima àquela alcançada pelo modelo produzido sem dados sintéticos.

A tGAN apresentou os piores resultado em comparação com os outros métodos - com mais de 50% de dados sintéticos, ele obteve a menor eficácia entre os quatro métodos comparados. Em contrapartida, a healthGAN apresentou os melhores resultados, sempre tendo a maior efetividade. No entanto, para todas as GANs analisadas, à medida que a quantidade de dados sintéticos utilizada na geração do modelo cresce, observamos uma degradação da efetividade do modelo

resultante. No pior caso temos a tGAN a qual apresenta uma Macro F1 de 67.8 considerando 90% de dados sintéticos, frente à Macro F1 de 73.9 com 100% de dados reais. Como melhores métodos, temos a GANOriginal e a healthGAN, ambas obtendo uma Macro F1 de 71.0, muito próxima à obtida com o uso apenas dos dados reais. Por outro lado, os modelos GAN não foram capazes de gerar um ganho estatístico em relação ao modelo sem dados sintéticos, apenas empates ou perdas estatísticas. Em resumo, podemos afirmar que os dados sintéticos produzidos pelos melhores métodos GAN são de boa qualidade para fins de treinamento de modelos de predição, emulando efetivamente os dados reais, porém sem capacidade de produzirem novos padrões úteis para melhoria dos modelos preditivos. Nesse sentido, podemos concluir que apesar de efetivos, existe espaço para melhorar os métodos GAN correntes para geração de dados de saúde sintéticos.

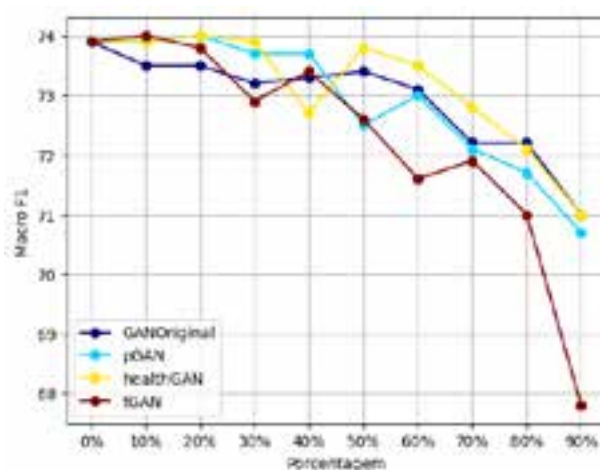


Figura 1 –Macro F1 dos modelos preditivos treinados com dados das quatro GANs.

4. DESAFIOS E APRENDIZADOS

Este estudo apresentou uma avaliação comparativa de GANs para a geração de dados sintéticos na criação de modelos de predição de UTI. Foram avaliadas quatro estratégias de geração de dados capazes de treinar classificadores efetivos. Mesmo com grandes quantidades de dados sintéticos em comparação a dados reais, as perdas de efetividade foram pequenas. Existe porém espaço para melhorias na geração de dados sintéticos no cenário da saúde.

5. REFERÊNCIAS

1. Andrade CMV, Paiva B, Morais G, et al. Predição de Necessidade de Internação em Terapia Intensiva e Custos de Tratamento de Pacientes com COVID-19. Anais do I Simpósio CI-IA Saúde da UFMG. Set 2023.
2. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. Advances in Neural Information Processing Systems. 27th ed. Curran Associates, Inc.; 2014.
3. Venugopal R, Shafqat N, Venugopal I, et al. Privacy preserving generative adversarial networks to model electronic health records. Neural Netw. 2022;153:339-348.
4. Yale A, Dash S, Dutta R, et al. Generation and evaluation of privacy preserving synthetic health data. Neurocomputing. 2020;416:244-255.
5. Xu L, Veeramachaneni K. Synthesizing tabular data using generative adversarial networks. arXiv preprint arXiv:1811.11264. 2018.
6. Marcolino MS, Ziegelmann PK, Souza-Silva MVR, et al. Clinical characteristics and outcomes of patients hospitalized with COVID-19 in Brazil: Results from the Brazilian COVID-19 registry. International Journal of Infectious Diseases. 2021 Jun 1;107:300–10.

Bioética Aplicada ao Uso da Inteligência Artificial na Assistência à Saúde: Análises e Perspectivas

Lívia Felberg (aluno)¹, Mizael A. Rodrigues (aluno)¹, Giovanna M. Vilas Boas (aluno)¹, Delcimar R. Silva (aluno)¹, Elisa Tuler Albergaria (orientador)¹, Cristiane Dias (orientador)¹, Zilma Reis (orientador)¹

¹CI-IA, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

liviafelberg@gmail.com, mizaelrodrigues20014@gmail.com, giovannamvb@gmail.com, dellcimarribeiro@aluno.ufsj.edu.br, etuler@gmail.com, profacristianedias@gmail.com, zilma.medicina@gmail.com

Resumo: A Inteligência Artificial (IA) tem revolucionado a área da saúde, trazendo avanços significativos. No entanto, os impactos decorrentes da aplicação dessa tecnologia exigem uma reflexão ética aprofundada. Este estudo revisa a literatura recente e identifica as atuais preocupações éticas associadas ao uso da IA na saúde. Os principais questionamentos abordam questões de privacidade e segurança dos dados, estigmatização, responsabilidade na implementação de IA, formação ética e capacitação contínua para profissionais de saúde. A partir dessa avaliação, conclui-se que uma reflexão crítica e contínua sobre o uso da IA é essencial para garantir a aplicação responsável e ética dessa tecnologia.

Abstract. Artificial Intelligence (AI) has revolutionized the healthcare sector, leading to significant advancements. However, the impacts arising from the application of this technology necessitate a profound ethical reflection. This study aims to review the recent literature and identify current ethical concerns associated with the use of AI in healthcare. The main concerns address issues of data privacy and security, stigmatization, responsibility in AI implementation, ethical training, and continuous education for healthcare professionals. This evaluation underscores the necessity of a critical and ongoing reflection on the use of AI to ensure its responsible and ethical application.

Palavras-chave: Inteligência Artificial; Saúde Digital; Bioética.

Nome do projeto: Centro de Inovação e Inteligência Artificial para a Saúde

1. CONTEXTO E MOTIVAÇÃO

A Inteligência Artificial (IA) para a Saúde tem se estabelecido como uma ferramenta transformadora, promovendo avanços na acurácia diagnóstica, personalização dos tratamentos e eficiência operacional dos sistemas de saúde (1). Embora a IA possa trazer importantes contribuições, sua aplicação também levanta discussões significativas sobre as implicações éticas envolvidas. Bioética, definida como o campo de estudo das implicações éticas, sociais e legais de práticas e avanços nas ciências da vida (2), é fundamental para superar esses desafios. O uso de grandes volumes de dados sensíveis, a automação de decisões que podem impactar diretamente a vida dos pacientes e a necessidade de assegurar equidade e justiça no acesso a essas tecnologias são aspectos que exigem uma análise ética cuidadosa (3).

1.1 Objetivo

Explorar as interseções entre IA e bioética, apresentando uma análise sobre preocupações éticas recentes e relevantes relacionadas ao uso da IA aplicada à saúde.

2. ATIVIDADES PRINCIPAIS

- Pesquisa em base de dados científicas;
- Seleção de artigos, considerando os objetivos da revisão;

- Leitura e análise dos artigos considerados mais relevantes;
- Levantamento das principais preocupações bioéticas no contexto de IA para a saúde.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

3.1 Metodologia:

A revisão da literatura contou com apoio de cientista da informação da biblioteca Baeta Vianna-UFMG. A estratégia de busca utilizando os termos “ethics”, “bioethics”, “ethics, clinical”, “ethics, medical”, “clinical ethics”, “medical ethics”, “artificial intelligence” e “AI (Artificial Intelligence)”, com o uso dos operadores booleanos “AND” e “OR”. A base utilizada foi a Pubmed com limite temporal para os últimos 5 anos. O processo de seleção de artigos está detalhado na [Figura 1](#). Identificação: Foram identificados 41 artigos por meio dos critérios de inclusão: disponíveis em texto completo diretamente relacionados a questões éticas e ao uso da IA na saúde. Triagem: Após ler os títulos e resumos, avaliadores independentes selecionaram, de acordo com a relevância, nove artigos para leitura na íntegra. Seleção: Uma professora responsável revisou os nove artigos. Após discussão com a equipe, foram selecionados cinco artigos (4,5,6,7,8) que mais se alinharam com os critérios estabelecidos. Leitura: Dois estudantes de iniciação científica realizaram a leitura detalhada dos artigos. Eles utilizaram fichas de leitura padronizadas para coletar dados sobre questões éticas discutidas em cada trabalho.

Análise qualitativa: Compilamos e estratificamos as principais preocupações éticas, utilizando um método de categorização para organizar os conceitos bioéticos discutidos.

3.2 Resultados:

As principais preocupações bioéticas identificadas foram:

Privacidade e Segurança dos Dados: Todos os artigos analisados destacaram a importância de proteger a privacidade e a segurança dos dados dos pacientes ao utilizar IA (4,5,6,7,8). A coleta, armazenamento e uso de dados sensíveis necessitam de governança clara, políticas robustas e regulamentação rigorosa para evitar abusos e vazamentos. Exemplos incluem a privacidade e estigmatização de pacientes com transtornos mentais e a proteção de informações clínicas nos sistemas de saúde digitais.

Equidade no Acesso às Tecnologias: É necessário garantir que todos os indivíduos, independentemente de sua localização geográfica ou condição socioeconômica, tenham acesso igualitário aos benefícios da IA (5). Os artigos (5,8) discutiram a rotulagem e estigma social, bem como a necessidade de inclusão de diversas populações. Na telemedicina, notam-se preocupações relacionadas à implementação ampla da IA, sendo esta a garantia da equidade. Para isso, menciona-se a necessidade de monitoramento da exclusão tecnológica (5).

Transparência e Explicabilidade: Métodos de interpretação pós-hoc e procedimentos de verificação de viés são necessários para aumentar a transparência e a confiança nos sistemas de IA. Os artigos destacaram a importância de tornar os modelos de “caixa preta” mais compreensíveis (7,8). Os modelos de IA devem ser compreensíveis tanto para os profissionais de saúde quanto para os usuários, para que as decisões possam ser bem informadas e claramente entendidas. É urgente a necessidade de combater o analfabetismo digital dos profissionais de saúde para melhor aproveitar as novas tecnologias. Isso requer uma formação técnica, legal e ética que garanta uma implementação segura (6). Além disso, os pacientes têm preocupações quanto à confiabilidade dos diagnósticos assistidos por IA, que podem ser mitigadas com maiores garantias dos programadores quanto à privacidade dos dados. A falta de transparência compromete a autonomia do paciente e prejudica sua confiança no sistema (4).

Confiabilidade e Validação dos Algoritmos: É uma preocupação constante, com necessidade de validação rigorosa em diagnósticos e tratamentos. Problemas de interpretabilidade e validade em modelos de aprendizado de máquina foram mencionados, assim como a necessidade de validação dos algoritmos para garantir precisão e segurança (4,7).

Impacto na Qualidade do Cuidado em Saúde: A importância de melhorar a qualidade do cuidado através de IA foi reiterada em vários artigos (5,6,7,8). A IA pode influenciar tanto positiva quanto negativamente a qualidade do cuidado em saúde. Melhorar a precisão dos diagnósticos e a personalização dos tratamentos foi um ponto positivo identificado, mas também há novos riscos se a IA não for implementada adequadamente (5,6).

Formação e Capacitação Contínua dos Profissionais de Saúde: Essencial para a implementação segura e eficaz da IA. Superar barreiras culturais e garantir uma adoção ética e eficaz das novas tecnologias foi amplamente discutido (4,7,8). Exemplos incluem a formação ética para evitar estigmatização e respeitar a autonomia dos pacientes, e a necessidade de capacitação em competências digitais (5,6). Além disso, ressalta-se que a resistência cultural interfere no alcance de uma IA responsável, pois as atitudes e expectativas da sociedade desempenham um papel determinante na aceitação e regulamentação da IA (4,8).

4. DESAFIOS E APRENDIZADOS

Como aprendizado, compreendemos a complexidade dos fatores a serem avaliados na implementação desta tecnologia, destacando a importância de considerar impactos sociais e éticos, além de técnicos, para garantir uma adoção responsável e benéfica da IA.

A análise dos 5 artigos revelou preocupações éticas consistentes que devem ser cuidadosamente consideradas para garantir a implementação responsável e ética da IA na assistência à saúde. Entretanto, durante a revisão da literatura, enfrentamos o desafio da escassez de estudos prévios que abordam detalhadamente a discussão ética do uso da IA para a saúde, o que limitou a construção de um panorama mais robusto dos impactos que envolvem o uso da tecnologia. Há uma necessidade urgente de mais pesquisas sobre IA responsável e ética, especialmente considerando preocupações com privacidade e resistência cultural (4,8).

Diferentes organizações prezam por transparência, responsabilidade, controlabilidade, segurança, orientação de valor ético, privacidade, segurança, risco e assistência ao usuário (1,7). Os modelos de IA para a saúde devem gerar benefícios sociais, evitar a criação ou reforço de preconceitos injustos, ser desenvolvidos e testados para garantir segurança, ser responsáveis perante os indivíduos, incorporar princípios de privacidade desde a concepção e manter elevados padrões de excelência científica (1). Além disso, a IA deve ser entendida apenas como um meio e não como um fim na transformação da saúde, com a prioridade sempre sendo colocada nas pessoas. Somente assim poderemos assegurar que a revolução tecnológica na saúde realmente beneficie a sociedade, promovendo um futuro mais justo, ético e humano.

5. REFERÊNCIAS

Nossas referências se encontram no seguinte [Link](#), respeitando a formatação proposta.

Breast Cancer Subtypes Multi-Layer Classification for Worst Prognosis

Leandro Y. S. Okimoto¹, Eduardo F. Nakamura, Fabíola G. Nakamura¹

¹Universidade Federal do Amazonas – IComp Instituto de Computação, Manaus, AM, Brasil

okimoto@icomp.ufam.edu.br, nakamura@icomp.ufam.edu.br, fabiola@icomp.ufam.edu.br

Abstract. Multi-layer classification methods are increasingly crucial for tackling complex classification problems and enhancing accuracy, particularly when data is insufficient for constructing advanced deep learning models. In the medical domain, gene expression data provides a promising alternative for classifying breast cancer subtypes. These datasets, reflecting the molecular state of cells, often have limited sample sizes. Accurate identification of breast cancer subtypes is essential to avoid aggressive and unnecessary treatments and to ensure personalized treatment plans. This study utilizes two breast cancer gene expression datasets: a training dataset with 1512 samples and a test dataset with 515 samples, both encompassing 9403 genes. We evaluate classification accuracy using Recursive Feature Elimination and SVM classification. Our results demonstrate that the proposed method effectively predicts the two worst prognoses, Basal and Her2, with precisions of 0.99 and 0.89, respectively. Additionally, our approach tracks misclassified samples through multi-layer classification, enabling the identification of incorrectly labeled samples for further investigation. This ability to track and analyze misclassified samples is vital for refining classification models and improving overall diagnostic accuracy.

Keywords: Gene expression; Breast Cancer Subtypes; Machine Learning;

Project name: AmazonIA - Aplicação de Modelos de Aprendizagem de Máquina e Aprendizagem Profunda para Análise de Dados sobre Doenças Tropicais Negligenciadas na Amazônia

1. CONTEXT AND MOTIVATION

Breast cancer, a leading cause of cancer-related deaths among women, is classified into hormone receptor-positive (Luminal A, Luminal B, HER2-enriched) and hormone receptor-negative (triple-negative or basal-like) groups (1). Accurate classification of these subtypes is crucial for personalized medicine, providing significant prognostic and predictive information essential for patient treatment. Misclassifying subtypes with worse prognoses as those with better prognoses (Figure 1), or vice versa, can lead to suboptimal treatment strategies (2). Ensuring precise classification is vital for appropriate treatment plans and improved patient outcomes. This study aims to address this need by using gene expression data and classification methods to enhance diagnostic accuracy and support to improve specific subtypes.

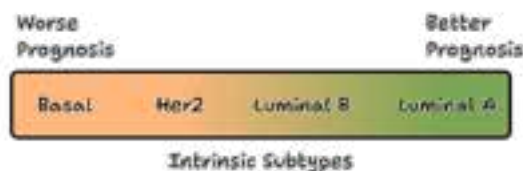


Figure 1. Worst to Best Prognosis for Breast Cancer Subtype.

1.1 Objective

This study aims to improve the classification accuracy of breast cancer subtypes with the worst prognoses by introducing a multi-layer support vector machine (SVM) classification approach. It focuses on optimizing feature selection techniques using Recursive Feature Elimination (RFE) to enhance the precision of subtype identification, thereby supporting more effective personalized treatment plans.

2. MAIN ACTIVITIES

The activities included in this project are:

- Curation and preprocessing of the dataset obtained from online sources;
- Construction, experimentation, and refinement of model hyperparameters;
- Analysis of the results obtained and performance comparison with baselines;
- Writing of the final report, which led to its submission.

3. DEVELOPMENT OF THE WORK (METHODOLOGY AND RESULTS)

Our proposed approach prioritizes the classification of breast cancer subtypes based on prognostic order, such as Basal, Her2, Luminal B, and Luminal A. This method integrates data preprocessing, feature selection, and machine learning techniques. We normalize and divide the ACES and TCGA datasets into training and testing sets (3). The training dataset undergoes a multi-layer process where subtypes are sorted, and gene selection and model fitting are performed. Each layer produces selected genes and a trained model for its subtype classification. The testing dataset is then used for predictions, updating results for each subtype, and tracking misclassified samples. This approach provides selected genes for each subtype for further study and documents misclassified and unclassified samples. It offers flexibility in selecting different gene selectors and models for each layer.

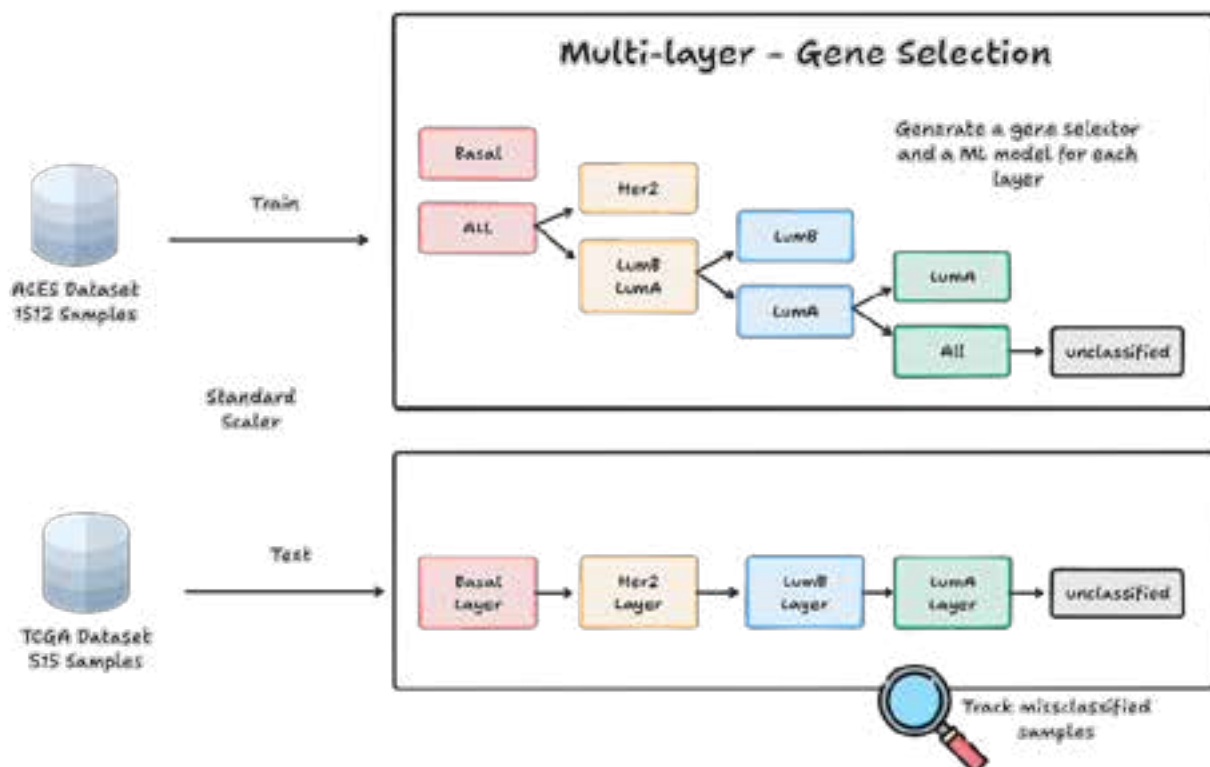


Figure 2 – Proposed approach.

The results shown in Table 1 indicate that the proposed method performed better for the Basal and Her2 subtypes, while the multiclass classifier excelled for the LumA subtype. Both classifiers showed similar performance for the LumB subtype, though there is potential for improving precision and F1- scores of our proposed method. These findings suggest that a tailored approach for each subtype could enhance overall accuracy and classification performance for the.

Class	Metric	Multiclass	Proposed Multi-layer	Better Classifier
Basal	Precision	0.95	0.99	Proposed Multi-layer
	Recall	1.00	1.00	Tie
	F1-Score	0.97	0.99	Proposed Multi-layer
Her2	Precision	0.84	0.89	Proposed Multi-layer
	Recall	0.88	0.86	Multiclass
	F1-Score	0.86	0.88	Proposed Multi-layer
LumA	Precision	0.98	0.97	Multiclass
	Recall	0.88	0.81	Multiclass
	F1-Score	0.92	0.88	Multiclass
LumB	Precision	0.81	0.78	Multiclass
	Recall	0.91	0.91	Tie
	F1-Score	0.86	0.84	Multiclass

Table 1. Comparison of Multiclass and Proposed Multi-layer Classifiers.

4. CHALLENGES AND LEARNINGS

One of the challenges of this research was the curse of dimensionality, avoiding deep learning approaches due to their high data and computational demands. So, we employed traditional machine learning techniques and feature selection methods to reduce the dimensionality of gene expression data. Additionally, we carefully balanced classifier parameters to enhance the performance for subtypes with the worst prognosis while maintaining acceptable results for those with better prognosis. This experience underscored the importance of feature selection and parameter balancing in managing high-dimensional data efficiently.

5. REFERENCES

1. Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M.C.U., Voduc, D., Speers, C.H., Nielsen, T.O., Gelmon, K.: Metastatic behavior of breast cancer subtypes. *Journal of clinical oncology* 28(20), 3271–3277 (2010)
2. Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., Shi, B.: Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research* 5(10), 2929–2943 (2015).
3. Staiger, C., Cadot, S., Györfy, B., Wessels, L.F., Klau, G.W.: Current composite-feature classification methods do not outperform simple single-genes classifiers in breast cancer prognosis. *Frontiers in genetics* 4, 289 (2013)

Characterizing and Understanding Temporal Shifts in Healthcare Data Using the DIS Methodology

Bruno B.M. Paiva (doutorando)¹, Guilherme Fonseca², Guilherme D. Bianco¹,
Jussara M. de Almeida¹, Milena S. Marcolino¹, Leonardo Rocha², Marcos André Gonçalves¹

¹Universidade Federal de Minas Gerais, ²Universidade Federal de São João del-Rei

contacting author: brunobarbosa.mpaiva@gmail.com

Summary. This study introduces the DIS (Detection, Initial Characterization, Semantic Characterization) methodology to analyze and understand temporal shifts in healthcare data. By applying this novel methodology to the Brazilian COVID-19 Registry and Medical Information Mart for Intensive Care (MIMIC-IV) datasets, the research demonstrates its effectiveness in identifying significant data shifts (drifts) and understanding their impacts. The DIS methodology shows promise in enhancing predictive models and improving healthcare outcomes.

Keywords: Temporal Data Shifts; Healthcare Analytics; Data Drift Detection.

Nome do projeto: Predição de Desfechos Clínicos e Econômicos por Meio de Representações Semânticas Multi-modais de Pacientes Resilientes a Drifts Temporais (Marcos André Gonçalves)

1. CONTEXT AND MOTIVATION

Our work addresses the challenges posed by temporal data shifts in healthcare, which can significantly affect the accuracy of predictive models and patient outcomes. Temporal shifts occur due to various factors, such as changes in data collection practices, patient behaviors, and the introduction of new treatments or technologies. Understanding these shifts is crucial for maintaining the reliability of healthcare predictions and improving patient care.

1.1 OBJECTIVE

To introduce and validate a novel methodology named DIS (Detection, Initial Characterization, Semantic Characterization) for analyzing and understanding temporal shifts (6) in healthcare data. This methodology aims to systematically detect, characterize, and better understand the reasons behind changes in distributions of healthcare data over time. By applying DIS to diverse healthcare datasets, such as the Brazilian COVID-19 Registry (1) and the MIMIC-IV dataset (2), the study seeks to demonstrate the methodology's effectiveness and generality for identifying patterns and trends crucial for decision-making, resource allocation, and improving the accuracy of predictive algorithms in healthcare.

2. MAIN ACTIVITIES

DIS is designed to support the systematic analysis of temporal shifts in healthcare data. The methodology aims to detect, characterize, and understand the reasons behind temporal data drifts. We applied DIS to two case studies: the Brazilian COVID-19 Registry (1) and the Medical Information Mart for Intensive Care IV (MIMIC-IV) dataset (2). Our main activities involved:

- Modeling: including the design and development of the DIS methodology tailored for analyzing temporal shifts in healthcare data.
- Group Discussions and Collaboration: Conducted regular meetings and discussions with interdisciplinary teams to refine the methodology and ensure its applicability to various healthcare datasets.

- Data Collection and Preprocessing: Prepared datasets from the Brazilian COVID-19 Registry (1) and MIMIC-IV (2), ensuring data quality and consistency for analysis.
- Experiment Preparation and Execution: Designed and executed experiments to apply and validate the methodology.

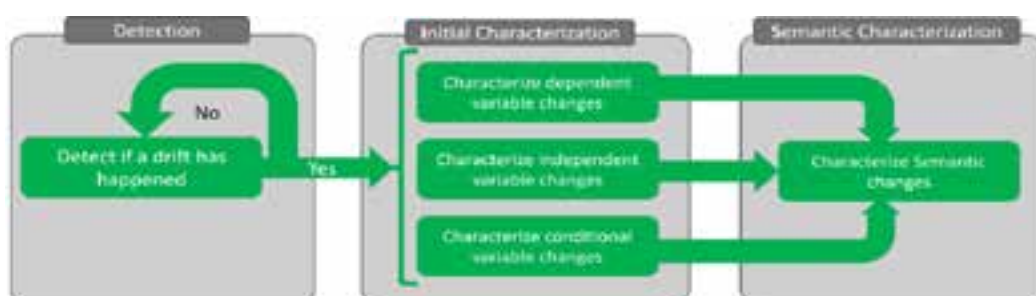
3. METHODOLOGY AND RESULTS

DIS was developed by means of a structured process involving the conceptualization of three main steps: Detection, Initial Characterization, and Semantic Characterization. Our study was approved by the National Research Ethics Committee, under the number CAAE: 30350820.5.0000.0008, with a waiver for the application of informed consent. A short summary of the methodology is shown in Figure 1. The Detection step employs metrics such as Jensen-Shannon divergence (3) and Autoencoder reconstruction error (4) to identify significant data drifts. The Initial Characterization step provides a global analysis of changes in data distribution and feature significance over time using techniques such as feature importance and Pearson correlation. The Semantic Characterization step leverages NLP-inspired techniques to understand the local context of changes by treating healthcare events as semantic units. This comprehensive methodology was tested on two distinct datasets—the Brazilian COVID-19 Registry and the MIMIC-IV dataset—to validate its effectiveness in identifying and explaining temporal shifts, thereby demonstrating flexibility and utility in diverse healthcare contexts.

The differing temporal granularities of these datasets allowed us to test various strategies within the Semantic Characterization step. The high temporal granularity of MIMIC-IV facilitated the use of NLP-inspired techniques to model sequences of healthcare events as semantic units, enabling detailed exploration of evolving relationships between events and outcomes. Conversely, the low temporal granularity of the Brazilian COVID-19 Registry required focusing on population-level semantic units, treating each patient’s data as a single event. This approach provided insights into how shifts in patient demographics and disease patterns influenced overall outcomes during the pandemic. The DIS methodology’s ability to analyze both high and low temporal granularity data underscores its utility in diverse healthcare contexts, offering valuable insights into the dynamics of healthcare data shifts and their implications for patient outcomes, ultimately aiding in the development of robust predictive models and informed healthcare policies.

For instance, our semantic analyses revealed a trend towards decreased mortality in both datasets. In the MIMIC-IV dataset, we discovered a decrease in correlation between cancer tokens and death, as well as between external causes and death. The former trend is possibly related to homecare practices where palliative care patients are increasingly being sent home to die, while the latter could be related to a reduction in iatrogenic effects. In the COVID-19 dataset, we observed a drifting pattern in the age of dying patients, indicating a potential correlation between this trend and vaccination efforts.

Figure 1 – The DIS Methodology



4. CHALLENGES AND LESSONS LEARNED

The main challenges in this work included handling the high variability and complexity of healthcare data, which often required customized approaches for different datasets. Temporal shifts in data distributions posed significant difficulties in maintaining the accuracy and reliability of predictive models. For instance, we were able to explore important data drifts in the COVID-19 dataset, including the reduction in overall mortality and the changes in the ages of dying patients. Similarly, we examined data drifts in the MIMIC-IV dataset, uncovering a reduction in overall mortality and a decrease in the similarity between cancer disease codes and lethality. From these conclusions, we can hypothesize on the causes, such as potential policy changes towards home care for palliative care cancer patients, leading to more deaths occurring at home.

Additionally, integrating various metrics and techniques, such as Jensen-Shannon divergence and NLP-inspired methods, into a cohesive framework demanded meticulous planning and validation. Our methodology is also suited for the exploration of entity embeddings, which allows us to uncover relationship changes that happen over time in those datasets. The interdisciplinary nature of the project required effective collaboration and communication among team members from diverse backgrounds to refine the methodology and interpret the results accurately. Ultimately, the goal is to enhance patient outcomes by providing deeper insights into the factors driving temporal changes in healthcare data.

5. REFERENCES

1. Marcolino MS, Ziegelmann PK, Souza-Silva MV et al. Clinical characteristics and outcomes of patients hospitalized with COVID-19 in Brazil: Results from the Brazilian COVID-19 registry. *International Journal of infectious diseases*. 2021 Jun 1;107:300-10.
2. Johnson A, Bulgarelli L, Pollard T, et al. R. MIMIC-IV. *PhysioNet*. At: <https://physionet.org/content/mimiciv/1.0/> (accessed Aug. 23, 2021). 2020 Sep:49-55.
3. Menéndez ML, Pardo JA, Pardo L, Pardo MC. The jensen-shannon divergence. *Journal of the Franklin Institute*. 1997 Mar 1;334(2):307-18.
4. Jaworski M, Rutkowski L, Angelov P. Concept drift detection using autoencoders in data streams processing. In *International Conference on Artificial Intelligence and Soft Computing 2020 Oct 7* (pp. 124-133). Cham: Springer International Publishing.
5. Paiva BBM, Rocha LC, Almeida JM, et al. DIS: A New Natural Language Processing Inspired Methodology to Investigate Temporal Shifts (Drifts) in Healthcare Data. *Journal of Medical Internet Research*. Accepted for publication.
6. Fernando Mourão, Leonardo Rocha, Renata Braga Araújo, et al. Understanding temporal aspects in document classification. *WSDM 2008*: 159-170

Combinação de Dados de Inquéritos Populacionais para Aprimorar a Predição dos Indicadores de Saúde por meio de IA Using the DIS Methodology

Fabiano Célio, Beatriz R. G. Barbosa, Samuel Norberto Alves, Thania M. T. Rezende Faria,
Marisa A. Vasconcelos, Crizian S. Gomes, Marcos A. Gonçalves, Jussara M. Almeida,
Regina Bernal, Laís Santos de Magalhães Cardoso, Larissa Fortunato Araújo,
Airton Tetelbom Stein, Deborah Carvalho Malta

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

Contacting author: beatriz.reisgb@gmail.com

Resumo. Este estudo visa apoiar estratégias de inteligência artificial (IA) ao unir duas bases de dados provenientes de dois grandes estudos em saúde realizados em Belo Horizonte: o Inquérito de Saúde da Região Metropolitana de Belo Horizonte (RMBH) e o Vigitel. Essa integração visa subsidiar algoritmos aplicados ao estudo de prevalência de doenças crônicas e seus fatores de risco. Os dados, provenientes de amostras probabilísticas e entrevistas telefônicas, foram filtrados temporalmente (2010) e espacialmente (Belo Horizonte). As variáveis foram recodificadas para garantir coesão entre os estudos. A concatenação resultou em uma amostra composta de 9190 participantes, permitindo estimativas mais precisas de fatores de risco. As duas bases isoladas, bem como sua combinação, foram usadas para gerar estimativas de prevalência de fumantes em pequenas áreas, sabidamente agrupamentos de setores censitários da cidade. A combinação dos dados aumentou significativamente o poder estatístico, resultando em diminuição da variabilidade e dos intervalos de confiança das estimativas.

Abstract. This study aims to support artificial intelligence (AI) strategies by combining two major studies on the city of Belo Horizonte. Data from two health surveys were integrated: the Health Survey of the Metropolitan Region of Belo Horizonte (RMBH) and Vigitel. This integration aims to support AI algorithms applied to the study of chronic diseases and their risk factors. The data, derived from probabilistic samples and telephone interviews, were filtered temporally (2010) and spatially (Belo Horizonte). The variables were recoded to ensure cohesion between the studies. The concatenation resulted in a sample of 9190 participants, allowing for more accurate estimates of risk factors. Each dataset in isolation as well as the combination of them were tested in a study of estimates of prevalence of smokers in small areas, notably groupings of city census tracts. The combination of both datasets increased statistical power, resulting in decreased variability and confidence intervals of the estimates.

Palavras-chave: Doenças Crônicas, Fatores de Risco, Amostras Probabilísticas, Modelos Preditivos

Nome do projeto: Uso de modelos preditivos de Inteligência Artificial na prevenção de doenças crônicas não transmissíveis (Deborah Malta)

1. CONTEXTO E MOTIVAÇÃO

Os inquéritos de saúde de base populacional podem ser conduzidos tanto por telefone fixo quanto presencialmente. Embora utilizem amostras probabilísticas, a operacionalização da coleta de dados encontra obstáculos relacionados, por exemplo, à segurança e à dificuldade de acesso, que tendem a favorecer o viés de seleção. As bases existentes apresentam, portanto, limitações que comprometem a acurácia das estimativas dos indicadores de saúde quando tentamos extrapolar os resultados para pequenas áreas. No caso da Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (Vigitel), por exemplo, ao se analisar a distribuição das entrevistas territorialmente, em pequenas áreas, há agrupamentos de setores censitários com

poucas ou mesmo ausência de entrevistas e estes coincidem com áreas classificadas como de maior vulnerabilidade social (1).

Pensando nos múltiplos desafios relacionados às análises de desigualdades em saúde em pequenas áreas, torna-se cada vez mais importante utilizar ferramentas avançadas, como a inteligência artificial (IA), em estudos de base populacional. No entanto, a eficácia da IA depende da disponibilidade de bases de dados robustas e bem estruturadas. Nesse sentido, a integração de bases de distintos inquéritos de saúde pode contribuir para aumentar o poder estatístico das análises, permitindo gerar estimativas mais precisas e confiáveis para pequenas áreas (2).

1.1 Objetivo

Avaliar o efeito da combinação de dois grandes inquéritos populacionais realizados em Belo Horizonte sobre a precisão das estimativas de indicadores de saúde e suas consequências na geração de populações sintéticas utilizando ferramentas de IA.

2. ATIVIDADES PRINCIPAIS

As atividades realizadas foram: solicitação de acessos às bases de dados; revisão dos dicionários de dados e questionários; análises de qualidade e recodificação de variáveis; integração das bases; documentação do processo; testes de integridade; análise de precisão e margem de erro.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Os dados foram oriundos de dois grandes inquéritos de base populacional. O Inquérito da Região Metropolitana de Belo Horizonte (RMBH) utilizou uma amostragem probabilística estratificada em dois estágios, abrangendo setores censitários do IBGE e domicílios urbanos da RMBH. Participaram 5.798 domicílios, totalizando 7.438 entrevistas em 2010 na cidade de Belo Horizonte (3). O segundo conjunto de dados refere-se à Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (Vigitel), com dados de Belo Horizonte, MG, no ano de 2010, englobando 1841 entrevistas (3). O setor censitário foi adicionado utilizando o Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) do Censo 2010 e permitiu a associação de cada número de telefone a um setor censitário através do linkage das bases. O Índice de Vulnerabilidade Social (IVS) foi utilizado para geração de pequenas áreas. O IVS é uma medida síntese, calculada com base em variáveis do Censo Demográfico de 2010, que captura fatores determinantes em saúde e classifica os setores censitários em quatro níveis de risco (4). Para este estudo, foram utilizadas 9 pequenas áreas, definidas como clusters de setores agrupados por níveis de vulnerabilidade derivados da clusterização do IVS original, da seguinte maneira: 1-BA-0: Baixo Risco, 2-BA-1: Baixo Risco, 3-MD-0: Médio Risco, 4-MD-1: Médio Risco, 5-EL-0: Elevado Risco, 6-EL-1: Elevado Risco, 7-EL-2: Elevado Risco, 8-ME-0: Muito Elevado Risco e 9-ME-1: Muito Elevado Risco. Foram calculadas as prevalências de fumantes entre indivíduos com 18 anos de idade ou mais e respectivas margens de erro, antes e depois da união das bases, para cada um dos nove clusters do IVS. As variáveis dos dois inquéritos foram recodificadas e os bancos de dados foram ajustados para um mesmo quantitativo de variáveis. A biblioteca Pandas da linguagem Python foi utilizada para realizar as operações necessárias à concatenação das bases de dados. A integração dos dados do RMBH e Vigitel (2010) totalizou 9190 entrevistas e resultou em menores margens de erro comparadas àquelas dos estudos isoladamente. Como exemplo, temos o estrato do IVS “muito elevado” (ME-0), cujas margens de erro foram reduzidas de 0.168 (RMBH) e 0,14 (Vigitel) para 0.11. A tabela 1 mostra as prevalências de fumantes e suas respectivas margens de erro dos estudos Vigitel, RMBH e Inquéritos Combinados por clusters. Destaca-se que o RMBH teve maior amplitude das margens de erro, enquanto os Inquéritos Combinados apresentaram maior precisão.

Aspectos éticos: Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais sob parecer nº 6.538.883.

Tabela 1 - Prevalência de fumantes para cada um dos inquéritos analisados e para a combinação final dos dados e a respectiva margem de erro.

cluster hierárquico	Inquéritos		
	RMBH (n=7438)	VIGITEL (n=1841)	COMBINADOS (n=9190)
1-BA-0	0,15 ± 0,015	0,14 ± 0,033	0,15 ± 0,014
2-BA-1	0,13 ± 0,020	0,13 ± 0,037	0,13 ± 0,018
3-MD-0	0,16 ± 0,018	0,12 ± 0,036	0,15 ± 0,016
4-MD-1	0,16 ± 0,017	0,13 ± 0,033	0,16 ± 0,015
5-EL-0	0,16 ± 0,071	0,23 ± 0,087	0,20 ± 0,056
6-EL-1	0,17 ± 0,068	0,27 ± 0,124	0,20 ± 0,060
7-EL-2	0,16 ± 0,030	0,15 ± 0,055	0,15 ± 0,026
8-ME-0	0,35 ± 0,168	0,20 ± 0,143	0,28 ± 0,113
9-ME-1	0,18 ± 0,072	0,12 ± 0,076	0,15 ± 0,053

4. DESAFIOS E APRENDIZADOS

A integração dos inquéritos de saúde da RMBH e Vigitel em Belo Horizonte demonstrou ser uma abordagem factível e benéfica para o aumento da precisão da estimativa do indicador de saúde investigado. O aumento no tamanho amostral e no poder estatístico contribuirão para o uso dos dados para treinamento de modelos de IA para: (1) gerar estimativas mais precisas e (2) gerar dados sintéticos representativos utilizando, por exemplo GANs (Generative Adversarial Networks)(5).

5. REFERÊNCIAS

As referências foram condensadas no link a seguir:

https://docs.google.com/document/d/1C69jN8svIH6sapVt4HA1WBZO60kml_9KsSGZtla3JQc/edit?usp=sharing

Resumo Expandido - Como a Inteligência Artificial está transformando o ambiente hospitalar: Breve relato da experiência da implantação de IA em serviço de farmácia clínica em hospital de alta complexidade

Melina Silveira Naves¹, Bruno Nascimento Moreira¹

¹Santa Casa BH, Belo Horizonte, MG, Brasil.

melinanaves@santacasabh.org.br, brunomoreira@santacasabh.org.br

Resumo. Este estudo visa mostrar como a implementação do sistema de Inteligência Artificial no serviço de farmácia clínica da Santa Casa BH, um hospital de alta complexidade com 1200 leitos, trouxe impactos positivos para a instituição. O objetivo da I.A. é aumentar a segurança do paciente, otimizar processos clínicos e reduzir custos operacionais através da automação da análise de prescrições médicas. A metodologia envolve a comparação dos dados de antes e após sua adoção, e a avaliação da farmacoeconomia e impacto, entre janeiro 2023 e Junho 2024. Os resultados do uso da IA mostram o impacto de alcançar a avaliação de 100% das prescrições com escore alto realizadas no hospital, melhorando a eficiência dos farmacêuticos clínicos, a qualidade do atendimento e reduzindo significativamente os custos diretos dos medicamentos. O projeto destaca desafios como a necessidade de expandir o serviço, treinar a equipe e lidar com a resistência à mudança. As intervenções farmacêuticas realizadas resultaram em uma economia direta de R\$1.373.971,26. Conclui-se que a adoção de ferramentas de IA pode agregar valor significativo à prática clínica, promovendo a segurança do paciente e a eficiência operacional. Este trabalho reafirma o compromisso da Santa Casa BH com a inovação e a excelência no cuidado à saúde.

Palavras-chave: Inteligência Artificial, Farmácia Clínica, Segurança do Paciente, Inovação.

Links: <https://santacasabh.org.br/> ; <https://santacasabh.org.br/orix-lab/> ; <https://noharm.ai/>;

1. CONTEXTO E MOTIVAÇÃO

A implementação eficaz de sistemas de saúde que promovam a segurança do paciente, otimizem os processos clínicos e reduzam custos operacionais é uma prioridade para instituições comprometidas com a excelência no cuidado. A Santa Casa BH, um hospital de alta complexidade com 1200 leitos, introduziu o sistema de IA NoHarm.ai para aprimorar a qualidade dos serviços oferecidos pelos farmacêuticos clínicos. Este trabalho relata a experiência da implantação dessa tecnologia inovadora.

O processo de auditoria manual de prescrições médicas na Santa Casa BH até outubro de 2022 foi realizado manualmente, processo lento, sujeito a erros e com necessidade de grande contingente de força humana para processar os mais de 31.000 itens prescritos diariamente na Santa Casa BH. Essa auditoria é essencial para garantir a segurança medicamentosa do paciente, além de ter como consequência a farmacoeconomia, controlando prescrições de alto custo, auditando tempo de tratamento de antibióticos, evitando interações medicamentosas, entre outros parâmetros essenciais para a segurança do paciente. Fazendo necessário o uso de ferramentas para digitalizar e automatizar este processo.

A Noharm.ai utiliza algoritmos de inteligência artificial e machine learning para analisar grandes volumes de dados de prontuários eletrônicos, prescrições médicas, histórico de saúde do paciente, resultados de exames e informações sobre medicamentos. O sistema cruza esses dados com informações de bases de dados farmacológicas, como interações medicamentosas, contra indicações e alertas de segurança. A partir dessa análise, a Noharm.ai identifica possíveis erros de prescrição, como doses inadequadas, interações medicamentosas perigosas, alergias e contraindicações. A plataforma oferece uma interface intuitiva para que os profissionais de saúde

possam visualizar os alertas e recomendações, facilitando a tomada de decisão e contribuindo para a segurança e qualidade do atendimento ao paciente.

1.1 OBJETIVO

O objetivo é apresentar a contribuição de ferramenta de IA hospitalar para a segurança e eficiência dos processos de auditoria de prescrições médicas, detalhando as atividades realizadas e os resultados obtidos.

2. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este estudo analisa o impacto da implementação da IA NoHarm.ai no serviço de farmácia clínica da Santa Casa BH. Foram avaliados dois cenários distintos: (1) Comparação da média dos resultados dos últimos três meses sem a utilização da IA (agosto/setembro/outubro de 2022) com o primeiro mês após sua implementação (novembro de 2022); (2) Avaliação contínua dos números de prescrições avaliadas, o número de vidas impactadas e a economia direta gerada pela IA ao longo do período de janeiro de 2023 a junho de 2024. A metodologia envolveu a coleta de dados periódicos sobre o número de prescrições médicas, as intervenções farmacêuticas realizadas e a economia gerada. Todos os dados iniciais foram validados junto a equipe de controladoria e financeira da instituição.

A Tabela 1 compara os dados antes e depois da implementação da IA NoHarm.ai na Santa Casa BH. Antes da implementação, com o processo ainda sendo executado manualmente, o número de intervenções realizadas em média por mês foi de 66, gerando uma economia média de R\$8.032,00. A economia é uma consequência das intervenções realizadas. Em novembro de 2022, após a introdução da IA, a farmácia clínica atingiu o dobro de intervenções realizadas, passamos a ter no primeiro mês 209, o aumento nas intervenções triplicou a economia gerada no mesmo cenário. A IA não apenas melhorou a segurança do paciente, mas também contribuiu para uma redução significativa nos custos operacionais. A ferramenta demonstrou um desempenho superior em termos de custo-benefício e impacto clínico, posicionando-se como uma ferramenta indispensável na farmácia clínica moderna. Esses resultados evidenciam a eficácia da IA em aumentar a eficiência das intervenções farmacêuticas e melhorar a gestão de recursos hospitalares.

Tabela 1 – Tabela 1. Farmacoeconomia com as intervenções antes e depois da implantação de I.A. em 50 leitos de CTI.

Mês	Intervenções	Farmacoeconomia
Ago/Set/Out	66/média mensal	R\$ 8.032,00/média mensal
Novembro (pós IA)	209/mês	R\$22.465,00/mês

Fonte: Santa Casa BH 2024

A figura 1 mostra a relação entre a economia gerada, o número de intervenções realizadas e o número de vidas impactadas na Santa Casa BH durante o período de janeiro de 2023 a junho de 2024. Observa-se um aumento contínuo na economia gerada, que está diretamente relacionado ao aumento das intervenções realizadas pelos farmacêuticos clínicos. A implementação da IA permitiu ampliar significativamente o número de intervenções realizadas, passando de 612 prescrições analisadas e processadas em Janeiro de 2023 para 3.970 em Janeiro de 2024, e 5.807 em Junho de 2024, gerando uma farmacoeconomia de R\$57.316,30 para R\$97.640,67 em junho 2024. Essa correlação positiva indica que o uso da IA não só aumentou a eficiência das intervenções, mas também contribuiu para uma gestão mais eficaz dos recursos hospitalares, resultando em uma economia substancial.

Figura 1 – Gráfico Farmacoeconomia vs Intervenções vs Vidas Impactadas durante o período de Janeiro 2023 a Junho 2024



Fonte: Santa Casa BH 2024

3. DESAFIOS E APRENDIZADOS

Durante a implementação da NoHarm.ai, enfrentamos diversos desafios, incluindo a necessidade de expandir o serviço, treinar adequadamente a equipe e superar a resistência à mudança. Estes obstáculos sublinham a importância de um planejamento estratégico detalhado e de um programa contínuo de treinamento para garantir a aceitação e a eficácia da tecnologia inovadora. No entanto, a experiência proporcionou aprendizados valiosos sobre a integração de tecnologias de ponta em um ambiente hospitalar complexo. A resistência inicial foi gradualmente superada através de uma comunicação clara dos benefícios tangíveis da IA e do envolvimento ativo da equipe em todo o processo de implementação. Estes insights são cruciais para futuras implementações de IA em outros contextos clínicos, destacando a necessidade de uma abordagem holística que inclua planejamento, treinamento e gestão de mudanças.

O Órix.lab, setor de inovação da Santa Casa BH, promove anualmente o Santa Ideia, competição onde os colaboradores são incentivados a propor ideias para: (1) Reduzir custos e/ou aumentar receitas ou (2) Melhorar a experiência do cliente. A implantação deste projeto recebeu o prêmio de “Santa Ideia 2022 - Redução de Custos e/ou Aumento de Receitas” Através dessa competição pela equipe de farmácia clínica.

4. REFERÊNCIAS

1. NoHarm.ai. (2024). Plataforma de Inteligência Artificial para Análise de Prescrições Médicas. Disponível em: <https://noharm.ai/>. Acessado em: 22/07/2024.
2. Santa Casa BH. (2024). Instituição de Saúde de Alta Complexidade. Disponível em: <https://santacasabh.org.br/>. Acessado em: 22/07/2024.
3. Santa Casa BH - Órix Lab. (2024). Laboratório de Inovação em Saúde. Disponível em: <https://santacasabh.org.br/orix-lab/>. Acessado em: 22/07/2024.

Comparison between DCGAN and PGGAN in Synthesizing Magnetic Resonance Images of Frontotemporal Dementia

Caio dos Santos Felipe, Dieine Estela Bernieri Schiavon, Thatiane Alves Pianoschi Alva,
Carla Diniz Lopes Becker

Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil.

caio.felipe@ufcspa.edu.br, dieineb@ufcspa.edu.br, thatiane@ufcspa.edu.br,
carladiniz@ufcspa.edu.br

Abstract. Recent advances in artificial intelligence in medicine have highlighted data scarcity and ethical and privacy issues associated with its use. This study aims to develop and compare methodologies for producing synthetic brain magnetic resonance images with frontotemporal dementia diagnosis, using generative systems as a solution to these challenges. We employed and compared two architectures: DCGAN (Deep Convolutional Generative Adversarial Network) and PGGAN (Progressive Growing of GANs). From a limited database, both models were able to produce synthetic images, with PGGAN demonstrating superior performance in both quantitative and qualitative assessments compared to DCGAN. Images generated by PGGAN presented SSIM of 0.2194 and NMI of 0.9762 compared to real images, outperforming DCGAN which achieved 0.2127 and 0.8059. This study contributes to the development and comparison of generative models for ethical production of synthetic medical images. The results obtained potentially facilitate future research and the development of artificial intelligence-assisted diagnostic systems in neurology.

Keywords: Frontotemporal Dementia; Synthetic Medical Imaging; Generative Adversarial Networks.

Project Name: APLICAÇÃO DE TÉCNICAS DE DEEP LEARNING PARA ANÁLISE DE IMAGENS MÉDICAS DE TUMORES CEREBRAIS.

1. CONTEXT AND MOTIVATION

Frontotemporal dementia (FTD) is a group of neurodegenerative conditions affecting the frontal and temporal lobes, altering behavior, personality, and linguistic abilities (1). Early diagnosis is crucial, and magnetic resonance imaging (MRI) plays a fundamental role in this process (2). However, the development of artificial intelligence (AI)-assisted diagnostic systems faces challenges due to the scarcity of labeled MRI data, attributed to the difficulty in obtaining FTD images (2), high costs, and ethical concerns.

Generative Adversarial Networks (GANs), especially Deep Convolutional Generative Adversarial Networks (DCGANs) and Progressive Growing Generative Adversarial Networks (PGGANs), emerge as promising solutions for generating high-quality synthetic images (3). Their application in synthesizing MRI images for FTD can overcome data limitations, providing valuable resources for training and validating automated diagnostic algorithms.

1.1 Objective

The main objective of this study is to develop, compare, and evaluate both quantitatively and qualitatively two GAN models, DCGAN and PGGAN, capable of generating synthetic brain MRI images with characteristics consistent with the diagnosis of FTD.

2. MAIN ACTIVITIES

The main activities included data collection and preparation, implementation of GAN architectures, data augmentation, model training, and quantitative and qualitative evaluation of the synthetic images generated by each model. For the quantitative analysis, we used two primary metrics: the Structural Similarity Index (SSIM) and Normalized Mutual Information (NMI).

3. MATERIALS AND METHODS

The study used data from the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI/NIFD) (4), limiting itself at this stage to 1,133 random images to simulate the use of generative models with reduced data. T1-weighted images of the brain volume were selected, focusing on sagittal slices between indices 75 and 85. The dataset includes 528 control images, 319 of non-fluent progressive aphasia, 154 of other FTD variants, 121 of the behavioral variant, and 11 of semantic dementia. The image resolution was set to 256x256. Data collection and sharing for this project was funded by the Frontotemporal Lobar Degeneration Neuroimaging Initiative (National Institutes of Health Grant R01 AG032306). The FTLDNI is coordinated through the University of California, San Francisco, Memory and Aging Center. FTLDNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

3.2 DCGAN

For the DCGAN, data augmentation was applied with random translations of up to 5 pixels and Gaussian noise (standard deviation 0.01) to combat mode collapse. Training process occurred over 300 epochs (and took 1 hour and 52 minutes), with a batch size of 8, an initial learning rate (LR) of 0.0002, and a 128-dimensional noise vector. We implemented exponential decay (ED) of the LR and used the Adam optimizer for both the generator and discriminator. Additionally, we incorporated Self-Attention, ResBlock, and Minibatch Discrimination techniques to improve the model's quality and stability.

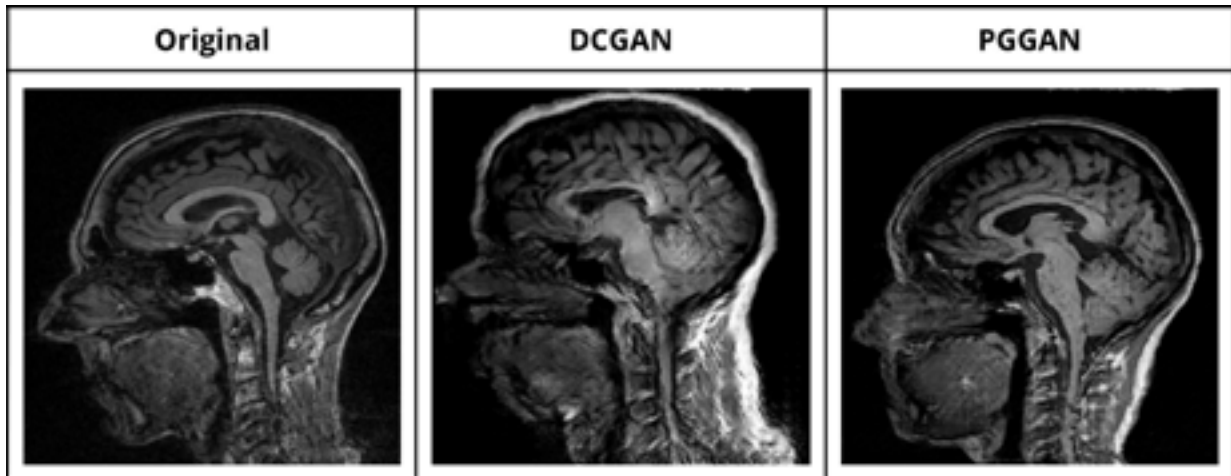
3.3 PGGAN

For the PGGAN, images were divided into 6 resolutions, from 8x8 to 256x256 pixels, with progressive training (with 1 hour and 45 minutes) using 30,000 to 80,000 images per resolution. We implemented ED of the LR and used the Adam optimizer for both the generator and discriminator. Additionally, we incorporated Self-Attention, Pixel Normalization, ResBlock, and Minibatch Discrimination techniques to improve the model's quality and stability. We utilized Wasserstein Loss with Gradient Penalty for more stable training and better convergence.

4. RESULTS AND DISCUSSIONS

The images generated by the DCGAN maintained the overall structure but exhibited distortion and loss of fine details, especially in the brainstem and cerebellum. The PGGAN, on the other hand, demonstrated superior fidelity to the original images, better preserving the contours and anatomical structures, although with slight smoothing of details. This visual comparison indicated that the PGGAN performed better in generating synthetic brain magnetic resonance images, producing results closer to anatomical reality and potentially more relevant for FTD diagnosis. Figure 1 presents a comparison between the original images and those generated by both models. In the quantitative analysis, we calculated the SSIM and NMI between the generated and real images. The DCGAN achieved an SSIM of 0.2127 and an NMI of 0.8059, while the PGGAN achieved an SSIM of 0.2194 and an NMI of 0.9762. The nearly identical SSIM values indicate moderately low structural similarity in both models, suggesting room for improvement. However, the high NMI values, especially for the PGGAN, demonstrated good capture of mutual information from the real images.

Figure 1 – Comparison between original images, images generated by the DCGAN model, and images generated by the PGGAN model.



Source: Authors

5. CONCLUSION

The project faced significant challenges, providing valuable learning experiences. The DCGAN struggled to produce small details in the generated images, often failing to accurately reproduce fine details, and suffered with mode collapse. The PGGAN, on the other hand, required optimization to adapt to GPU memory limitations but demonstrated fewer occurrences of mode collapse. These challenges highlighted the complexity of generating high-quality synthetic medical images, especially with limited datasets, and emphasized the need for a multifaceted approach, combining advanced machine learning techniques with specialized medical knowledge to overcome the limitations of generative models.

6. REFERENCES

1. Magrath Guimet N, Zapata-Restrepo LM, Miller BL. Advances in Treatment of Frontotemporal Dementia. *The Journal of Neuropsychiatry and Clinical Neurosciences* [Internet]; 2022 Oct 1;34(4):316–27. doi: 10.1176/appi.neuropsych.21060166.
2. Antonioni A, Raho EM, Lopriore P, et al. Frontotemporal dementia, where do we stand? A narrative review. *Int J Mol Sci.* 2023;24(14):11732. doi: 10.3390/ijms241411732.
3. Singh NK, Raza K. Medical image generation using generative adversarial networks: a review. In: Patgiri R, Biswas A, Roy P, editors. *Health informatics: a computational perspective in healthcare. Studies in Computational Intelligence*, vol 932. Singapore: Springer; 2021. p. 109-131. doi: 10.1007/978-981-15-9735-0_5.
4. Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI) Database [Internet]. San Francisco: University of California, San Francisco, Memory and Aging Center; 2010. Available from: <http://memory.ucsf.edu/research/studies/nifd>

Concepção de dataset de epítomos lineares de células B de organismos do gênero *Trypanosoma* para treinamento de algoritmos baseados em aprendizado de máquina

B. C. Russi¹, R. S. Moreira², P. D. C. Cabrera¹, S. C. Cazella³

¹Universidad Tecnológica del Uruguay, Rivera, Uruguay.

²Instituto Federal de Santa Catarina, Gaspar, SC, Brasil.

³Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil.

bruna.bruna@estudantes.utec.edu.uy, renato.simoes@ifsc.edu.br,
pablo.cuna@utec.edu.uy, silvioc@ufcspa.edu.br

Resumo. Introdução: O estudo dos epítomos é fundamental para a vacinologia e a medicina personalizada. Ferramentas de bioinformática e aprendizado de máquina oferecem abordagens promissoras para prever epítomos a partir de dados proteômicos. No entanto, identificar epítomos relevantes entre os preditos é um desafio devido à necessidade de considerar características bioquímicas. **Objetivo:** Criar um dataset que reúne dados sobre epítomos lineares de células B de organismos do gênero *Trypanosoma* incorporando características bioquímicas obtidas nos software BepiPred-3.0 e EpiBuilder-1.0. **Método:** Sintetizar conjunto de dados de epítomos de *Trypanosoma* do Immune Epitope Database (IEDB) e de proteínas revisadas no UniProt Knowledgebase (UniProtKB) para obter epítomos válidos e sintetizados. **Resultados:** Concepção de um dataset integrando dados de epítomos válidos do IEDB e sintéticos (com base no UniProtKB), e realizado experimento utilizando a ferramenta Orange Data Mining para análise de padrões, tendo sido o dataset disponibilizado no repositório aberto Zenodo. **Conclusão:** o dataset pode contribuir com a Ciência Aberta, permitindo que pesquisadores possam utilizá-lo em suas pesquisas.

Palavras-chave: Dataset; Epítomos lineares; *Trypanosoma*; Aprendizagem de máquina.

Abstract. Introduction: The study of epitopes is crucial for vaccinology and personalized medicine. Bioinformatics tools and machine learning offer promising approaches to predict epitopes from proteomic data. However, identifying relevant epitopes among predictions remains challenging due to the need to consider biochemical characteristics. **Objective:** To create a dataset that gathers data on linear B-cell epitopes from *Trypanosoma* organisms incorporating biochemical features obtained from BepiPred-3.0 and EpiBuilder-1.0 software. **Method:** Synthesize a dataset of *Trypanosoma* epitopes from the Immune Epitope Database (IEDB) and proteins reviewed in the UniProt Knowledgebase (UniProtKB) to obtain valid and synthesized epitopes. **Results:** Development of a dataset integrating valid epitope data from IEDB and synthesized epitopes (based on UniProtKB), and an experiment conducted using the Orange Data Mining tool for pattern analysis, with the dataset made available on the open repository Zenodo. **Conclusion:** The dataset can contribute to Open Science by enabling researchers to use it in their studies.

Keywords: Dataset, Linear epitopes; *Trypanosoma*; Machine learning.

Link: <https://zenodo.org/records/11388652>

1. CONTEXTO E MOTIVAÇÃO

O sistema imunológico é o mecanismo biológico cuja função fisiológica é reconhecer e eliminar antígenos, como vírus, bactérias, protozoários ou outras moléculas estranhas, através da interação entre células de defesa e esses antígenos (Abbas; Lichtman; Pillai, 2021). O reconhecimento do antígeno pelos linfócitos B e T ocorre através de receptores específicos que identificam regiões do antígeno chamadas de epítomos. Desse modo, o estudo de epítomos é crucial para a vacinologia e a medicina personalizada (Passos; Mwangi; Kapczinski, 2016). A identificação de epítomos em dados biológicos de grande escala (big data) é um desafio devido à natureza complexa de dados ômicos

e à diversidade de antígenos, sendo necessário o uso de ferramentas de bioinformática. Nesse contexto, o uso de aprendizado de máquina tem sido uma abordagem *in silico* promissora para a busca e predição de epítomos em dados proteômicos. Além disso, esta abordagem é estratégica para o estudo de Doenças Tropicais Negligenciadas - DTNs como a doença de Chagas, causada pelo *T. cruzi*, com 30 mil casos/ano e cerca de 14 mil óbitos/ano no Brasil – razão pela qual o país lançou um pacto nacional de combate à doença (Ministério da Saúde, 2022) e outras estratégias via Brasil Saudável (Decreto Nº 11.908/2024). Apesar disso, identificar epítomos de qualidade entre os epítomos preditos também é um desafio, pois requer a consideração de características bioquímicas não presentes nessas bases de dados, exigindo o processamento isolado dos dados em ferramentas de bioinformática para a sua posterior análise.

1.1 Objetivo

Desenvolver um dataset contendo epítomos lineares de células B de organismos do gênero *Trypanosoma* contendo as suas respectivas características bioquímicas.

2. ATIVIDADES PRINCIPAIS

- Coletar epítomos lineares de *Trypanosoma* no banco de dados IEDB;
- Sintetizar um conjunto de peptídeos baseados nas proteínas do UniProtKB;
- Pré-processar os dados coletados;
- Gerar um dataset integrando dados de epítomos válidos e sintéticos;
- Realizar experimentos com algoritmos de aprendizado de máquina utilizando a ferramenta Orange Data Mining e o dataset.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Foi realizado o download dos epítomos de diversas espécies de *Trypanosoma* no Immune Epitope Database - IEDB (Vita et al., 2018) e de todas as proteínas revisadas da base de dados UniProt Knowledgebase - UniProtKB (Magrane; Consortium, 2011) ambos no formato CSV. O arquivo CSV da base IEDB passou por um pré-processamento utilizando o editor de planilhas LibreOffice Calc 24.2.5 para filtrar apenas epítomos lineares de células B entre 10 a 30 aminoácidos. Em seguida, o arquivo CSV foi processado no software BepiPred-3.0 via BioLib (Clifford et al., 2022) resultando em um arquivo FASTA com a predição de epítomos.

Do mesmo modo, para criar epítomos artificiais, foi desenvolvido um código-fonte Java para selecionar 2.500 proteínas randomizadas a partir do arquivo CSV obtido no UniProtKB gerando uma substring de cada contendo entre 10 a 30 aminoácidos, mantendo o padrão de tamanho dos epítomos do IEDB. Estas substrings foram persistidas em um único arquivo FASTA, que também foi processado no software BepiPred-3.0 via BioLib resultando em outro arquivo FASTA.

Em seguida, os arquivos foram processados no EpiBuilder-1.0 via BioLib (Moreira et al., 2022), resultando em um arquivo TSV de scores em cada execução. Este arquivo foi então utilizado para a etapa final de pré-processamento utilizando SQL via software DBeaver Community Edition 24.0.0, de modo a rotular os epítomos sintéticos como “UNIPROT_RANDOM” e os epítomos validados em “IEDB_Tryp”, além de agrupá-los gerando uma média de suas características bioquímicas além de outros tratamentos dos dados. Foi criada uma view para este dataset, posteriormente exportando-a em formato TSV utilizando a funcionalidade nativa de exportação de dados do software DBeaver Community Edition. O tamanho total do dataset é de 4.288 linhas, sendo 1.993 epítomos sintéticos e 2.295 epítomos válidos.

Para a validação do dataset, foi montado um workflow na ferramenta Orange Data Mining 3.37.0 (Demsar et al., 2013), permitindo aplicar Principal Component Analysis (PCA) onde se observou a diferenciação entre as duas classes conforme esperado. Além disso, foi realizada a execução

dos algoritmos de floresta aleatória, árvore de decisão, KNN (K-Nearest Neighbors), SVM (Support Vector Machine) e rede neural também usando Orange. Para avaliar a performance, foram utilizadas as métricas de AUC, F1, Prec, Recall e Matriz de Confusão que são plotadas dentro do Orange. Todos os modelos conseguiram diferenciar as duas classes, havendo melhor performance com a árvore de decisão neste dataset, que foi disponibilizado no repositório aberto Zenodo (European Organization For Nuclear Research; OpenAIRE, 2013) sob DOI 10.5281/zenodo.11388651. O estudo aqui reportado fez uso de base de dados pública, dispensando dessa forma avaliação por comitê de ética em pesquisa.

4. DESAFIOS E APRENDIZADOS

O dataset criado mostrou-se adequado e atendeu a proposta, e nos experimentos realizados utilizando Orange Data Mining, foi possível diferenciar as classes “UNIPROT_RANDOM” e “IEDB_Tryp”. Apesar disso, são reconhecidas as limitações da síntese aleatória de peptídeos, que pode introduzir vieses nos resultados ao não capturar completamente a diversidade e complexidade dos epítomos de organismos de *Trypanosoma*, incluindo sequências que não existem na natureza. Cientes destas limitações, este estudo serviu como ponto de partida para futuras pesquisas que busquem desenvolver estratégias de síntese de peptídeos mais precisas, com o objetivo de minimizar vieses e aumentar a relevância biológica dos dados sintéticos – o que contribui para avanços na qualificação de epítomos lineares de células B preditos e na pesquisa das doenças causadas por organismos do gênero *Trypanosoma*.

5. REFERÊNCIAS

1. [ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. Imunologia Básica - Funções e Distúrbios do Sistema Imunológico. 6aed. \[S. l.\]: GEN Guanabara Koogan, 2021.](#)
2. [CLIFFORD, J. N. et al. BepiPr ed -3.0: Improved B-cell epitope prediction using protein language models. Protein Science, \[s. l.\], v. 31, n. 12, p. e4497, 2022.](#)
3. [DEMSAR, J. et al. Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research, \[s. l.\], v. 14, p. 2349–2353, 2013.](#)
4. EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH; OPENAIRE. Zenodo. [S. l.]: CERN, 2013. Disponível em: <https://www.zenodo.org/>.
5. [MAGRANE, M.; CONSORTIUM, U. UniProt Knowledgebase: a hub of integrated protein data. Database, \[s. l.\], v. 2011, n. 0, p. bar009–bar009, 2011.](#)
6. MINISTÉRIO DA SAÚDE. Pacto Nacional para a Eliminação da Transmissão Vertical de HIV, Sífilis, Hepatite B e Doença de Chagas como Problema de Saúde Pública. Brasília, DF: Ministério da Saúde, 2022. Disponível em: <https://bvsmms.saude.gov.br>.
7. [MOREIRA, R. S. et al. EpiBuilder: A Tool for Assembling, Searching, and Classifying B-Cell Epitopes. Bioinformatics and Biology Insights, \[s. l.\], v. 16, p. 117793222210952, 2022.](#)
8. [PASSOS, I. C.; MWANGI, B.; KAPCZINSKI, F. Big data analytics and machine learning: 2015 and beyond. The Lancet Psychiatry, \[s. l.\], v. 3, n. 1, p. 13–15, 2016.](#)
9. [VITA, R. et al. The Immune Epitope Database \(IEDB\): 2018 update. Nucleic Acids Res, \[s. l.\], v. 46, n. D1, p. D339–D343, 2018.](#)

Convolutional Neural Network for Clinical Diagnosis Support of Spondylolisthesis Through Imaging: Expanded Abstract

William Moraes da Silva (aluno)¹, Silvio César Cazella (co-orientador)¹,
Rafaela Soares Rech (orientadora)¹

¹Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil.

william.moraes@ufcspa.edu.br, silvioc@ufcspa.edu.br, rafaela.rech@ufcspa.edu.br

Abstract. Low back pain is a multifactorial condition that affects up to 65% of the population annually and up to 84% at some point in life, resulting in significant costs for healthcare systems and a decline in quality of life. One of the common pathologies is spondylolisthesis, which can be diagnosed through imaging exams, usually opting for X-rays due to cost-effectiveness. The aim of the research is to implement a Convolutional Neural Network (CNN) algorithm to enhance the diagnosis of spondylolisthesis. As a methodology, a CNN was developed using the VGG16 architecture. Images were selected from a public database collected from the King Abdullah University Hospital (KAUH), divided into three groups for training (80%), testing (16%), and validation (4%), respectively. For the model structures, the convolutional bases are frozen, and dense layers are added. For the evaluation of the architectures, prediction graphs, confusion matrix, and Grad-CAM are used. To date, an accuracy of 89.58% has been achieved, with a loss rate of 0.318, recall of 92.59%, precision of 89.29%, F-1 score of 0.9091, and area under the ROC curve of 0.9436.

Keywords: Deep Learning; Convolutional Neural Networks; Medical Imaging Diagnosis; Spondylolisthesis.

1. CONTEXT AND MOTIVATION

Spondylolisthesis is characterized by the adjacent and caudal displacement of one vertebra over another, usually caused by biomechanical alterations that generate significant shear forces leading to vertebral fractures. This condition can cause pain and dysfunction in the lumbar spine (1). It poses a significant problem due to its impact on individuals' functional capacity, as lower back pain results in high costs for healthcare systems and society due to work absenteeism and the need for prolonged treatments (2). Diagnostic evaluation of spondylolisthesis often involves the use of imaging exams, such as X-rays, preferred for their cost-effectiveness, accessibility, and lower radiation exposure for patients (3). However, analyzing these images is challenging as it requires precise interpretation by specialists who are overwhelmed by the increasing number of exams (3). To address this challenge, researchers have explored developing machine learning algorithms, particularly Convolutional Neural Networks (CNNs), to assist in identifying pathologies through image analysis (4). CNNs are particularly suitable for image classification tasks due to their ability to automatically extract relevant features from images and adjust the weights of their layers during training, improving diagnostic accuracy (4). Therefore, this research aims to develop a convolutional neural network for identifying spondylolisthesis in lumbar radiographs, motivated by the potential to expedite patient referrals for appropriate treatments through rapid and accurate diagnosis.

1.1 Objective

The objective of this research is to develop a Convolutional Neural Network, based on pre-trained networks, to assist in diagnosing spondylolisthesis by evaluating lumbar X-rays.

2. MAIN ACTIVITIES

The research activities carried out so far include a theoretical foundation with a review of the existing literature on spondylolisthesis and the fundamentals of Convolutional Neural Networks (CNN) and their applications in medical imaging diagnosis. Additionally, a search was conducted in public

radiological image databases to select the data to be trained. For the model development, data augmentation techniques were initially applied, and the data were divided into training, testing, and validation sets. Finally, a CNN algorithm was developed, which is being evaluated and readjusted according to the results obtained so far.

3. WORK DEVELOPMENT (METHODOLOGY AND RESULTS)

The development of the Convolutional Neural Network is being carried out on the Google Colab 1.0.0 platform, using Ubuntu 22.04.3, Python 3.10.12, and TensorFlow 2.15.0.

The public image database by Fraiwan, Audat, and Manasreh (5), which contains lumbar X-ray images of 388 patients from King Abdullah University Hospital, was chosen. These patients are categorized into three diagnostic groups: 71 without pathologies (normal), 79 with spondylolisthesis, and 188 with scoliosis. Images from patients classified as normal and with spondylolisthesis were selected, resized to 224x224, and data augmentation techniques were applied, resulting in a more substantial dataset with 143 normal images and 159 spondylolisthesis images. The images were converted to numerical arrays and randomly divided into three groups: 241 images for training (80%), 48 for testing (16%), and 13 for validation (4%).

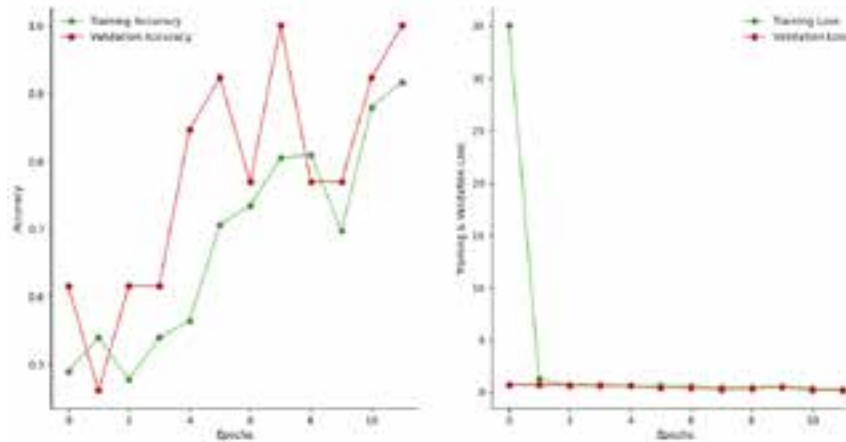
The Keras library was used to create a sequential model in a simple pipeline, incorporating the VGG16 architecture. The last fully connected layer was configured as false, followed by a flatten layer to convert it into a 1D output vector. An intermediate dense layer with 128 units and ReLU activation function was then added, followed by a dropout layer of 50% to avoid overfitting and a final output layer with a single binary classification unit and Sigmoid activation function.

The model was trained using the training and validation sets over 12 epochs, divided into batches of 32 samples. Additionally, three Keras callbacks were utilized: TensorBoard, to monitor training in real-time; ModelCheckpoint, to save the best model; and ReduceLROnPlateau, to adjust the learning rate.

After training, the test set was used to evaluate the network with the following partial results:

- **Confusion Matrix:** The CNN presented 18 true negatives, 3 false positives, 2 false negatives, and 25 true positives, indicating a high prediction accuracy.
- **Recall (Sensitivity):** 92.59% of actual positive cases were identified.
- **Precision:** About 89.29% of positive predictions were correct.
- **F1-Score:** The score of 0.9091 indicates a good balance between precision and recall.
- **Area Under the ROC Curve:** The value of 0.9436 indicates the model has an excellent capacity to distinguish between positive and negative classes, with a high true positive rate and a low false positive rate.
- **Accuracy:** Figure 1 shows that the model started at 50%, indicating initial learning, and increased throughout the training epochs, reaching 89.58% correct predictions. Some fluctuations suggest potential generalization issues or data variability.
- **Loss Rate:** The loss started high but saw a steep decline in subsequent epochs, reaching 0.318. This decline indicates rapid network learning, and the proximity to zero indicates minimal error during training, as seen in Figure 1.

Figure 1 – Epochs vs. Training and Validation Accuracy/Loss



Source: the authors

4. CHALLENGES AND LESSONS LEARNED

The biggest challenge of this research was finding a public image database with diagnosed lumbar X-rays. Once located, there was still the adversity of having a limited amount of data available, which was overcome with data augmentation techniques. A major lesson learned was the importance of using transfer learning, which significantly sped up the training process and improved performance, along with the role of hyperparameter tuning, where finding the optimal configuration for learning rates, batch sizes, and other parameters through experimentation was very important for achieving the best results. The model still needs adjustments but already shows promising results, such as an accuracy close to 90% and a loss rate close to zero.

5. REFERENCES

1. Martins DE, Puertas EB, Wajchenberg M. Clínica da coluna vertebral. São Paulo: Atheneu; 2014. 303 p.
2. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J.* 2008;8(1):8–20.
3. Fortes CGMNB, Gargioni LA, Frast RV, Zardo E de A, Severo CM. Manifestações clínicas da hérnia discal lombar. *Acta méd (Porto Alegre).* 2013;[7]-[7].
4. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA.* 2016 Dec 13;316(22):2353–4.
5. Fraiwan M, Audat Z, Manasreh T. A dataset of scoliosis, spondylolisthesis, and normal vertebrae X-ray images. 2022 Jan 17 [cited 2024 Jun 20];1. Available from: <https://data.mendeley.com/datasets/xkt857dsxk/1>.

Convolutional Neural Networks in the Diagnosis of Attention Deficit Hyperactivity Disorder in Children Using Electroencephalogram Signals: A Systematic Review

Wanessa Almeida Luna¹, Daniela Lopes Vieira¹, Stephanie Senna Cotta Queiroz de Oliveira¹, Ana Beatriz da Cruz Souza¹, Alan Cristian Marinho Ferreira¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

lunawanessa1@gmail.com, acristianff@gmail.com

Abstract. Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder that manifests in childhood, characterized by difficulties in skill acquisition, retention, and application. To enhance the precision and reduce bias in ADHD diagnosis, electroencephalography (EEG) signals combined with artificial intelligence (AI), particularly Convolutional Neural Networks (CNN), presents a promising, non-invasive approach. Therefore, this systematic review, conducted according to PRISMA guidelines, aimed to evaluate the diagnostic performance of CNN in recognizing patterns from EEG data. A total of 16 studies were retrieved from PubMed, Embase, Cochrane, and Scopus in July 2024. Among the analyzed articles, 7 studies assessed CNN combined with input enhancement techniques, 4 evaluated CNN with output optimization, and 7 compared CNN with other Deep Learning methods. Overall, results indicated that CNN, particularly with pre- and post-processing techniques, achieves high diagnostic accuracy, although it requires larger training datasets. Nonetheless, other Machine Learning techniques outperformed CNN.

Keywords: Attention Deficit Disorder with Hyperactivity; Child; Electroencephalography; Artificial Intelligence; Neural Networks, Computer; Deep Learning.

1. INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder whose symptoms appear early in childhood and are characterized by inattention, hyperactivity, and impulsivity (1). Currently, the diagnosis is clinical and based on medical evaluations that can, sometimes, be biased, compromising the diagnosis. An alternative to aid diagnosis and reduce errors is the use of imaging exams such as electroencephalogram (EEG), which is a non-invasive, inexpensive method with high precision. Associated with the use of imaging exams, the application of artificial intelligence (AI) in clinical practice presents a promising perspective for improving diagnostic accuracy. AI has already been used in healthcare to optimize administrative processes and assist in disease diagnosis (2). In the context of ADHD diagnosis, Convolutional Neural Networks (CNN), a deep learning algorithm, excel at recognizing patterns and hierarchizing input data characteristics through convolutional layers with applied filters.

2. OBJECTIVE

Evaluate the performance and accuracy of Convolutional Neural Network (CNN) models for the detection and diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) in young children by analyzing signals recorded by electroencephalogram (EEG).

3. METHODS

This is a systematic literature review, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (3). Using the PICO strategy, the following guiding question was formulated: "How effective and accurate are Convolutional Neural Network (CNN) models in the early childhood detection and diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) from electroencephalogram (EEG) data analysis?". A search key was created using Boolean operators

and the following standardized descriptors: “Attention Deficit Disorder with Hyperactivity”, “Child”, “Diagnosis”, “Electroencephalograph”, “Artificial Intelligence”, “Deep Learning” and “Convolutional Neural Network”. The search key was applied to the PubMed, Embase, Cochrane, and Scopus databases during July 2024. The inclusion criteria were full-text articles containing the pertinent terms and published between 2019 and 2024 in English, Portuguese, and Spanish. Thus, articles not addressing the terms of interest, published outside the defined period or in other languages, along with literature reviews and conference abstracts were excluded. After retrieving the studies, they were imported into the Rayyan platform for duplicate removal and screening to define the sample for this study. Two reviewers independently screened the titles and abstracts, selecting studies that met the inclusion criteria. In the second screening phase, full texts were reviewed. Disagreements regarding inclusion or exclusion were resolved by consensus. Google Sheets (Google LLC, 2024) was used for data tabulation and analysis.

4. RESULTS

Six articles were retrieved from PubMed, 19 from Embase, 0 from Cochrane, and 12 from Scopus, totaling 37 publications. After applying the eligibility criteria, 20 articles proceeded to full reading, of which 4 were excluded after complete reading. Thus, this review evaluated 16 articles focused on developing and validating different CNN-based techniques for detecting EEG signal patterns that allow the diagnosis of ADHD in early childhood patients. The references of the studies included in this systematic review are listed in Appendix A ([APPENDIX A - References of Included Studies](#)).

5. DISCUSSION

CNN is a deep learning algorithm based on neural network architecture, known for its ability to hierarchize characteristics from raw data. This systematic review sought to analyze the main categories of CNNs and their accuracies, which differ regarding the techniques coupled for preprocessing and postprocessing. Furthermore, this methodology was compared with other Machine Learning (ML) strategies also used in detecting and diagnosing ADHD in children.

Initially, several CNN variations have been investigated, each utilizing specific preprocessing techniques to enhance data quality. According to He et al. (4) and Feng et al. (5), specialized EEG processing architectures produced varying model accuracies ranging from 75% to 95%: EEGNet, ShallowConvNet, DeepConvNet, and ConvMixer-ECA. Chen et al.(6) introduced a DL framework integrating EEG-based brain network analysis, using a connectivity matrix as CNN input, whereas Torghabeh et al. (7) utilized Granger Causality Indices (GCI) to create Color-Coded Images for CNN analysis, enhancing sensitivity, specificity, and accuracy in distinguishing ADHD from Healthy Controls (HC). Alsharif et al. (8) combined CNN with Bidirectional Long Short-Term Memory (BiLSTM), and with GRU-Transformer, achieving accuracies of 94.50% and 95.59%, respectively. In addition, Moghaddari et al. (9) demonstrated a CNN achieving 98.48% accuracy in subject-based tests; while Mafi et al. (10) reported two distinct CNN architectures: 4D CNN and 6D CNN, achieving, respectively, 98.56% and 98.85% accuracies in processing EEG connectivity tensors.

In a second analysis, output enhancement models associated with CNN are also highlighted. A focus of postprocessing is improving the computational interface for healthcare professionals, thus, Nouri et al. (11) combined CNN with the Layer-wise Relevance Propagation (LPR), a technique that explains the neural network’s conclusions to improve interpretability. Furthermore, Wang et al. (12), Bakhtyari et al. (13) and Chugh et al. (14) studied the CNN-LSTM combination to encode spatiotemporal representations of EEG recordings, where CNN-extracted features are used as input for LSTM (Long Short-Term Memory Networks). These aforementioned studies found that CNN-LSTM (98% accuracy) is one of the most robust methods for diagnostic interpretation.

Furthermore, seven studies evaluated isolated ML techniques. He et al.(4), Alsharif et al.(8) and Chen et al. (15) demonstrated that while CNN excelled in learning capacity, its accuracy was lower

than SVM, except when combined with GRU-Transformer. Jahani et al. (16) found that ResNet captured complex EEG features better than CNN. Chang et al. (17) emphasized that LSTM significantly outperformed CNN in classifying children with ADHD and neurotypicals due to its ability to capture temporal variations in EEG signals while CNN focuses on spatial patterns. Also, Ahmadi et al. (18) examined a CNN-based tool that distinguishes ADHD subtypes with 99.46% accuracy and emphasized the innovation of using EEG over commonly used imaging techniques like fMRI and MRI. Additionally, Dong et al. (19) used CNN as part of a multitask model (RO-MLT) to identify ADHD as a comorbidity in Autism Spectrum Disorder (ASD), achieving an 11.07% improvement over traditional DL techniques.

Therefore, despite the promising results, several studies face common limitations like data scarcity and lack of diversity in samples, which hinder models from capturing the variability in clinical conditions. Hence, to prevent overfitting, model architectures are often simplified, compromising accuracy and robustness. Moreover, challenges such as overlapping EEG characteristics, difficulties in obtaining high-quality data, and the imbalance in specific data make it difficult to differentiate ADHD subtypes and genders, limiting clinical applicability. Thus, to address these issues, strategies like collaborative multicenter data collection, transfer learning, synthetic data generation, integration of multiple data modalities, and external dataset validation are crucial for enhancing the generalizability, accuracy and clinical robustness of AI models in ADHD diagnosis.

6. CONCLUSION

Based on the high-precision results predominantly found in the studies, Convolutional Neural Network (CNN) is a highly effective DL machine technique for applying AI to diagnose ADHD in children. However, it is worth highlighting that more specialized architectures and the association with preprocessing and postprocessing methods are responsible for achieving even more precise results. It is also worth noting that there are already other techniques with superior performance to CNN, and the main limitation for the persuasiveness of these techniques' experimental results is the restriction in data samples. Therefore, the importance of the present review relies on highlighting the validation of machine learning techniques in the accurate and early diagnosis of ADHD, as well as emphasizing the need for larger and more diversified data sets for training these AI models.

7. REFERENCES

1. Joseph HM, Lorenzo NE, Fisher N, Novick DR, Gibson C, Rothenberger SD, et al. Research Review: A systematic review and meta-analysis of infant and toddler temperament as predictors of childhood attention-deficit/hyperactivity disorder. *J Child Psychol Psychiatry* 2023;64:715–35. <https://doi.org/10.1111/jcpp.13753>.
2. Pressman SM, Borna S, Gomez-Cabello CA, Haider SA, Haider CR, Forte AJ. Clinical and Surgical Applications of Large Language Models: A Systematic Review. *J Clin Med* 2024;13:3041. <https://doi.org/10.3390/jcm13113041>.
3. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;n160. <https://doi.org/10.1136/bmj.n160>.

Data Mining associated with Electrocardiogram for atrial fibrillation Diagnosis: an integrative review

Victor Hugo Bartholomeu Gardino^{1,2,3}, Leonardo Onodera de Andrade⁴,
Brenda Guedes dos Santos⁴, Daiana Cristina Machado⁴, Adriano Lages dos Santos⁵

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; ²Fundação de Pesquisa e Ensino em Cirurgia, Belo Horizonte, MG, Brazil; ³Laboratório Interdisciplinar de Investigação Médica-UFMG, Belo Horizonte, MG, Brazil; ⁴Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brazil; ⁵Instituto Federal de Minas Gerais, Itabirito, MG, Brazil.

victorbartholomeu@med.grad.ufmg.br, adrianocomp@gmail.com

Abstract.

Objective: To analyze the use of data mining associated with electrocardiogram (EKG) for the diagnosis of atrial fibrillation. **Methods:** An integrative literature review conducted across three electronic portals in 2022. Fifteen original research articles in English, published between 2018 and 2022, were included, all of which featured the use of data mining for the analysis of electrocardiograms for the diagnostic purpose of atrial fibrillation. **Results:** The study highlighted the use of data mining combined with machine learning for the development of constant monitoring and early diagnosis systems for atrial fibrillation. Additionally, trends within this practice were noted, such as the use of classification functions, Convolutional Neural Network algorithms, and mobile devices. **Conclusion:** The association of data mining with electrocardiogram for the diagnosis of atrial fibrillation emerges as a potential solution to meet the demands of the current medical scenario. Although still in the development phase and with limitations in applicability, it shows potential for future use. **Keywords:** data mining, electrocardiogram, EKG, atrial fibrillation

1. CONTEXT AND MOTIVATION

Atrial fibrillation (AF) is the most common arrhythmia in the world, affecting about 33.5 million people in 2010, though this number is likely underestimated due to undiagnosed asymptomatic and transient cases (1). Associated with increased cardiovascular disease risk, such as stroke, AF is a major global health concern (2). Diagnosis relies on electrocardiogram (ECG), which records heart electrical activity, with the P wave representing atrial depolarization, the QRS complex ventricular depolarization, and the T wave ventricular repolarization. Recognizing cardiac physiology helps identify both normal and pathological ECG tracings, including AF (3). However, detecting imminent AF is challenging due to its short duration and asymptomatic nature, necessitating constant ECG monitoring and more effective diagnostic methods (4,5).

Data mining, part of Knowledge Discovery in Databases (KDD), involves exploring databases to uncover patterns from existing records (6). This process includes data cleaning, selection, transformation, and applying algorithms to find patterns (6,7). In healthcare, data mining aids diagnosis under challenging conditions, such as limited infrastructure or time constraints, by creating predictive models for early disease detection. In this context, data mining using Machine learning algorithms, a branch of artificial intelligence, excel in analyzing large data volumes with minimal human intervention (6). Thus, data mining offers a promising approach for diagnosing AF through continuous ECG analysis, even in transient or asymptomatic cases.

1.1 Objective

This literature review aims to analyze the use of data mining associated with electrocardiogram examination in the detection of atrial fibrillation.

2. MAIN ACTIVITIES

The article was initially developed as part of an elective course at the Federal University of Health Sciences of Porto Alegre, where students conducted a literature review on atrial fibrillation and data mining. They received study materials for independent study kindly provided by Professor Silvio Cazella. Later, while studying digital health at the Federal University of Minas Gerais, the lead author was advised by Professor Zilma Reis to submit the work. For that, he searched for orientation, and found Professor Adriano Lages dos Santos, who kindly mentored the work.

3. DEVELOPMENT OF THE STUDY (METHODS AND RESULTS)

Firstly, this bibliographic review study does not require approval from any research ethics committee, as it does not involve human subjects.

Besides that, methodology wise, between October and November of 2022, a non-systematic search was carried out through PubMed, Scopus and ScienceDirect databases. Using the keywords (“data mining” or data-mining) AND (ecg OR electrocardiogram OR ekg) AND (atrial fibrillation), 27 articles were initially found. Of these, 5 were duplicates, resulting in 22 articles. Out of these, 5 were excluded because they were not available in full text and 5 did not meet the research focus. Thus, 12 articles were assessed for eligibility, of which 9 answered the review question and were included in the final corpus of the study.

After the result analysis, it was observed that the use of machine learning or deep learning was unanimous in the application of the data mining function, making the concepts practically interchangeable within the scope of this study.

Moreover, the application of machine learning in the atrial fibrillation diagnosis process generally requires three phases: data preprocessing, feature extraction, and data mining function application (classification or prediction in the analyzed models) (8). In this context, data preprocessing was crucial for various aspects of the analyzed model, particularly concerning the reduction of interference and artifacts in ECG measurements. Among the technologies used, the Butterworth noise reduction filter was the most frequent, present in 33% of the models.

This trend can be explained by the move towards using mobile devices and less complex physical setups compared to traditional methods (which involve multiple electrodes and require the patient to be at rest on a hospital bed), aiming to enable continuous and automatic monitoring outside the hospital environment (8-13). In this context, the use of mobile measurement devices for ECG, also known as “wearables,” is limited to single-lead examination, which is less complex but more susceptible to interference and artifacts than the traditional 12-lead setup (8,9,11,12,13).

The intention to reduce the influence of artifacts on the examination was also crucial factor in selecting the wave to be analyzed. In atrial fibrillation detection, it is observed that cardiac anomalies cause changes in the atrial and ventricular activities of ECG signals, where the P waves are replaced by a series of F waves and the RR intervals (differences between adjacent R waves) are irregular (8,9,12). With this in mind, the RR interval emerged as the preferred segment of the ECG for analysis, accounting for 57.14% of the algorithms. This is because atrial activity characteristics used to detect AF (absence of the P wave along with the appearance of the F wave) are more susceptible to interference and artifacts due to their lower amplitude compared to the R wave (8,9). For this reason, studies using these parameters require more extensive data preprocessing steps (9,12).

Among the algorithms, 6 of 9 models used convolutional neural network (CNN) architectures, a deep learning algorithm that processes images to recognize patterns through multiple layers of filters identifying features like shapes, colors, and lines. This preference is due to CNNs being

one of the classification algorithms requiring minimal data preprocessing and therefore being the predominant data-mining function in the studies (4,9,14,15).

Regarding the limitations of data mining with ECG, challenges include the various stages of data processing and extraction, which can introduce errors and affect classifier performance (8,9). Moreover, AF detection capabilities do not easily transfer to other arrhythmias, limiting the applicability when other cardiac diseases are present (8,13). Additionally, 77% of studies used the same database, the MIT-BIH Atrial Fibrillation database, and model performance was not extensively tested on other datasets, potentially reducing effectiveness (4,8,9,10,11,12,16).

Lastly, computational costs were an advantage of the analyzed models due to the ongoing efforts to reduce costs through algorithm improvements (12,13,16) and the ease of integration with low-cost wearable devices (9,10,11). Performance was also high, with the lowest recorded accuracy at 81% (13) and 5 out of 9 models achieving accuracy above 90% (9,10,8,16).

In conclusion, the analyzed articles demonstrated that, despite the specific circumstances that allow its use, the application of data mining to electrocardiogram examination considerably aids in the detection of atrial fibrillation and that this approach can indeed be associated with mobile devices to create continuous monitoring systems for early detection of atrial fibrillation. Thus, although still under development, the association of data mining with ECG for the diagnosis of atrial fibrillation appears promising and potentially capable of meeting the demands that originated its study.

4. CHALLENGES AND LESSONS

The main challenge faced through was the need to study of informatics concepts in order to integrate both medical and computational fields, which required autonomous studying, attending college classes from the Biomedical Informatics course at Universidade Federal de Ciências da Saúde de Porto Alegre and other related events, like Bioinformatics Winter Course at Universidade Federal de Minas Gerais. Moreover, finding professionals who work in the intersection of medical and computational fields was somewhat challenging as well, since it is an emerging area in Brazil. Thankfully, the lack in numbers was compensated by their receptiveness. Now, the challenge is to produce an updated review including the new studies from the last two years.

5. ACKNOWLEDGMENTS

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) through a scientific initiation scholarship awarded to Victor Hugo Bartholomeu Gardino.

6. REFERENCES

Link to all references:

https://docs.google.com/document/d/14kjVrDweY6lhAR4vxwE8tfW_abOD8f4Ps00_WvgcMIk/edit?usp=sharing

Desenvolvimento de tecnologia para o monitoramento da fluência de leitura

Pedro Henrique F. Oliveira¹, Ramiro Noronha Reis Ribeiro¹, Isa Mourão Carvalho¹,
Débora Marques de Miranda^{1,2}, Marco A. Romano-Silva^{1,2}, Letícia Correa Celeste³,
Luciana Mendonça Alves (orientadora)^{1,2}

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

²INCT - NeurotecR - Faculdade de Medicina da UFMG

³Universidade de Brasília – UnB, Brasília - DF

pedrohfernandes2001@gmail.com, ramironrribeiro@gmail.com, isamouraoc@gmail.com,
debora.m.miranda@gmail.com, romanosilva@gmail.com, lumendoncaalves@gmail.com

Resumo. A fluência de leitura é um importante indicador do desenvolvimento da leitura e da qualidade de ensino e é composta por três componentes: a velocidade, a acurácia e a expressividade. Sua avaliação e monitoramento são práticas relevantes e normalmente realizadas de modo manual, o que dificulta análises em larga escala. Com o auxílio da tecnologia, é possível facilitar essa análise. O aplicativo Lepic é um exemplo de tecnologia gratuita, disponível para dispositivos Android, que facilita a análise da velocidade de leitura e da acurácia. Seu desenvolvimento é realizado com o auxílio de ferramentas que buscam gerar um instrumento intuitivo, prático e coerente, possibilitando o uso do aplicativo em ambiente clínico e escolar. Para um aluno de computação, desenvolver uma tecnologia para uma área completamente diferente pode ser um grande desafio, a interdisciplinaridade se faz necessária. No entanto, trata-se de um processo enriquecedor, visto que é possível observar o impacto do software desenvolvido.

Abstract: Reading fluency is an important indicator of reading development and teaching quality and is composed of three components: speed, accuracy, and expressiveness. Its assessment and monitoring are relevant practices and are usually carried out manually, which makes large-scale analyses difficult. With the help of technology, this analysis can be facilitated. The Lepic app is an example of a free technology available for Android devices that simplifies the analysis of reading speed and accuracy. Its development is assisted with the help of tools aimed at creating an intuitive, practical, and coherent instrument, making it useful in clinical and educational context. For a computing student, developing a technology for a completely different field can be a significant challenge, necessitating interdisciplinarity. However, it is an enriching process, as it allows for observing the impact of the developed software.

Palavras-chave: Leitura; Monitoramento; Aplicativo; Fluência de leitura.

Nome do projeto: Desenvolvimento de tecnologia para o monitoramento e aprimoramento da leitura.

1. CONTEXTO E MOTIVAÇÃO

A fluência de leitura é a habilidade de ler com velocidade adequada, precisão e expressividade (1). É um componente essencial para a formação de um leitor hábil e um importante indicador do desenvolvimento da leitura e da qualidade do ensino(2,3). Seu monitoramento e avaliação são auxiliados por três aspectos: a acurácia, a velocidade e a expressividade. A acurácia se trata da precisão na decodificação correta das palavras, medida pela quantidade de palavras lidas corretamente por minuto (PCPM); a velocidade, ou taxa de fluência, refere-se ao número de palavras lidas em um minuto (PPM) e a expressividade se relaciona à demonstração de atitudes e emoções durante a leitura, utilizando ritmo e entonação apropriados(3).

A avaliação desses aspectos é feita a partir da amostra de leitura em voz alta e já foi utilizada em estudos nacionais(4). A amostra permite a análise das medidas de palavras por minuto (PPM)

e palavras corretas por minuto (PCPM), por meio das respectivas fórmulas: $PPM = (\text{Número de palavras lidas} \times 60 \text{ segundos}) / \text{tempo total de leitura (em segundos)}$; $PCPM = (\text{Número de palavras lidas corretamente} \times 60 \text{ segundos}) / \text{tempo total de leitura (em segundos)}$. Essa metodologia de avaliação, normalmente, é feita manualmente, o que dificulta a realização de análises em larga escala.

Apesar das evidências científicas reforçarem a importância do monitoramento da fluência leitora, esta prática não é ainda frequentemente utilizada no Brasil. Levantamos como hipóteses para esta situação a falta de capacitação dos educadores sobre o tema, ou mesmo a dificuldade (e falta de praticidade) em se conduzir tais práticas. Portanto, tecnologias que facilitem esta importante prática de monitoramento da leitura são urgentes e essenciais para a educação brasileira.

1.1 Objetivo: Apresentar o processo de desenvolvimento e resultados de um aplicativo direcionado à análise de fluência leitora.

2. ATIVIDADES PRINCIPAIS

O projeto consiste no desenvolvimento de um aplicativo mobile, intitulado de “Lepic”, que atenda às necessidades encontradas na manipulação de dados obtidos no dia a dia clínico de pacientes fonoaudiológicos ou mesmo de alunos em ambiente escolar. O aplicativo atua possibilitando o monitoramento e avaliação histórica de pacientes (persistência de dados), assim como em análises gráficas e cálculos estatísticos que fornecem informações baseadas em dados coletados pelas próprias funcionalidades do sistema.

A princípio, o Lepic foi projetado para análise semiautomática da leitura do paciente, entretanto, existem mudanças já planejadas para incorporar uma inteligência artificial para análise automática dos erros e acertos deste. Para isso, foi escolhido o Wav2Vec2, um modelo para ASR (Automatic Speech Recognition), o qual será treinado com uma base de dados dos pesquisadores, os quais coletam dados para essa finalidade desde 2017. Além disso, para aprimorar o treinamento do modelo, será utilizada a base de dados CORAA-ASR, que é amplamente reconhecida por conter gravações de fala em português.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Instituição, sob número CAAE 35589420.5.0000.5149. O aplicativo está sendo desenvolvido em linguagem Dart, utilizando o framework Flutter (framework de desenvolvimento multiplataforma - web, desktop, android, iOS), com o auxílio de ferramentas como Visual Studio Code, Android Studio, Xcode, Git, GitHub, entre outras.

O desenvolvimento do software ocorre com constantes atualizações e melhorias de código, de forma que sua publicação se mantenha atualizada (seguindo as atualizações de mercado) e sua utilização pelo público possibilitada, de maneira gratuita (o aplicativo está disponível para download na Play Store, para dispositivos Android). Assim sendo, o software apresenta bons resultados aplicados em situações reais - conforme figura 1, que apresenta parte do fluxo do usuário que realiza a avaliação da fluência leitora de um paciente.

Figura 1 – Prints da tela do aplicativo com representações de: a) exemplo do gráfico com o monitoramento de 3 leituras; b) tela do texto com marcação dos erros para o cálculo da acurácia; c) caixa de descrição do tipo de erro para análise qualitativa.



Fonte: Aplicativo Lepic

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do aplicativo Lepic, há o constante desafio de criar uma ferramenta intuitiva, prática e coerente, que utilize de boas práticas de engenharia de software, como visibilidade de erros, arquitetura de projetos, minimização de cliques e a necessidade de aplicativos externos. É, ainda, necessário possibilitar uma correta experiência do usuário, com a preocupação de confeccionar funcionalidades e telas descritivas. Dessa forma, tendo em vista que o público-alvo e as aplicações do software não envolvem diretamente a área de tecnologia (a computação é utilizada como meio), o fluxo de usabilidade deve ser claro e intuitivo, de forma que impossibilite complicações de uso para usuários tecnologicamente inexperientes.

Dada sua aplicabilidade, o software desenvolvido é um nítido exemplo da utilização da informática em saúde, refletindo positivamente na sociedade. No referido projeto, o aluno de computação tem a oportunidade de estar em contato com conceitos que vão além de sua área de atuação, sendo necessário utilizar seus conhecimentos de maneira interdisciplinar. Ademais, é possível visualizar, de maneira prática e aplicada, o impacto do software desenvolvido, uma vez que o aplicativo é utilizado para facilitar processos que, outrora, eram realizados de forma manual.

5. REFERÊNCIAS

1. National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instructions. (2000). Technical Report. USA: National Institute of Child Health & Human Development.
2. Alves LM, Carvalho IM, Santos LF dos, Ribeiro G de L, Freire L de SC, Martins-Reis V de O, et al.. Reading fluency during the COVID-19 pandemic: a longitudinal and cross-sectional analysis. *Arq Neuro-Psiquiatr* [Internet]. 2022Oct; 80(10):994–1003. Disponível em: <https://doi.org/10.1055/s-0042-1758446>
3. Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, 45(2), 230–251. Disponível em: <https://doi.org/10.1598/rrq.45.2.4>
4. Celeste, L. C., Pereira, E. S., Pereira, N. R. R., & Alves, L. M. (2018). Parâmetros prosódicos de leitura em escolares do segundo ao quinto ano do ensino fundamental. In *CoDAS* (Vol. 30, p. e20170034). Sociedade Brasileira de Fonoaudiologia.

Educação em Saúde: O uso de Inteligência Artificial e Ambientes Imer- sivos para apoiar a Aprendizagem

Anrafel Fernandes Pereira^{1,2}, Gabriel Maranduba Littleton Lage¹

¹Universidade de Vassouras, Vassouras, RJ, Brasil.

²Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

anrafel.pereira@univassouras.edu.br, gabriel.lage@univassouras.edu.br

Resumo. O estudo discute a aplicação de Inteligência Artificial Generativa e Ambientes Virtuais Imer-
sivos, como o metaverso, na educação em saúde. O estudo é um relato de experiência
vivenciado em uma disciplina optativa de Inteligência Artificial aplicada em Saúde na Universidade
de Vassouras. Os alunos participaram de atividades práticas que incluíam a utilização de IA e o
metaverso para criar cenários de aprendizagem dinâmicos e discussões sobre o metaverso e o uso
de IA para gerar casos clínicos detalhados. A integração dessas tecnologias oferece simulações
realísticas, promovendo um aprendizado progressivo e seguro para os discentes. O estudo destaca
os desafios técnicos e financeiros, mas enfatiza os benefícios potenciais para a educação e a
prática na área da saúde.

Palavras-chave: Inteligência Artificial; Metaverso; Educação em saúde.

1. CONTEXTO E MOTIVAÇÃO

No campo das ciências da saúde, enfrentamos desafios inéditos diariamente, mas, em contrapartida,
também celebramos descobertas revolucionárias. No contexto educacional, um tópico tem se
destacado nas discussões e debates acadêmicos: a Inteligência Artificial (IA) (Donato et al., 2023;
Oliveira & Pinto, 2023). A Inteligência Artificial (IA) pode ser definida como uma subárea da informática
que busca desenvolver sistemas capazes de executar tarefas que, anteriormente, exigiriam a
inteligência humana. Esta definição abrange desde funções simples, como o reconhecimento de
padrões, até operações complexas, como diagnósticos médicos e tomada de decisões clínicas
(Meskó, & Görög, 2020).

Outro exemplo emblemático dos avanços tecnológicos digitais na área da saúde é a utilização do
metaverso. Através da criação de cenários em ambientes virtuais, por exemplo, é possível apoiar
a preparação de procedimentos complexos e simulações realísticas em saúde, proporcionando
práticas mais seguras e realistas. A IA Generativa oferece diversas oportunidades, entre elas, pode
contribuir gerando casos clínicos detalhados, permitindo o ajuste da complexidade conforme a
experiência do usuário e promovendo um aprendizado progressivo.

A integração dessas tecnologias oferece feedback em tempo real e personalização do ensino,
aprimorando habilidades clínicas. Apesar dos desafios técnicos e de investimento, essa abordagem
inovadora amplia o alcance da educação, reduz custos e prepara melhor os profissionais de saúde
para situações reais, marcando um avanço significativo na área médica (RIGBY, 2019).

2. ATIVIDADES PRINCIPAIS

Este trabalho retrata um relato de experiência de atividades práticas, inseridas no contexto de
educação em saúde, vivenciado na disciplina optativa intitulada Inteligência Artificial aplicada a
Saúde ofertada no curso de Medicina da Universidade de Vassouras no semestre de 2024.1. O

estudo emprega o uso de ambientes virtuais imersivos, também conhecidos como metaverso
e a Inteligência Artificial Generativa (IAG) como ferramentas para apoiar a criação de cenários
de aprendizagem dinâmicos e imersivos. As atividades práticas realizadas envolveram: (i) a
apresentação sobre o que é o metaverso; (ii) discussões em sala de aula sobre oportunidades

de utilização do metaverso na saúde; e (iii) a apresentação de um desafio prático, que é descrito com detalhes na seção seguinte. É importante ressaltar que os conceitos e as oportunidades para utilização da Inteligência Artificial na saúde já haviam sido apresentados e discutidos previamente com os alunos em sala de aula.

Este trabalho explora como o metaverso pode ser utilizado para criar cenários de simulações e como a IAG pode apoiar a criação de casos clínicos detalhados para serem trabalhados nesses ambientes virtuais imersivos. Evidências mostram que essas simulações permitem que estudantes e profissionais de saúde pratiquem procedimentos, tomem decisões críticas e experimentem diferentes situações clínicas sem os riscos associados a treinamentos em ambientes reais (Morgado, 2022).

3. DESENVOLVIMENTO

Após a realização dos primeiros passos apresentados na seção anterior, o passo três consistiu na proposta de um desafio prático que deveria ser realizado apoiado na utilização de ambientes imersivos de aprendizagem, tais como o metaverso. Para isso, a ferramenta Spatio.io foi apresentada e explorada juntamente com os alunos em sala de aula. Vários espaços foram apresentados, entre eles um espaço desenvolvido no Spatio para a Universidade de Vassouras¹.

O desafio constituiu em duas etapas. A primeira etapa envolveu, essencialmente a descrição de uma prática da área da saúde que pudesse ser realizada no metaverso. Para isso, os alunos foram orientados a descrevê-la, indicar o espaço em que a prática pudesse ser realizada (link do ambiente imersivo na ferramenta Spatial.io), e ainda discutir os benefícios e as oportunidades que a realização desta prática no metaverso pudesse trazer para a área da saúde, bem como as limitações e dificuldades percebidas, aspectos éticos, entre outros pontos que considerassem relevantes. Para a segunda etapa foi solicitado que os alunos pudessem escolher alguma ferramenta de IAG para geração de textos, tais como o ChatGPT, Gemini, entre outras, para apoiá-los na geração de casos clínicos detalhados que pudessem ser realizados no ambiente que eles descreveram para a primeira etapa. Para isso, os alunos foram orientados a escrever prompts para apoiá-los na geração destes cenários.

Os desafios foram realizados em duplas ou trios. Entre as práticas propostas pelos grupos estão: (i) a realização de consultas médicas por telemedicina através do metaverso; (ii) a criação de cenários para a discussão de casos clínicos por uma equipe multidisciplinar; (iii) simulação de procedimentos cirúrgicos, por exemplo, através de um espaço no metaverso que retrata uma sala cirúrgica, profissionais e estudantes podem praticar cirurgias, revisando ou praticando o procedimento antes de iniciá-la de fato; (iv) aulas práticas de anatomia; (v) atendimentos com psiquiatras ou psicólogos em que o paciente não precise sair de casa e possa estar imerso em uma sala digital e relaxante que melhor o agrade com o uso de um óculos de realidade virtual, por exemplo; (vi) conferência médica; entre outros. A proposta foi que com a geração de casos clínicos utilizando IAG, os alunos pudessem, a partir de suas experiências, ampliar os seus conhecimentos no contexto de seu interesse e ainda, perceberem as oportunidades práticas de utilização de tecnologias como o metaverso e a IAG na educação em saúde.

4. DESAFIOS E APRENDIZADOS

Atualmente, a complexidade técnica ainda é um dos principais obstáculos, pois integrar o metaverso e a IA exige recursos tecnológicos e necessidades técnicas específicas. Além disso, o desenvolvimento

¹ Espaço da Universidade de Vassouras desenvolvido no Metaverso: <https://encurtador.com.br/HyUMY>.

de ambientes virtuais realistas e algoritmos de IA avançados requerem financiamento substancial para construção, manutenção e atualização contínua, representando um desafio financeiro considerável. A adoção e aceitação dessas novas tecnologias também apresentaram dificuldades, uma vez que estudantes e profissionais de saúde podem mostrar resistência inicial, mas que podem ser superadas através de workshops, treinamentos e demonstrações práticas, como foi o caso do relato vivenciado na disciplina optativa do curso de Medicina da Universidade de Vassouras. Superar esses desafios pode proporcionar aprendizados valiosos para toda a comunidade. A colaboração interdisciplinar é essencial, destacando a importância de reunir especialistas de diferentes áreas, tais como tecnologia e saúde, para enfrentar os problemas e garantir o sucesso destas práticas. Por fim, integrar novas tecnologias na educação em saúde requer uma visão ampla, considerando todos os aspectos técnicos e humanos envolvidos, reforçando a importância de uma abordagem holística e iterativa.

5. REFERÊNCIAS

- Donato, H., Escada, P., & Villanueva, T. (2023). A transparência da ciência com o ChatGPT e as ferramentas emergentes de inteligência artificial: como se devem posicionar as revistas científicas médicas. *The Transparency of Science with ChatGpt and the Emerging Artificial Intelligence Language Models: Where Should Medical Journals Stand*.
- Meskó, B., & Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. *NPJ digital medicine*, 3(1), 126.
- Morgado, L. (2022). Ambientes de aprendizagem imersivos. *Video Journal of Social and Human Research*, 102-116.
- Oliveira, L., & Pinto, M. (2023). A inteligência artificial na educação: ameaças e oportunidades para o ensino-aprendizagem.
- Rigby, M. J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics*, 21(2), 121-124.

Elucidativa: usando modelos de linguagem para explicar resultados de exames complementares

Luana Cruz Machado¹, Rafael Petri Pinheiro¹, Felipe Sahb Furtado²

¹Faculdade de Medicina - Universidade Federal do Rio de Janeiro (UFRJ),
Rio de Janeiro, RJ, Brasil. ²Amparo, São Paulo, SP, Brasil.

{luana,rafael, felipe}, luanacruzmachado@hotmail.com, rafaelpetri@ufrj.br, felipe@amparo.com.br

Resumo. No Brasil, cerca de dois bilhões de exames são realizados anualmente, muitos dos quais pacientes têm dificuldade em interpretar, podendo levar a perdas de seguimento e desconhecimento de sua saúde. Modelos de linguagem (LLM), como o GPT-4, têm sido explorados para interpretar e explicar exames, promovendo a autonomia do paciente. Este projeto utiliza reconhecimento óptico de caracteres (OCR) com o GPT-4 para extrair e simplificar laudos médicos, tornando as informações mais acessíveis aos pacientes. Embora seja uma prova de conceito, são necessários estudos qualitativos para validar a precisão das explicações e avaliar se os pacientes retêm melhor as informações após receberem os laudos simplificados.

Palavras-chave: Exames médicos, Modelos de linguagem, GPT-4, OCR, Inovação, Saúde digital.

Nome do projeto: Projeto Elucidativa: IA para Simplificação de Laudos Médicos

Links: O aplicativo está disponível em <https://ai4good.streamlit.app/> e o vídeo de demonstração pode ser acessado em <https://shorturl.com/8iLV>.

1. CONTEXTO E MOTIVAÇÃO

No Brasil, exames complementares são frequentemente solicitados para direcionar o tratamento clínico. Apenas no sistema de saúde suplementar, 1,1 bilhão de exames foram realizados em 2022, de acordo com o Mapa Assistencial da Saúde Suplementar (AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR, 2023). Normalmente, os resultados de exames laboratoriais, de imagem ou patológicos, como biópsias, são entregues diretamente ao paciente - seja fisicamente ou por meio de meios digitais - para que sejam apresentados ao médico assistente que os solicite posteriormente. Esse fluxo apresenta pontos de fragilidade que podem comprometer o seguimento do cuidado. A interpretação equivocada dos resultados de exames pelos pacientes, por exemplo, pode ocasionar perda do acompanhamento (CAVALCANTE; REIS, 2021), o que é especialmente prejudicial em casos de exames de rastreamento como Papanicolau, Mamografia e Colonoscopia. Isso pode ocorrer porque a linguagem utilizada nos resultados de exames é técnica, repleta de jargões e complexos, o que dificulta sua compreensão por pessoas sem instrução específica na área da saúde.

Entendendo que esse problema é importante e potencialmente solucionável de maneira escalável com o uso de tecnologias desenvolvidas recentemente, como os grandes modelos de linguagem (LLM) fundacionais, propôs-se o desenvolvimento da ferramenta Elucidativa. Acredita-se que, ao simplificar os resultados de laudos médicos para linguagem comum, abre-se caminho para maior autonomia e assertividade dos pacientes em relação às condições de saúde. Isso reforça os ideais de humanização e centralidade no paciente, permitindo decisões informadas e compartilhadas.

1.1 Revisão da Literatura

Diversas iniciativas têm explorado a aplicação de LLMs na saúde. Estudos como os de Cadamuro et al. (2023) e Chow et al. (2023) mostram o potencial de ferramentas como ChatGPT na interpretação de exames. A Elucidativa se diferencia ao focar na simplificação de laudos para pacientes leigos, utilizando OCR combinado com GPT-4.

1.2 Objetivo

Desenvolver uma ferramenta que utilize OCR e GPT-4 para interpretar e explicar resultados de exames médicos de forma acessível, promovendo a compreensão e a tomada de decisões informadas pelos pacientes.

2. ATIVIDADES PRINCIPAIS

Durante a fase inicial da Elucidativa, utilizamos Design Thinking para definir os requisitos. Entrevistamos 5 pacientes, 2 estudantes de medicina e 2 médicos para entender as necessidades dos usuários. A partir do estudo da jornada do usuário, concluímos que o melhor formato de entrada seria a imagem, pois os exames laboratoriais são frequentemente compartilhados em formato físico (IMPRESSOS) ou digital (PDF). Desenvolvemos uma arquitetura modular com funcionalidades centrais como OCR, processamento em modelos de linguagem (LLM) e geração de explicações simplificadas. Utilizamos Python e as bibliotecas Streamlit, EasyOCR e a API do GPT-4 da OpenAI para implementar uma aplicação web. Os insights gerados na etapa de Design Thinking levaram à definição de funcionalidades centrais, como a remoção óptica de texto de imagens (OCR), o envio desse texto para processamento em modelos de linguagem (LLM) e a geração de explicação simplificada.

[Figura 1. Interface de usuário do aplicativo](#)

[Figura 2. Desenho esquemático do funcionamento da aplicação:](#) O texto extraído é exibido ao usuário para conferência enquanto o GPT-4 transforma o laudo em um texto explicativo para leigos no plano de fundo. Por fim, essa explicação simplificada é exibida na tela.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Testes preliminares de usabilidade e satisfação foram conduzidos com 7 usuários, todos com ensino médio completo, que demonstraram alta satisfação com a aplicação. No link disponibilizado ao público, 55 visitantes únicos foram registrados até o momento. Usuários são sempre instruídos a procurar seu médico antes de tomar qualquer outra atitude em resposta ao resultado da interpretação do exame pelo aplicativo.

[Figura 3. Visualização de analytics do aplicativo no streamlit:](#) A imagem mostra a seção de analytics da Elucidativa hospedada no Streamlit. Na seção “App viewers”, mostra que o aplicativo teve 55 visualizações únicas. Abaixo, uma lista dos visitantes mais recentes, com nomes como “Abstruse Torte”, “Spiraling Brownie”, preservando o anonimato dos usuários, e a última vez que acessaram o aplicativo.

[Figura 4: Exemplo de laudo padrão de exame de imagem e a respectiva explicação fornecida ao usuário, dado o laudo apresentado:](#) A imagem demonstra exemplos de tomografia computadorizada e a explicação fornecida pelo aplicativo. Informações que permitissem a identificação do paciente foram ocultadas.

[A figura 5. Demonstra um ultrassom de artérias carótidas com tradução e a explicação fornecida pelo aplicativo. Informações que permitissem a identificação do paciente foram ocultadas:](#) A primeira imagem é a visualização do ultrassom, destacando a artéria carótida com medições de velocidade do fluxo sanguíneo (PSV e EDV). A segunda imagem mostra a interface do aplicativo Elucidativa no Streamlit, exibindo a tradução concluída de um laudo médico original para um formato simplificado e de fácil compreensão.

4. DESAFIOS E APRENDIZADOS

A Elucidativa demonstrou ser uma ferramenta promissora, mas desafios como a integração de dados complexos e a personalização das explicações ainda precisam ser superados. A escalabilidade da tecnologia permite oferecer proteção acessível a todos os resultados de exames, beneficiando os usuários com uma compreensão mais profunda e fundamentada de sua saúde. Estudos qualitativos futuros avaliarão a precisão das explicações e o impacto na retenção de informações pelos pacientes. Está planejado um estudo qualitativo para avaliar a acurácia das explicações fornecidas pelo modelo de linguagem, conforme a avaliação de médicos, e se há de fato aumento da retenção de informação pelos pacientes, medindo o entendimento dos resultados antes e depois da apresentação do laudo simplificado. Entre outros aprimoramentos futuros, estão previstos deploy em ambiente com suporte a GPU para melhoria de desempenho, filtros para garantir o processamento apenas de imagens válidas, uso de datasets no treinamento dos modelos de linguagem e emprego de técnicas de Retrieval Augmented Generation (RAG).

5. REFERÊNCIAS

Agência Nacional de Saúde Suplementar. Planos de saúde realizaram 1,8 bilhão de procedimentos em 2022.

Cadamuro J, et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. *Clinical Chemistry and Laboratory Medicine*. 2023;61(7):1158-1166.

Cavalcante GHO, Reis GJ dos. Avaliação do seguimento de lesões precursoras de câncer do colo do útero – uma revisão bibliográfica. *Pesquisa e Ensino em Ciências Exatas e da Natureza*. 2021;5:25.

Chow JCL, et al. Developing an AI-assisted educational chatbot for radiotherapy using the IBM Watson Assistant platform. *Healthcare (Basel)*. 2023;11(17):2417.

Mokmin NAM, Ibrahim NA. The evaluation of chatbot as a tool for health literacy education among undergraduate students. *Education and Information Technologies*. 2021;26(5):6033-6049.

Estratégia para Padronização de Dados em Imagens Mamográficas para Análises por Radiômica

Érica M. Policarpo¹, Orlando Rodrigues Jr.¹, Mário O. Menezes¹

¹Instituto de Pesquisas Energéticas e Nucleares – IPEN - CNEN
mo.menezes@gmail.com, ericamauriciopolicarpo@gmail.com

Resumo. A mamografia digital é considerada o método com maior sensibilidade e especificidade para o rastreamento do câncer de mama. A evolução dos sistemas de aquisição das imagens médicas para o formato digital, possibilitou o auxílio de técnicas e ferramentas computacionais na detecção e identificação de possíveis patologias. Radiômica, uma dessas técnicas, faz a extração de dados quantitativos dos fenótipos de imagens radiológicas, visando auxiliar tanto no diagnóstico como no acompanhamento da evolução do tratamento. Este trabalho apresenta uma metodologia para a padronização e análises por radiômica, de dados gerados de imagens mamográficas adquiridas em equipamentos de diferentes fabricantes, que seja de fácil reprodução, através do uso de softwares livres e de um simulador padrão para mamografia digital.

Abstract. Digital mammography is considered the method with the highest sensitivity and specificity for breast cancer screening. The evolution of digital medical imaging technologies has enabled computational tools to assist in the analysis and identification of potential pathologies. Radiomics involves the extraction of quantitative data from radiological image phenotypes, aimed at aiding both diagnosis and treatment monitoring. This study presents a methodology for standardizing and conducting radiomic analyses of mammographic image data acquired with equipment by different vendors, designed to be easily reproducible using open-source software and a standard digital mammography simulator.

Palavras-chave: Radiomics; mamografia digital; processamento de imagem; imagens médicas.

Nome do projeto: Desenvolvimento De Uma Metodologia Para A Radiômica Em Mamografia Digital.

1. CONTEXTO E MOTIVAÇÃO

O câncer de mama é o tipo mais comum em mulheres, e é o responsável pela maioria das mortes por câncer (1). A introdução de tecnologias de aquisição de imagens médicas no formato digital, possibilitou que ferramentas computacionais auxiliem na análise e identificação de possíveis patologias (2). Radiomics ou radiômica, é a extração de dados quantitativos dos fenótipos de imagens radiológicas, com a proposta de auxiliar tanto no diagnóstico como no acompanhamento à resposta ao tratamento do câncer (3). As características radiômicas são divididas em classes fundamentadas em análises estatísticas baseadas em histograma, textura e forma. Porém ainda não possui uma padronização totalmente caracterizada em todo seu processo para imagens mamográficas, por exemplo, em como essas imagens devem ser adquiridas e processadas, pois os valores calculados das várias características podem sofrer variação entre si, quando as imagens são adquiridas com protocolos de aquisição variados e equipamentos de fabricantes diferentes (4,5).

1.1 Objetivo

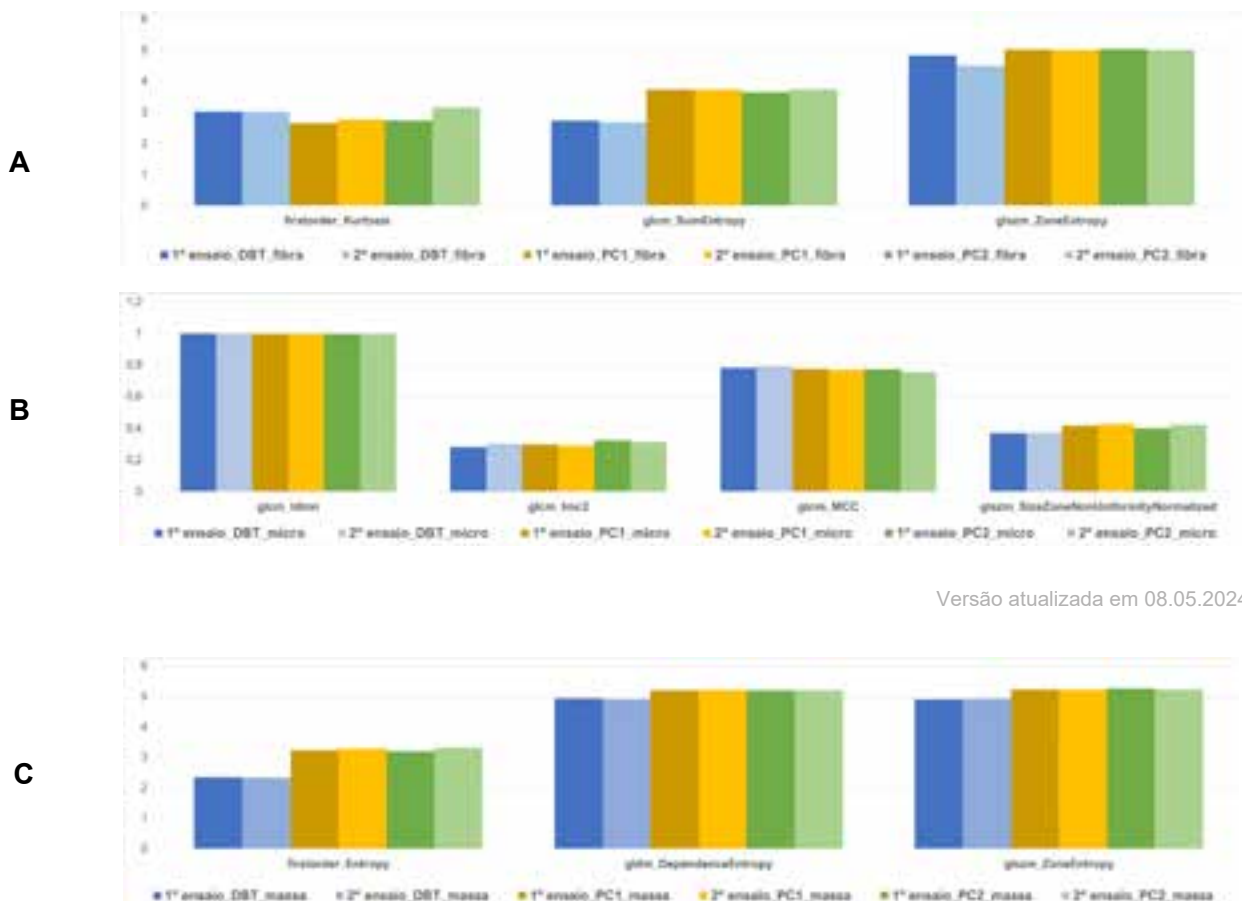
Desenvolver uma estratégia para a padronização e o planejamento inicial de análises por radiômica em imagens de mamografia digital.

2. METODOLOGIA E RESULTADOS

Foram utilizados 3 equipamentos de mamografia digital, sendo dois com tecnologia contador de fótons e um com tecnologia tomossíntese. Para a aquisição das imagens foi utilizado o simulador padrão do Colégio Americano de Radiologia (ACR). O simulador ACR possui Test Objects com os três principais achados patológicos mamográficos de interesse: Fibras, Microcalcificações e Massa. No total foram adquiridas 30 imagens em 2D do simulador ACR para a análise, sendo 5 imagens em cada equipamento, divididas em 2 ensaios de aquisição, com um intervalo de um ano entre os ensaios, para avaliar a reprodutibilidade e a estabilidade da padronização. Nenhuma imagem de ser humano foi utilizada nesta pesquisa. Foram realizados processamentos pós aquisição nas imagens do simulador ACR, de reamostragem do tamanho do pixel, discretização das intensidades de tons de cinza e segmentação das Regiões de Interesse (ROI) utilizando o software livre 3D Slicer, que possibilita também carregar o plugin Radiomics, com a biblioteca PyRadiomics, para o cálculo e extração das características radiômicas. Para as análises estatísticas, foram utilizados os recursos e bibliotecas do Python, executando no Jupyter Notebook.

Após a padronização e segmentação das ROIs dos Test Objects nas imagens, o plugin calcula e extrai 108 características radiômicas. Aquelas que apresentaram um coeficiente de variação <15% foram consideradas como as que melhor representaram os Test Objects nesse tipo de imagem. A Figura 1 apresenta algumas das características para os 3 Test Objects que apresentaram boa reprodutibilidade e estabilidade para os valores calculados: **A** - Test Object **Fibra**; **B** - Test Object **Microcalcificação**, e **C** - Test Object **Massa**.

Figura 1 – Comportamento de características radiômicas após a padronização das imagens.



Fonte: O autor.

3. CONCLUSÃO

A metodologia aplicada para a padronização das imagens apresentou resultados satisfatórios, possibilitando extrair e analisar dados radiômicos de imagens adquiridas com diferentes equipamentos de mamografia digital; também foi possível identificar as seguintes características com boa estabilidade e reprodutibilidade: Fibra: Energy; 10 Percentile; 90 Percentile; Kurtosis; Sum Entropy; IDMN; Zone Entropy; RLNU; GLNU. Microcalcificação: 10 Percentile; 90 Percentile; Energy; MCC; Sum Entropy; IDMN; IMC2; Zone Entropy; SZNU. Massa: 10 Percentile; 90 Percentile; Entropy; Kurtosis; Energy; Median; Dependence Entropy; IDMN; Zone Entropy; GLNU.

4. DESAFIOS E APRENDIZADOS

A aluna iniciou o projeto com compressão limitada sobre a complexidade do uso das ferramentas do software 3DSlicer, e em como extrair e analisar os dados radiômicos das imagens. No desenvolvimento da pesquisa, a aluna demonstrou capacidade em superar os desafios encontrados através de pesquisas bibliográficas, treinamentos e cursos adicionais para aprimorar seu conhecimento em teoria da formação e processamento da imagem digital, em linguagem de programação e análise de dados.

5. REFERÊNCIAS

1. Caleffi M, Ribeiro RA, Filho DLD, Ashton-Prolla P, Bedin AJ, Skonieski GP, et al. A model to optimize public health care and downstage breast cancer in limited-resource populations in southern Brazil. (Porto Alegre Breast Health Intervention Cohort). BMC Public Health. dezembro de 2009;9(1):83.
2. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imaging. 22 de agosto de 2016;3(3):034501.
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. fevereiro de 2016;278(2):563–77.
4. Ger RB, Zhou S, Chi PCM, Lee HJ, Layman RR, Jones AK, et al. Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies. Sci Rep. 29 de agosto de 2018;8(1):13047.
5. Yip SS, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol. 7 de julho de 2016;61(13):R150–66.

HOMOGENiSE: Um método para organizar o conhecimento de dados em análise quali-quantitativa

Evaldo de Oliveira da Silva¹, Marcello Peixoto Bax², Auxiliatrice Caneschi Badaró¹

¹Centro Universitário UniAcademia, Juiz de Fora, MG, Brasil.

²Universidade Federal de Minas Gerais (PPGGOC/UFMG), Belo Horizonte, MG, Brasil.

evaldosilva@uniacademia.edu.br, bax@eci.ufmg.br, auxiliatricebadaro@uniacademia.edu.br

Resumo. A estratificação de risco (ER) em saúde mental permite a avaliação de sinais e sintomas pontuados de acordo com o grau de gravidade do paciente. Equipes de atendimento usam a ER para classificar o nível de risco que orienta as intervenções em tratamentos psicoterapêuticos e que produzem dados qualitativos relevantes. Esses dados produzidos abrem caminho para a análise quali-quantitativa a fim de evidenciar a prática clínica. No entanto, a formalização da produção dos dados, a reutilização do conhecimento das variáveis e a visualização de dados mesclados, apresentam-se como desafios. Este trabalho descreve o método HOMOGENiSE, que formaliza a produção de dados qualitativos e quantitativos para organizar o conhecimento e integrar esses dados usando grafos de conhecimento. Ele é baseado em técnicas de web semântica e inteligência artificial, e visa superar os desafios da análise quali-quantitativa. O HOMOGENiSE foi aplicado em um estudo de caso que realizou análises quali-quantitativas de dados psicoterapêuticos com a colaboração de psicólogos.

Palavras-chave: Saúde mental; Métodos mistos; Ontologia; Web semântica.

Abstract. Risk stratification (RE) in mental health allows the assessment of signs and symptoms scored according to the degree of severity. Care teams use RE to classify a risk level that guides interventions in psychotherapeutic treatments that produce relevant qualitative data. Generating qualitative and quantitative data paves the way for quali-quantitative analysis to highlight clinical practice. However, this analysis presents challenges, such as the formalization of data production, the reuse of knowledge of variables, and the visualization of merged data. Thus, this work describes the HOMOGENiSE, a method that formalizes producing qualitative and quantitative data to organize knowledge and integrate these data using knowledge graphs. It is grounded on semantic web and artificial intelligence techniques to help overcome qualitative and quantitative analysis challenges. HOMOGENiSE was applied in a case study that performed quali-quantitative analysis of psychotherapeutic data with collaboration of psychologists.

Keywords: Mental Health; Mixed Methods; Ontology; Semantic web.

1. CONTEXTO E MOTIVAÇÃO

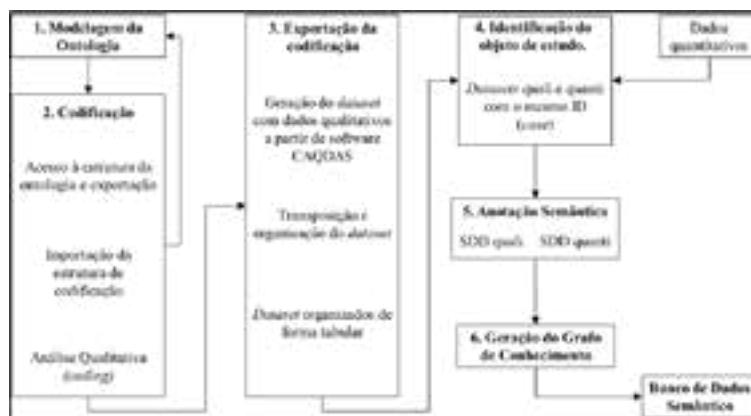
A ER é um processo estatístico que apoia tratamentos ajustado ao risco atribuído ao paciente (1). Em saúde mental a ER é aplicada por questionários que agrupam questões por transtornos mentais, usados para avaliar sinais e sintomas pontuados de acordo com o grau de gravidade, a fim de determinar um nível de risco. Em tratamentos psicoterapêuticos a ER é usada em casos clínicos para identificar o risco em transtornos de ansiedade e depressão (2). Os dados quantitativos formam a base para os protocolos de intervenção desenvolvidos nas sessões terapêuticas, que fornecem dados qualitativos relevantes. A produção de dados quantitativos e qualitativos oportuniza o uso de métodos mistos (ou análise quali-quantitativa) como técnica de pesquisa, visando evidenciar a prática clínica. A análise quali-quantitativa visa mesclar os dados em um mesmo objeto de estudo. Porém, a formalização da produção dos dados, o compartilhamento do conhecimento das variáveis, a reutilização do conhecimento de dados qualitativos e a visualização dos dados combinados, apresentam-se como desafios em análise quali-quantitativa (3). Técnicas derivadas da web semântica e inteligência artificial podem auxiliar a superar esses desafios. Ontologias são geralmente usadas como estruturas para descrever o conhecimento contido em coleções de dados e podem facilitar a

transferência de informações, preparação de dados, desambiguação de conceitos e preservação semântica das variáveis do estudo (4). A anotação semântica permite que especialistas conectem dados com conceitos estruturados em ontologias, abrindo caminho para explorar novos conhecimentos por meio de grafos de conhecimento (GCs). GCs organizam dados semanticamente anotados para analisar e filtrar informações em resposta a questões de pesquisa (5). Eles também facilitam a visualização gráfica de dados e têm o potencial de aprimorar análises quali-quantitativas. Deste modo, a Seção 2 deste trabalho descreve as etapas do método HOMOGENiSE (*metHod for Ontological Modeling Of knowledGE iN quali-quantitativE reSEarch*)¹ (6). O método formaliza a produção de dados quali-quantitativos, representando e organizando o conhecimento sobre eles, integrando-os por meio de GCs e mesclando-os em um mesmo objeto de estudo. A Seção 3 descreve como o método foi aplicado em um estudo de caso com dados de tratamentos psicoterápicos. Finalmente, a Seção 4 apresenta os desafios e aprendizados.

2. ETAPAS DO MÉTODO HOMOGENiSE

O método é composto por sete etapas. A etapa inicial é a Descrição do Minimundo, centrada na identificação do problema, que permite elaborar questões de competência (QC) que orientam as análises quali-quantitativas. A Definição dos Objetivos é a etapa que instiga o desenvolvimento das etapas posteriores, e sugere elaborar um planejamento para aplicar o método. A terceira etapa de Concepção e Desenvolvimento (CD), especifica atividades representadas na Figura 1.

Figura 1 – Etapa de Concepção e Desenvolvimento.



Fonte: SILVA (2023) (6).

A etapa de CD é executada por especialistas que modelam a ontologia para representar o conhecimento sobre os dados. A estrutura da ontologia é acessada para codificar dados qualitativos por meio de CAQDAS (Computer Aided Qualitative Data Analysis Software), que automatiza a importação da estrutura de conceitos da ontologia e exportação dos dados qualitativos em formato tabular, compatível com o formato dos dados quantitativos. A identificação do objeto de estudo estabelece o mesmo ID em ambos datasets. Em seguida, a anotação semântica dos dados permite associar os conceitos da ontologia às colunas dos dados dos datasets, utilizando a técnica SDD (Semantic Data Dictionary) (5). O script `sdd2rdf` escrito em Python ingere os SDDs (quali e quanti) com os dados anotados semanticamente para gerar GCs que podem ser armazenados em banco de dados semânticos. A etapa seguinte do método visa demonstrar como fragmentos dos GCs podem responder às QCs. A etapa de avaliação usa indicadores que mensuram o tempo de execução do método e o percentual de reutilização da codificação. A etapa de divulgação permite compartilhar os resultados com os interessados nas análises dos dados. A última etapa visa controlar a evolução dos artefatos produzidos por meio de uma matriz de rastreabilidade.

¹ <https://github.com/evaldo/homogenise>

3. APLICAÇÃO DO MÉTODO HOMOGENISE

O HOMOGENISE é resultado de pesquisa realizada no Programa de Pós-Graduação em Gestão e Organização do Conhecimento da UFMG (PPGGOC/UFMG) (6). A pesquisa aplicou o método em um estudo de caso que permitiu realizar análises quali-quantitativas de dados psicoterápicos. Psicólogos atuaram como especialistas de domínio em Terapia Cognitivo-Comportamental (TCC). Dados sintéticos extraídos de artigos científicos forneceram informações baseadas em ER, tratamento de ansiedade, questionários de casos clínicos e protocolos de intervenção encontrados na TCC (7) (8). A execução do estudo de caso apoiou a modelagem de uma nova ontologia de domínio chamada ONTRISCAL (*Ontology of Risk Stratification in Mental Health*)². Os tipos de dados provenientes de tratamentos psicoterápicos, usados como domínio de aplicação do estudo de caso, e integrados em um único grafo de conhecimento, aprimoraram as análises quali-quantitativas dos casos clínicos.

4. DESAFIOS E APRENDIZADOS

O método estabelece que a codificação deve utilizar a estrutura da ontologia de domínio, uma dificuldade adicional para profissionais que não dominam ontologias. A exportação da estrutura da ontologia para ser usada na codificação foi realizada manualmente, criando uma dificuldade técnica adicional para obter dados qualitativos. Observou-se que a tarefa de codificação demanda tempo para gerar coleções de dados com qualidade, e se o texto estiver inadequado (ou desorganizado em relação à sequência às sessões terapêuticas) a codificação pode gerar ambiguidades e enviesar a análise quali-quantitativa. Outro desafio é a compreensão das técnicas para análise quali-quantitativa, ou separadamente, análise qualitativa e quantitativa, e em como aplicá-las com o suporte de GCs.

5. REFERÊNCIAS

1. Miller CC, Reardon MJ, Safi HJ. Risk stratification: a practical guide for clinicians. Cambridge, UK ; New York, NY: Cambridge University Press; 2001.
2. Paiano M, Lopes FP, Miguel MEGB, Salci MA. Percepção dos enfermeiros sobre estratificação de risco em saúde mental e as ações de enfermagem. Rev Saúde Pública Paraná. 2019;2(1):68–79.
3. Khankeh H, Ranjbar M, Khorasani-Zavareh D, Zargham-Boroujeni A, Johansson E. Challenges in conducting qualitative research in health: A conceptual paper. Iran J Nurs Midwifery Res. 2015;20(6):635–41.
4. Gruber TR. A translation approach to portable ontology specifications. Knowl Acquis. 1993;5(2):199–220.
5. Rashid SM, Chastain K, Stingone JA, McGuinness DL, McCusker JP. The Semantic Data Dictionary Approach to Data Annotation & Integration. SemSci ISWC. 2017;2017.
6. Silva E de O, others. Homogenise: método para modelagem ontológica do conhecimento em pesquisas quali-quantitativas. 2023;
7. Melo ALC, Lourenço LM. Terapia cognitivo-comportamental no tratamento de um caso de transtorno de ansiedade generalizada: um relato de caso. Psicol PT Internet. 2020;1–23.
8. Remor EA. Tratamento psicológico do medo de viajar de avião, a partir do modelo cognitivo: caso clínico. Psicol Reflex E Crítica. 2000;13(1):205–16.

² <https://biportal.bioontology.org/ontologies/ONTRISCAL>

Identificação do risco neonatal com algoritmos de inteligência artificial embarcados em um dispositivo: Premie-teste - Inovação

Jannine dos Santos Nascimento (aluno)¹, Gabriela Silveira Neves (aluno)¹,
Zilma Silveira Nogueira Reis (orientador)¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

ninne.Nascimento@hotmail.com

Resumo. Reconhecer a prematuridade de acordo com o grau de maturidade pulmonar é fundamental para atender às necessidades imediatas dos recém-nascidos. Para isso, um novo dispositivo foi desenvolvido por meio de uma abordagem inovadora para estimar a idade gestacional (IG) e prever a maturidade do pulmão, com base nas propriedades fotobiológicas da pele do recém-nascido. O Premie-test é um dispositivo óptico portátil que reúne dados capturados por um sensor que entra em contato com a pele do recém-nascido, bem como de outras variáveis clínicas para prever o prognóstico neonatal através de algoritmos XGboost. A inovação tecnológica possibilita avaliar, de forma indireta, a maturidade pulmonar neonatal por meio da avaliação da maturidade da pele, através de suas propriedades fotobiológicas. Com isso, tornou-se possível o uso de uma ferramenta, acessível e de fácil manuseio, para auxiliar na tomada de decisão clínica ao nascimento, através de avaliação não-invasiva.

Palavras-chave: Inteligência artificial; equipamentos e suprimentos; idade gestacional; prematuridade; fenômenos fisiológicos da pele.

Nome do projeto: Detecção da prematuridade através da interação entre a luz e a pele neonatal: a validação do Premie-test. Número **CAAE: 81347817.6.1001.5149**.

1. CONTEXTO E MOTIVAÇÃO

O Premie-test é uma inovação tecnológica portátil desenvolvida para avaliar a maturidade da pele de um recém-nascido, através das propriedades fotobiológicas do tecido e do processamento de um algoritmo de aprendizado de máquina sobre o sinal de dispersão de luz adquirido na pele para prever o prognóstico neonatal precoce relacionado à prematuridade (1)(2). Este novo dispositivo de reflectômetro de pele automatizado, em combinação com variáveis clínicas, como peso ao nascer e uso de corticosteróide antenatal, foi capaz de prever a idade gestacional e pode ser útil quando a informação é duvidosa ou desconhecida (3). Modelos preditivos multivariáveis associaram a reflectância da pele com parâmetros clínicos fáceis de obter, no cenário de nascimento (4). A identificação precoce de recém-nascidos prematuros de acordo com o grau de maturidade pulmonar continua sendo uma meta relevante para os sistemas de saúde, já que a intervenção precoce para o manejo do desconforto respiratório no recém-nascido pode significar a diferença entre a sobrevivência e, possivelmente, a redução da mortalidade (5). Com o advento de uma ferramenta tecnológica acessível e de fácil utilização, que pode ser integrada na prática clínica atual, surge um caminho promissor para a avaliação indireta da maturidade pulmonar.

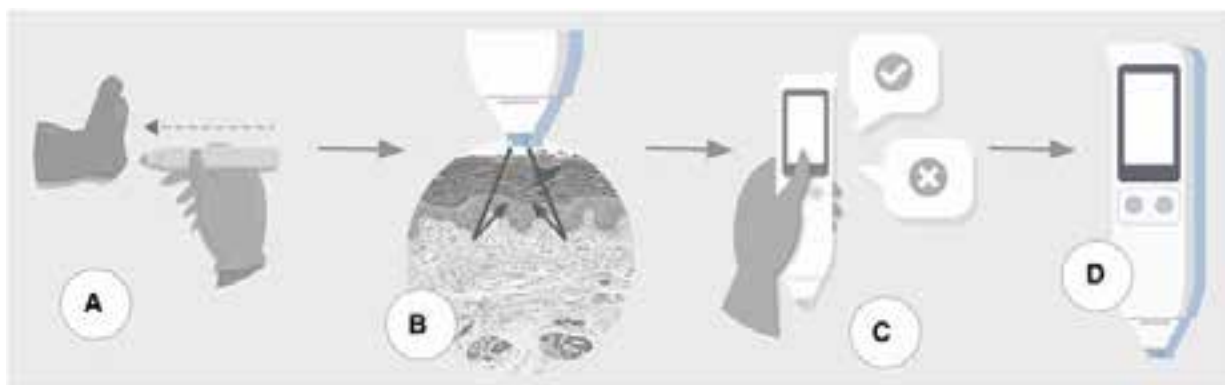
1.1 Objetivo

O objetivo da nova tecnologia é analisar a viabilidade do uso da reflectância da pele do recém-nascido para estimar a idade gestacional no nascimento de forma não-invasiva. Ao avaliar a capacidade de detectar prematuridade, é também possível prever o prognóstico neonatal precoce relacionado à prematuridade, e com isso avaliar a associação entre a maturidade pulmonar neonatal e a maturidade da pele, através de suas propriedades fotobiológicas.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

O Preemie-test já provou segurança e eficácia na capacidade de estimar a idade gestacional ao tocar a pele do recém-nascido, com acurácia de 91,4% em estudos prévios com 781 recém-nascidos. A variabilidade intra-observador foi de 1,97% (IC 95% 1,84% a 2,11%) e interobservador de 2,6% (IC 95%: 2,1% a 3,1%) (5). A nova tecnologia também provou através de um estudo caso-controle onde foram selecionados 640 recém-nascidos, sendo 470 (73,4%) do grupo com Síndrome do Desconforto Respiratório (SDR) e 170 (26,6%) controle, que a reflectância da pele foi associada à SDR, Exp(B): 0,982 (IC 95%: 0,979 a 0,985), modelo R² 0,632, P<0,001 (6). Além disso, um modelo preditivo promissor de SDR foi desenvolvido, onde de cada 100 recém-nascidos avaliados, 90 foram classificados corretamente como de alto risco ou não para SDR até 72 horas de vida (1). Sendo portanto, um estudo pioneiro a desvendar o fenômeno de associação entre a maturidade da pele e do pulmão, permitindo uma avaliação não-invasiva durante o período pós-natal.

Figura 1 - Etapas do processo de teste. (A) O dispositivo toca a pele de um recém-nascido. (B) O sensor adquire a maturidade da pele avaliando as propriedades fotobiológicas do tecido medindo as porções de reflexão do feixe de luz incidente na pele. (C) O usuário insere dados clínicos. (D) O processador de dados usa algoritmos de aprendizado de máquina para estimar a idade gestacional e a maturidade pulmonar.



Fonte: arquivo dos pesquisadores.

4. DESAFIOS E APRENDIZADOS

O desenvolvimento da tecnologia inovadora exigiu a integração de diversas áreas como Medicina, Física e Computação, criando um ambiente de pesquisa complexo. Os principais desafios incluíram a necessidade de uma linguagem comum entre as áreas de conhecimento, a unificação de métodos de pesquisa e a gestão eficiente de grandes volumes de dados de fontes diversas. A colaboração interdisciplinar foi crucial para enfrentar os desafios da saúde e da tecnologia contemporâneas. No Brasil, a disseminação de tecnologia enfrenta obstáculos como um sistema regulatório rigoroso para ensaios clínicos, infraestrutura de pesquisa desigual, diversidade étnica, variações na assistência médica e dificuldades de financiamento. Superar esses desafios demandou uma colaboração eficaz entre pesquisadores, patrocinadores, reguladores e instituições de saúde, além de um planejamento cuidadoso da pesquisa.

Este trabalho foi apoiado pela Fundação Bill & Melinda Gates, Grant Number OPP1128907 Contract. Isso não altera nossa adesão às políticas da ONE sobre compartilhamento de dados e materiais.

5. REFERÊNCIAS

1. Reis ZSN, et al. Respiratory distress syndrome prediction at birth by optical skin maturity assessment and machine learning models for limited-resource settings: a development and

- validation study. *Front Pediatr.* 2023 Nov 15;11:1264527. doi: 10.3389/fped.2023.1264527. PMID: 38054190; PMCID: PMC10694507.
2. Reis ZSN, et al. Prematurity detection evaluating interaction between the skin of the newborn and light: protocol for the preemie-test multicentre clinical trial in Brazilian hospitals to validate a new medical device. *BMJ Open.* 2019 Mar 5;9(3):e027442. doi: 10.1136/bmjopen-2018-027442. PMID: 30842119; PMCID: PMC6429838.
3. Vitral GLN, et al.. Gestational age assessed by optical skin reflection in low-birth-weight newborns: Applications in classification at birth. *Front Pediatr.* 2023 Mar 28;11:1141894. doi: 10.3389/fped.2023.1141894. PMID: 37056944; PMCID: PMC10086374.
4. Reis ZSN, et al. . A cross-sectional study. *PLoS One.* 2017 Sep 20;12(9):e0184734. doi: 10.1371/journal.pone.0184734. PMID: 28931040; PMCID: PMC5607181.
5. Reis ZSN, et al Newborn Skin Maturity Medical Device Validation for Gestational Age Prediction: Clinical Trial. *J Med Internet Res.* 2022 Sep 7;24(9):e38727. doi: 10.2196/38727. PMID: 36069805; PMCID: PMC9494223.
6. Neves GS, et al. Assessment of Skin Maturity by LED Light at Birth and Its Association With Lung Maturity: Clinical Trial Secondary Outcomes. *JMIR Biomed Eng.* 2023 Dec 25;8:e52468. doi: 10.2196/52468. PMID: 38875690; PMCID: PMC11041497.

In silico screening of p-coumaric acid derivatives with antileishmanial activity using methods coupled to Artificial Intelligence

T. M. Machado¹, F. Guilhon-Simplicio¹

¹Faculty of Pharmaceutical Sciences, Federal University of Amazonas, Manaus, AM, Brazil.

tallita.machado@yahoo.com.br, guilhon_simplicio@ufam.edu.br

Abstract. Leishmaniasis are a group of neglected tropical diseases caused by different trypanosomatid parasites of the genus *Leishmania*, being endemic in several countries and affecting 12 to 15 million people worldwide. In view of the few and obsolete therapeutic options, this study sought to plan, through in silico studies, derivatives of p-coumaric acid that had potential antileishmanial activity and that presented an ideal toxicological profile for drug candidates. Thus, 10 derivatives were planned, where D1 to D5, D7, D10 and p-coumaric acid obtained a complementarity profile with the enzyme pteridine reductase 1, and D2, D4-D6 and D8 obtained a complementarity profile with the enzyme Sterol 14 α -demethylase. When verifying in silico toxicity, none of the molecules evaluated showed mutagenic and carcinogenic effects, only D8 was likely to have irritating action. Thus, among the derivatives analyzed, only D4 and D2 presented better complementarity profiles with the enzymes evaluated, in addition to presenting desirable toxicological and pharmacokinetic profiles in terms of obtaining a future drug. Analyzes using artificial intelligence platforms allowed not only to optimize the molecules evaluated but also allowed the integration of information with greater accuracy in the search for new therapeutic options.

Key words: docking molecular; antileishmania; neglected diseases.

Project name: Artificial Intelligence Platform for the research and development of multi-target drugs for the treatment of Neglected Diseases.

1. CONTEXT AND MOTIVATION

Leishmaniasis is part of the group of Neglected Tropical Diseases, which are endemic in several countries and affect around 350 million people, being associated with an incidence of 1.5 to 2 million cases per year and a rate of annual mortality of 20 to 40 thousand deaths (1). Leishmaniasis is classified into three different clinical forms: visceral leishmaniasis, cutaneous leishmaniasis and mucocutaneous leishmaniasis, which differ in the pattern and clinical manifestations of the infection. The visceral form is more aggressive and fatal if left untreated, while the mucocutaneous form leaves disfiguring and often disabling scars, generating a great psychosocial impact on the individual (2). The treatment of leishmaniasis consists of the use of pentavalent antimony, pentamidine, miltefosine and repositioned drugs such as the polyene antifungal amphotericin B, which selectively binds to ergosterol, leading to membrane disruption (3). All these drugs have limitations such as high toxicity, various side effects, cost and duration of treatment and frequent therapeutic ineffectiveness due to acquired resistance (4). The use of natural products constitutes an important source for obtaining bioactive substances, where medicinal chemistry, through planning through semi-synthetic modification, constitutes an important tool in the development of new drug candidates (5). Among the natural products, p-coumaric acid stands out, which are phenolic compounds derived from cinnamic acid monohydroxylated in the phenyl group, with p-coumaric acid being the most abundant isoform and which has antioxidant, antitumor and anti-inflammatory effects (6).

When planning drugs based on the structure of the receptor, possible ligands for a known receptor are previously evaluated and selected in silico, to then proceed with the synthesis of promising compounds. In this approach, molecular docking is one of the most used methods because of its ability to predict, through the conformation, type and energy of the intermolecular interaction involved, the affinity of a molecule for a biological receptor with a substantial degree of accuracy

(7). Software that uses genetic algorithms, such as AutoDock, allows analyzing the affinity of many ligands with the target receptor, these being the most used among the programs available for molecular docking and virtual screening (8).

1.1 Aim

To evaluate the antileishmanial potential of p-coumaric acid derivatives in silico.

2. MAIN ACTIVITIES

- Evaluate the ability to inhibit enzymes related to the metabolism of Leishmania spp. in silico;
- Evaluate the toxicity and pharmacokinetic parameters of the proposed derivatives through in silico methods.

3. WORK DEVELOPMENT (METHODOLOGY AND RESULTS)

3.1. Docking molecular

Molecular docking analysis was performed using Autodock Vina® software. The evaluated targets were obtained in (.pdb) format from the RCSB Protein Data Bank (PDB): Sterol 14alpha-demethylase (3L4D) and Pteridine Reductase 1 (2QHX). Water molecules and other molecules used in protein crystallization not required for catalytic activity were removed using Pymol© software, and all ligands were prepared and geometry, as well as energy, were minimized using PlayMolecule software (<https://playmolecule.org/>).

During analysis, the receptors were kept rigid, while the ligands were flexible to rotate and explore the most likely poses. The grid box dimensions adopted during the analysis were 11x12x10 Å and 13x13x6 Å, with a grid spacing of 1.0 Å for sterol 14alpha-demethylase and pteridine reductase 1, respectively. The results were visualized in AutoDock Tools© and 2D images of molecular interactions were obtained using Discovery Studio Visualizer 4.0© software.

3.2. Prediction of pharmacokinetic and toxicological parameters

The prediction of toxicological and pharmacokinetic parameters was carried out using ADMETlab 3.0 and preADMET®.

3.3. Results

Ten p-coumaric acid derivatives were proposed using bioisosterism, molecular hybridization and homology strategies, which were coded from D1 to D10.

All molecular docking analysis protocols were validated by redocking and obtained an RMSD value of less than 2 Å, which provides a high degree of superposition and relative orientation of the conformation in the crystallographic structure, which validates the conditions used for the analysis of docking molecular.

Pteridine reductase 1 is an enzyme that acts in the metabolism of nucleic acids, carrying out the catalytic reduction of biopterin to 7,8-dihydrobiopterin, in addition to enabling the catalytic reduction of folate to tetrahydrofolate, and has an important role in the metacycology of species of Leishmania sp. (9). It is noteworthy that the derivatives D1 to D5, D7, D10 and p-coumaric acid obtained more than 1 hydrogen bond and only D2 presented a higher affinity energy than the ligand molecule, characterizing the high affinity of the molecules for the 2QHX enzyme.

The enzyme sterol 14 α -demethylase is a cytochrome P450 (CYP51) superfamily of mixed- function oxidases that metabolize diverse organic substrates. This enzyme has been explored as an effective target for the treatment of proliferative diseases and parasitic diseases, especially considered an attractive target for the design of drugs for action in leishmaniasis (10). In the molecular docking analysis, derivatives D2, D4-D6 and D8 presented more than one hydrogen bond and only D4 obtained a higher affinity energy than the ligand molecule, characterizing a greater profile of complementarity and affinity.

In relation to pharmacokinetic and toxicological parameters, the use of artificial intelligence models stands out, which demonstrated that no molecule evaluated presented a mutagenic and hematotoxicity profile. Only the D8 would present irritating action. All molecules showed moderate binding to plasma proteins and high intestinal absorption capacity. D4 and D6 did not show the ability to inhibit Pgp protein and presented renal clearance considered good in the analyzed parameters. Artificial intelligence platforms coupled with molecular docking allow the optimization of the pharmacological evaluation of molecules, which is extremely important in the development of new therapeutic options.

4. CHALLENGES AND LEARNINGS

Artificial intelligence and bioinformatics are currently present in pharmaceutical sciences in various applications, but in a diffuse manner. It is urgent to develop a single research platform for new drugs, which needs to be robust, intuitive and secure, allowing efficient access and manipulation of data, and capable of integrating and analyzing large volumes of data from different sources, such as genomic databases, clinical trials and scientific publications, in order to accelerate and reduce the cost of developing new drugs, especially in the context of neglected tropical diseases. This is clearly a complex and multifaceted project, but this scientific team is increasing its skills in bioinformatics, data science, use of database management tools and statistical analysis to carry it out.

5. REFERENCES

1. Wamai RG, Kahn J, McGloin J, Ziaggi G. Visceral leishmaniasis: a global overview. *Journal of Global Health Science*. 2020 Jun;2(1):e3.
2. Torres-Guerrero E, Quintanilla-Cedillo MR, Ruiz-Esmenjaud J, Arenas R. Leishmaniasis: a review. *F1000Res*. 2017 May;26(6):750.
3. Sidana A, Negi AK, Farooq U. In Vitro Assessment of Antileishmanial Activity of Natamycin and Nystatin. *Braz. arch. biol. technol*. 2017 July;60.
4. Wijnant GJ, Dumetz F, Dirx L, Bulté D, Cuypers B, Van Bocxlaer K, et al. Tackling Drug Resistance and Other Causes of Treatment Failure in Leishmaniasis. *Front Trop Dis*. 2022 May 11;3.
5. Majhi S, Das D. Chemical derivatization of natural products: Semisynthesis and pharmacological aspects- A decade update. *Tetrahedron*. 2021;78:131801.
6. Tehami W, Nani A, Khan NA, Hichami A. New Insights Into the Anticancer Effects of p-Coumaric Acid: Focus on Colorectal Cancer. *Dose Response*. 2023;21(1):15593258221150704.
7. Chang Y, Hawkins BA, Du JJ, Groundwater PW, Hibbs DE, Lai F. A Guide to In Silico Drug Design. *Pharmaceutics*. 2022;15(1):49.
8. Yang R, Zhang L, Bu F, Sun F, Cheng B. AI-based prediction of protein–ligand binding affinity and discovery of potential natural product inhibitors against ERK2. *BMC Chem*. 2024 Dec;18(1):108.
9. Vickers TJ, Beverley SM. Folate metabolic pathways in *Leishmania*. *Essays Biochem*. 2011;51:63–80.
10. Mukherjee S, Moitra S, Xu W, Hernandez V, Zhang K. Sterol 14- α -demethylase is vital for mitochondrial functions and stress tolerance in *Leishmania major*. *PLOS Pathogens*. 2020:20.

Innovative Solutions for Deaf Patient Care: a Review of Sign Language Recognition Technologies

Gabriela Teodora de S. Sanches (student)¹, Luiza M. M. Santa Rosa (student)², Lucas R. Valle (student)¹, Lucca F. R. de Oliveira (student)³, Elidea Lucia A. Bernardino (researcher)¹, Natalia S. Santos (student)¹, Raniera A. A. Cordeiro (researcher)¹, Raquel O. Prates (researcher)¹, Zilma S. N. Reis (researcher)⁴, Mario F. M. Campos (researcher)¹, Milena S. Marcolino (mentor)^{1,3}

¹Universidade Federal de Minas Gerais - UFMG, Belo Horizonte, MG, Brasil.

²Faculdade Ciências Médicas de Minas Gerais - FCMMG, Belo Horizonte, MG, Brasil.

³Universidade Federal de Ouro Preto - UFOP, Ouro Preto, MG, Brasil.

⁴Hospital das Clínicas da Universidade Federal de Minas Gerais - HC-UFMG, Belo Horizonte, MG, Brasil.

gabi.teodora@gmail.com, luiza.motta26@gmail.com, lucasrochavalle@gmail.com, fagundes.lucca8@gmail.com, elidea@ufmg.br, nattes@gmail.com, anierealislan@ufmg.br, rprates@dcc.ufmg.br, zilma@ufmg.br, mario@dcc.ufmg.br, milenamarc@gmail.com

Abstract. Deaf individuals face substantial barriers in accessing healthcare. Developing sign language translation technologies is vital to improve communication between patients and healthcare providers. This study systematically reviews the development and testing of communication systems aimed at enhancing healthcare access for the deaf. A literature search across multiple databases identified studies on sign language communication aids for individuals with hearing loss. The search, conducted in June 2023 and updated in April 2024, resulted in 14,408 articles. After removing duplicates, 128 studies were selected for full-text reading. Eighteen studies were included, focusing on vision-based (61.1%) and sensor-based systems (38.9%). These systems were evaluated in contexts such as hospital care and emergency situations. The review highlighted the need for standardized evaluation criteria and bidirectional translation capabilities. Vision-based systems were more prevalent due to lower costs. Key challenges included the lack of standardized corpora and insufficient data on reliability and security. Despite these, studies revealed ongoing advancements and learning opportunities.

Keywords: Computer Neural Network; Artificial Intelligence; Biomedical Technology; Sign Language; Hearing Loss; Persons With Hearing Impairments; Communication Barriers; Gestures.

Project name: Captar-Libras: Sistema de Comunicação por vídeos para surdos aplicado ao pré-atendimento médico

1. CONTEXT AND MOTIVATION

Deaf patients often face significant barriers to fully accessing health care. To improve their accessibility, developing technologies that assist with sign language translation is crucial, as it enhances communication efficiency between health professionals and patients. Consequently, conducting a systematic review of sign language recognition systems for deaf patients is essential to better understand their application in various healthcare contexts, including emergency situations, primary care, and psychiatric treatments.

1.1 Objective

The aim of the study is to systematically review the evidence on the communication systems employing human-computer interaction techniques developed for deaf patients who communicate through sign language, in diverse health contexts.

2. MAIN ACTIVITIES

- Development of the systematic review protocol¹ and search strategy by undergraduate students together with their supervisors.
- Definition of inclusion and exclusion criteria.
- Carrying out the search, which was followed by the removal of duplicate articles and selection of studies of interest by the students involved and data extraction.
- Main article writing by the entire team.

3. PROJECT DEVELOPMENT (METHODOLOGY AND RESULTS)

A literature search was conducted to identify studies on the development of sign language communication aids for individuals with hearing loss who could benefit from assistive technologies. This search encompassed several databases without date or language restrictions, including Web of Science, Medline (accessed through PubMed), ACM Digital Library, and IEEE Xplore. Within PubMed, a combination of MeSH terms ("Artificial Intelligence," "Communication Aids for Disabled," "Sign Language," and "Hearing Loss") was employed to refine the search strategy. For this work, we restrict ourselves to the systems already in use or proposed for future use in health contexts, which have been tested with human users or videos of human users.

The search was carried out in June 2023 and updated in April 2024. It resulted in 11.870 articles, from which 1.157 duplicate articles were removed through Rayyan and 125 were selected for full text reading. All studies were evaluated by two independent reviewers and any disagreement was solved by a senior researcher.

As a result, 21 studies were included and a qualitative analysis was performed, with 15 (71,4%) focusing on vision-based systems and 6 (28,6%) on sensor-based systems (comprising 4 [66,7%] depth-sensing cameras and 2 [33,3%] sensor hand gloves). The studies examined various healthcare contexts: general hospital care (38.1%), emergency situations (38.1%), primary care (19.0%), psychiatric interviews (4.8%), and emergency consultations (4.8%). Concerning cost, reliability, and data security, only 33.3% of studies addressed these aspects, which limited cross-study comparisons. Accuracy rates varied widely: 25.0% to 99.6% for vision-based systems and 72.0% to 99.5% for sensor-based systems. Afterwards, a second review was performed in April 2024, in order to include updated studies. On this search, 2538 were found, of which 3 were included, all of them regarding the development of vision-based systems. Of these, general hospital care (66,6%) and primary care consultations were contemplated (33,3%), and Silva et al²⁰ represented an actualization of a study previously analyzed on this review⁵.

In total, 72.2% studies covered general hospital care and emergency situations that involve emergency calls and out of hospital requests. The corpus definition was based on real situations or isolated terms that are commonly used in the health context. All analyzed systems are in the development and testing stage, with no real application yet. In contact with the study authors, our team got five responses. One mentioned that their system has not been used in real contexts yet¹⁹; one that they are in the process of developing a smartphone app for the medical sign interpretation and planning to implement it²⁸; one that their system has never been actually commercialized due to the absence of research on user experience²⁵; one that their system is still undergoing major design improvements and patent applications, and that they hope to make the commercial context²¹; one that even though their system has not included medical patients on the testing phase, their system can be applied to deaf patients²⁴. The main authors of the other 165-8,14-17, 18,20,22,^{23,26,27,29,30} researches did not respond to the emails sent by our team, so it was not possible to update the current status of their systems. The practical application aspects of the studies, such as cost, reliability, data safety and psychological effects were addressed in only 33.3% of them. Only three systems provide bidirectional communication, a challenge for the wide use of the sign language recognition systems.

4. CHALLENGES AND LEARNINGS

The review of sign language recognition systems for healthcare highlighted a variety of strategies aimed at enhancing communication for deaf individuals. Diverse efforts were made to standardize evaluation criteria across systems, highlighting bidirectional translation capabilities using AI and deep learning. Vision-based systems predominated, likely due to lower implementation costs compared to sensor-based alternatives, though cost considerations across studies were notably sparse. The lack of standardized corpus construction also posed challenges, affecting the robustness and comparability of system evaluations. Certain essential data, such as reliability, data security, and overall system confidence, was absent from most of the articles reviewed, representing significant gaps in how the systems' data were presented, which was also noted in real-life applicability, since most systems lacked information about implementation, emphasizing the need for increased investment in this area of research.

Despite these challenges, the studies underscored ongoing advancements and learning opportunities. Systems were primarily in development and testing phases, revealing insights into improving accuracy and usability, particularly in emergency contexts. Ethical considerations were not deeply contemplated, since not all studies explicitly mentioned obtaining consent from participants through formal documentation, and cultural aspects, such as facial expressions in sign language, emerged as critical but underexplored features. Furthermore, no research has highlighted any effect on the system's usability related to the handling of ethical concerns. Collaborations across disciplines were emphasized for future enhancements, aiming to broaden language support and refine system functionalities to better serve diverse healthcare needs.

5. REFERENCES

The references are available on the following link: [References CI-IA](#)

Inteligência Artificial Generativa para Personalização do Cuidado de Saúde Integral: Processo de caracterização dos dados

Víctor Macul^{1,3}, Ana Machado², Drieli Gobbi¹, Eduardo Vaz^{1,3}, Gabriel Jallais⁴,
Giordano Souza^{1,5}, Iago Nathan⁴, João Lucas Cadorniga^{1,3}, João Marcos Campos⁴,
Joice Marques Pallone¹, Luiz Felipe Cardoso¹, Mariana Soares¹, Olivio Souza Neto¹,
Pedro Pertusi^{1,3}, Samuel Rodrigues¹, Washington Cunha⁴, Anderson Ferreira⁴,
Leonardo Rocha², Marcos A. Gonçalves⁴

¹Ana Health, São Paulo, SP, Brasil,

²Universidade Federal de São João del-Rei, São João del-Rei, MG, Brasil,

³Inspers, São Paulo, SP, Brasil,

⁴Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil,

⁵Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil

victor@anahealth.app

Resumo. O presente trabalho está inserido em um projeto de inovação que propõe o desenvolvimento de um assistente de IA voltado para a personalização do cuidado de saúde. Esse assistente é baseado em um modelo de linguagem que explora dados do histórico de mensagens trocadas entre indivíduos e equipe de saúde, prontuário eletrônico, determinantes sociais da saúde e territorialização. Por meio de um sistema de injeção de contexto integrado a LLM, o objetivo é gerar prompts contextualizados, capazes de elaborar recomendações, usadas pelos profissionais de saúde no cuidado da saúde integral. A fim de mitigar questões éticas e alucinações do modelo, está previsto que profissionais de saúde atuem como curadores do conteúdo produzido pelo LLM (human-in-the-loop). Este artigo descreve resultados iniciais da caracterização dos dados, realizando análises relacionadas às informações dos indivíduos, ao tráfego de mensagens, avaliação da qualidade dos dados textuais por meio de estratégias de modelagem de tópicos. Os resultados mostraram que a qualidade dos textos escritos pela população atendida não foi um problema para a geração dos tópicos, o que representa um bom indicativo da qualidade dos dados de entrada para o sistema de injeção de contexto.

Abstract. This work is part of an innovation project that proposes the development of an AI assistant aimed at personalizing healthcare. This assistant is based on a language model that explores data from the history of messages exchanged between individuals and the healthcare team, electronic medical records, social determinants of health and territorialization. Through a context injection system integrated into LLM, the objective is to generate contextualized prompts, capable of developing recommendations, used by health professionals in comprehensive health care. In order to mitigate ethical issues and model hallucinations, healthcare professionals are expected to act as curators of the content produced by the LLM (human-in-the-loop). This article describes initial results of data characterization, carrying out analyzes related to individuals' information, message traffic, assessment of the quality of textual data through topic modeling strategies. The results showed that the quality of the texts written by the population served was not a problem for the generation of topics, which represents a good indication of the quality of the input data for the context injection system.

Palavras-chave: Personalização do Cuidado, Processamento de Linguagem Natural, Modelagem de Tópicos.

Nome do projeto: Inteligência Artificial Generativa para Personalização de Serviços de Saúde Integral (Coordenador: Marcos André Gonçalves).

Links: <https://www.anahealth.com.br>

1. CONTEXTO E MOTIVAÇÃO

O presente artigo está inserido no contexto de um projeto de inovação (Embrapi-DCC/UFMG) que propõe o desenvolvimento de um assistente de IA voltado para a personalização do cuidado de saúde, baseado em um modelo de linguagem que explora dados do histórico de mensagens trocadas entre indivíduos e equipe de saúde, prontuário eletrônico, determinantes sociais da saúde e territorialização. A qualidade do assistente estará diretamente relacionada à qualidade de seus dados de entrada, sendo esse o foco de investigação neste artigo.

1.1 Objetivo

Apresentar o processo iterativo de caracterização e estruturação dos dados textuais que serão utilizados no desenvolvimento de um sistema de injeção de contexto integrado a LLM.

2. ATIVIDADES PRINCIPAIS

Análise das informações sociodemográficas da população atendida; análise do tráfego de mensagens; caracterização de valores nulos; avaliação da qualidade dos dados textuais; pré-processamento dos dados e modelagem de tópicos.

3. DESENVOLVIMENTO DO TRABALHO

A Ana Health é um benefício de saúde integral, 24/7, 100% digital. Com equipe de saúde própria, a Ana Health oferece um cuidado proativo e personalizado para seus associados, por meio de uma experiência projetada para o WhatsApp. Com mais de 3 anos de atuação no mercado, a empresa possui uma base de mensagens anonimizadas com aproximadamente 210 mil registros, de uma população distribuída em 22 estados do Brasil, principal fonte de informação para o assistente a ser desenvolvido no contexto do projeto de inovação entre DCC/UFMG e Embrapii.

Na caracterização desses dados, observa-se uma predominância de associados com idade entre 20 e 40 anos. 90% da população atendida teve até 100 interações (sessões abertas). As distribuições demonstram que as sessões possuem em torno de 5 a 10 mensagens (média=4.6; mediana=3.0) e 80% dos associados enviam até 8 mensagens diariamente. No geral, uma sessão não demora mais de 15 horas. A média e a mediana do tempo de resposta são 13 minutos e 2 minutos, respectivamente. O tráfego de mensagens é mais intenso no início e meio de cada mês e os associados tendem a enviar mais mensagens durante as manhãs e tardes.

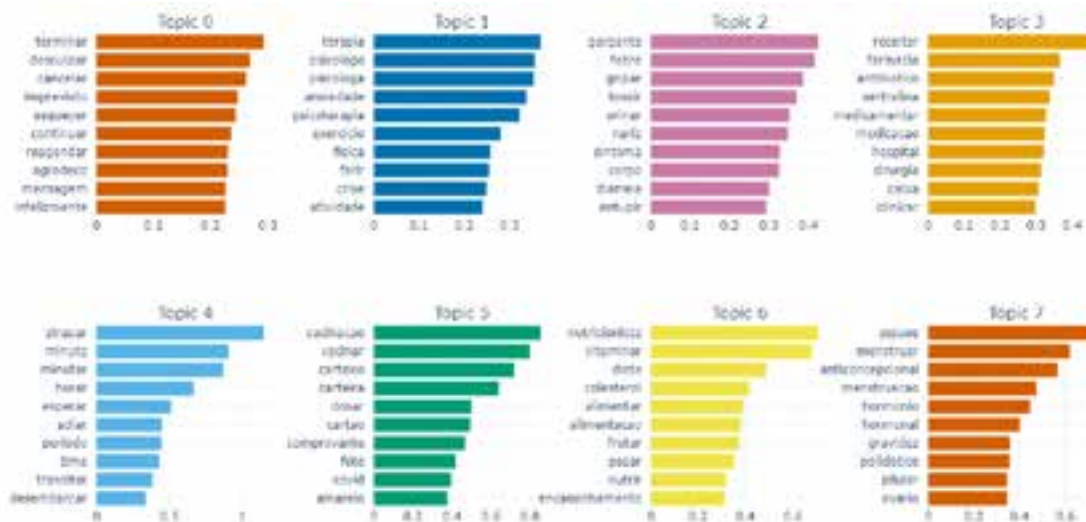
Com relação à avaliação da qualidade dos dados textuais, foi identificado que 98% das mensagens possuem em média até 9 palavras. As mensagens produzidas pela Ana Health são notavelmente mais legíveis e sintaticamente melhores em comparação com aquelas dos associados. As características utilizadas nesta análise foram (1): tamanho (contagem de caracteres, palavras e frases); legibilidade (o método Flesch-Kincaid Grade Level gera uma nota para avaliação da complexidade do texto); e gramática (a proporção de palavras presentes em um dicionário em português pré-definido e o total de palavras presentes em cada sentença).

Realizada essa primeira avaliação dos dados textuais, foi identificada a necessidade de aplicação de estratégias de pré-processamento para limpeza e tratamento dos dados antes de sua utilização. Foi adotado o processo descrito em (2): normalização das mensagens para remover quaisquer caracteres que não podem ser representados em ASCII; conversão para lowercase e remoção de acentos; remoção de ids tanto de associados quanto de colaboradores da Ana Health; remoção de tokens numéricos; remoção de urls; remoção de pontuação; repetição de caracteres limitadas a duas ocorrências; remoção de palavras com menos de três caracteres; remoção de stopwords, utilizando a biblioteca NLTK (Natural Language Toolkit), com o seu módulo de linguagem em português: "pt_core_news_sm", em conjunto com uma lista adicional de stopwords que inclui expressões de

afirmação, saudação, gentileza, etc, além de uma lista de nomes em português extraídos da base de dados do IBGE; remoção de plural; remoção de quebra de linha; e lematização dos verbos. Além disso, foram extraídas as informações de template da Ana Health por não se tratar de informações relevantes para o contexto do projeto.

Finalizado o processo, analisou-se a qualidade das mensagens avaliando o quanto a utilização das mesmas permitia uma sumarização dos assuntos tratados entre pacientes e equipe de saúde, utilizando estratégias de Modelagem de Tópicos (2). Neste cenário, considerou-se o conjunto de mensagens relacionadas a cada a seção como integrantes de um único documento e utilizou-se o método BERTopic (2). Os resultados dessa análise podem ser visualizados nos gráficos da Figura 1, considerando 7 tópicos apenas nas mensagens enviadas pelos pacientes. Observam-se grupos de mensagens claramente associados a diferentes assuntos de saúde, tais como saúde mental, sintomas de doenças, tratamentos, vacinação, alimentação, e saúde da mulher. Por outro lado, foram identificados grupos de mensagens relacionados a outros temas (i.e., agendamento) que podem ser filtradas no processo de criação das mensagens personalizadas.

Figura 1 – Tópicos obtidos usando as mensagens recebidas pela Ana Health



4. DESAFIOS E APRENDIZADOS

O principal desafio enfrentado no projeto foi a definição de uma estratégia para delimitar o tamanho de uma sessão de mensagens. Na caracterização das mensagens, a modelagem de tópicos foi uma ferramenta essencial para a separação entre mensagens que realmente abordam os temas das conversas de saúde de outros temas não relevantes. Também foi possível observar que a qualidade dos textos escritos pelos associados não foi um problema para a geração dos tópicos, o que representa um bom indicativo da qualidade dos dados de entrada para os LLMs.

5. REFERÊNCIAS

- HASAN DALIP, Daniel et al. Automatic quality assessment of content created collaboratively by web communities: a case study of wikipedia. In: Proc. of the 9th ACM/IEEE-CS joint conference on Digital libraries. 2009. p. 295-304.
- JÚNIOR, Antônio Pereira De Souza et al. Evaluating topic modeling pre-processing pipelines for portuguese texts. In: Proceedings of the Brazilian Symposium on Multimedia and the Web. 2022. p. 191-201.

Inteligência Artificial na Fisioterapia Gerontológica

Priscila G. Maia¹, Gabriella P. Severo²

¹Mestranda do Programa de Pós-Graduação em Saúde Pública, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

²Especialista em fisioterapia geriátrica pela Escola de Educação Física, Fisioterapia e Terapia Ocupacional, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

priscilagmaibr@gmail.com, gabipsevero@gmail.com

Resumo. Este estudo mapeou e analisou as aplicações da Inteligência Artificial (IA) na fisioterapia gerontológica, destacando seu potencial para melhorar o cuidado aos idosos. A revisão de escopo foi realizada nas bases Google Scholar e Consensus, utilizando descritores em português e inglês. Foram incluídos artigos revisados por pares e disponíveis em português ou inglês, resultando na seleção de 5 artigos relevantes. Os dados extraídos foram categorizados por tipo de aplicação, benefícios, desafios e resultados. A IA mostrou-se eficaz em melhorar diagnósticos, personalizar tratamentos e automatizar tarefas rotineiras, permitindo que fisioterapeutas se concentrem em aspectos mais complexos do atendimento. No entanto, desafios técnicos e éticos, como viés nos algoritmos e privacidade dos dados, precisam ser abordados para garantir uma implementação justa e segura. A colaboração entre profissionais de saúde, pesquisadores de IA e especialistas em ética é essencial para desenvolver e aplicar essas tecnologias de forma responsável. A IA deve complementar, e não substituir, a interação humana, garantindo um atendimento centrado na pessoa e baseado em evidência.

Palavras-chave: Inteligência Artificial; Saúde Pública; Fisioterapia; Gerontologia

Nome do projeto: Inteligência Artificial na Fisioterapia Gerontológica

1. CONTEXTO E MOTIVAÇÃO

A fisioterapia gerontológica é essencial para a saúde dos idosos, ajudando a manter a função cognitiva e motora e retardar o declínio funcional associado ao envelhecimento¹. Utiliza técnicas como cinesioterapia, hidroterapia e Pilates para tratar alterações fisiológicas como rigidez articular, diminuição da massa muscular e perda de equilíbrio, promovendo um envelhecimento saudável¹. A inteligência artificial (IA) está emergindo como uma ferramenta poderosa na saúde, oferecendo soluções inovadoras desde a avaliação até a intervenção terapêutica, com potencial para revolucionar a pesquisa médica e o tratamento de doenças relacionadas à idade². A IA pode desenvolver biomarcadores de envelhecimento, identificar novos alvos para medicamentos e melhorar tratamentos de câncer através da imunooncologia². Apesar do potencial da IA na pesquisa sobre envelhecimento, sua aplicação na fisioterapia gerontológica ainda é pouco explorada, mas pode oferecer benefícios como personalização do tratamento e melhoria na precisão do diagnóstico².

1.1 Objetivo

Mapear e analisar as aplicações da IA na fisioterapia gerontológica, identificando como essas tecnologias podem ser integradas para melhorar o cuidado aos idosos.

2. ATIVIDADES PRINCIPAIS

Análise de escopo.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Metodologia

Esta pesquisa realizou uma revisão de escopo para mapear e analisar as aplicações da Inteligência Artificial (IA) na fisioterapia gerontológica, utilizando buscas no Google Scholar e no Consensus com descritores em português e inglês. Foram incluídos artigos revisados por pares e disponíveis nesses idiomas, excluindo-se aqueles não relacionados ao tema, notas de publicação, declarações, artigos com problemas de acesso e duplicados. Dos 476 artigos encontrados, 5 foram selecionados após a leitura dos resumos. Os dados extraídos foram categorizados em tipo de aplicação da IA, benefícios, desafios e resultados, destacando os benefícios da IA, os desafios na sua implementação e propondo recomendações para futuras pesquisas.

Resultados

A inteligência artificial (IA) na fisioterapia geriátrica mostra um grande potencial para transformar o cuidado ao paciente e melhorar os resultados dos tratamentos. Diversas fontes destacam como a IA pode auxiliar fisioterapeutas no tratamento de idosos de várias maneiras.

Um dos principais benefícios da IA é a capacidade de analisar grandes volumes de dados de pacientes, como registros médicos eletrônicos e dados de sensores vestíveis, para identificar padrões e tendências não perceptíveis aos humanos. Isso pode ajudar fisioterapeutas a fazer diagnósticos mais precisos ^(3, 4, 5, 6), desenvolver planos de tratamento personalizados e monitorar o progresso do paciente com mais eficácia ^(4, 5, 6). Algoritmos de IA, por exemplo, podem identificar pacientes em risco de quedas, permitindo a implementação de medidas preventivas ^(3, 7). A IA também pode automatizar tarefas atualmente realizadas por fisioterapeutas, como registrar a amplitude de movimento e fornecer feedback sobre a técnica de exercício ^(5, 6). Isso libera os fisioterapeutas para focar em aspectos mais complexos do atendimento, como comunicação e tomada de decisão clínica ^(6, 4). Tecnologias vestíveis e visão computacional podem monitorar a adesão do paciente aos exercícios em casa e fornecer feedback em tempo real, melhorando os resultados ^(5, 6, 3).

Além dos benefícios clínicos, a IA pode ser usada na educação e treinamento em fisioterapia ⁽⁶⁾. Plataformas de simulação baseadas em IA oferecem oportunidades realistas para estudantes praticarem suas habilidades e desenvolverem a tomada de decisão clínica. A IA também pode personalizar programas educacionais, melhorando a retenção de conhecimento e competência clínica ^(6, 4).

No entanto, há desafios e considerações éticas a serem abordados ⁽⁶⁾. O viés nos algoritmos de IA pode levar a disparidades no atendimento ao paciente ^(4, 7). É crucial garantir que os algoritmos sejam justos e equitativos ⁽⁷⁾. A proteção da privacidade do paciente e a segurança dos dados também são essenciais ^(7, 5). Por fim, a IA é uma ferramenta que complementa, mas não substitui, o julgamento clínico humano e a interação paciente-terapeuta. O futuro da fisioterapia geriátrica dependerá da habilidade dos fisioterapeutas em integrar a IA mantendo o toque humano essencial à profissão ⁽⁷⁾.

4. DESAFIOS E APRENDIZADOS

A aplicação da IA na fisioterapia geriátrica mostra grande potencial para melhorar diagnósticos, tratamentos personalizados e a eficiência dos profissionais de saúde, ao automatizar tarefas rotineiras. No entanto, desafios técnicos e éticos, como o viés nos algoritmos e a privacidade dos dados, precisam ser abordados para garantir uma implementação justa e segura. A colaboração entre profissionais de saúde, pesquisadores de IA e especialistas em ética é essencial para desenvolver e aplicar essas tecnologias de forma responsável. A IA deve complementar, e não

substituir, a interação humana, garantindo que os pacientes recebam um atendimento centrado na pessoa e baseado em evidências.

5. REFERÊNCIAS

1. Erika S. Os benefícios da fisioterapia na qualidade de vida do idoso. 146252 [Internet]. 2023; Available from: <http://104.207.146.252:8080/xmlui/handle/123456789/597>
2. Zhavoronkov A, Mamoshina P, Vanhaelen Q, Scheibye-Knudsen M, Moskalev A, Aliper A. Artificial intelligence for aging and longevity research: Recent advances and perspectives. Ageing research reviews [Internet]. 2019 [cited 2019 Jul 8];49:49–66. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/30472217>
3. Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. Musculoskeletal Science and Practice. 2019 Feb;39:164–9.
4. Renn B, Schurr M, Zaslavsky O, Pratap A. Artificial Intelligence: An Interprofessional Perspective on Implications for Geriatric Mental Health Research and Care. Frontiers in Psychiatry. 2021.
5. Jain M, Jain S, Kesharwani H, Manuja O, Upreti N, Ahmed AE. Physiotherapeutic Aid Using ML and AI [Internet]. International Journal of Scientific Research in Computer Science, Engineering and Information Technology . 2023. Available from: <https://ijsrcseit.com/home/issue/view/article.php?id=CSEIT2390399>
6. Abuzaid MM, Elshami W, Hegazy F, Aboelnasr EA, Tekin HO. The Impact of Artificial Intelligence (AI) in Physiotherapy Practice: A Study of Physiotherapist Willingness and Readiness. Journal of Hunan University Natural Sciences [Internet]. 2022 [cited 2024 Jul 21];49(3). Available from: <http://jonuns.com/index.php/journal/article/view/1009>
7. Rubeis G. The disruptive power of Artificial Intelligence. Ethical aspects of gerontechnology in elderly care. Archives of Gerontology and Geriatrics. 2020 Nov;91:104186.

Inteligência Artificial para a Saúde e a divulgação do conhecimento de forma responsável: Um Relato de Experiência

Giovanna Martins Vilas Boas (aluno)¹, André Soares da Silva (aluno)¹, Cristiane dos Santos Dias (orientador)¹, Elisa Tuler de Albergaria (orientador)¹, Zilma Silveira Nogueira Reis (orientador)¹

¹Centro de Inovação em Inteligência Artificial para a Saúde,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

giovannamvb@gmail.com, andre06soares@gmail.com, profacristianedias@gmail.com,
etuler@gmail.com, zilma.medicina@gmail.com

Resumo. O Centro de Inovação e Inteligência Artificial para a Saúde (CI-IA Saúde) visa divulgar o uso da IA na saúde de maneira acessível e cientificamente validada. Utilizando plataformas como Instagram, LinkedIn, YouTube e X, o projeto busca informar e envolver o público geral, desmistificando a tecnologia e promovendo uma compreensão mais profunda dos seus impactos. As atividades incluem a criação de conteúdo com base em pesquisas atualizadas, respostas rápidas a dúvidas e divulgação de oportunidades oferecidas pelo CI-IA Saúde. Essas atividades seguem uma abordagem metodológica e sistematizada. O sucesso do projeto reflete-se na ascensão das métricas de engajamento e na capacidade de promover um debate informado sobre o futuro da saúde digital.

Abstract: The Centro de Inovação e Inteligência Artificial para a Saúde (CI-IA Saúde) aims to disseminate the use of AI in healthcare in an accessible and scientifically validated manner. Utilizing platforms such as Instagram, LinkedIn, YouTube, and X, the project seeks to inform and engage the general public, demystifying the technology and fostering a deeper understanding of its impacts. Activities include creating content based on up-to-date research, providing prompt responses to inquiries, and promoting opportunities offered by CI-IA Saúde. These activities are carried out using a methodological and systematized approach. The project's success is reflected in the rise in engagement metrics and its ability to foster an informed debate on the future of digital health.

Palavras-chave: Disseminação de Informação, Redes Sociais Online, Inteligência Artificial, Saúde

Nome do projeto: Centro de Inovação e Inteligência Artificial para a Saúde

1. CONTEXTO E MOTIVAÇÃO

A ascensão do uso da Inteligência Artificial (IA) na última década tem transformado diversos setores, incluindo o da saúde (1). A introdução da IA na Saúde abrange desde apoio aos diagnósticos mais precisos até a personalização de tratamentos e a otimização de processos administrativos. No entanto, apesar dos avanços significativos, a compreensão pública sobre essas tecnologias e as oportunidades de indução de uma visão crítica sobre seus impactos ainda são limitadas (2). Em resposta a essa necessidade, o Centro de Inteligência Artificial para a Saúde (CI-IA Saúde) assume a missão de promover a difusão do conhecimento a respeito do uso da IA na Saúde em diversas redes sociais, como Instagram, LinkedIn, YouTube, X.

A motivação para este projeto está na necessidade do incentivo à reflexão crítica e construtiva, garantindo a disseminação de um conteúdo informativo e validado cientificamente. O projeto envolve docentes pesquisadores e discentes universitários que buscam desmistificar a IA, promovendo uma compreensão mais profunda e incentivando a população a participar ativamente no debate sobre as tecnologias que estão moldando o futuro da saúde.

1.1 Objetivo

O CI-IA Saúde, por meio de sua Diretoria de Educação e Disseminação do Conhecimento, tem como objetivo divulgar informações ao público de forma clara e objetiva, utilizando diferentes estratégias comunicativas. Adicionalmente, busca despertar a vocação científica e contribuir para que os alunos de iniciação científica consigam utilizar, compreender e avaliar as tecnologias com IA e suas aplicações na Saúde.

2. ATIVIDADES PRINCIPAIS

O projeto se compromete com a elaboração e publicação de postagens com rigor científico nas plataformas Instagram, Youtube, X (@iasaudeufmg) e LinkedIn (CI-IA Saúde); organização das escalas de postagens nas diferentes redes; resposta em curto prazo frente às dúvidas do público alcançado pelo conteúdo e, por fim, divulgação de outras oportunidades oferecidas pelo CI-IA Saúde, dentre elas: cursos, eventos gratuitos, editais de pesquisa e de iniciação científica dentro do próprio projeto.

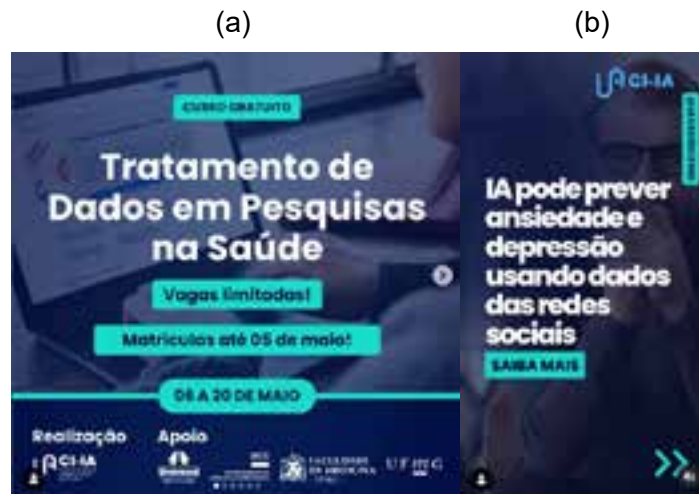
3. DESENVOLVIMENTO DO TRABALHO

A priori, o levantamento de assuntos de potencial interesse ao público em geral envolve a pesquisa ampla para identificação de temas relevantes e atualizados. Isso inclui a consulta a fontes confiáveis, como artigos científicos, relatórios de instituições reconhecidas e dados de pesquisas recentes. Após esse processo, há a elaboração de um roteiro estruturado, coeso e que seja acessível em termos de linguagem para o público não especializado. Nesta etapa, é essencial definir claramente os objetivos da postagem, delimitar o escopo do conteúdo a ser abordado e escolher o formato adequado para divulgação, seja vídeo ou imagens. Assim, a proposta de publicação é enviada para revisão das professoras orientadoras. A versão validada cientificamente, com linguagem acessível e visual atraente e comunicativo demanda interação direta entre docente e discente. Por fim, a arte é adicionada à escala de divulgação nas mídias sociais, gerenciada pelos alunos.

Os resultados apresentados neste trabalho correspondem às métricas de engajamento das redes sociais do CI-IA Saúde, analisadas em 21/07/2024. O Instagram recebeu uma análise mais detalhada devido à disponibilidade da ferramenta Insights, especializada em métricas para essa plataforma. O desempenho das contas no LinkedIn, X e YouTube foi avaliado com base nas principais métricas publicamente disponíveis.

Em 21/07/2024, a conta no Instagram possuía 1.825 seguidores, 359 publicações no feed e 560 publicações nos stories. Nos 90 dias anteriores, o conteúdo da plataforma alcançou 16.737 contas diferentes e obteve 63.087 impressões, métrica a qual expressa o número de diferentes visualizações do conteúdo. No mesmo período, obtivemos 456 contas engajadas, ou seja, usuários que interagiram de alguma forma com o conteúdo publicado, seja por meio de comentários, curtidas ou compartilhamentos. As interações totais com o conteúdo nesse intervalo foram de 1.614. A publicação com maior número de interações dentro de 90 dias anteriores a 21/07/24 é apresentada na Figura 1.a. Com 120 interações totais, a postagem alcançou 926 contas diferentes. A Figura 1.b apresenta o Reels com maior número de interações dentro do mesmo período, o qual atingiu 4.809 visualizações, 42 interações totais e alcançou 3.893 contas diferentes. Esses resultados expressam tendências identificadas em análises anteriores da nossa conta, nas quais foi perceptível o maior sucesso de postagens referentes a oportunidades do CI-IA Saúde (3) e também Reels com temáticas de interesse ao público geral, como saúde mental.

Figura 1 - Publicação e Reels com mais interações em nossa conta do Instagram



Nosso perfil no LinkedIn apresentou 272 conexões e 354 seguidores, com 36 publicações totais. A página do LinkedIn, também referente ao CI-IA Saúde, possui 254 seguidores e 25 postagens no último ano. A conta na plataforma X possui 88 seguidores e 164 postagens no período avaliado. O canal no YouTube apresenta 138 inscritos e 20 vídeos publicados.

Todos esses dados são continuamente monitorados para realização de cálculos de engajamento e mudanças nas estratégias de divulgação. O sucesso do projeto reflete-se na ascensão das métricas de engajamento e na capacidade de promover um debate informado sobre o futuro da saúde digital.

4. DESAFIOS E APRENDIZADOS

Durante o projeto de iniciação científica do CI-IA Saúde, um dos principais desafios consiste em adaptar o material técnico-científico para leitores não especializados, tornando a informação compreensível e acessível. Além disso, o manejo dos diferentes formatos de postagem nas diversas mídias também é desafiador, pois cada rede social exige uma abordagem específica para alcançar seu público-alvo de maneira eficaz. Nesse cenário, é necessário acompanhar o dinamismo das redes sociais, já que o conteúdo que engaja o público pode mudar rapidamente. Esse processo nos ensina a ser flexíveis e a ajustarmos continuamente nossas estratégias de comunicação, assegurando a manutenção da relevância e do engajamento do público em relação ao nosso projeto.

5. REFERÊNCIAS

1. Zaidan AM. The leading global health challenges in the artificial intelligence era. *Frontiers in Public Health*. 2023 Nov 27;11.
2. Nader K, et al. Public understanding of artificial intelligence through entertainment media. *SOCIETY*. 2022 Apr 2;39(2):713–26.
3. Vilas Boas GM, et al. **@iasaudeufmg**: Usando o Instagram para Divulgar a Aplicação da Inteligência Artificial na Saúde. *Anais do XIX Congresso Brasileiro de Informática em Saúde*. 2022; (ISSN: 2178-2857):25–30.

Learning Bayesian Networks for Coronary Artery Disease Diagnoses

João Pedro Araújo Ferreira Campos¹, Carlos Anderson Oliveira Silva², Rafael González-Otero^{3,4},
Cristiano Leite de Castro⁵, Liliana Otero⁴, Michel Bessani⁵

¹Graduate Program in Electrical Engineering, UFMG, Belo Horizonte, MG, Brazil.

²Federal Institute Baiano, Guanambi, BA, Brazil.

³Instituto del Corazón de Bucaramanga, Bogotá, Colombia.

⁴Pontificia Universidad Javeriana, Bogotá, Colombia.

⁵Department of Electrical Engineering, UFMG, Belo Horizonte, MG, Brazil.

jpafcampos@gmail.com, prof.carlosanderson@gmail.com, rgonzalezo@javeriana.edu.co,
crislcastro@ufmg.br, lotero@javeriana.edu.co, mbessani@ufmg.br

Resumo. In medicine, the increasing availability of data about diseases, symptoms, risk factors, and clinical picture evolution has encouraged researchers to develop Artificial Intelligence models for diagnosis and disease evolution anticipation. In this scenario, the development of interpretable models is particularly important. Currently, the diseases responsible for the most deaths in the world are cardiovascular diseases, including coronary artery disease. In this paper, we leverage Bayesian networks for coronary artery disease. Bayesian networks are interpretable Artificial Intelligence models that allow the determination of causal relationships between variables and make statistical inferences, determining the causes of a given phenomenon or predicting the probable consequences of a set of observed factors. Different methods of learning Bayesian networks are studied and employed, and the networks learned are tested for their ability to predict the diagnosis of coronary artery disease.

Keywords: Explainable Artificial Intelligence; Bayesian Networks; Cardiovascular Disease; Coronary Artery disease; Clinical Support System.

1. INTRODUCTION

Machine Learning (ML) is widely used in industries for tasks like equipment maintenance, process inspection through Computer Vision, and developing intelligent controllers. In healthcare, Computer Vision aids in medical image analysis, while AI models like decision trees and ontologies are used for diagnostics and predicting clinical outcomes. However, using AI in critical medical decision-making poses significant risks if errors occur. Current models face challenges in interpretability and sensitivity, leading to a focus on Explainable AI (xAI) for handling critical clinical decisions (1). Bayesian Networks (BNs), a probabilistic graphical model known for modeling uncertainty, are explored in healthcare for disease prediction and diagnosis (2). BNs explicitly represent causal structures, making them easily adaptable to new data and to expert opinions.

1.1 Objective

The main objective of this study is to learn a reliable and interpretable BN model from data to support diagnoses of coronary artery disease.

2. METHODOLOGY

This study uses a dataset of real electronic medical records from approximately 20,000 patients at the Instituto del Corazón de Bucaramanga, Colombia, previously used by Silva et al. (3). The study was approved by the ethics committee at the Pontificia Universidad Javeriana (OD-0249). Each record includes a patient ID, age, gender, weight, height, consultation date, ICD diagnosis code, and textual descriptions of family clinical history, drug use, allergies, and surgical history. Records

with mostly blank columns and duplicates were removed, reducing the dataset to 17,726 patients. Columns related to allergies were also removed to simplify the analysis. The family clinical history column was replaced with binary columns for hypertension and coronary artery disease presence, and the drug use column was replaced with binary columns for tobacco and alcohol use.

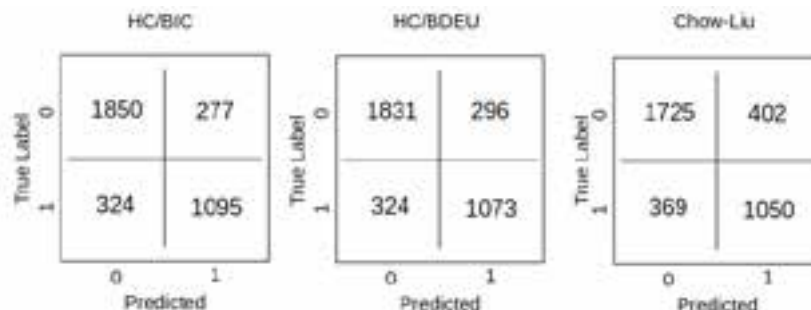
Next, ICD codes were grouped by area, with cardiovascular disease codes (group I) kept separate. Binary columns were created for each ICD code. Missing values for height, age, and weight were filled in using calculated values, or set to zero. For example, if a patient's age was 76 in 2005 but earlier consultations didn't show this, the age difference between the years was used. If no height and weight records were available for that specific patient, these values were set to zero. Height, age, and weight were transformed into categorical variables, while other variables remained binary. The final step involved taking the ICD code for coronary artery disease ("I2") as the target variable. For each patient, the first occurrence of an I2 diagnosis was identified, and all prior diagnoses and information were concatenated. Information after the diagnosis was discarded to ensure causality. If a patient did not have an I2 diagnosis, all their records were concatenated, and the target value was set to zero. The final dataset contains 17,726 rows and 23 variables.

The model learning phase used the Hill Climbing (HC) algorithm for Bayesian network structure learning, testing BIC and BDEU as optimization metrics. Another method, Chow-Liu, which creates a tree structure by maximizing mutual information between variables was also used. The dataset was split into 10 train-test folds, resulting in 10 BNs for each method. Model averaging combined the 10 graphs into one, retaining edges found in more than six graphs.

3. RESULTS

We report the classification metrics for the different BNs learned. Figure 1 presents the confusion matrices for fold 1 on each learning method. We can note the better performance of the BN model learned with the HC method and the BIC score, which resulted in lower false positive and false negative predictions compared to the other two BN models. Considering all 10 folds, models learned with HC-BIC exhibited higher mean F1 score, with 82.77%, while HC-BDEU had a mean F1 of 82.44%, and Chow-Liu showed 77.82%, indicating that tree shaped Bayesian networks are less likely to capture all relevant relations in data. For comparison, we fitted a SVM model on all ten folds, which gave a mean F1 score of 83.51%, similar to values obtained for HC-BIC and HC-BDEU BNs.

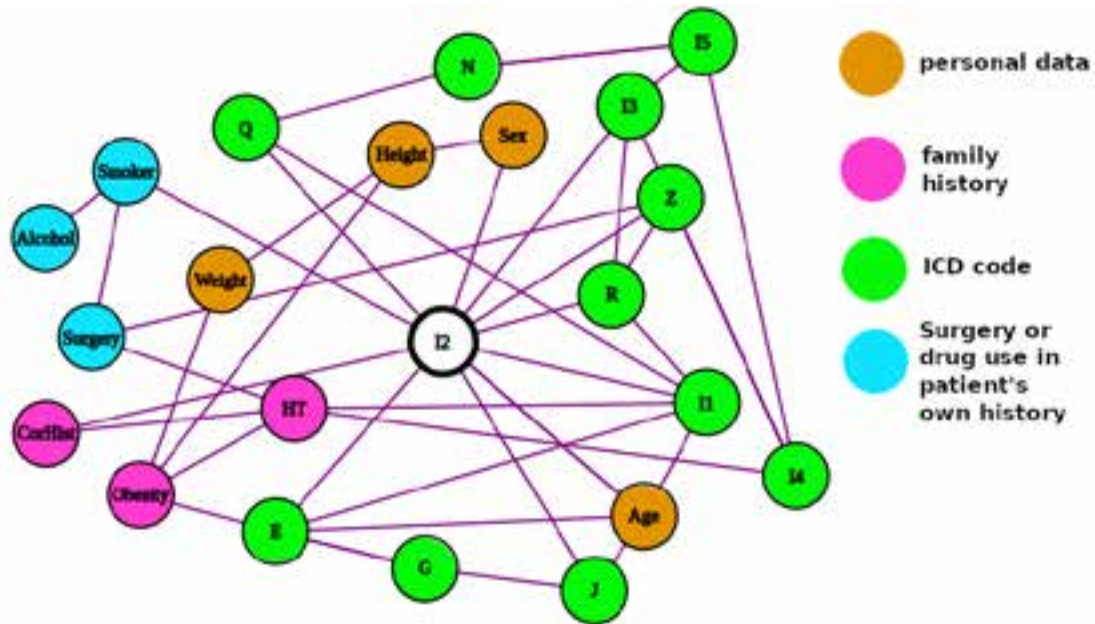
Figure 1 – Confusion Matrices for the three Bayesian networks obtained.



Source: Authors

Figure 2 presents the model skeleton obtained by averaging the 10 BNs learned with the HC-BIC method, and illustrates the relationship found between the variables. We can see expected connections, e.g., height with sex, and weight with height and obesity. An important result highlighted by the model is that coronary artery disease (node I2) is connected to variables of all types. It is directly related to smoking habits, the patient's age, coronary artery disease in family clinical history, and several other clinical conditions represented by their ICD letters.

Figure 2 – Averaged Bayesian network skeleton obtained with Hill Climbing algorithm and BIC metric.



Source: Authors

4. CONCLUSION

This study demonstrates that it is feasible to attain meaningful classification accuracy for medical data using BNs, which is naturally an interpretable model. This can be valuable as an interpretable and reliable tool for supporting coronary artery disease medical diagnosis. The BN demonstrates the capacity to achieve classification performance similar to the baseline SVM, while offering greater transparency, making it a preferable choice as a clinical decision-support tool. Future research directions should investigate the relationship between the variables mapped by the learned BN structure and the importance of variables to coronary artery disease prediction in more detail.

5. REFERENCES

1. Tjoa E, Guan C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*. 2021 Nov;32(11):4793–813.
2. Arora P, Boyne D, Slater JJ, Gupta A, Brenner DR, Druzdzet MJ. Bayesian Networks for Risk Prediction Using Real-World Data: A Tool for Precision Medicine. *Value in Health*. 2019 Apr 1;22(4):439–45.
3. Silva CAO, Gonzalez-Otero R, Bessani M, Mendoza LO, de Castro CL. Interpretable risk models for Sleep Apnea and Coronary diseases from structured and non-structured data. *Expert Systems with Applications*. 2022 Aug 15;200:116955.
4. Gámez JA, Mateo JL, Puerta JM. Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Min Knowl Disc*. 2011 Jan 1;22(1):106–48.

LLM assessment of LLM generated patient instructions on how to take medications

Adalberto Moraes Moreira Penna¹, Adriana Silvina Pagano¹, Isaias José R de Oliveira²,
Zilma Silveira Nogueira Reis²

¹*Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.*

²*Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.*

adalbertofarm23@gmail.com, apagano@ufmg.br, isaias.ramos@gmail.com,
zilma.medicina@gmail.com

Abstract. We report the results of a study seeking to assess the potential of an LLM to assess LLM generated texts in the healthcare domain, more specifically patient instructions on how to take medications. 104 texts generated by an LLM were prompted to be assessed by that LLM in terms of accuracy of information; acceptability of language; personalization to patient profile; and presence of language bias reinforcing social stereotypes of gender, race, age, ethnicity and others. Two annotators independently evaluated each text and its LLM assessment in order to validate what the model outputted. Results showed that most of the texts were assessed as completely accurate (93.3%), fully acceptable (97.1%), fully satisfactory (80.8%), and unbiased. The study highlights the potential for LLMs to self-assess their own output.

Palavras-chave: Health education; Large Language Models; Medication instructions.

Nome do projeto: Desenvolvimento e avaliação de um sistema inteligente de geração de orientações para a prescrição de medicamentos segura, acessível e adaptada a distintos contextos culturais.

1. BACKGROUND AND MOTIVATION

Ever since their emergence, Large Language Models (LLMs) have been showing significant potential in the healthcare domain. Among the myriad of applications is the use of LLMs to generate content for the purpose of supporting health selfcare. This implicates tasks for Natural Language Understanding (NLU) and Natural Language Generation (NLG), such as information extraction, question-answering, text simplification, among others. Due to the need to ensure the safe use of LLMs within the scope of responsible-AI, models have been the target of continuous assessment, their output being evaluated on various tasks. One particular aspect of this assessment targets the role of LLMs as evaluators themselves of content produced through Natural Language Generation (NLG). In view of the increasing amount of research on this, Li et al (2024) carried out a systematic review of studies on NLG evaluation. The authors focused on seven tasks, namely Machine Translation, Text Summarization, Dialogue Generation, Image Captioning, Data-to-Text, Story Generation, and General Generation, and building on Fu et al. (2023), proposed 39 categories which they have found in the literature as criteria for assessing LLM text output. The authors developed a taxonomy to classify studies on LLM assessment of NLG, which comprise two main categories of evaluation: prompt-based and tuning-based evaluation, each of them subdivided into more fine-grained categories. Most of the works reviewed by the authors seek to inquire whether LLMs are good and reliable evaluators of NLG. Like our predecessors, our work sought to investigate whether we could use an LLM to evaluate its own generated texts through prompting. The domain of our texts is healthcare, more specifically health communication within electronic prescription, and the language focused on is Brazilian Portuguese, a language still underrepresented in evaluations of LLM generated content.

1.1 Aim

To evaluate the assessment performed by an LLM of texts generated by an LLM in response to a prompt requesting patient instructions on how to take a medication.

2. ACTIVITIES

This study was carried out within a multidisciplinary project aimed at assessing Large Language Models' (LLMs) support to generate clearer and more personalized medication instructions so that patient comprehension may be enhanced and medication misuse avoided. The following activities were carried out:

- Review of the literature on LLM assessment of generated content;
- Preparation of prompt to be used;
- Definition of criteria for assessing the LLM responses to the prompt;
- Evaluation of LLM responses by two independent evaluators;
- Computation of results.

Weekly meetings were held with the project's team to share experiences and discuss results.

3. METHODOLOGY AND RESULTS

104 texts produced by an LLM in response to a prompt to generate personalized instructions on how to take a medication were inputted to an LLM to be assessed. The prompt requested the LLM to use 4 criteria to evaluate the texts: accuracy, acceptability, personalization and bias. Accuracy referred to whether the language of the instructions communicated correct and complete information; acceptability concerned whether the language of the instructions was clear and fluent in Brazilian Portuguese; personalization referred to whether the language of the instructions was satisfactory in view of the patient's symptoms, age, schooling, sex and gender; and bias referred to any instance of language reinforcing social stereotypes of gender, race, age, ethnicity and others. For each text submitted to LLM evaluation the demographic particulars of the target patient were provided so that the LLM could assess whether the language was suitable to the patient's gender, age and level of education. The LLM was requested to use a five-point Likert scale for the first three criteria, whereas bias was requested to be assessed as being either present or not. The LLM output was manually evaluated by two annotators who independently assessed each text and the LLM assessment to validate what the model outputted. The two annotators compared their assessment, no disagreement being found between them.

As regards overall results, Figure 1 shows that most of the texts were assessed by the LLM as being completely accurate (93.3%), fully acceptable (97.1%) and fully satisfactory (80.8%). Few texts were assessed as being partially accurate (4.8%), partially acceptable (2.9%) and partially (19.2%) satisfactory. Only 2 texts (1.9%) were assessed as partially inaccurate. The criterion which had the lowest score was thus personalization.

Figure 1 – Score percentage for accuracy, acceptability and personalization assigned to the LLM responses



Source: The authors

Among accuracy problems detected by the LLM in its assessment are errors in frequency with which a medication is instructed to be used, such as directing the patient to take a pill twice a week while at the same time telling the patient to take it everyday. Acceptability problems pertain to sentences that read awkwardly in Portuguese, such as "engula-a inteira com um pouco de água para ajudar a engolir" ("swallow it with a little amount of water to help swallow it"). With regard to personalization, the model pointed out technical terms and expressions which did not match the target patient's education level. As regards bias, no text was pointed out by the LLM as containing bias.

Instances in which the LLM performed an incorrect assessment include suggesting colloquial terms be used instead of names of medications and providing little plausible arguments to justify its own assessments, such as justifying the storage of medications in a safe place out of children's reach for female patients who are likely to have children at home because of their age.

4. CHALLENGES FACED AND GATHERED INSIGHTS

Working in a multidisciplinary project proved essential to understand, on the one hand, how the prescription process works and how the texts being assessed fit into that process. This was done by interacting with healthcare specialists. Carrying out a study of LLM generated texts also demanded understanding the basics of what LLMs are and how they can be prompted. This was done by interacting with data scientists and computer specialists.

In terms of language assessment, a big challenge was to understand the categories used in the literature and how they are implemented.

The study highlights the potential for LLMs to self-correct their own output. More rounds will be needed to assess the impact of optimizing self-correcting prompts in obtaining useful and qualified texts on medication use that promote understanding of how a patient should use their medications.

5. REFERENCES

1. Li Z, Xu X, Shen T, Xu C, Gu JC, Tao C. Leveraging large language models for NLG evaluation: A survey. arXiv preprint arXiv:2401.07103. 2024 Jan 13.
2. Fu J, Ng SK, Jiang Z, Liu P. GPTscore: Evaluate as you desire. arXiv preprint arXiv:2302.04166. 2023 Feb 8.

Mapeando ontologias oncopediátricas com foco em tumores do Sistema Nervoso Central: Uma Revisão de Escopo

Mell A. Matsuda^{1,2}, Mariana C. Pereira^{1,2,3}, Marialva Sinigaglia^{1,3} e Sílvio C. Cazella²

¹Instituto do Câncer Infantil, Porto Alegre, RS, Brasil.

²Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil

³Instituto Nacional de Ciência e Tecnologia em Biologia do Câncer Infantil e Oncologia Pediátrica - INCT BioOncoPed, Porto Alegre, RS, Brasil

mell.matsuda@ufcspa.edu.br, marianacp@ufcspa.edu.br, msinigaglia@ici.org,
silvioc@ufcspa.edu.br

Resumo. A dispersão de informações e a diversidade terminológica na saúde trazem desafios para a gestão e estruturação do conhecimento. Ontologias surgem como solução promissora para melhorar a recuperação e processamento de dados. Este projeto relata o mapeamento de ontologias oncopediátricas com foco em tumores do Sistema Nervoso Central (SNC), visando identificar lacunas nas ontologias existentes e propor um modelo ontológico que facilite a integração, padronização e recuperação de dados, aprimorando o acompanhamento e visualização sistêmica de pacientes. A metodologia inclui revisão de escopo e de literatura. O protocolo de revisão foi registrado, e um formulário de extração de dados foi criado para capturar informações relevantes. A NeOn Methodology e a ferramenta Protégé foram selecionadas para modelagem. A revisão mostrou escassez de estudos recentes e falta de especificidade em metodologias e ferramentas. A experiência ressalta a importância da coleta e divulgação de informações detalhadas para facilitar a replicação e aprimoramento de trabalhos na área.

Abstract. Dispersion of information and terminological diversity in healthcare pose challenges for knowledge management and structuring. Ontologies offer a promising solution to enhance data retrieval and processing. This project reports on mapping pediatric oncology ontologies, focusing on Central Nervous System (CNS) tumors. It aims to identify gaps in current ontologies and propose a model to improve data integration, standardization, and retrieval, enhancing monitoring and systemic visualization of patients. The methodology includes scope and literature review. A scope review protocol was registered, and a data extraction form was developed to capture relevant information. NeOn Methodology and Protégé tool were selected for ontology modeling. The review revealed a lack of recent studies and specificity in methodologies and tools. This highlights the importance of detailed information collection and dissemination to support replication and improvement in the field.

Palavras-chave: Interoperabilidade; Oncologia pediátrica; Ontologia; Sistema Nervoso Central.

Nome do projeto: Proposta de ontologia de domínio de pacientes oncopediátricos visando à aplicação em sistemas inteligentes

Links: pubmed.ncbi.nlm.nih.gov; bvsalud.org; ieeexplore.ieee.org; osf.io/8a3ep; parsif.ai; protege.stanford.edu.

1. CONTEXTO E MOTIVAÇÃO

A dispersão da informação e a diversidade de terminologias na saúde são grandes desafios para a estruturação e gestão do conhecimento. Essa diversidade exige métodos sofisticados para a recuperação de dados. Ontologias, nesse contexto, têm se mostrado promissoras para melhorar o acesso, recuperação e processamento de informações. Diante disso, torna-se crucial investigar os modelos ontológicos de domínio que foram desenvolvidos para bancos de dados oncopediátricos, especialmente considerando que o câncer é a principal causa de morte por doença entre crianças

e adolescentes no Brasil e em países desenvolvidos (1). A complexidade do cuidado abrangente e contínuo a pacientes oncopediátricos requer a colaboração de equipes multidisciplinares, frequentemente dispersas geograficamente e com centros de referência nas capitais. Essa situação gera um grande volume de registros clínicos dispersos, contribuindo para a falta de interoperabilidade, a ausência de padronização e a dificuldade de integração de dados e acompanhamento clínico (2). Além disso, muitos sistemas terminológicos não foram projetados para automação, o que agrava esses desafios.

Neste cenário, as ontologias aparecem como uma solução potencial para os problemas de interoperabilidade, padronização e recuperação de dados. A Portaria nº 2.073/2011 do Ministério da Saúde (3) incentiva o uso de ontologias, terminologias e classificações comuns para a definição e representação de conceitos relacionados à saúde, facilitando a harmonização de dados e a melhoria do atendimento clínico de acordo com diretrizes nacionais. Assim, as ontologias podem facilitar a harmonização dos dados e melhorar a qualidade do atendimento ao proporcionar uma visão mais clara e integrada dos casos clínicos ao mesmo tempo que respondem a uma orientação nacional.

1.1. Objetivo

Relatar o mapeamento das ontologias existentes no contexto oncopediátrico, com foco em tumores do Sistema Nervoso Central, a fim de identificar lacunas e propor um modelo ontológico de domínio que facilite a integração, padronização e recuperação de dados, melhorando assim o acompanhamento e a visualização sistêmica dos pacientes.

2. ATIVIDADES PRINCIPAIS

O projeto visa modelar uma ontologia de domínio para o contexto da oncopediatria. Para isso, foi crucial revisar a literatura existente sobre ontologias na área, a fim de compreender o estado atual das pesquisas e identificar lacunas. Também foram analisados modelos ontológicos capazes de especificar relações terminológicas e conceituais dentro da oncopediatria. A escolha das ferramentas adequadas para a modelagem foi essencial para garantir a eficiência no desenvolvimento.

Adicionalmente, foram estudados dados de saúde de registros institucionais e hospitalares, com o objetivo de entender melhor as informações que precisavam ser integradas à ontologia. Esse estudo detalhado assegura que a ontologia seja completa e útil tanto para o acompanhamento clínico quanto para a pesquisa em oncopediatria, promovendo a padronização e a interoperabilidade entre diferentes sistemas de saúde.

As principais atividades realizadas incluem: 1) Identificar descritores para definir um protocolo de revisão de escopo, listando termos e dados clínicos pertinentes; 2) Definir o protocolo de revisão para realizar buscas estruturadas; 3) Estruturar um formulário para extração de dados da literatura, interpretando os resultados e identificando ferramentas; 4) Selecionar uma metodologia e ferramenta para modelar a ontologia; 5) Apoiar o desenho de fluxo e hipóteses de um banco de dados para estudos epidemiológicos, orientando o esquema lógico com base na ontologia e dados disponíveis.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

O desenvolvimento do projeto para modelagem de uma ontologia de domínio em oncopediatria foi conduzido de forma detalhada e estruturada. Inicialmente, foi realizada uma revisão de escopo nos repositórios PubMed, BVS e IEEE Xplore, visando identificar descritores relevantes e termos clínicos específicos à oncopediatria. O protocolo de revisão foi formalizado e registrado no Open Science Framework, assegurando uma abordagem padronizada. Para organizar as informações

coletadas, um formulário de extração de dados foi desenvolvido no Parsifal, facilitando a análise de ferramentas e metodologias identificadas. Com base nos resultados da revisão, a metodologia NeOn e a ferramenta Protégé foram selecionadas, reconhecidas pela robustez e capacidade de estruturar relações terminológicas de forma eficaz. A decisão de focar inicialmente em tumores do Sistema Nervoso Central (SNC) se justifica por sua prevalência na pediatria, representando 20% das neoplasias infantis, sendo o tipo de tumor sólido mais comum em crianças (4).

Contudo, a escassez de estudos recentes pode limitar a replicação e evolução de pesquisas futuras. A implementação da ontologia em ambientes clínicos enfrenta barreiras como a integração com sistemas legados e a adaptação às rotinas hospitalares já estabelecidas. A solução para esses desafios passa pelo desenvolvimento de estratégias de interoperabilidade, automação de processos e capacitação de equipes. Além disso, a validação da ontologia através de estudos de caso em cenários clínicos reais pode facilitar a adoção dessa tecnologia, contribuindo para a melhoria da padronização e integração de dados em oncologia pediátrica. Este trabalho foi registrado na Comissão de Pesquisa da Universidade Federal de Ciências da Saúde de Porto Alegre (nº 590-2023).

4. DESAFIOS E APRENDIZADOS

Foram encontrados desafios referentes à escassez de estudos recentes específicos para a oncopediatria, o que dificultou a identificação de métodos apropriados, pois não havia referências atualizadas suficientes para fundamentar a escolha das melhores práticas e ferramentas. Além disso, a falta de detalhes nas publicações sobre metodologias e ferramentas utilizadas na modelagem de ontologias limitou a capacidade de avaliar e selecionar as abordagens mais eficazes. A integração de Inteligência Artificial e a automação de processos, no entanto, destacam-se como soluções promissoras para superar essas lacunas, acelerando a padronização e facilitando a interoperabilidade entre sistemas. A experiência ressalta a importância da coleta e divulgação de informações detalhadas para facilitar a replicação e o aprimoramento de trabalhos na área. Os resultados gerados pela revisão de escopo foram incluídos em “Domain Ontology Proposition for Central Nervous System Tumors in Pediatric Patients” (5).

5. AGRADECIMENTOS

Fundação de Amparo à Pesquisa do Rio Grande do Sul, Projeto FAPERGS 23/2551-0001184-0 e Instituto do Câncer Infantil.

6. REFERÊNCIAS

1. Instituto Nacional de Câncer - INCA. Câncer infantojuvenil. 2018. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-infantojuvenil>. Acesso em 22/07/2024.
2. Teixeira LMD, Almeida MB. Aspectos ontológicos e epistemológicos em terminologias clínicas: em busca de interoperabilidade semântica no ambiente clínico. *Encontros Bibli.* 2019. May 6;24(55):1–21.
3. Brasil. Ministério da Saúde. Gabinete do Ministro. Portaria nº 2.073, de 31 de agosto de 2011.
4. Instituto Nacional de Câncer - INCA. Tumores do sistema nervoso central. 2023. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/infantojuvenil/especificos/tumores-do-sistema-nervoso-central>. Acesso em 22/07/2024.
5. Pereira M, Matsuda M, Lanferdini L, Sinigaglia M, Cazella S. Domain Ontology Proposition for Central Nervous System Tumors in Pediatric Patients. In: *Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS 2024)*. v. 3565. p. 1-8 (aceito para publicação).

Melhorando qualidade de modelos preditivos de custo hospitalar por enriquecimento descritivo e geração de dados sintéticos

Guilherme Fonseca², Ana Machado², Gabriel Prenassi², Ricardo Cardoso³, Fernanda Lana¹,
Claudio Andrade¹, Milena S. Marcolino¹, Jussara M. Almeida¹, Leonardo Rocha²,
Marcos A. Gonçalves

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG,

²Universidade Federal de São João del-Rei, São João del-Rei, MG,

³Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.

guilhermefonseca8426@aluno.ufsj.edu.br

Resumo. Nosso estudo é focado em custos hospitalares de pacientes com covid-19, em que parte significativa desses custos está associada à internação em unidade de terapia intensiva (UTI). Estudos anteriores revelaram que os modelos preditivos de UTI para pacientes com covid-19 apresentaram limitações de efetividade. Diante desse cenário, o objetivo do presente estudo é investigar duas estratégias para melhorar a efetividade de modelos de predição de custo por internação hospitalar: 1) enriquecimento descritivo, por meio da introdução de novas características que descrevem a condição médica dos pacientes no momento da internação; e 2) geração de dados sintéticos, utilizando Generative Adversarial Networks (GANs). Os resultados mostraram que o enriquecimento descritivo levou a melhorias significativas na qualidade da predição de UTI, atingindo um Macro F1 de 83%. Em contraste, a geração de dados sintéticos não resultou em melhorias. Observamos uma grande interseção entre os casos de classificação errônea com dados reais e sintéticos, indicando que as características desses pacientes incorretamente classificados precisam ser estudadas mais detalhadamente em estudos futuros.

Abstract. Our study is focused on hospital costs of patients with covid-19, in which a significant part of these costs is associated with admission to the intensive care unit (ICU). Previous studies revealed that ICU predictive models for patients with covid-19 had low effectiveness. Given this scenario, the objective of this study is to investigate two strategies to improve the effectiveness of cost prediction models per hospital stay by means of: 1) descriptive enrichment, through incorporation of new characteristics that describe the medical condition of patients at the time of admission; and 2) generation of synthetic data, using Generative Adversarial Networks (GANs). Results reveal that descriptive enrichment significantly improved the quality of UTI prediction, reaching a MacroF1 of 83% while synthetic data generation did not produce improvements. We observed a large intersection between misclassification cases with real and synthetic data, indicating that the characteristics of these incorrectly classified patients need further investigation.

Palavras-chave: modelos de predição de custo; unidade de terapia intensiva; covid-19.

Nome do projeto: Predição de Desfechos Clínicos e Econômicos por Meio de Representações Semânticas Multi-modais de Pacientes Resilientes a Drifts Temporais (Marcos André Gonçalves)

1. CONTEXTO E MOTIVAÇÃO

Durante boa parte da pandemia de covid-19 os sistemas de saúde foram sobrecarregados. Além da importância da vida das pessoas, uma questão que surgiu foi a avaliação do impacto da pandemia na viabilidade econômica dos hospitais (i.e. custos). Em análises de custos hospitalares de pacientes com covid-19 realizadas pelo nosso grupo, identificamos que parte significativa desses custos estava associada à internação na unidade de terapia intensiva (UTI) (1). Entretanto, o modelo para predição de UTI criado em (1) possuía uma baixa efetividade (i.e. melhor resultado: 58,5% de MacroF1 para um dos hospitais), o que atribuímos a duas causas principais: baixo número de pacientes (i.e., 733 pacientes) e baixo número de variáveis descritivas (i.e., 7). No presente estudo

avaliamos duas estratégias para melhorar a capacidade preditiva desse modelo.

1.1 Objetivo

Investigar duas estratégias para melhorar a eficácia do modelo de predição de necessidade de Unidade de Terapia Intensiva (UTI): 1) incorporação de novas variáveis que descrevem a situação dos pacientes no momento à admissão hospitalar; e 2) geração de dados sintéticos por meio de Generative Adversarial Networks (GANs), aumentando a quantidade de dados para o aprendizado do modelo e conseqüentemente a descoberta de novos padrões interessantes de inferência.

2. ATIVIDADES PRINCIPAIS

Integração de bases de dados para a obtenção de novos atributos, assim como o estudo e a implementação de GANs para geração de dados sintéticos. O estudo foi aprovado pela Comissão Nacional de Ética em Pesquisa, sob o número CAAE: 30350820.5.0000.0008, com dispensa de aplicação de termo de consentimento livre e esclarecido.

3. DESENVOLVIMENTO DO TRABALHO

Foi utilizado um repositório de dados de duas coortes retrospectivas de pacientes com covid-19 (1, 2). A primeira base inclui informações completas de custos hospitalares de 733 pacientes, admitidos entre os meses de março a agosto de 2020 em 4 hospitais públicos/acadêmicos e um privado, localizados no Rio Grande do Sul e em Pernambuco. Ela contém 7 atributos relacionados a cada paciente (idade; sexo; hipertensão arterial; obesidade; diabetes; doença renal crônica; doença pulmonar crônica). A segunda é o "Registro hospitalar multicêntrico nacional de pacientes com covid-19" (2) base com 21559 pacientes admitidos em 41 hospitais de 5 estados, que contém 820 atributos coletados. Primeiro, buscando enriquecer a descrição dos pacientes, fizemos o pareamento entre as duas bases de dados, identificando todos os pacientes presentes na primeira coleção com seu respectivo registro na base 2, o que acrescentou um grande número de variáveis descritivas reais. Para evitar o vazamento de informações (i.e. informações relacionadas à presença na UTI) durante o treinamento do modelo preditivo, utilizamos apenas variáveis que foram avaliadas em tempo de admissão hospitalar, resultando em 90 variáveis que incluem as sete do estudo anterior. Essa coleção enriquecida foi dividida em treino e teste (50% cada), realizando a classificação com 10 repetições utilizando o classificador XGBoost. A MacroF1 do modelo foi avaliada usando as partições de teste.

Tabela 1 – Macro F1 na predição de UTI e com 95% de intervalo de confiança.

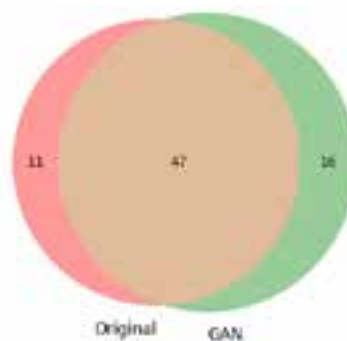
Hospital	Variáveis 1	Variáveis 2	Variáveis 1 +GAN	Variáveis 1 +3 GAN	Variáveis 2 +GAN	Variáveis 2 +3 GAN
Todos	53.0(1.1)	81.9(1.3)	52.6(1.4)	53.8(1.9)	81.5(1.1)	81.3(1.3)
I	58.5(2.9)	75.4(3.8)	54.1(3.6)	57.5(5.0)	75.1(4.3)	74.0(2.8)
II	50.5(5.1)	78.1(4.0)	48.1(6.1)	47.1(5.8)	77.3(4.4)	77.9(2.3)
III	54.4(4.1)	78.6(2.8)	53.6(4.3)	50.2(3.3)	77.8(3.2)	78.5(2.7)
IV	52.0(3.5)	80.0(1.7)	49.1(3.2)	47.4(3.7)	76.5(3.4)	75.9(2.4)
V	49.2(2.2)	81.6(4.0)	50.9(3.4)	51.7(3.6)	79.3(4.1)	79.3(4.3)

A Tabela 1 apresenta os resultados dos modelos preditivos considerando cada um dos 5 hospitais separadamente (linhas I a V) e todos eles juntos (linha "todos"). Os modelos foram treinados com as variáveis do trabalho anterior (coluna "variáveis 1") e com o novo conjunto enriquecido de variáveis (coluna "variáveis 2"). Os resultados mostram que o novo conjunto de variáveis permitiu um aumento significativo na efetividade do modelo para todos os hospitais, alcançando melhoria de até 65.8 pontos percentuais na MacroF1 (49.2 para 81.6 no hospital V).

Na segunda abordagem, utilizamos a pGAN (3), para aumentar, sinteticamente, o número de pacientes em nossa coleção. Nossa hipótese é que o baixo número de amostras (pacientes) é um fator limitador da capacidade de aprendizado dos modelos. Esta GAN, treinada na partição de treinamento, gera instâncias que, junto com os dados originais, são utilizadas no processo de treinamento do modelo. Utilizamos a pGAN para gerar instâncias que dobravam (colunas “+GAN” na Tabela 1) e triplicavam (colunas “+3 GAN” na Tabela 1) o total de pacientes no conjunto de treino. Na Tabela 1, temos que, em ambos os conjuntos de variáveis, independentemente do número de novas instâncias (pacientes sintéticos) gerados, a eficiência dos classificadores se manteve estatisticamente equivalente ao modelo gerado apenas com os dados originais.

Para compreender tais resultados, investigamos, a partir de Diagramas de Venn, os erros de classificação do modelo original e do modelo acrescido de dados sintéticos. Por limitações de espaço, apresentamos na Figura 1 apenas o diagrama do modelo treinado com todos os hospitais comparando a base original com a base contendo o triplo de dados gerados pela pGAN. Observamos que há uma grande interseção entre os erros dos dois modelos, ou seja, com ou sem dados sintéticos ambos os modelos tendem a errar a classificação dos mesmos pacientes, o que parece ser um forte indício que tenhamos atingido o limite preditivo da tarefa. Em trabalhos futuros iremos avaliar detalhadamente as características desses pacientes, identificando o que induz os modelos a errarem a sua predição de necessidade de UTI.

Figura 1 - Diagrama de Venn dos erros de classificação dos modelos com e sem dados sintéticos.



4. DESAFIOS E APRENDIZADOS

Este trabalho estende (1) ao investigar duas estratégias para melhoria da efetividade de um modelo preditivo de admissão de pacientes em UTI. Com a inclusão de novas variáveis descritivas, conseguimos obter uma melhora de até 65.8 pontos percentuais de efetividade. Em contrapartida, a geração de dados sintéticos com GANs não foi capaz de produzir melhorias nos modelos preditivos. Investigação dos casos de falha de predição é o próximo passo do estudo.

5. REFERÊNCIAS

1. Andrade CMV, Paiva B, Morais G, Cardoso R, Marcolino M, et al. Predição de Necessidade de Internação em Terapia Intensiva e Custos de Tratamento de Pacientes com COVID-19. Anais do I Simpósio CI-IA Saúde da UFMG. Set 2023.
2. Marcolino MS, Ziegelmann PK, Souza-Silva MVR, Nascimento IJB, Oliveira LM, et al. Clinical characteristics and outcomes of patients hospitalized with COVID-19 in Brazil: Results from the Brazilian COVID-19 registry. International Journal of Infectious Diseases. 2021 Jun 1;107:300–10.
3. Venugopal R, Shafqat N, Venugopal I, Tillbury BM, Stafford HD, Bourazeri A. Privacy preserving Generative Adversarial Networks to model Electronic Health Records. Neural Netw. Set 2022;153:339-48.

Método Automatizado de Registro de Métricas do Instagram: resumo expandido - Iniciação Científica

André Felipe de Andrade Santos (aluno)¹, André Soares da Silva (aluno)¹,
Raíssa Silva Guimarães Gomes Magalhães (aluno)¹, Cristiane Dias (coorientador)¹,
Zilma Silveira Nogueira Reis (orientador)¹

¹Centro de Inovação em Inteligência Artificial para a Saúde, Universidade
Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

andresants@ufmg.br, andreso@ufmg.br, raissasggm@ufmg.com, cristianedias@ufmg.br,
zilma@ufmg.br

Resumo. O CI-IA busca estar presente nas mídias sociais como forma de promover a disseminação do conhecimento sobre o assunto. Nesse sentido, o perfil no Instagram “Inteligência Artificial para a Saúde” é um dos principais meios digitais utilizados pelo Centro no contexto de divulgação científica. O estudo demonstra o uso de um método automatizado para registrar as métricas de engajamento no perfil citado. Para isso, utilizou-se um script e a plataforma Google Sheets para extração automática dos dados referentes no Instagram. O script foi capaz de realizar a extração e viabilizar o cálculo de engajamento de forma automatizada.

Abstract. The Centro de Inovação em Inteligência Artificial para a Saúde (CI-IA) seeks to be on social media in order to promote the dissemination of knowledge on the subject. In this regard, the Instagram profile "Inteligência Artificial para a Saúde" (Artificial Intelligence for Health) is one of the main social medias used by CI-IA to disseminate scientific knowledge. This study shows the use of an automated method to record engagement metrics on the aforementioned profile. To accomplish that, a script and the Google Sheets platform were used for automatic extraction of data from Instagram. As a result, the script successfully extracted data and enabled the automated calculus of engagement.

Palavras-chave: Engajamento; Interação Social; Inteligência Artificial.

Nome do projeto: Centro de Inovação em Inteligência Artificial para a Saúde

Links: <https://www.instagram.com/iasaudeufmg/>

1. CONTEXTO E MOTIVAÇÃO

O objetivo do perfil no Instagram “Inteligência Artificial para a Saúde” é promover a divulgação científica, fazendo uso de uma linguagem que seja acessível ao público externo ao ambiente acadêmico. Dessa maneira, almeja-se democratizar o conhecimento sobre o tema, de modo que sejam transmitidas informações confiáveis e baseadas em ciência sobre o uso da IA na área da saúde. O engajamento informacional é definido como o ato de interagir e compartilhar informações entre usuários conectados a uma rede social (1), sendo um dos parâmetros para analisar e quantificar a eficiência da comunicação com o público. Nesse contexto, o registro rotineiro de variáveis, como o número de seguidores em razão do tempo, é necessário para avaliar o engajamento no perfil. O uso de métodos automatizados para registrar as métricas faz-se necessário para uma otimização da análise desses dados, visto que a plataforma Instagram disponibiliza dados limitados que não permitem o cálculo de maneira longitudinal.

1.1 Objetivo

Este estudo tem como objetivo demonstrar um método de registro da interação social aos conteúdos publicados no perfil do CI-IA Saúde na rede social Instagram, de modo automatizado e eficiente

para analisar o engajamento.

2. ATIVIDADES PRINCIPAIS

A partir de um modelo livre disponível no YouTube, um script foi desenvolvido para a plataforma “Google Sheets” para extração automática de dados do Instagram. Semanalmente, é feita a captura dos dados e o cálculo do engajamento, obtido pela razão entre o número de likes somado ao dobro do número de comentários, multiplicado por 100 e o número de seguidores (2).

3. DESENVOLVIMENTO DO TRABALHO

A plataforma Google Sheets permite a utilização de extensões, as quais podem adicionar funcionalidades extras às planilhas já criadas. Dentre essas extensões, há a chamada Apps Script que utiliza códigos de programação escritos em JavaScript. Tal funcionalidade tem a capacidade de trabalhar com APIs, que são um conjunto de regras e protocolos que permitem a integração de sistemas. A API chamada GraphQL realiza a integração de informações do Instagram com as planilhas do Google.

Informando o nome da página da planilha, o código de ID referente ao perfil que deseja extrair os dados, o cookie de autenticação de um usuário e as variáveis que deseja computar, pode-se realizar solicitação para essa API. Desse modo, pode-se obter uma tabela com o número de seguidores, curtidas, postagens e outros dados que permitem a monitorização e cálculo de engajamento da página. Uma versão simplificada do código para a contagem de seguidores está disponibilizada no link abaixo:

Figura 1 – QR code de acesso ao código em JavaScript para registro de seguidores do Instagram



Fonte: Próprio autor

Para o código criado ser realizado de maneira automatizada, a extensão Apps Script possui a funcionalidade de criar acionadores, os quais podem ser colocados em razão do tempo, seja diário, semanal ou mensal. Assim, pode-se criar um acionador para extração de dados automática de um determinado perfil do Instagram.

Figura 2 – Resultados da extração automática de métricas do instagram realizadas semanalmente do perfil @iasaudeufmg

Data	Seguidores	Diferença de Seguidores	Número de Posts	Likes Totais	Comentários Totais	Taxa de Engajamento	Diferença da Taxa de Engajamento
30/04/2024	1634	0	321	423	23	2,27%	0
30/04/2024	1635	1	321	425	23	2,28%	0,01%
07/05/2024	1663	28	324	379	23	2,01%	-0,27%
14/05/2024	1674	11	327	315	18	1,64%	-0,36%
21/05/2024	1712	38	330	288	17	1,48%	-0,17%
28/05/2024	1729	17	331	264	17	1,35%	-0,13%
04/06/2024	1738	9	334	262	16	1,33%	-0,02%
17/06/2024	1747	9	341	258	16	1,31%	-0,03%
18/06/2024	1747	0	342	232	13	1,17%	-0,14%
25/06/2024	1753	6	346	256	13	1,28%	0,11%
02/07/2024	1772	19	350	244	13	1,21%	-0,07%
09/07/2024	1791	19	353	249	13	1,22%	0,01%
16-07-2024	1816	25	357	269	12	1,29%	0,07%
23-07-2024	1832	16	360	248	11	1,18%	-0,11%

Fonte: Próprio autor

Essa informação possui grande importância para monitoramento de crescimento da interação das mídias e engajamento do perfil no Instagram. Permite-se assim, o entendimento da eficiência das estratégias de divulgação e sua interação com o público, podendo avaliar o interesse pelas postagens de acordo com o seu tema. Além disso, propor melhorias na divulgação acerca do conhecimento a respeito da Inteligência Artificial aplicada à Saúde.

4. DESAFIOS E APRENDIZADOS

O principal desafio diz respeito à ampla gama de conhecimentos aplicados no estudo. Como acadêmicos do curso de medicina, o contato direto com o pensamento computacional e com a divulgação científica em linguagem simplificada, mesmo que desafiadores, ampliam a visão da comunicação em saúde. Os dados obtidos precisam ser interpretados e entendidos para que tenham utilidade prática, desde a seleção da sua captura até as estratégias de confecção e publicação das postagens. Dentre os aprendizados, destaca-se o contato direto com a linguagem de programação e do conhecimento a respeito de sistemas automatizados, o qual viabiliza a busca de estratégias mais eficazes para a divulgação científica.

5. REFERÊNCIAS

1. Terra CF, Carvalho E. Street art: das margens da sociedade para a comunicação organizacional: o grafite no ambiente digital como técnica de branding para relacionamento com o público consumidor metropolitano. Em: Anais do 14o congresso internacional de relações públicas e comunicação Salvador: ALARP. 2014.
2. Liang S, Wolfe J. Getting a Feel of Instagram Reels: The Effects of Posting Format on Online Engagement. Journal of Student Research. 30 de novembro de 2022;11.

Métodos de Aprendizagem de Máquina para Previsão da Quantidade Semanal de Casos de Dengue*

Giovanni E. Zanardo¹, Éfren L. Souza^{1,2}, Fabíola G. Nakamura¹, Eduardo F. Nakamura¹

¹Instituto de Computação - Universidade Federal do Amazonas (UFAM)

²Instituto de Engenharia e Geociências - Universidade Federal do Oeste do Pará (UFOPA)

{gio.zanardo, fabiola, nakamura}@icomp.ufam.edu.br, efren.souza@ufopa.edu.br

Resumo. Arboviroses são doenças provocadas por arbovírus, sendo estes, vírus transmitidos pela picada de artrópodos hematófagos. Atualmente, a dengue é a arbovirose mais prevalente do mundo, sendo responsável por cerca de 100 milhões de casos/ano em uma população de risco de 2,5 a 3 bilhões de seres humanos. As arboviroses transmitidas pelo *Aedes aegypti* e *Aedes albopictus* estão entre os principais problemas de saúde pública, sendo a dengue a mais proeminente. O manejo epidemiológico da dengue requer preparação avançada; assim, prever a quantidade de casos em uma região específica pode auxiliar nas estratégias de prevenção e controle do processo epidêmico. Com isso em vista, este estudo avalia a eficácia de técnicas estatísticas clássicas e métodos de aprendizado de máquina na previsão de casos de dengue a partir de dados geográficos de San Juan, Porto Rico. Para isso, selecionamos características usando a matriz de correlação cruzada com o número total de casos semanais de dengue, que foram posteriormente filtrados por transformações wavelet. O modelo de Regressão Linear, utilizando níveis de precipitação e vegetação filtrados pela wavelet symmlet (sym20), mostrou o melhor desempenho nas métricas MAE, R-quadrado, MAPE, RMSE e BIAS.

Palavras-chave: Dengue; Aprendizado de Máquina; Wavelets.

Nome do projeto: AmazonIA - Aplicação de Modelos de Aprendizagem de Máquina e Aprendizagem Profunda para Análise de Dados sobre Doenças Tropicais Negligenciadas na Amazônia

1. CONTEXTO E MOTIVAÇÃO

O número de casos de dengue aumentou significativamente nas últimas duas décadas, com a OMS documentando um crescimento de dez vezes entre 2000 e 2019, chegando a 5,2 milhões de casos**. Em 2019, um pico histórico ocorreu com casos em 129 países, resultando em mais de 3,18 milhões de casos, 28.208 casos graves e 1.823 mortes, evidenciando uma taxa de letalidade de 0,06%. Apesar de uma queda temporária entre 2020 e 2022 devido à pandemia de COVID-19, 2023 registrou um ressurgimento alarmante, estendendo a transmissão para regiões previamente não afetadas. Nas Américas, a incidência cresceu drasticamente, exemplificado pelo aumento de 1,5 milhão de casos nos anos 1980 para 17,5 milhões na década de 2010. Diferenças climáticas e fatores demográficos, como natalidade e mobilidade, influenciam a sazonalidade da doença, complicando a previsão de surtos. Dessa forma, métodos baseados em aprendizado de máquina emergem como ferramentas fundamentais para modelar esses complexos padrões multifatoriais da dengue, prevendo sua incidência semanal por região, destacando a necessidade urgente de medidas preventivas e tratamentos eficazes frente à mortalidade associada à doença.

*Uma versão estendida desse trabalho foi publicada no SEMISH 2024 [4].

** World Health Organization. Dengue - Global situation [Internet]. 2023 [cited 2023 Dec 29].

1.1 Objetivo

O estudo visa explorar a eficiência preditiva de uma gama de modelos, abrangendo tanto técnicas estatísticas clássicas quanto métodos rasos e profundos de aprendizado de máquina. Entre os modelos estatísticos clássicos avaliados estão o Random Walk (RW), Holt's linear trend method

(SEH), Simple Exponential Smoothing (SES), Holt-Winters (ES) e ARIMA. Paralelamente, investigamos arquiteturas de aprendizado profundo e modelos rasos regressivos, incluindo Linear Regression (LR), Support Vector Regression (SVR), Random Forest Regression (RFR), Gradient Boosting Regressor (GBR) e XGBoost, visando comparar o desempenho individual de cada modelo na predição de número de casos semanais de dengue. Além da comparação entre modelos, verificamos estratégias eficientes para a reformatação de conjuntos de dados constituídos por séries temporais, adequando-os para aplicação em modelos supervisionados de aprendizado de máquina, bem como selecionamos as características mais relevantes e determinamos os modelos mais eficazes para a predição do número de casos semanais de dengue em San Juan (Porto Rico).

2. ATIVIDADES PRINCIPAIS

No decorrer do projeto, as atividades desenvolvidas consistiram em:

- Curadoria e pré-processamento do dataset obtido em bases on-line;
- Construção, experimentação e refinamento de hiperparâmetros dos modelos;
- Análise dos resultados obtidos e comparação do desempenho com baselines;
- Escrita de relatório final que promoveu a publicação do artigo.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Inicialmente, os dados ausentes são preenchidos através de interpolação polinomial. Em seguida, aplica-se Z-score na coluna de casos semanais e mantém-se apenas as instâncias com valores de Z-score maior que -3 e menor que 3. Esta abordagem identifica valores aberrantes (outliers) ao excluir pontos que estão a mais de três desvios padrão da média. As séries temporais são então analisadas quanto à estacionariedade usando o Teste Aumentado de Dickey-Fuller. No pré-processamento, também aplicamos transformadas wavelet para dividir cada série em componentes de alta e baixa frequência, com o nível de decomposição ajustado para capturar detalhes ou tendências gerais. Para o aprendizado supervisionado, os dados são reformatados

[1] visando abordagens de predição one-step, multi-step e multi-output, onde as janelas temporais são construídas usando apenas a coluna de casos semanais de dengue (univariada) ou múltiplas características (multivariada). Para a abordagem multivariada, dentre os 24 atributos presentes no dataset, escolhemos com base na análise da matriz de correlação cruzada entre a coluna de casos totais e os atributos restantes, visando obter a combinação de características com maior poder preditivo. Finalmente, o dataset pré-processado é preparado para treinamento e teste, dividindo-o na proporção 80/20. Ao término, avaliamos o desempenho dos modelos através de desempenho em métricas e visualização gráfica. Nossos modelos apresentaram uma redução de 1,90 em RMSE quando comparados ao baseline que utiliza algoritmos genéticos no pré-processamento e um ensemble de modelos (ANN, SVR e CNN) para predição [3]. Ademais, o modelo LR (sym20) excedeu os resultados de Panja et al. [2] em termos de MAE e RMSE, com reduções de 2,94 e 3,89, respectivamente. Interessantemente, o desempenho do modelo profundo CNN-LSTM-LR replicou os resultados de LR (sym20), mas adicionou maior viés.

4. DESAFIOS E APRENDIZADOS

A principal limitação deste estudo é a falta de generalização do modelo mais eficaz, que foi testado apenas com dados de San Juan, e o número reduzido de instâncias pode ter beneficiado modelos menos complexos. A alta dimensionalidade gerada pela estratégia de janelamento também comprometeu o desempenho de modelos como SVR e árvores de regressão, resultando na "maldição da dimensionalidade". Para trabalhos futuros, propomos explorar técnicas de pré-processamento, como PCA adaptado para séries temporais, e avaliar arquiteturas de codificador-decodificador com atenção autorregressiva, como Transformers, buscando melhorar a predição

de dengue e expandir a aplicação dos modelos a mais datasets para ampliar a generalização dos resultados obtidos.

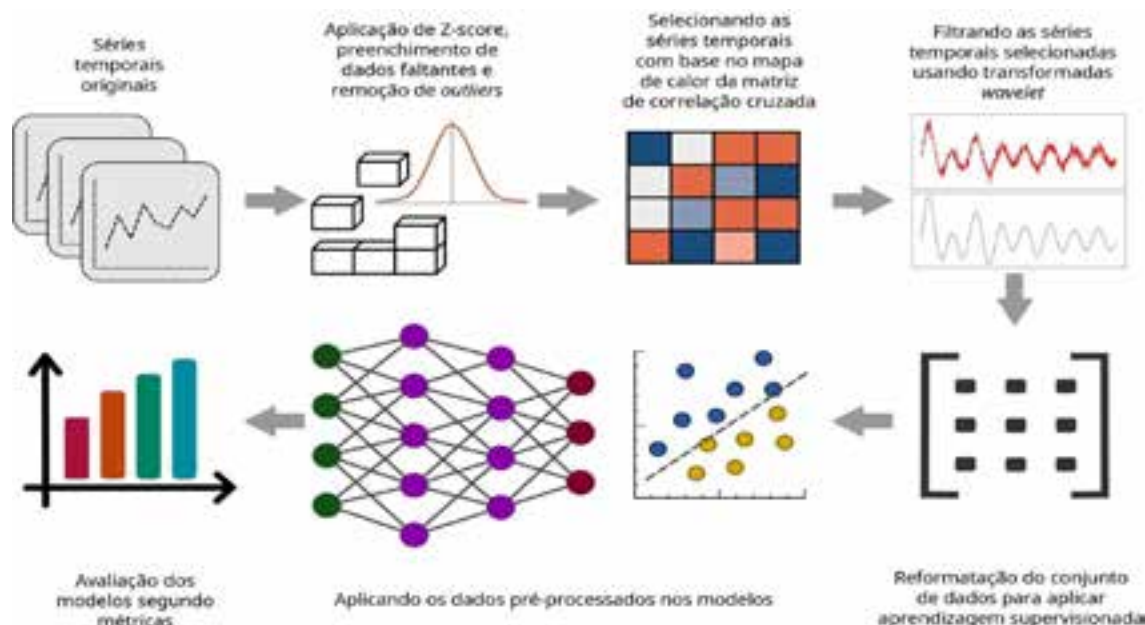


Figura 1. Esquema geral da experimentação [4].

5. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES-PROEX) - Código de Financiamento 001. Este trabalho foi parcialmente financiado pela Fundação de Amparo à Pesquisa do Estado do Amazonas – FAPESAM – por meio do projeto POSGRAD 2024/2025. O trabalho também conta com o apoio do Centro de Inovação em Inteligência Artificial para a Saúde (CIIA-Saúde), Proc. 2020/09866-4 - FAPESP/ MCTIC/CGI.

6. REFERÊNCIAS

1. Benidis K, Rangapuram SS, Flunkert V, et al. Deep Learning for Time Series Forecasting: Tutorial and Literature Survey. *ACM Computing Surveys*. 2022;55(6):1-36.
2. Panja M, Chakraborty T, Nadim SS, et al. An ensemble neural network approach to forecast Dengue outbreak based on climatic condition. *Chaos Solitons Fractals*. 2023;167:113124.
3. Shaikh MSG, SureshKumar DB, Narang D. Development of optimized ensemble classifier for dengue fever prediction and recommendation system. *Biomed Signal Process Control*. 2023;85:104809.
4. Zanardo G, Souza É, Nakamura F, Nakamura E. Uma Comparação entre Métodos baseados em Aprendizado de Máquina para inferir número de casos semanais de Dengue. In: *Anais do 51º Seminário Integrado de Software e Hardware*; 2024; Brasília/DF.

Métodos de Estimação de Expressões Gênicas de Câncer de Mama Baseados em Correlação*

Beatriz A. Rodrigues¹, Rayol M. Neto¹, Fabiola F. Nakamura¹, Eduardo F. Nakamura¹

¹Universidade Federal do Amazonas-UFAM, Manaus, AM, Brasil.

{beatriz.albuquerque,fabiola,nakamura,rayol}@icomp.ufam.edu.br

Resumo. Os dados de expressão gênica geralmente sofrem de problemas de valor perdido devido a uma variedade de razões experimentais. Em bases de dados de câncer de mama, a análise subsequente e a classificação de subtipos podem sofrer fortemente com dados ausentes, sendo assim é primordial tratar esse problema. Várias abordagens para estimação desses valores em dados de expressão gênica foram desenvolvidas, mas a tarefa é difícil devido a fatores como a existência ou não de uma estrutura de correlação nos dados e à alta dimensionalidade (número de genes x número de amostras) dos dados. Nesta pesquisa, desenvolvemos quatro métodos, para tratar valores ausentes em expressões gênicas de câncer de mama, que lidam com a alta dimensionalidade dos dados realizando a seleção de genes que melhor caracterizam o câncer de mama, a partir do uso de informações de correlação entre genes. Os métodos principais mostraram melhores desempenhos nas métricas MAE e RMSE em comparação aos métodos baselines propostos.

Palavras-chave: Câncer de Mama; Expressão Gênica; Aprendizado de Máquina.

Nome do projeto: AmazonIA - Aplicação de Modelos de Aprendizagem de Máquina e Aprendizagem Profunda para Análise de Dados sobre Doenças Tropicais Negligenciadas na Amazônia.

1. CONTEXTO E MOTIVAÇÃO

O câncer é uma das doenças que mais mata pessoas no mundo e a variância dos seus tipos apresenta uma discrepância de mortes por gênero (4). No Brasil, o câncer de mama é o mais comum entre as mulheres, com 66.280 novos casos e 17.825 óbitos em 2020 (4). Existem quatro subtipos moleculares principais de câncer de mama: Basal, Her2, Luminal A e Luminal B. Os subtipos Basal e Her2 têm prognósticos piores, enquanto Luminal A e Luminal B estão associados a prognósticos melhores devido a terapias direcionadas eficazes (6). O avanço tecnológico permitiu o aumento dos estudos sobre o câncer de mama, resultando em diversos repositórios públicos com dados de expressões gênicas de tumores, que indicam a presença ou ausência de determinados genes/proteínas no tecido cancerígeno. No contexto dos bancos de dados públicos, entender como os genes se relacionam e o funcionamento dos sistemas biológicos é possível, mas complexo, devido à presença de valores ausentes nos dados genômicos, o que pode levar a resultados tendenciosos e conclusões inválidas (5). Nas bases de dados de expressão de câncer de mama, a ausência de dados é ocasionada por motivos como poeira ou arranhões no slide, imagem corrompida ou resolução insuficiente (7). A estimativa desses valores ausentes é crucial para melhorar a precisão na caracterização e classificação dos subtipos de câncer de mama, possibilitando a criação de modelos mais eficientes e realistas e, assim, determinando melhores tratamentos para cada subtipo.

* Uma versão estendida desse trabalho foi publicada no SEMISH 2024 [1].

1.1 Objetivo

O objetivo principal deste trabalho é o desenvolvimento de métodos de estimação de expressões gênicas de câncer de mama baseados em correlação para seleção de genes que permitam selecionar os genes completos mais correlacionados aos genes com dados incompletos, a serem estimados por subtipo de câncer de mama, através da utilização de algoritmo supervisionado de

Aprendizado de Máquina (AM), para gerar bases de dados mais completas e precisas.

2. ATIVIDADES PRINCIPAIS

Para alcançar o objetivo definido anteriormente, os seguintes passos foram realizados:

- Identificação do Estado-da-arte relacionado ao tema de pesquisa;
- Coleta e avaliação de Bases de Dados de câncer de mama;
- Caracterização das bases de dados com identificação de distribuição de valores dos dados, identificação de valores faltantes, dimensão (genes x amostras);
- Desenvolvimento de métodos para tratamento de dados faltantes em bases individuais;
- Avaliação dos métodos por meio de métricas adequadas;
- Publicação de resultados em veículo de qualidade.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

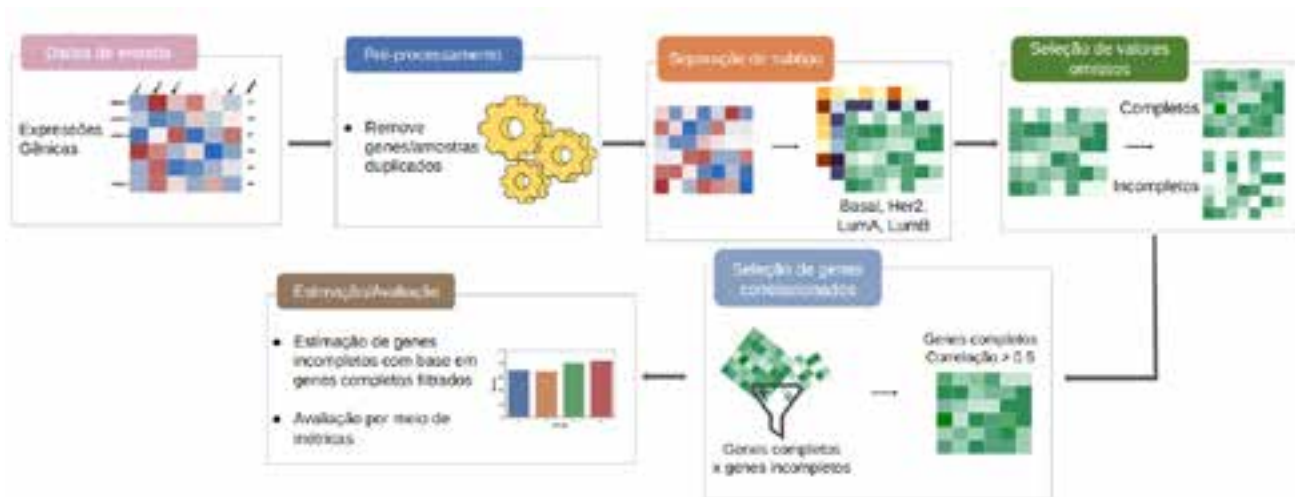
O método começa com o pré-processamento dos dados, removendo valores duplicados. Após concluir esse passo, fazemos a separação dos subtipos de câncer de mama. Nesta pesquisa resolvemos trabalhar, inicialmente, com o subtipo Basal, por ser um dos de pior prognóstico. Em seguida, os genes são separados em dois conjuntos de dados contendo, respectivamente, genes com dados ausentes e genes sem dados ausentes. Aplica-se, então, a correlação entre essas bases buscando um conjunto de genes sem dados ausentes mais significativo para estimação dos genes com valores ausentes. Além disso, para melhorar a seleção de genes relevantes, são aplicadas as técnicas Eliminação Recursiva de Atributos (RFE) e Adição Recursiva de Atributos (RFA) (6). Para realizar a estimação, dividimos nosso conjunto de dados em treino e teste, fazendo uso da abordagem Leave-One-Out Bootstrap, pelo fato de fornecer uma medida menos tendenciosa do RMSE de teste. A partir disso, utilizamos o algoritmo Regressão de Vetor de Suporte (SVR) (2) para estimar os valores ausentes. Para avaliar nosso método, utilizamos apenas os genes presentes na lista PAM50 aceitos como representativos para a caracterização do câncer de mama e que é considerado o conjunto referencial de genes para diferenciar os subtipos (6). Extraímos esses genes do conjunto de dados de proteínas do Clinical Proteomic Tumor Analysis Consortium (CPTAC) (3), o qual fornece maior amplitude analítica, pois usa espectrometria de massa para analisar os proteomas de amostras de tumor TCGA anotadas pelo genoma. O desempenho dos métodos propostos foram medidos por meio do Erro Médio Absoluto (MAE) e do Erro Quadrático Médio (RMSE), que são indicadores de desempenho (KPIs) muito bons para medir a precisão da previsão. Foram desenvolvidos 4 métodos de estimação, sendo dois deles métodos baselines, descritos como RFE46 Pam50 e RFA46 Pam50. Do ponto de vista da estimação, pode-se observar que os métodos não baselines, denominados RFECA Pam50 e RFACA Pam50, alcançaram a maior precisão na estimativa de todos os genes. De modo geral, o método RFECA Pam50 apresentou o melhor desempenho. Quatro genes foram estimados, sendo eles: FOXA1, KIF2C, ORC6 e TMEM45B. Respectivamente, para cada um deles o método RFECA Pam50 teve uma redução em RMSE de 1,24; 0,13; 0,15; e 0,26 comparado ao melhor dos baselines.

4. DESAFIOS E APRENDIZADOS

A principal limitação deste estudo é encontrar bases de dados de expressões gênicas com uma quantidade de amostras proporcional a quantidade de genes ou com um número de amostras maior do que a quantidade de genes, sendo assim treinar os modelos de Aprendizado de Máquina torna-se uma tarefa desafiadora. Dessa forma, precisamos pensar em estratégias adequadas para lidar com o problema da alta dimensionalidade dos dados e garantir a qualidade da precisão da nossa estimação. Para trabalhos futuros, pretendemos verificar como se sai a estimação ao utilizar todos os subtipos da base e investigar se os resultados das métricas tendem a se tornar ainda mais precisos. Além disso, queremos estender a abordagem proposta estimando não apenas genes com dados ausentes do PAM50 como também os demais genes com dados ausentes da base de

dados. Ademais, temos interesse em verificar o quanto a estimação de dados ausentes contribui para a tarefa de classificação dos genes.

Figura 1 – Esquema geral da solução proposta.



Fonte: Própria.

5. REFERÊNCIAS

- Rodrigues BA, Neto RM, Nakamura FF, Nakamura EF. Um método de Estimação de Expressões Gênicas de Câncer de Mama com Base em Correlação. In: Seminário Integrado de Software e Hardware (SEMISH). SBC; 2023. 107–18.
- Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support Vector Regression Machines. In: Advances in Neural Information Processing Systems. MIT Press; 1996.
- Edwards NJ, Oberti M, Thangudu RR, et al. The CPTAC Data Portal: A Resource for Cancer Proteomics Research. J Proteome Res. 2015;14(6):2707-2713.
- INCA. Instituto nacional do câncer - Estatística. 2021. Disponível em: http://www.nlm.nih.gov/bsd/uniform_requirements.html. Acessado em 12/08/2022.
- Kang, H.. The prevention and handling of the missing data. Korean journal of anesthesiology. 2013; 64(5): 402–406.
- Mendonça-Neto R, Reis J, Okimoto L, Fenyö D, Silva C, Nakamura F, et al. Classification of breast cancer subtypes: A study based on representative genes. Journal of the Brazilian Computer Society. 2022; 28(1):59–68.
- Tan AC, Gilbert D. Ensemble machine learning on gene expression data for cancer classification. Appl Bioinformatics. 2003;2(3 Suppl):S75-S83.

Modelagem de ontologia de domínio com foco em tumores do Sistema Nervoso Central em crianças e adolescentes

Mariana C. Pereira^{1,2,3}, Mell A. Matsuda^{1,2}, Lucas N. Lanferdini^{1,2}, Marialva Sinigaglia^{1,3} e Sílvio C. Cazella²

¹Instituto do Câncer Infantil, Porto Alegre, RS, Brasil.

²Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil

³Instituto Nacional de Ciência e Tecnologia em Biologia do Câncer Infantil e Oncologia Pediátrica - INCT BioOncoPed, Porto Alegre, RS, Brasil

marianacp@ufcspa.edu.br, mell.matsuda@ufcspa.edu.br, lucas.lanferdini@ufcspa.edu.br, msinigaglia@ici.org, silvioc@ufcspa.edu.br

Resumo. As instituições de saúde encaram o desafio de lidar com a dispersão da informação e, gradativamente, buscam alternativas para melhorar esta situação. Nos últimos anos as ontologias têm sido utilizadas para modelar a representação do conhecimento, favorecendo a recuperação da informação. Este resumo apresenta um projeto em execução que objetiva propor uma ontologia de domínio para a visualização sistêmica do paciente oncopediátrico com tumor de Sistema Nervoso Central. A metodologia envolve a triagem de fichas clínicas, protocolos clínicos e diretrizes terapêuticas, com a participação de profissionais da área. A NeOn Methodology guiará os processos de construção da ontologia, que será modelada no Protégé. Através da NeOn são apresentados nove cenários que colaboram para a criação de ontologias e redes ontológicas, sendo estes flexíveis a diferentes combinações. Espera-se que o modelo proposto não só represente o conhecimento relacionado aos pacientes oncopediátricos com tumores do Sistema Nervoso Central, como contribua para a expansão das ontologias direcionadas ao público pediátrico e à área da saúde.

Abstract. Health institutions face the challenge of dealing with the dispersion of information and are gradually looking for alternatives to improve this situation. In recent years ontologies have been used to model the representation of knowledge, favoring information retrieval. This summary presents an ongoing project that aims to propose a domain ontology for the systemic visualization of pediatric cancer patients with Central Nervous System tumors. The methodology involves screening clinical records, clinical protocols and therapeutic guidelines, with the participation of professionals in the field. NeOn Methodology will guide the ontology construction processes, which will be modeled in Protégé. Through NeOn, nine scenarios are presented that contribute to the creation of ontologies and ontological networks, which are flexible to different combinations. It is expected that the proposed model will not only represent knowledge related to pediatric oncology patients with tumors of the Central Nervous System, but also contribute to the expansion of ontologies aimed at the pediatric public and the health area.

Palavras-chave: Oncologia pediátrica; Ontologia; Sistema Nervoso Central.

Nome do projeto: Proposta de ontologia de domínio de pacientes oncopediátricos visando à aplicação em sistemas inteligentes

Links: <https://protege.stanford.edu/>

1. CONTEXTO E MOTIVAÇÃO

Embora sejam a primeira causa de morte por doença em crianças e adolescentes no Brasil (1), os tumores pediátricos são considerados raros quando comparados aos adultos, representando cerca de 3% dos tumores malignos mundiais. Em razão disso, os estudos em oncologia pediátrica ainda são pouco explorados e são insuficientes os centros especializados no cuidado à criança e

ao adolescente com câncer no Brasil. Por consequência, o paciente oncopediátrico normalmente precisa se deslocar às capitais brasileiras para dar andamento ao seu tratamento clínico, o que favorece o cenário de dispersão de registros clínicos, que traz desafios para a atuação dos profissionais e para a padronização e integração das informações (2). Desta forma, a proposição de uma ontologia de domínio é relevante para conduzir a gestão de dados e do conhecimento, sobretudo ao se considerar o impacto que estas informações apresentam para a qualidade dos registros em saúde e para a tomada de decisão baseada em evidências.

1.1 Objetivo

O objetivo deste resumo é relatar os processos adotados para a proposição de uma ontologia de domínio de pacientes oncopediátricos visando à aplicação em Sistemas Inteligentes.

2. ATIVIDADES PRINCIPAIS

As atividades realizadas incluem: 1) Mapear e pré-processar variáveis de fichas clínicas de um estudo epidemiológico em oncopediatria; 2) Desenvolver um dicionário de dados com base em parâmetros internacionais; 3) Realizar revisão de escopo e mapear ontologias para reuso, bem como consultar especialistas e fontes consolidadas no domínio da oncopediatria; 4) Descrever o fluxo, objetivos e hipóteses de um banco de dados para estudos epidemiológicos com pacientes oncopediátricos; 5) Especificar as relações terminológicas e conceituais entre os domínios oncopediátricos; e 6) Propor uma ontologia de domínio para pacientes oncopediátricos com foco no cenário dos tumores de Sistema Nervoso Central (SNC).

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A Figura 1 apresenta o método de modelagem escolhido para a construção da ontologia. As atividades principais descritas na seção anterior serão apresentar a seguir.

Figura 1 – Método de modelagem



Fonte: adaptado de Suárez-Figueroa et. al (2012)

As atividades 1 e 2 foram realizadas através das fichas clínicas do estudo “Acompanhar para transformar: um olhar integrado para o câncer infantojuvenil a longo prazo, considerando aspectos clínicos, psicológicos e sociais”, CAAE 52044221.8.1001.5327, disponibilizadas pelo Instituto do Câncer Infantil (ICI). As fichas foram enviadas em branco, sem quaisquer dados de pacientes. Por não envolver seres humanos, este estudo não é apreciado em Comitê de Ética Pesquisa, sendo aprovado pela Comissão de Pesquisa da Universidade Federal de Ciências da Saúde de Porto Alegre, nº 201/2022, e pelo Comitê de Projetos de Pesquisa do ICI, nº 2022-0108. As ontologias para reuso estão sendo pesquisadas no repositório Bioportal e através de uma revisão de escopo. Para a construção das atividades 5 e 6, estão em uso a NeOn Methodology (3) e a ferramenta

Protégé. A escolha por abordar apenas tumores do SNC decorreu da necessidade de restringir o escopo de abrangência da ontologia. Estes tumores estão entre as neoplasias mais prevalentes na pediatria e são o grupo de tumores sólidos mais frequentes em crianças, representando 20% de todas as neoplasias na infância (4). Em (5) encontram-se mais detalhes sobre essa pesquisa.

4. DESAFIOS E APRENDIZADOS

Os desafios referem-se a escassos estudos recentes direcionados à oncopediatria, dificultando a identificação dos métodos mais apropriados para a modelagem da ontologia. Nesse sentido, precisou-se buscar informações em estudos voltados para a oncologia adulta e, ainda, para o campo da saúde de forma ampla. Ademais, destaca-se que o domínio da oncopediatria sofre alterações com o avanço científico, de forma que algumas das ontologias encontradas não contemplam todas as atualizações do domínio. No futuro, espera-se testar a ontologia finalizada e que o aprendizado adquirido favoreça a ampliação da ontologia para todo o cenário da oncologia pediátrica, beneficiando seu uso para as áreas de saúde, ensino e políticas públicas.

5. AGRADECIMENTOS

Fundação de Amparo à Pesquisa do Rio Grande do Sul, Instituto do Câncer Infantil e Instituto Ronald McDonald.

6. REFERÊNCIAS

1. Instituto Nacional de Câncer - INCA. Câncer infantojuvenil [Internet]. 2018. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-infantojuvenil>. Acesso em 16/07/2024.
2. Teixeira LMD, Almeida MB. Aspectos ontológicos e epistemológicos em terminologias clínicas: em busca de interoperabilidade semântica no ambiente clínico. *Encontros Bibli*. 2019. May 6;24(55):1–21. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2019.e57996>. Acesso em 16/07/2024.
3. Suárez-Figueroa MC, Gómez-Pérez A, Motta E, Gangemi A. *Ontology Engineering in a Networked World*. Springer eBooks. Springer Nature; 2012. Disponível em: <https://link.springer.com/book/10.1007/978-3-642-24794-1>. Acesso em 16/07/2024.
4. Instituto Nacional de Câncer - INCA. Tumores do sistema nervoso central [Internet]. 2023. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/infantojuvenil/especificos/tumores-do-sistema-nervoso-central>. Acesso em 16/07/2024
5. Pereira M, Matsuda M, Lanferdini L, Sinigaglia M, Cazella S. Domain Ontology Proposition for Central Nervous System Tumors in Pediatric Patients. In: *Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS 2024)*. v. 3565. p. 1-8 (artigo aceito para publicação)

Predição de desfechos em saúde em pequenas áreas

Crizian S. Gomes¹, Thania M.T. R. Faria¹, Regina T. I. Bernal¹, Laís S. M. Cardoso¹,
Larissa F. Araújo², Samuel N. Alves¹, Beatriz R. G. Barbosa¹, Fabiano A. Célio¹, Juliana B. Souza¹,
Marcos A. Gonçalves¹, Jussara M. Almeida¹, Airton T. Stein³, Jorge G. V. Meléndez¹,
Deborah C. Malta¹

¹Universidade Federal de Minas Gerais,

²Universidade Federal do Ceará,

³Universidade Federal de Ciências da Saúde de Porto Alegre

Contacting authors: criziansaar@gmail.com

Resumo. Este estudo procurou estimar a prevalência e a distribuição espacial do consumo recomendado de frutas e hortaliças (FH) e a prática de atividade física no tempo livre (AFTL) em pequenas áreas da cidade de Belo Horizonte, utilizando métodos de estimação direta para pequenas áreas e algoritmos de aprendizado de máquina não supervisionado. Utilizaram-se os dados do Vigitel - Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico, de Belo Horizonte, MG, georreferenciados. Aplicou-se o algoritmo de agrupamento K-Means para identificar 9 clusters de vulnerabilidade, derivados do Índice de Vulnerabilidade da Saúde (IVS), que foram considerados mais homogêneos (isto é, com menor variabilidade) quando comparados aos 4 níveis originais de IVS. As prevalências dos desfechos foram estimadas por clusters e considerando os períodos de 2009 a 2013 (T1) e 2014 a 2018 (T2). Para cada desfecho, foram identificados os clusters com maiores prevalências bem como sua evolução nos dois períodos e avaliado a precisão das estimativas dos indicadores.

Abstract. This study aimed to estimate the prevalence and spatial distribution of recommended intake of fruits and vegetables (FV) and leisure-time physical activity (LTPA) in small areas of Belo Horizonte, using direct small area estimation methods and unsupervised machine learning algorithms. Georeferenced data from the Surveillance System for Risk and Protective Factors for Chronic Diseases by Telephone Survey of Belo Horizonte, MG, were used. The K-Means clustering algorithm was applied to identify 9 vulnerability clusters derived from the Health Vulnerability Index (HVI), which were considered more homogeneous (i.e., with lower variability) when compared to the original 4 levels of HVI. The prevalence of outcomes was calculated by clusters for the periods from 2009 to 2013 (T1) and 2014 to 2018 (T2). For each outcome, we identified the clusters with higher prevalences and their evolution over the two periods and the precision of the indicator estimates was assessed.

Palavras-chave: Análise de pequenas áreas; Fatores de risco; Algoritmos de aprendizado.

Nome do projeto: O uso de modelos preditivos de inteligência artificial na prevenção de doenças crônicas não transmissíveis (Deborah Carvalho Malta)

1. CONTEXTO E MOTIVAÇÃO

As Doenças crônicas não transmissíveis (DCNT) e seus fatores de risco têm sido monitorados no país por meio de sistemas de informação e de grandes inquéritos nacionais, que são importantes para gerar evidências que orientem políticas públicas, programas e intervenções em saúde (1). Esses inquéritos, contudo, em vista dos modelos de amostragem empregados, permitem que sejam analisados indicadores segundo grandes unidades de análise, como o país e suas regiões e Unidades Federadas, desagregando-se a informação a até, no máximo, as capitais. Desse modo, a investigação de indicadores de saúde fica limitada à análise de estimativas para grandes áreas, perdendo o grau de detalhamento dos atributos e mascarando as desigualdades existentes devido à heterogeneidade dos territórios maiores (2). Dessa forma, torna-se necessário avançar no conhecimento das distribuições de eventos de saúde em pequenas áreas, a fim de apoiar políticas públicas para redução de desigualdades em saúde.

Os métodos de estimação em pequenas áreas vêm sendo desenvolvidos para gerar estimativas com boa precisão para níveis mais desagregados, como municípios, distritos sanitários, setores censitários e outros arranjos de agregação destas unidades (3). No entanto, vários são os desafios para estimar indicadores em saúde em pequenas áreas, como o número amostral reduzido ou a ausência de observações em algumas áreas, principalmente entre as mais vulneráveis, o que pode levar a menor precisão das estimativas.

Nesse sentido, o uso da inteligência artificial apresenta-se como uma ferramenta importante para melhorar a precisão das estimativas dos indicadores de saúde em pequenas áreas, permitindo assim avançar no monitoramento dos fatores de risco e de proteção à saúde, bem como de desfechos de saúde, como as DCNT, tornando mais otimizada e eficiente a vigilância epidemiológica (4).

1.1 Objetivo

Estimar a prevalência e a distribuição espacial do consumo recomendado de frutas e hortaliças (FH) e da prática de atividade física no tempo livre (AFTL) em pequenas áreas de Belo Horizonte, utilizando métodos de estimação direta para pequenas áreas e algoritmos de classificação não supervisionada.

2. ATIVIDADES PRINCIPAIS

As atividades realizadas foram: organização e georreferenciamento do banco de dados do Vigitel; estimativas dos desfechos em grupos de vulnerabilidade utilizando método direto para estimativa em pequenas áreas; análise do padrão de ocorrência espacial das estimativas dos desfechos.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Trata-se de estudo ecológico que utilizou dados do Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (Vigitel) referente a Belo Horizonte, MG, entre 2009 e 2018. A base de dados do Vigitel foi georreferenciada utilizando os endereços dos entrevistados fornecidos por empresas telefônicas. Posterior vinculação com o Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) do Censo 2010 permitiu a associação de cada número de telefone a um setor censitário. Utilizou-se dois desfechos neste estudo: consumo recomendado de frutas e hortaliças (FH)¹ e prática de atividade física no tempo livre (AFTL)².

Para este estudo, utilizou-se como “pequenas áreas” 9 clusters de setores agrupados em níveis de vulnerabilidade, derivados do Índice de Vulnerabilidade da Saúde (IVS). O IVS é um indicador composto que associa variáveis socioeconômicas e ambientais e permite a análise das características de grupos populacionais residentes nos setores censitários. No cálculo original do IVS, os 3.830 setores de Belo Horizonte classificados foram agrupados em quatro grupos de risco à saúde denominados de “baixo risco” (BA) (1.330 setores), “médio risco” (MD) (1.460 setores), “elevado risco” (EL) (737 setores) e “muito elevado risco” (ML) (303 setores) (5). Neste estudo, os grupos originais foram subdivididos em 9 clusters utilizando-se um algoritmo de clustering hierárquico (6). Esta abordagem de reagrupamento foi proposta em outro trabalho (7) visando reduzir a variabilidade dos valores de IVS associados a cada cluster e, indiretamente, melhorar as estimativas de prevalência de desfechos por cluster. As prevalências dos desfechos e os respectivos intervalos de confiança de 95% (IC95%) foram calculados para cada cluster utilizando método direto para estimativa em pequenas áreas, que considera somente as variáveis do desenho da amostra (2), para dois períodos (T1: 2009 à 2013; T2: 2014 à 2018). Também foi analisada a autocorrelação espacial, com base no Índice Local de Moran (LISA). Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais sob parecer nº 6.538.883.

¹ Consumo de FH em pelo menos cinco dias da semana e quando a soma das porções consumidas totalizava 5 ou mais por dia.

² Prática semanal de pelo menos 150 minutos de atividade física de intensidade moderada ou pelo menos 75 minutos de atividade física de intensidade vigorosa.

Entre 2009 e 2018 foram entrevistados 19.041 indivíduos e, destes, 16.019 foram georreferenciados e fizeram parte do estudo. Ao analisar as prevalências para Belo Horizonte, observa-se aumento tanto no consumo recomendado de FH (27,5% em T1 para 31,7% em T2) quanto na AFTL (34,8% em T1 para 39,9% em T2) entre os períodos. No entanto, ao analisar por clusters, verifica-se grande variabilidade entre eles, sendo as prevalências para ambos os desfechos maiores nos clusters de baixo risco (Tabela 1). Observa-se boa precisão da estimativa de consumo de FH para Belo Horizonte (margem de erro de 1,1 em T1 e de 1,35 em T2). Contudo, ao analisar por cluster, tem-se grande variabilidade das margens de erro (T1: 2,5 (BA-0) a 8,2 (ME-0); T2: 3,0 (BA-0) a 8,8 (ME-0)). Os resultados são semelhantes para o indicador de AF (dados não mostrados). Destaca-se que a variabilidade na margem de erro pode ser explicada pelo tamanho da amostra, visto que os clusters de maior vulnerabilidade apresentam menor número de entrevistas (dados não mostrados).

Clusters	Consumo de frutas e hortaliças		Atividade física no tempo livre	
	2009 à 2013	2014 à 2018	2009 à 2013	2014 à 2018
	% (IC 95%)	% (IC 95%)	% (IC 95%)	% (IC 95%)
BA-0	31,3 (28,8 - 33,8)	35,8 (32,8 - 38,8)	39,2 (36,5 - 42,0)	46,9 (43,7 - 50,1)
BA-1	40,1 (37,0 - 43,2)	45,2 (41,0 - 49,4)	45,5 (42,3 - 48,7)	52,8 (48,6 - 57,0)
MD-0	27,2 (24,7 - 29,7)	33,3 (30,3 - 36,3)	37,4 (34,6 - 40,2)	41,0 (37,9 - 44,1)
MD-1	25,9 (23,8 - 28,1)	27,3 (24,8 - 29,8)	31,2 (28,9 - 33,6)	35,5 (32,8 - 38,3)
EL-0	20,7 (16,4 - 25,0)	25,6 (20,2 - 31,0)	34,1 (28,9 - 39,4)	31,4 (25,7 - 37,2)
EL-1	16,7 (11,2 - 22,2)	22,3 (16,0 - 28,6)	25,6 (18,7 - 32,5)	36,1 (28,3 - 44,0)
EL-2	25,0 (21,5 - 28,4)	30,4 (26,2 - 34,6)	28,4 (24,7 - 32,2)	35,2 (30,8 - 39,5)
ME-0	22,8 (14,6 - 31,1)	22,8 (13,9 - 31,6)	28,4 (19,3 - 37,6)	34,4 (24,2 - 44,6)
ME-1	15,4 (10,9 - 20,0)	24,9 (18,5 - 31,3)	23,7 (18,0 - 29,4)	31,8 (24,7 - 38,9)
Total	27,5 (26,4 - 28,6)	31,7 (30,4 - 33,1)	34,8 (33,6 - 36,0)	39,9 (38,5 - 41,3)

Nota: BA: baixo; MD: médio; EL: elevado; ME: muito elevado riscos. IC95%: intervalo de confiança de 95%.

A análise de dependência espacial indicou a existência de autocorrelação (I de Moran AFTL = 0,747; I de Moran FH = 0,746) e de agrupamentos espaciais “alto-alto” e “baixo-baixo” ($p < 0,05$), que sugeriram maior concentração de baixas prevalências de prática de AFTL e de consumo de FH nos locais mais periféricos da cidade e maiores prevalências nas regiões centrais.

4. DESAFIOS E APRENDIZADOS

Destaca-se a complexidade de estimar os desfechos em áreas com reduzido número amostral (coincidente com áreas de maior vulnerabilidade), consequência de aspectos metodológicos inerentes às técnicas de amostragem dos inquéritos de saúde nacionais. Dessa forma, torna-se necessário avançar em métodos para produzir estimativas válidas e com boa precisão, a fim de se obter um desenho de política, programa e intervenção em saúde que considere as particularidades de diferentes territórios e subgrupos populacionais, garantindo a equidade prevista no Sistema Único de Saúde.

5. REFERÊNCIAS

As referências foram condensadas no link a seguir: <https://docs.google.com/document/d/1nHPjVIAPW1I5hxMntWZQ-2UsPAimamOkJ-rtmgDpbIE/edit>

Predição do uso de serviços de Urgência e Emergência em Pelotas-RS, 2021-2023

Felipe M. Delpino¹, Bruno P. Nunes²

^{1,2}Universidade Federal de Pelotas, Rio Grande do Sul, RS, Brasil.

fmdsocial@outlook.com, nunesbp@gmail.com

Resumo. Os serviços de urgência e emergência são essenciais para a Rede de Urgência e Emergência no Brasil. Não obstante, enfrentam, historicamente, superlotação e uso inadequado. Este trabalho teve como objetivo desenvolver modelos de Machine Learning para prever a utilização de serviços em uma cidade do Sul do Brasil, categorizando-os em serviços públicos e privados. Foram testados três algoritmos (XGBoost, LightGBM e Catboost) com dados sociodemográficos, comportamentais e de saúde dos participantes. O estudo EAI PELOTAS foi aprovado pelo comitê de ética em pesquisa da Faculdade de Enfermagem da Universidade Federal de Pelotas, sob CAAE número 39096720.0.0000.5317. As análises foram realizadas na linguagem de programação Python. Utilizou-se variáveis da linha de base do estudo para prever o uso dos serviços até um ano após a primeira entrevista. A amostra analítica foi composta por 3.460 pessoas. Dessas, 18,4% usaram serviços públicos (principalmente o Pronto-Socorro municipal) e 12,7% utilizaram serviços privados. O modelo Catboost apresentou os melhores resultados de AUC-ROC (0,682 e 0,789 para serviços públicos e privados, respectivamente). Os resultados preliminares sinalizam o potencial para uso dos modelos na tomada de decisões e a complexidade do desenvolvimento de modelos preditivos equitativos e de alta performance para a predição de serviços de urgência e emergência.

Palavras-chave: Machine Learning; Urgência; Emergência.

Nome do projeto: Multimorbidade e procura por serviços de urgência e emergência em Pelotas-RS: predição a partir de análises de inteligência artificial (EAI PELOTAS?).

Abstract. Urgent and emergency services are essential to the Urgent and Emergency Care Network in Brazil. However, they have historically faced overcrowding and inadequate use. The aim of this study was to develop Machine Learning models to predict the use of services in a city in southern Brazil, categorizing them into public and private services. Three algorithms (XGBoost, LightGBM, and Catboost) were tested with participants' sociodemographic, behavioral, and health data. The EAI PELOTAS study was approved by the research ethics committee of the Faculty of Nursing of the Federal University of Pelotas, under CAAE number 39096720.0.0000.5317. The analyses were carried out in the Python programming language. Variables from the study's baseline were used to predict the use of services up to one year after the first interview. The analytical sample consisted of 3,460 people. Of these, 18.4% used public services (mainly the municipal emergency room), and 12.7% used private services. The Catboost model showed the best AUC-ROC results (0.682 and 0.789 for public and private services, respectively). The preliminary results indicate the potential for using the models in decision-making and the complexity of developing equitable, high-performance predictive models for predicting urgent and emergency services.

Keywords: Machine Learning; Urgency; Emergency.

Project name: Multimorbidity and demand for urgent and emergency services in Pelotas-RS: prediction based on artificial intelligence analysis (EAI PELOTAS?).

1. CONTEXTO E MOTIVAÇÃO

No Brasil, a Rede de Urgência e Emergência (RUE) foi criada para melhorar a atenção em saúde. Os serviços de urgência e emergência enfrentam desafios como superlotação e uso inadequado, muitas

vezes relacionados com o tratamento de doenças crônicas ¹. Estudos mostram que multimorbidade e fatores sociodemográficos influenciam a procura por esses serviços ². Tecnologias como machine learning (ML) são promissoras para prever a utilização desses serviços, o que pode auxiliar na tomada de decisões através da priorização de casos, direcionamento de pacientes para outros níveis de atenção e melhoria do cuidado.

1.1 Objetivo

Desenvolver dois modelos de ML para prever a utilização de serviços de urgência e emergência, um para serviços públicos e outro para privados, entre moradores de uma cidade do Sul do Brasil.

2. ATIVIDADES PRINCIPAIS

Os dados foram coletados no estudo "Multimorbidade e procura por serviços de urgência e emergência em Pelotas-RS". Foram incluídos moradores com 18 anos ou mais. A amostra inicial foi de 5722 indivíduos, com 3460 acompanhados um ano depois. Variáveis sociodemográficas, comportamentais e de saúde foram utilizadas como preditoras.

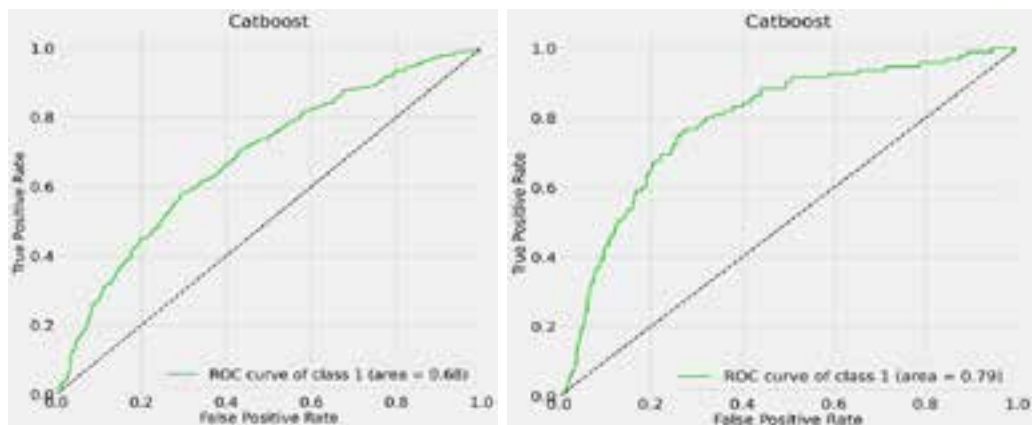
3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Trata-se de um estudo longitudinal de base populacional realizado em residências de adultos e idosos da área urbana de Pelotas, Rio Grande do Sul, Brasil. Foram incluídos moradores com idade igual ou superior a 18 anos. O cálculo do tamanho de amostra envolveu diferentes desfechos do estudo, resultando em 5722 entrevistas na linha de base e 3460 no acompanhamento (amostra analítica) ³.

Para a predição da utilização dos serviços de urgência e emergência no período de 12 meses subsequentes à entrevista, foi utilizada a variável: "nos últimos 12 meses, quantas vezes o senhor(a) procurou algum serviço de urgência e emergência como pronto socorro e Unidade de Pronto Atendimento?". As variáveis preditoras incluíram informações sociodemográficas, comportamentais e de saúde, totalizando 40 variáveis. O desenvolvimento dos modelos foi feito na linguagem de programação Python. O estudo foi aprovado pelo comitê de ética em pesquisa da Faculdade de Enfermagem da Universidade Federal de Pelotas (CAAE 39096720.0.0000.5317).

Dos 3460 participantes avaliados, 18,4% usaram serviços públicos (principalmente o Pronto Socorro municipal) e 12,7% utilizaram serviços privados. Os algoritmos XGBoost, LightGBM e Catboost foram testados pois apresentam bom desempenho para dados tabulares. A avaliação dos modelos foi realizada por meio da área sob a curva ROC (AUC) e outras métricas, como brier score, precisão, recall e F1-Score. O Catboost apresentou o melhor desempenho, com AUC de 0.682 para serviços públicos e 0.789 para serviços privados (Figura 1). LightGBM apresentou AUC de 0.63 para serviços públicos e 0.768 para privados e XGBoost 0.624 para públicos e 0.734 para privados. A precisão do modelo Catboost para os serviços públicos foi de 0.578 e para os privados foi de 0.143; o Recall para os públicos foi de 0.365 e para os privados 0.011. As principais variáveis preditoras para os modelos públicos foram utilização do serviço de urgência na linha de base, idade, posse de plano de saúde e busca por serviços de saúde; para os privados foram posse de plano de saúde, escolaridade, idade e tabagismo.

Figura 1 – Curva ROC mostrando o valor de AUC do modelo Catboost de classificação da utilização de serviços públicos (à esquerda) e privados (à direita) de urgência e emergência em Pelotas, Rio Grande do Sul



Fonte: Criado pelo próprio autor

4. DESAFIOS E APRENDIZADOS

Os resultados mostraram potencial para a predição de serviços de urgência e emergência, tanto públicos quanto privados, podendo ser implementados no contexto clínico e colaborando na tomada de decisão. No entanto, são necessários mais dados para o aprimoramento dos modelos, especialmente no contexto privado, além da exploração de novas variáveis e análise de subgrupos. Com essas melhorias, torna-se possível utilizar modelos capazes de identificar usuários, tanto do setor público quanto privado, com necessidade de utilização dos serviços. Por fim, os modelos precisam ser pensados em termos de equidade, garantindo que todos possam se beneficiar de forma justa.

5. REFERÊNCIAS

1. Carret MLV, Fassa AG, Domingues MR. Prevalência e fatores associados ao uso inadequado do serviço de emergência: Uma revisão sistemática da literatura. *Cad Saude Publica*. 2009;25(1):7– 28.
2. Gaulin M, Simard M, Candas B, Lesage A, Sirois C. Combined impacts of multimorbidity and mental disorders on frequent emergency department visits: A retrospective cohort study in Quebec, Canada. *CMAJ*. 2019;191(26).
3. Mendes Delpino FI, Munhoz Figueiredo LI, Krolow Costa ÂI, Carreno II, Nascimento da Silva LI, Duarte Flores AI, et al. Emergency department use and Artificial Intelligence in Pelotas: design and baseline results. *Revista Brasileira de Epidemiologia* [Internet]. 2023 Mar 10 [cited 2023 Mar 12];26:e230021. Available from: <http://www.scielo.br/j/rbepid/a/kbtnFrZmCzJtSHWMNbWsB6J/?lang=en>

Prescrições Médicas Apoiadas por LLMs: Um Estudo Comparativo entre ChatGPT 4-o e LLaMa 3 8B

Vinícius Araújo Basílio¹, Antônio Pereira², Davi Reis², Isaias José Ramos de Oliveira¹, Leonardo Rocha², Zilma Silveira Nogueira Reis¹

¹ Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

² Universidade Federal de São João Del Rei, Belo Horizonte, MG, Brasil.

forblood@ufmg.br, antoniopereira@aluno.ufsj.edu.br, davireisjesus@aluno.ufsj.edu.br,
isaias-ramos@ufmg.br, lcrocha@ufsj.edu.br, zilma@ufmg.br

Resumo. Este estudo comparou a qualidade das orientações ao paciente geradas por dois Grandes Modelos de Linguagem (Large Language Models - LLMs): um proprietário, GPT 4, e outro aberto, Llama 3 8B, para prescrições médicas no sistema PEC eSUS APS. Cada LLM gerou 104 prescrições de orientação sobre casos específicos. Essas foram comparadas às orientações cuidadosamente prescritas por médicos utilizando a similaridade (i.e. similaridade de cosseno) entre as suas respectivas representações vetoriais (i.e. word embeddings). Os resultados mostraram que ambos os modelos apresentaram alto índice de similaridade com as orientações humanas, com médias globais de 0,9142 para o LLaMa e 0,9131 para o GPT 4. Nos cinco melhores casos, o LLaMa obteve 0,9568 e o GPT 4 0,9579; nos cinco piores, 0,7943 e 0,7947, respectivamente. A análise estatística revelou estatística t de 0,2043 e $p = 0,8383$, indicando que as diferenças entre os modelos não foram estatisticamente significativas. A avaliação qualitativa apontou que o GPT 4 gerou respostas mais detalhadas e articuladas, enquanto o Llama 3 apresentou falhas em instruções críticas. Considerando a aplicação do trabalho no PEC eSUS APS, que é um sistema governamental, é essencial o uso de uma LLM aberta.

Abstract. This study compared the quality of patient guidance generated by two Large Language Models (LLMs): a proprietary model, GPT-4, and an open-source model, Llama 3 8B, for medical prescriptions in the PEC eSUS APS system. Each LLM generated 104 prescriptions of guidance on specific cases. These were compared to the carefully prescribed guidance by physicians using similarity metrics (i.e., cosine similarity) between their respective vector representations (i.e., word embeddings). The results showed that both models achieved a high degree of similarity with the human-guided instructions, with overall averages of 0.9142 for Llama and 0.9131 for GPT-4. In the five best cases, LLaMa scored 0.9568 and GPT-4 0.9579; in the five worst cases, 0.7943 and 0.7947, respectively. Statistical analysis revealed a t-statistic of 0.2043 and $p = 0.8383$, indicating that the differences between the models were not statistically significant. Qualitative evaluation indicated that GPT-4 generated more detailed and articulate responses, while Llama 3 exhibited flaws in critical instructions. Considering the application of this work in the PEC eSUS APS, which is a government system, the use of an open-source LLM is essential.

Palavras-chave: Inteligência Artificial, Prescrição Médica, PEC eSUS APS, Representações Vetoriais, Similaridade de Textos.

Nome do projeto: Desenvolvimento e avaliação de um sistema inteligente de geração de orientações para a prescrição de medicamentos segura, acessível e adaptada a distintos contextos culturais (IA-Polis)

1. CONTEXTO E MOTIVAÇÃO

No Brasil, é essencial que as orientações para prescrições médicas sejam claras, devido à diversidade linguística e cultural do país, que pode afetar a comunicação entre médicos e pacientes. Orientações não qualificadas em prescrição de medicamentos podem levar ao seu uso incorreto, gerando custos adicionais para o sistema de saúde e colocando em risco a segurança dos pacientes

(1,2). O presente trabalho está inserido no contexto do projeto IA Polis, o qual planeja integrar Grandes Modelos de Linguagem (Large Language Models - LLMs) ao Sistema de Prontuário Eletrônico do Cidadão para Atenção Primária à Saúde, para dar clareza às orientações à pacientes ao gerar prescrições considerando sua personalização e adaptação cultural.

1.1 Objetivo

O objetivo principal deste trabalho é avaliar a qualidade das informações fornecidas por dois LLMs, um aberto (i.e. LLaMa 3) e outro proprietário (GPT 4), compará-las com orientações para uso de medicamentos e prescrições médicas validadas por profissionais de saúde e analisar pontos positivos e negativos dos resultados. Além disso, busca-se identificar áreas de melhoria nas respostas dos modelos de linguagem (LLMs) para aprimorar a Interface de Programação de Aplicações (API) do projeto IA Polis. A meta é otimizar o processo de prescrição digital no sistema PEC eSUS APS, com prescrições seguras e adaptadas ao contexto cultural e linguístico brasileiro.

2. ATIVIDADES PRINCIPAIS

Durante o desenvolvimento deste projeto, uma base de dados simulada foi criada com 104 prescrições de princípios ativos distintos, seguindo os padrões do PEC eSUS APS. Boas práticas de prescrição médica foram definidas (3) e utilizadas na engenharia de prompt para interações com os modelos. Em maio de 2024, o modelo ChatGPT 4-o gerou prescrições, enquanto o modelo de código aberto LLaMa 3 foi testado em julho de 2024, atendendo aos requisitos de segurança e privacidade. As prescrições geradas por ambos os modelos foram comparadas às orientações de médicos usando a similaridade (i.e. similaridade de cosseno) entre as suas respectivas representações vetoriais (i.e. word embeddings). O aluno de Iniciação Científica realizou análises estatísticas e qualitativas dos cinco melhores e piores casos, considerando precisão das dosagens, clareza das instruções e adequação cultural. Foi aplicado um teste t de Student para verificar diferenças significativas entre os desempenhos dos modelos.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A metodologia consistiu na coleta das prescrições geradas pelas LLMs e foram realizadas avaliações qualitativas e quantitativas dessas respostas. Na quantitativa, as prescrições, tanto as geradas pelos LLMs quanto as geradas por humanos, foram vetorialmente representadas por meio de word embeddings (i.e. Open AI text-embedding-3-large?). Em seguida, calculou-se a similaridade entre os vetores resultantes de cada prescrição gerada pelos LLMs e os vetores das prescrições humanas. Essa similaridade é definida em uma escala de 0 a 1, onde 0 indica textos totalmente diferentes e 1 indica textos idênticos. Assim, quanto mais próximo de 1 o resultado, melhor. O resultado dessa avaliação é apresentada na Tabela 1, acompanhada da análise estatística. Podemos observar que ambos os modelos, LLaMa 3 e GPT 4, obtiveram desempenhos de similaridade elevados e comparáveis quando comparados às prescrições médicas humanas. As médias gerais de similaridade foram de 0,9142 para o LLaMa e 0,9131 para o ChatGPT 4-o. Analisando os cinco melhores resultados, o LLaMa atingiu 0,9568 e o ChatGPT 4-o alcançou 0,9579. Nos cinco piores resultados, o LLaMa teve 0,7943 e o ChatGPT 4-o, 0,7947. A análise estatística revelou não haver diferença significativa entre as médias de similaridade das prescrições geradas pelos modelos. Com uma estatística t de 0,2043 e p igual a 0,8383 ($p > 0,05$), conclui-se que as variações observadas entre os modelos não são estatisticamente significantes, indicando que ambos produzem prescrições com níveis de similaridade comparáveis às prescrições médicas humanas, por esse método.

Modelo	Média de Similaridade	Melhores 5 Resultados	Piores 5 Resultados	Estatística t	p - valor
LLaMa 3 8B	0,9142	0,9568	0,7943	0,2043	0,8383
ChatGPT 4-o	0,9131	0,9579	0,7947	0,2043	0,8383

No entanto, a similaridade não é o único parâmetro de qualidade, sendo a avaliação humana essencial para analisar resultados, considerando nuances e contextos que algoritmos podem não captar. A técnica de similaridade entre representações vetoriais cria uma matriz que identifica proximidade semântica entre frases, como “tomar a pomada” e “passar a pomada”, mesmo não sendo literalmente idênticas. Essa técnica mede similaridade de significados, independentemente da variação na escrita. Assim, a percepção e o julgamento humano são fundamentais para avaliar clareza, precisão e adequação das respostas, complementando as métricas quantitativas.

Na análise qualitativa dos resultados extremos, observou-se que a qualidade dos textos gerados pelo ChatGPT-4 é mais próxima do ideal. Nos cinco melhores casos, o modelo conseguiu gerar prescrições médicas detalhadas com instruções precisas e completas sobre dosagem, frequência e administração, semelhantes às fornecidas por médicos, embora tenham faltado detalhes sobre armazenamento, uso pessoal e suspensão do tratamento. Nos cinco piores casos, o modelo apresentou inconsistências críticas na dosagem e falta de orientações importantes. Quanto ao LLaMa, nos cinco melhores casos, os textos incluíram informações essenciais para a administração segura e eficaz dos medicamentos, apesar da ausência de alguns detalhes adicionais. Nos cinco piores casos, a LLM apresentou deficiências em fornecer instruções detalhadas e precisas, com recomendações de dosagem incorretas ou incompletas e falta de orientações sobre armazenamento adequado e não compartilhamento de medicamentos.

4. DESAFIOS E APRENDIZADOS

O projeto IA Polis enfrentou desafios significativos, incluindo a compreensão das necessidades do SUS e a adaptação de LLMs para futuras implementações, assegurando a proteção de dados e integração com o Sistema PEC eSUS APS, visando fortalecer a adesão dos cidadãos ao tratamento(2,4). O aluno de Iniciação Científica explorou tecnologias de LLMs e Inteligência Artificial, além do campo da medicina, em uma colaboração multidisciplinar crucial para atender às normas de saúde e necessidades locais. A análise de tecnologias focou em garantir prescrições médicas claras, seguras e éticas. O objetivo final é a aplicação prática dos resultados no PEC eSUS APS, exigindo o uso de uma LLM aberta visando os requisitos de segurança e privacidade. Agradecimentos: o estudo foi realizado com recursos do CNPq-Inova Fiocruz, Fundação Bill & Melinda Gates. A autora ZNSR é bolsista CNPq 305837/2021-4.

5. REFERÊNCIAS

1. [Wrzosek N, Zimmermann A, Balwicki Ł. A Survey of Patients' Opinions and Preferences on the Use of E- Prescriptions in Poland. Int J Environ Res Public Health. 16 de setembro de 2021;18\(18\):9769.](#)
2. He K, Mao R, Lin Q, Ruan Y, Lan X, Feng M, et al. A Survey of Large Language Models for Healthcare: from Data, Technology, and Applications to Accountability and Ethics [Internet]. arXiv; 2024 [citado 2 de julho de 2024]. Disponível em: <http://arxiv.org/abs/2310.05694>
3. protocols.io [Internet]. [citado 9 de junho de 2024]. Safe directions for taking medicines. Disponível em: <https://www.protocols.io/view/safe-directions-for-take-medicines-dckh2ut6>
4. [Klepser D, Lanham A, Cochran G. Electronic prescriptions: opportunities and challenges for the patient and pharmacist. Adv Health Care Technol. fevereiro de 2016;1.](#)

Protótipo de Ferramenta para o Auxílio de Exercícios Fisioterapêuticos

Enzo Guarnieri¹, Mário Olímpio de Menezes¹

¹ Faculdade de Computação e Informática - FCI Universidade
Presbiteriana Mackenzie São Paulo, SP, Brasil

enzo.guarnieri@mackenzista.com.br, mario.menezes@mackenzie.br

Resumo. Lesões podem ser causadas por diversas razões, desde acidentes até condições médicas. Elas podem causar dor significativa, desconforto e podem dificultar o dia-a-dia dos pacientes em suas atividades rotineiras. Em tais casos, o tratamento com fisioterapia pode ser uma maneira efetiva de lidar com esses sintomas e de adquirir funcionalidade física completa. Em alguns casos, a prática de exercícios específicos de fisioterapia pode se estender por longos períodos, e após o início em uma clínica especializada, deve ser realizada pelo paciente em sua residência ou outro local de sua escolha, como academia, etc. Nesta situação, a ausência do profissional fisioterapeuta pode acarretar a execução errada dos movimentos, comprometendo sua eficácia e até se tornar prejudicial ao paciente. Neste trabalho apresentamos a proposta de uma ferramenta que auxilie o paciente na prática dos exercícios de fisioterapia prescritos pelo profissional em sua residência, utilizando visão computacional e inteligência artificial (aprendizado de máquina), de modo que a execução seja correta e traga benefícios ao paciente. Os resultados iniciais incluem a captura de pontos-chaves no movimento do paciente, que serão utilizados para o treinamento de um modelo de Machine Learning para indicação de possíveis erros na realização dos exercícios.

Palavras-chave: Fisioterapia; Reabilitação; Visão Computacional; Aprendizado de Máquina

Abstract. Injuries can be caused by a variety of reasons, from accidents to medical conditions. They can cause significant pain, discomfort, and make it difficult for patients to carry out their daily activities. In these cases, physiotherapy treatment can be an effective way to deal with these symptoms and regain full physical functionality. In some cases, the practice of specific physiotherapy exercises can last for long periods, and after starting in a specialized clinic, they must be performed by the patient at home or in another place of their choice, such as a gym, etc. The absence of a physiotherapist can result in incorrect execution of the movements, compromising their effectiveness and even becoming harmful to the patient. In this work, we present the proposal of a tool that assists the patient in practicing the physiotherapy exercises prescribed by the professional at home, using computer vision and artificial intelligence (machine learning), so that the execution is correct and brings benefits to the patient. Initial results include the capture of key points in the patient's movement, which will be used to train a Machine Learning model to indicate possible errors in performing the exercises.

Keywords: Physiotherapy; Rehabilitation; Computer Vision, Machine Learning

Nome do projeto: Explorações de técnicas e tecnologias de Inteligência Artificial, Ciência de Dados e Aprendizado de Máquina em domínios de interesse.

1. CONTEXTO E MOTIVAÇÃO

Lesões são mudanças anormais nos tecidos do corpo, advindas de injúrias ou doenças. Podem afetar quaisquer partes do corpo. Por exemplo, pacientes com lesões na coluna podem experimentar dificuldades no caminhar, para se inclinar ou levantar objetos. Pacientes com lesões musculoesqueléticas podem experimentar dor, rigidez e redução de amplitude de movimento. O tratamento fisioterápico atua no processo de recuperação de lesões musculares, posturais, fortalecimento muscular após eventos debilitadoras, como cirurgias, imobilizações prolongadas, etc. A fisioterapia também envolve a educação dos pacientes acerca de sua condição e também

equipando-os com ferramentas e estratégias para gerenciar os sintomas. Muitas vezes, o tratamento é iniciado em clínicas especializadas e depois o paciente continua a prática de alguns exercícios em seu domicílio ou outro local que não a clínica, de modo que já não conta com a presença do profissional.

O problema que esse projeto busca amenizar é a adesão ao exercício domiciliar na reabilitação, no contexto da fisioterapia e especialmente, sua execução correta. Segundo estimativas a taxa de não adesão pode chegar até 50% dos pacientes, o que pode ocasionar em um efeito prejudicial nos resultados clínicos. Entre os principais motivos para a não adesão está a percepção de barreiras do paciente, que muitas vezes acaba não aderindo de maneira satisfatória por motivos de se esquecer de fazer os exercícios, não ter tempo, além de falta de motivação (1).

1.1 Objetivo

O objetivo desse projeto será a produção de um protótipo de uma ferramenta que empregue visão computacional para o auxílio em exercícios de reabilitação com fisioterapia, visando a execução correta dos exercícios recomendados.

Pensando nisso foram propostos os seguintes objetivos específicos: 1º OE: Criação do banco de dados e de vídeos que serão usados para treinamento inicial e sistemas de suporte; 2º OE: Analisar, e posteriormente eleger, possíveis ferramentas/bibliotecas para o desenvolvimento do recurso de análise de movimento, de maneira que consiga dar feedback se o movimento está correto e como corrigi-lo; 3º OE: Criação de métricas quantitativas que avaliem o quão bem o exercício foi feito em relação à velocidade, e movimento esperado e desenvolver algum método para o acesso remoto a essas informações, e aperfeiçoar tudo o que foi feito anteriormente. Serão escolhidos entre dois e três exercícios específicos para o desenvolvimento do trabalho.

2. ATIVIDADES PRINCIPAIS

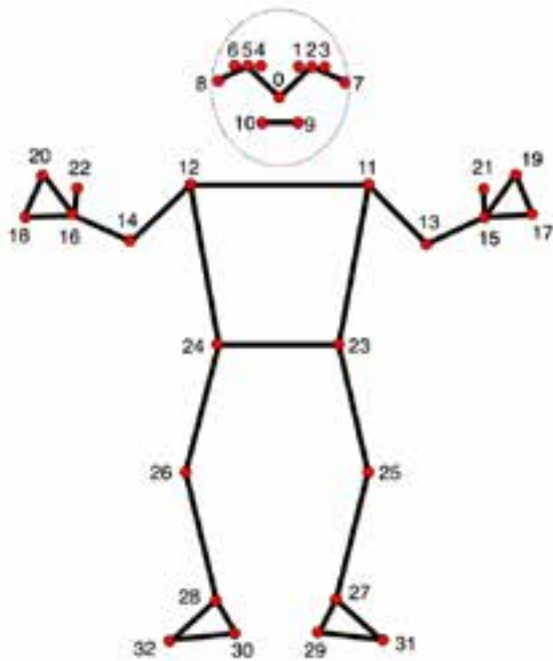
A primeira atividade desenvolvida no projeto foi a busca e identificação de datasets públicos que pudessem ser utilizados para o projeto; essa busca não deu resultado (será necessário aperfeiçoar os critérios de busca). Outra atividade realizada inicialmente foi a identificação de uma ferramenta (biblioteca) para a identificação de pose utilizando visão computacional. A primeira ferramenta identificada foi a OpenPose (2), e a outra foi a MediaPipe, do Google (3,4); a ferramenta MediaPipe foi escolhida em função da facilidade de uso e integração em protótipos de software mobile. Com essa ferramenta identificada, e após a primeiras avaliações, decidimos que ela seria utilizada no projeto.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Tendo em vista que não foram ainda identificados datasets públicos para uso no projeto, o próprio autor (aluno) foi utilizado nas gravações de vídeos para serem utilizados no desenvolvimento do trabalho. A ferramenta escolhida, o MediaPipe, recebe um frame e devolve um conjunto de pontos específicos identificados no corpo humano em três dimensões (X,Y,Z). A Figura 1 apresenta os pontos identificados pelo MediaPipe, 32 pontos ao todo.

Figura 1 – (A) Todos os pontos Identificados pelo MediaPipe; (B) Identificação dos Pontos no Paciente

Com os pontos identificados será desenvolvido um modelo de Machine Learning que indicará se os movimentos estão sendo executados corretamente, além de outras análises para o paciente e para o profissional (fisioterapeuta ou médico).



(A)



(B)

4. DESAFIOS E APRENDIZADOS

O desenvolvimento do trabalho tem proporcionado uma rica experiência de aprendizado das ferramentas de visão computacional e de machine learning, além das sugestões dos profissionais da área de fisioterapia que estão sendo intensificadas nesta fase final do trabalho.

5. REFERÊNCIAS

1. Argent R, Daly A, Caulfield B. Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence? *JMIR MHealth UHealth*. 1o de março de 2018;6(3):e8518.
2. Cao Z, Martinez GH, Simon T, Wei S, Sheikh YA. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans Pattern Anal Mach Intell*. 2019;
3. Lugaresi C, Tang J, Nash H, McClanahan C, Ubowaja E, Hays M, et al. arXiv.org. 2019 [citado 8 de maio de 2024]. MediaPipe: A Framework for Building Perception Pipelines. Disponível em: <https://arxiv.org/abs/1906.08172v1>
4. Kim JW, Choi JY, Ha EJ, Choi JH. Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model. *Appl Sci*. janeiro de 2023;13(4):2700.

RAGECAN: Framework de seleção automática de genes para classificação de amostras com câncer

Arthur Bindá Alves¹, Eduardo Freire Nakamura¹

¹ Universidade Federal do Amazonas, Manaus, AM, Brasil.

aba@icompu.ufam.edu.br, nakamura@icompu.ufam.edu.br

Resumo. O câncer é uma das doenças com o maior número de vítimas por ano, causando em 2020 quase 10 milhões de mortes. Por se tratar de uma doença com alto potencial de disseminação no corpo, o diagnóstico precoce desempenha um papel importante na realização de um tratamento eficaz. O avanço na coleta dos níveis de expressão gênica possibilitou o uso de técnicas de aprendizagem de máquina para a detecção precoce de câncer. No entanto, bases de expressão gênica apresentam um desequilíbrio entre o número de amostras e atributos, conhecido como “maldição da dimensionalidade”. Este problema é caracterizado pelo número maior de atributos em relação a quantidade de amostras, penalizando a aprendizagem de classificadores em razão da alta esparsidade, resultando em diagnósticos imprecisos e pouco confiáveis. Técnicas de redução dimensional podem ser utilizadas resultando em um novo conjunto de dados de dimensão menor, mais simples e representativo. Esta linha de pesquisa integra áreas da computação que buscam viabilizar o uso de classificadores em bases esparsas. Por se tratar de uma área de pesquisa que utiliza dados de expressão gênica para compreender a relação dos genes e o surgimento de câncer, pesquisadores de áreas da saúde também se beneficiam de avanços feitos pela computação na área de redução dimensional. Entretanto, etapas de desenvolvimento e ajuste de parâmetros permanecem um desafio para pesquisadores sem conhecimento específico na área. Neste trabalho, disponibilizamos um framework com código aberto que retira a complexidade de realizar esta etapa, integrando dez técnicas de redução dimensional, três classificadores e interface gráfica que simplifica o processo de interação com o usuário fora da área da computação.

Palavras-chave: Seleção de Atributos, Classificação de Câncer, Expressão Gênica

Nome do projeto: RAGECAN (Redução Automática de Genes para Classificação de Câncer)

Links: <https://github.com/artb/ragecan-app>

1. CONTEXTO E MOTIVAÇÃO

O exame de expressão gênica é uma técnica laboratorial que detecta o nível de atividade de milhares de genes em um determinado tecido. Por meio destes dados é possível analisar genes que estão diretamente associados ao surgimento do câncer. Atualmente, analisar dados de expressão gênica com algoritmos de aprendizagem de máquina é uma abordagem estado-da-arte com alta confiança para detecção precoce do câncer (1). Contudo, bases de dados de expressão gênica são caracterizadas pelo desbalanceamento entre o número de amostras e atributos, afetando ou podendo inviabilizar a aplicabilidade desta abordagem. O problema da maldição da dimensionalidade é alvo de estudos por parte da comunidade de computação para viabilizar o uso de classificadores em bases de dados com esta característica. A solução encontrada para este problema é reduzir a dimensionalidade através de técnicas como a seleção de atributos, que utiliza de métodos estatísticos para avaliar o subconjunto de genes com maior relação com o tipo de câncer analisado. Esta abordagem quando aplicada por pesquisadores de computação visa apenas maximizar o resultado obtido pelo classificador, contudo, quando nos referirmos à dados de expressão gênica, é possível notar que o produto desta etapa possui alto valor para pesquisadores da área da saúde, que possuem compreensão sobre o papel desempenhado por cada gene, potencializando novas descobertas a respeito da doença e principalmente o desenvolvimento de novos tratamentos e medicações que atingem especificamente os genes mais relevantes para um

tipo específico de câncer.

1.1 Objetivo

O objetivo deste trabalho é minimizar as barreiras do conhecimento relacionados à aplicação de métodos de redução dimensional no contexto de dados de expressão gênica. O foco está em aproximar pesquisadores da área da saúde como geneticistas e biomédicos do estado-da-arte de computação, mitigando desafios como o conhecimento específico da área, etapas de desenvolvimento do código e implementação de técnicas de otimização de parâmetros. Buscamos através desta ferramenta fomentar a colaboração entre profissionais de computação e saúde, potencializando avanços na compreensão e detecção precoce da doença.

2. ATIVIDADES PRINCIPAIS

No decorrer do projeto, as atividades desenvolvidas consistiram em:

- Desenvolvimento de um arcabouço automático para seleção dos genes mais relevantes;
- Desenvolvimento de uma interface gráfica que simplifica a interação com o usuário leigo;
- Integração de dez técnicas de redução dimensional com otimização de hiperparâmetros;
- Análise dos resultados obtidos pela solução proposta em comparação com baselines

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Para compor o framework RAGECAN selecionamos três classificadores para obtenção de resultados comparativos do impacto causado pelos métodos de redução dimensional. Os classificadores escolhidos são: SVM, Random Forest e Naive Bayes. As técnicas de redução dimensional implementadas pertencem a grupos de abordagens baseadas em filtro, wrappers e extração de atributos. As abordagens baseadas em filtro são: ANOVA, Info Gain, mRMR e Chi-Square. Os métodos baseados em wrappers são o SVM-RFE e RF-RFE. Por fim, os métodos baseados em extração de atributos são: PCA, Autoencoder, Denoising Autoencoder e Redes Siamesas. De forma direta o RAGECAN é responsável por realizar diversos testes com diferentes combinações de parâmetros e dimensão final e cada subconjunto selecionado é avaliado com um dos classificadores disponíveis. Ao final do número de execuções definida pelo usuário, o framework retorna um relatório detalhando os resultados obtidos e a configuração necessária para a obtenção destes resultados. Os resultados obtidos pelo framework foram melhores em 3 tipos diferentes de câncer quando comparados com o baseline.

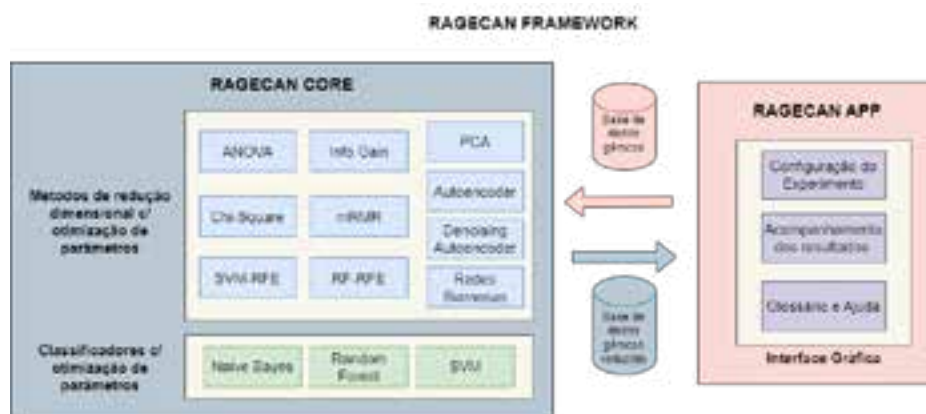


Figura 1 – Descrição geral das técnicas empregadas no Framework.

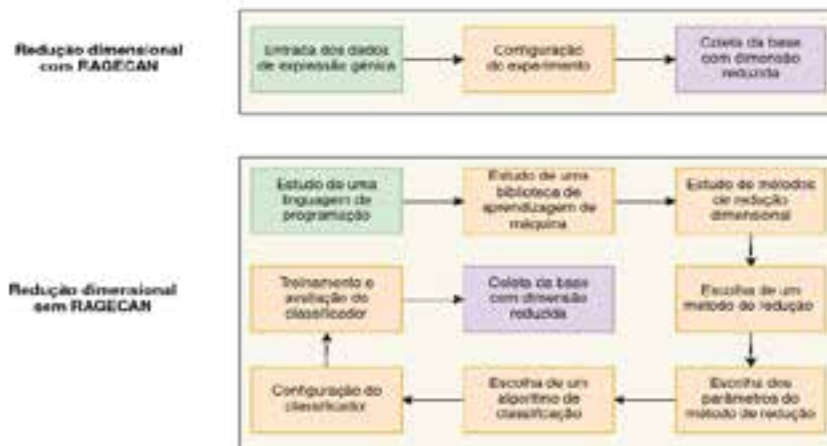


Figura 2 – Comparação das etapas de desenvolvimento de um pipeline de redução dimensional com e sem o framework RAGECAN.

4. DESAFIOS E APRENDIZADOS

As principais limitações deste trabalho estão relacionadas às limitações impostas pelos algoritmos de redução dimensional integrados, como por exemplo o método PCA que não é capaz de extrair representações em dimensão reduzida menores que a quantidade de amostras disponíveis para cada classe. Outro desafio está relacionado com o uso de métodos profundos como Autoencoders e Redes Siamesas, que acarretam em um alto custo computacional, inviabilizando o uso destas técnicas por usuários sem placa de vídeo com memória dedicada inferior à 2GB.

5. REFERÊNCIAS

- de Barros AC, Leite KR. Classificação molecular dos carcinomas de mama: uma visão contemporânea. *Revista Brasileira de Mastologia*. 2015 Oct 25;25(4):146-55.
- Feltes BC, Chandelier EB, Grisci BI, Dorn M. Cumida: An extensively curated microarray database for benchmarking and testing of machine learning approaches in cancer research. *Journal of Computational Biology*. 2019 Apr 1;26(4):376-86.

Random Forest para predição de risco cardiovascular em internos no processo de abstinência por uso de álcool e drogas

Pedro Elias P. Freire¹, Ana Clara B. Silva², Lucas Magalhães P. Carrara¹,
Chrystian A. Pereira (orientador)¹

¹ Departamento de Medicina, Universidade Federal de Lavras (UFLA), Lavras, MG, Brasil.

² Departamento de Agricultura, Universidade Federal de Lavras (UFLA), Lavras, MG, Brasil.

pedro.freire@estudante.ufla.br, ana.silva160@estudante.ufla.br, lucas.carrara@estudante.ufla.br,
chrystian.pereira@ufla.br

Resumo. O uso de substâncias psicoativas prejudica o convívio social e causa sérios impactos à saúde física, dentre os quais se destacam os efeitos sobre o sistema cardiovascular. Este estudo buscou relacionar variáveis clínicas de internos em reabilitação química com o risco cardiovascular, utilizando métodos de aprendizado de máquina. Foram analisados dados de 22 internos de uma Comunidade Terapêutica. Entre os fatores considerados estão as substâncias psicoativas de dependência, tempo de uso e abstinência, idade, peso, altura, IMC, glicemia capilar e razão cintura-quadril. O algoritmo empregado foi o Random Forest, e as métricas de análise de desempenho tiveram como resultado acurácia 0.680 (IC95% 0.495-0.865), precisão 0.833 (IC95% 0.580-1.086), sensibilidade 0.500 (IC95% 0.223-0.777), especificidade 0.867 (IC95% 0.633-1.100) e AUC de 0.991. Portanto, foi possível traçar uma relação entre os dados de dependência química, características individuais, glicemia e o risco cardiovascular derivado da razão cintura-quadril.

Palavras-chave: Aprendizado de máquina; Dependência química; Risco Cardiovascular.

Abstract. The use of psychoactive substances impairs social interactions and causes serious impacts on physical health, notably affecting the cardiovascular system. This study aimed to relate clinical variables of inpatients undergoing chemical rehabilitation with cardiovascular risk, using machine learning methods. Data from 22 inpatients of a Therapeutic Community were analyzed. The factors considered included the psychoactive substances of dependence, duration of use and abstinence, age, weight, height, BMI, capillary blood glucose, and waist-to-hip ratio. The algorithm employed was Random Forest, and the performance analysis metrics yielded an accuracy of 0.680 (95% CI 0.495-0.865), precision of 0.833 (95% CI 0.580-1.086), sensitivity of 0.500 (95% CI 0.223-0.777), specificity of 0.867 (95% CI 0.633-1.100), and an AUC of 0.991. Therefore, it was possible to establish a relationship between chemical dependency data, individual characteristics, metabolic profile, and cardiovascular risk derived from the waist-to-hip ratio.

Keywords: Machine Learning; Dependences, Chemical; Cardiovascular Risk.

1. CONTEXTO E MOTIVAÇÃO

Os parâmetros metabólicos podem ser considerados como os principais fatores para o desenvolvimento de alterações cardiovasculares, incluindo excesso de peso, hiperglicemia, sedentarismo e entre outros fatores. Alterações metabólicas associadas ao uso de substâncias psicoativas, intensificam consideravelmente os riscos cardiovasculares (1). No entanto, são poucos estudos que buscam avaliar a relação entre os parâmetros metabólicos e risco cardiovascular em populações de dependentes químicos em reabilitação, que se encontram em fase de abstinência. Ao longo dos últimos anos, foi possível observar um maior emprego de métodos de aprendizado de máquina (MAM) para relação entre variáveis clínicas e previsão de parâmetros que sejam úteis para o manejo do paciente (2).

1.1 Objetivo

Relacionar variáveis clínicas de internos em reabilitação química com o risco cardiovascular, através do método de aprendizado de máquina Random Forest.

2. ATIVIDADES PRINCIPAIS

Coleta dos dados e formação do dataset, pré-processamento dos dados, aplicação do MAM Random Forest e seleção de hiperparâmetros e, por conseguinte, avaliação das métricas de análise de desempenho do modelo proposto.

3. DESENVOLVIMENTO DO TRABALHO

Este trabalho consiste em um estudo de campo, realizado em uma Comunidade Terapêutica no município de Lavras, Minas Gerais, Brasil, abrangendo um espaço amostral de 22 internos.

A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos (COEP) da Universidade Federal de Lavras (UFLA) sob o parecer 6803699/CAAE: 77207923.9.0000.5148. Os dados selecionados para análise foram as substâncias psicoativas (SPA) citadas como causa de dependência pelos indivíduos em reabilitação neuropsicossocial, tempo de uso da SPA, idade, tempo de abstinência, glicemia, altura, peso, IMC e razão cintura-quadril. Assim, tais parâmetros foram usados para compor o dataset para a aplicação do método de aprendizado de máquina (MAM). O modelo selecionado foi Random Forest, implementado na linguagem Python, utilizando a biblioteca Scikit-learn.

A coleta de dados para a pesquisa foi realizada por meio de anamnese aplicada a cada um dos internos. Para identificar as substâncias psicoativas (SPA) mais utilizadas pelos indivíduos em reabilitação, foi perguntado sobre a dependência prévia de álcool, cocaína ou crack, tabaco e maconha. Os internos que relataram o uso de crack foram considerados no mesmo grupo de usuários de cocaína. Para cada uma das afirmações, foi atribuído um valor binário. Quanto ao tempo de uso, foi solicitado o tempo, em anos, desde o primeiro uso que levou à dependência daquela substância. Além disso, foram coletados dados sobre o tempo de abstinência (último uso) em meses e a idade, em anos. Já para avaliação antropométrica, foram coletadas as medidas de peso (Kg) e altura (m), por consequência, calculado o Índice de Massa Corpórea (IMC). Também obteve-se os valores, em centímetros, da circunferência da cintura e quadril, montando a razão cintura-quadril (RCQ). A glicemia capilar foi coletada ao acaso, por meio de um glicosímetro padrão (3). Para tais métricas, foi aplicado a média \pm desvio padrão (Tabela 1).

Para determinação do risco cardiovascular, foi utilizada a razão cintura-quadril (RCQ) tendo em vista a relação da idade com o valor obtido (3). Assim, para a formação do dataset no qual o MAM irá atuar teve-se a divisão dos previsores e a classe (como demonstrado na Tabela 1). Indivíduos com risco cardiovascular alto ou muito alto, após a análise da RCQ, foram considerados como positivos (valor 1), já para moderado ou baixo foram considerados como negativos (valor 0). Devido à utilização de dados binários e valores não binários, foi necessário usar a técnica de padronização dos dados.

Tabela 1: Divisão dos previsores e classe.

Previsores							Classe	
Substâncias Psicoativas em Reabilitação (vício prévio)	Tempo de Uso (anos)	Tempo de Abstinência (meses)	Idade (anos)	Peso (Kg)	Altura (m)	IMC (Kg/m ²)	Glicemia (mg/dL)	Razão cintura-quadril ALTO ou MUITO ALTO
Valores binários para cada SPA	15.77 \pm 6.50	3.38 \pm 3.95	40.36 \pm 9.96	81.22 \pm 12.81	1.74 \pm 0.07	27.02 \pm 3.69	83.45 \pm 16.18	Valores binários

Dentre as métricas para avaliar o desempenho do nosso modelo, usamos o cross-validation k-fold. Para nosso trabalho, utilizamos $k = 5$. Além disso, foi gerada a matriz de confusão, que mostra a taxa de verdadeiros positivos, verdadeiros negativos, verdadeiros previstos e falsos previstos. Portanto, foi possível montar as seguintes métricas: acurácia, precisão, especificidade e sensibilidade (Tabela 3). A matriz de confusão foi montada a partir da agregação de cada k-fold, sendo a soma dos valores individuais (soma das 5 matrizes criadas) (Tabela 2). A fim de avaliar o modelo de desempenho geral do MAM, foi obtida a Área Sob a Curva (Area Under Curve - AUC).

Tabela 2: Matriz de confusão agregada.

	Verdadeiros Positivos	Verdadeiros Negativos
Previsto Positivo	6	2
Previsto Negativo	5	9

Tabela 3: Métricas de análise de desempenho.

Acurácia	0.680 (IC95% 0.495-0.865)
Precisão	0.833 (IC95% 0.580-1.086)
Sensibilidade	0.500 (IC95% 0.223-0.777)
Especificidade	0.867 (IC95% 0.633-1.100)
AUC	0.991

A partir dos resultados demonstrados, é possível traçar uma relação entre os dados de dependência química, características individuais, glicemia e o risco cardiovascular derivado da RCQ. Para o nosso modelo, as métricas de precisão e especificidade aturam como bons avaliadores da predição na prevalência dos verdadeiros positivos e negativos, respectivamente. Assim, como um modelo preditivo, o algoritmo Random Forest atuou como um bom classificador (AUC 0.991) para o dataset fornecendo uma ferramenta com potencial interesse para conduta e manejo no processo de abstinência por álcool e outras drogas.

4. DESAFIOS E APRENDIZADOS

O principal desafio enfrentado foi na coleta dos dados em campo, devido a necessidade de precisão e padronização na coleta de cada parâmetro.

5. REFERÊNCIAS

1. Fragasso G, Margonato A, Spoladore R, Lopaschuk GD. Metabolic effects of cardiovascular drugs. *Trends in Cardiovascular Medicine*. 2019 Apr;29(3):176–87.
2. Chhetri B, Goyal LM, Mittal M. How machine learning is used to study addiction in digital healthcare: A systematic review. *International Journal of Information Management Data Insights*. 2023 Nov;3(2):100175.
3. Loureiro NS de L, Amaral TLM, Amaral CDA, Monteiro GTR, Vasconcellos MTL de, Bortolini MJS. Relationship between anthropometric indicators and risk factors for cardiovascular disease in adults and older adults of Rio Branco, Acre. *Revista de Saúde Pública*. 2020 Mar 11;54:24.

Segmentação automática da próstata em imagens médicas com deep learning por meio da técnica de federated learning

Luciana Silva Albuquerque de Melo (aluna)¹, Mário Olímpio de Menezes (orientador)¹

¹ Instituto de Pesquisas Energéticas e Nucleares – IPEN-CNEN, São Paulo, SP, Brasil.

lucianasmfl@gmail.com, mo.menezes@gmail.com

Resumo. A inteligência artificial (IA) e a ciência de dados têm crescido significativamente, especialmente na área da saúde, moldando o futuro da medicina e da assistência médica. No entanto, a aplicação de IA na saúde enfrenta desafios, especialmente relacionados à privacidade dos dados e à necessidade de grandes volumes de dados para treinar modelos eficazes. Este estudo visa desenvolver um método automático de segmentação da próstata utilizando IA e a técnica de federated learning (FL), que permite a colaboração entre instituições mantendo os dados localmente, sem compartilhamento direto. O método foi avaliado em dois cenários: centralizado e federado, com o cenário federado apresentando desempenho ligeiramente superior. Ainda assim, como o trabalho é preliminar, são necessários mais estudos para validar os resultados e explorar técnicas adicionais de proteção de privacidade no contexto do aprendizado federado.

Abstract. Artificial intelligence (AI) and data science have grown significantly, especially in the healthcare sector, shaping the future of medicine and medical care. However, the application of AI in healthcare faces challenges, particularly related to data privacy and the need for large volumes of data to train effective models. This study aims to develop an automatic method for prostate segmentation using AI and the technique of federated learning (FL), which allows collaboration between institutions while keeping data local, without direct sharing. The method was evaluated in two scenarios: centralized and federated, with the federated scenario showing slightly better performance. Nevertheless, as this is a preliminary work, further studies are needed to validate the results and explore additional techniques for privacy protection in the context of federated learning.

Palavras-chave: Inteligência artificial na saúde; Segmentação automática da próstata; Aprendizado federado.

1. CONTEXTO E MOTIVAÇÃO

A IA está revolucionando a assistência médica. Pesquisas indicam que ela pode desempenhar um papel tão competente quanto, ou até superior, ao dos seres humanos em tarefas relacionadas à saúde, como o diagnóstico de doenças (1).

O uso de exames de imagem para prevenção, diagnóstico e tratamento de tumores prostáticos é crescente, superando os métodos mais convencionais, como os exames de toque retal e o teste de antígeno prostático específico (PSA). A tarefa de segmentação da próstata que envolve a separação do volume prostático de outras estruturas presentes na imagem, é geralmente realizada manualmente, no entanto, é moroso e propenso a diferentes interpretações entre especialistas (2). Uma alternativa eficaz ao método manual é a exploração de métodos automáticos para a segmentação de imagens digitais (3). Essa abordagem pode não apenas acelerar o processo, bem como contribuir para resultados mais consistentes na identificação e delimitação da próstata em imagens médicas.

Contudo, o emprego da IA em saúde, especialmente na aplicação de algoritmos de machine learning (ML), enfrenta grandes desafios, destacando-se a preocupação com a privacidade dos dados, pois de acordo com a Lei Geral de Proteção a dados (Lei nº 13.709/2018), os dados médicos são considerados dados pessoais sensíveis. Além disso, a necessidade de treinamento com grandes volumes de dados agrava a complexidade deste cenário. Uma alternativa para superar esses

desafios é uma abordagem inovadora em ML, o aprendizado federado ou federated learning (FL). Essa técnica permite a colaboração no aprendizado mantendo os dados em suas respectivas fontes, eliminando a necessidade de compartilhamento e, assim, preservando a privacidade dos dados, além de melhorar a qualidade dos modelos (4).

1.1 Objetivo

Este estudo tem como objetivo desenvolver um modelo generalizável de ML utilizando a técnica de FL, capaz de realizar a segmentação automática de imagens de ressonância magnética da região da próstata. Ao integrar essa abordagem, pretende-se contribuir significativamente para o desenvolvimento de estratégias de tratamento mais precisas, como a radioterapia, que preservem os órgãos adjacentes, além de fornecer uma base para futuras pesquisas e contribuir para o avanço contínuo na segmentação de imagens médicas e aprendizado de máquina.

2. ATIVIDADES PRINCIPAIS

As principais atividades desenvolvidas foram:

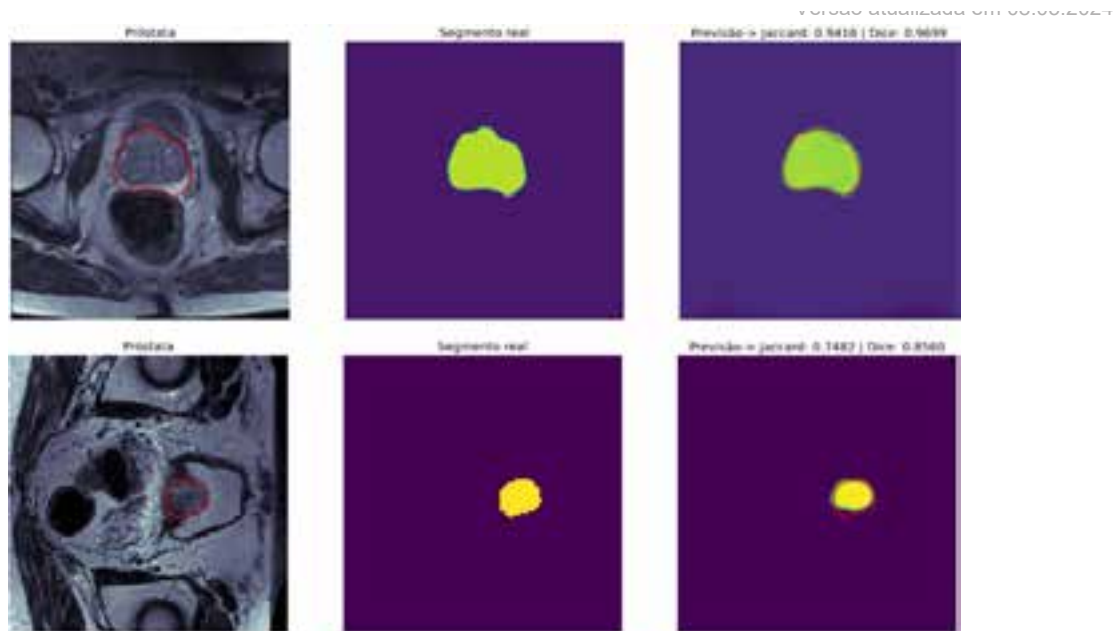
- Modelo de deep learning (DL) em ambiente centralizado: O treinamento inicial do modelo foi realizado utilizando dados centralizados, o que significa que todas as imagens utilizadas foram coletadas e armazenadas em um único local, seguindo o método tradicional de ML.
- Migração para um ambiente descentralizado: O fluxo de trabalho foi adaptado para um cenário descentralizado, utilizando o framework Flower para simular o FL.

Os próximos passos incluem a adição de novos datasets para aumentar a robustez do modelo, a integração de técnicas adicionais de proteção de privacidade, testar diferentes configurações de clientes e disponibilizar o modelo para uso com novas imagens.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este trabalho segue o ciclo de vida de um projeto de ciência de dados: 1- Definição do objetivo; 2- Coleta de 1.979 imagens reais de fontes públicas de dois datasets amplamente reconhecidos na área de segmentação de próstata, o PROMISE12 e o Task 05: Prostate. Destas, 1.380 imagens foram selecionadas por serem relevantes para o treinamento do modelo. 3- Foram desenvolvidos modelos de ML com a arquitetura U-Net, inicialmente treinados em um ambiente centralizado e depois migrados para um cenário descentralizado, simulando o FL com o framework Flower, envolvendo dois clientes. As configurações de clientes e treinamento estão sendo ajustadas para otimizar o desempenho e garantir a escalabilidade do modelo, visando expandi-lo para mais clientes e dados. 4- Os modelos foram avaliados utilizando as métricas Coeficiente de Jaccard e Dice Score. Os resultados preliminares mostraram índices médios de 0.7763 e 0.8613 no cenário centralizado, enquanto no federado os índices foram de 0.7872 e 0.8691, respectivamente. A Figura 1 demonstra a performance do modelo federado, evidenciando uma melhoria na precisão em segmentos maiores, destacando o potencial do FL para aprimorar a segmentação de imagens médicas. 5- Todas as etapas e resultados do projeto estão sendo devidamente documentados para garantir a replicabilidade e transparência dos processos adotados. 6- O modelo final será disponibilizado por meio de uma aplicação web, facilitando seu acesso para uso clínico e em pesquisas futuras.

Figura 1 – Exemplos de segmentação automática da próstata utilizando aprendizado federado



Fonte: O autor (2024)

4. DESAFIOS E APRENDIZADOS

Os principais desafios do projeto incluíram a necessidade de adquirir rapidamente conhecimentos técnicos sobre algoritmos de ML e FL, superada por meio de estudos intensivos, e a dificuldade em encontrar bases de dados públicas, parcialmente mitigada com o uso dos datasets PROMISE12 e Task 05: Prostate. Além disso, a integração de ferramentas e tecnologias enfrentou problemas de dependência entre bibliotecas, que foram solucionados pela migração para o Google Colaboratory.

5. REFERÊNCIAS

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/>. Acessado em 19/07/2024.
2. Ferreira JL, da Silva GLF, Reis ABS, Cavalcante AB, Silva AC, de Paiva AC. Segmentação Automática da Próstata em Imagens de Ressonância Magnética utilizando Redes Neurais Convolucionais e Mapa Probabilístico. In: *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, 18; 2018; Natal. Porto Alegre: Sociedade Brasileira de Computação; 2018. ISSN 2763-8952. doi: 10.5753/sbcas.2018.3683. Disponível em: <https://sol.sbc.org.br/index.php/sbcas/article/view/3683>. Acessado em: 19/07/2024.
3. Mahapatra D, Buhmann JM. Prostate MRI segmentation using learned semantic knowledge and graph cuts. *IEEE Trans Biomed Eng.* 2014 Mar;61(3):756-64. doi: 10.1109/TBME.2013.2289306. Epub 2013 Nov 6. PMID: 24235297. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6656821>. Acessado em: 19/07/2024.
4. Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep.* 2020;10:12598. Disponível em: <https://www.nature.com/articles/s41598-020-69250-1>. Acessado em 19/07/2024.

SEREIA – Document Store Exploration through Keywords

Ariel A. Afonso, Paulo R. Martins, Altigran S. da Silva

Universidade Federal do Amazonas, Manaus, AM, Brasil.

ariel.afonso@icomp.ufam.edu.br, paulo.martins@icomp.ufam.edu.br, alti@icomp.ufam.edu.br

Resumo. The adoption of document stores, such as MongoDB or CouchDB, has increased drastically in recent years. Part of this popularity can certainly be explained by their flexibility in loading, storing and retrieving semi-structured data on massive scales. However, adopting such systems presents challenges when exploring the data they store, since document structure may not follow a pattern and thus present complex hierarchical and nested structures that vary. In this work, we propose SEREIA, a system that facilitates data exploration in document stores through keyword search. We evaluated SEREIA using five datasets previously used in the literature, and the results achieved indicate that SEREIA is suitable for helping users in data exploration tasks by removing the burden of understanding the data organization of the stored documents and by automatically generating queries to explore data of interest.

Palavras-chave: information retrieval; data exploration; keywords, document stores.

Nome do projeto: AmazonIA - Aplicação de Modelos de Aprendizagem de Máquina e Aprendizagem Profunda para Análise de Dados sobre Doenças Tropicais Negligenciadas na Amazônia

Links: <https://github.com/bdri-ufam/sereia>

1. CONTEXTO E MOTIVAÇÃO

The adoption of document stores, such as MongoDB, has increased drastically in recent years. Several domains and many applications take advantage of this kind of system (1). Part of this popularity is undoubtedly due to their flexibility in loading, storing, and retrieving semi-structured data on massive scales (2, 3). This flexibility is mainly due to two aspects: first, document stores allow the creation of semi-structured documents with complex hierarchical and nested structures (2); second, they do not require a schema definition (4). In fact, they support a schema later approach, that is, they allow easier data ingestion without requiring documents to follow a predefined structure (3).

Dealing with these aspects adds a degree of complexity to stored data and incurs challenges that are noticed when data scientists need to retrieve information from these systems (5 – 7). Indeed, due to their lack of strict structure declarations, document stores require the user to have a deep understanding of the underlying structure and organization of the data when seeking information. Another challenge is that the query languages used by document stores are considerably more complex than, for instance, SQL. In fact, document stores challenge even experienced data scientists who want to analyze the data they store (8), resulting in extra time and effort when searching for relevant data rather than effectively analyzing it (6).

In this work, we propose SEREIA, a system that facilitates data exploration in document stores. More specifically, SEREIA takes as input a non-structured keyword-based query and generates a corresponding structured query written in the document stores' native query language. By doing so, we aim to help data analysts/scientists undertake their data analysis tasks without having to deal with the typical complexities of retrieving data from document stores. SEREIA avoids the need for users to have prior knowledge of the underlying structure and organization of the stored documents in addition to the syntax and operations of its native query language.

To illustrate the effectiveness of our method in document stores' exploratory tasks, we present a case study in the next section in which we compare exploratory queries built by an analyst with experience in MongoDB against queries generated by SEREIA. We provided a set of 18 keyword queries and asked him to generate structured queries to fulfill them. Then, we compared the manually generated queries with those generated with SEREIA.

2. ATIVIDADES PRINCIPAIS

The student activities involved: research on Document Stores exploration previous literature, keyword search literature review, implementation of an exploration tool for Document Stores, tool assessment through a case study.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

To illustrate the effectiveness of our method in helping analysts with exploratory tasks in document stores, in this section, we present a case study in which we compare exploratory queries built by an analyst with experience in MongoDB against queries generated by SEREIA. For this, we provided a set of 18 queries, described in plain English, for the analyst to generate MongoDB queries from them. We selected these queries from query sets proposed in previous works and also used in our experiments, namely, nine queries from the Yelp! query set and nine from the IMDb query set. For the sake of conciseness, we selected only 4 queries out of the total for each query set and show the results below.

Table 1 – Excerpt from the tasks for the document store exploration

Query Set	Task	Keywords	Analyst		SEREIA	
			Stages	Time	Stages	Time
Yelp!	Businesses rated with 5 stars by Bricola	Businesses rated 5 stars bricola	4	1.59	6	2.69
Yelp!	Businesses reviewed by Analucia	Businesses reviewed analucia	4	1.65	5	2.74
Yelp!	Reviews by Michelle for Italian Restaurants	Reviews michelle Italian restaurants	5	1.81	7	2.16
Yelp!	Italian Restaurants with 5 stars	Italian restaurants 5 stars	1	0.0001	3	0.29
IMDb	Movies with Denzel Washington	Movies denzel washington	3	0.19	8	0.58
IMDb	Movies with Will Smith	Movies will smith	3	0.09	8	0.62
IMDb	Lord of the Rings Movie	Lord rings movie	1	0.00009	4	0.07
IMDb	Forrest Gump movie	Forrest gump movie	1	0.00009	8	0.45

Table 1 shows all the exploration tasks (Task) provided to the analyst, along with the keywords we input to SEREIA to fulfill the task (Keywords). The table also presents comparisons of execution time (in seconds) and the number of query stages generated by the human analyst and SEREIA. First, the queries generated by the analyst and by SEREIA produced the same results for each task. Regarding execution time, in general, queries generated by SEREIA had a reasonable execution time compared to queries by the analyst. In the Yelp! query set, the average query execution time for analyst queries was 0.58 seconds, while queries generated by SEREIA took an average execution time of 0.93 seconds. In the IMDb dataset, the average query execution time for analyst queries was 0.05 seconds, while it took 0.31 seconds on average for queries generated by SEREIA. This time increase is related to the number of stages to query data from the document store. Since

SEREIA uses a generic algorithm to generate stages based on the underlying data organization, it may generate stages that could otherwise be optimized by an expert user. Nevertheless, the much more straightforward approach of providing keyword queries is more attractive than manually building and evaluating a MongoDB query.

It is also worth noting that the main difference between both is the time it takes to start exploring the database. When executing the proposed tasks, the analyst took several pauses to address query issues that arise when building MongoDB queries, such as query syntax and operators for each processing step. Additionally, the analyst tested each new stage against the database to assess the correctness of the items retrieved in the answer. When using SEREIA, only the set of keywords was necessary to retrieve the data and to begin the exploration of the database content effectively.

4. DESAFIOS E APRENDIZADOS

The main challenge was the scarcity in document store exploration available research, requiring an innovative solution to this context with few baselines for the objective of the proposed tool.

5. REFERÊNCIAS

1. Pramod, Sadalage, and Martin Fowler. NoSQL Distilled. 2013.
2. DiScala, Michael, and Daniel Abadi. "Automatic Generation of Normalized Relational Schemas from Nested Key-Value Data." Proceedings of the 2016 International Conference on Management of Data, 2016, pp. 295–310, <https://doi.org/10.1145/2882903.2882924>.
3. Tahara, Daniel, et al. "Sinew: A SQL System for Multi-Structured Data." Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 815–26. DOI.org (Crossref), <https://doi.org/10.1145/2588555.2612183>.
4. Chasseur, Craig, et al. "Enabling JSON Document Stores in Relational Systems." International Workshop on the Web and Databases, 2013, <https://api.semanticscholar.org/CorpusID:1381872>.
5. Terrizzano, Ignacio, et al. Data Wrangling: The Challenging Journey from the Wild to the Lake. 2015.
6. Castro Fernandez, Raul, et al. "Aurum: A Data Discovery System." 2018 IEEE 34th International Conference on Data Engineering (ICDE), IEEE, 2018, pp. 1001–12. DOI.org (Crossref), <https://doi.org/10.1109/ICDE.2018.00094>.
7. Sethi, Raghav, et al. "Presto: SQL on Everything." 2019 IEEE 35th International Conference on Data Engineering (ICDE), IEEE, 2019, pp. 1802–13. DOI.org (Crossref), <https://doi.org/10.1109/ICDE.2019.00196>.
8. Ben Hamadou, Hamdi, et al. "Schema-Independent Querying for Heterogeneous Collections in NoSQL Document Stores." Information Systems, vol. 85, Nov. 2019, pp. 48–67. DOI.org (Crossref), <https://doi.org/10.1016/j.is.2019.04.005>.

Subset of PAM50 Genes for Breast Cancer Subtypes Classification¹

Leandro Y. S. Okimoto¹, Eduardo F. Nakamura, Fabíola G. Nakamura¹

¹Universidade Federal do Amazonas – IComp Instituto de Computação, Manaus, AM, Brasil.

okimoto@icomp.ufam.edu.br, nakamura@icomp.ufam.edu.br, fabiola@icomp.ufam.edu.br

Resumo. In recent years, significant strides have been made in understanding the heterogeneity of breast cancer and its subtypes. However, analyzing the extensive genomic and proteomic data available today requires efficient tools. While the PAM50 gene signature is a successful prognostic gene set, its reliance on many genes presents cost and complexity challenges. This study addresses the need for efficient breast cancer subtype classification using a reduced gene set. We employ a “Subset of PAM50 Genes” method to randomly select smaller subsets from PAM50, evaluating their performance with metrics and a Support Vector Machine (SVM) classifier. Our findings show that certain reduced gene subsets can perform as well as or better than the full PAM50 gene signature. The identified gene subsets, consisting of 36 genes, have the potential to develop more cost-effective and streamlined diagnostic tools in breast cancer research and clinical settings.

Keywords: Gene expression; Breast Cancer Subtypes; Machine Learning;

Project name: AmazonIA - Aplicação de Modelos de Aprendizagem de Máquina e Aprendizagem Profunda para Análise de Dados sobre Doenças Tropicais Negligenciadas na Amazônia

1. CONTEXT AND MOTIVATION

Breast cancer, known for its heterogeneity, presents distinct subtypes with unique biological characteristics and treatment responses. The influx of genomic and proteomic data offers valuable insights, but efficient analysis requires robust computational tools. Breast cancer subtypes include luminal A, luminal B, and HER2-enriched for hormone receptor-positive, and triple-negative and basal-like for hormone receptor-negative (1). The PAM50 gene signature classifies these subtypes, aiding prognosis and treatment decisions. However, its reliance on 50 genes increases costs for clinical use and complexity, highlighting the need for more efficient classification methods using fewer genes without compromising accuracy (2).

1.1 Objective

The primary objective of this research is to explore the potential for accurately categorizing breast cancer subtypes using a reduced gene set derived from the PAM50 gene signature. By employing this gene select method, we will randomly select smaller subsets from PAM50 and evaluate their performance using the F1 Score and a Support Vector Machine (SVM) classifier. This approach aims to determine if a more compact gene set can maintain accuracy while simplifying the classification process.

2. MAIN ACTIVITIES

The activities included in this project are:

- Curation and preprocessing of the dataset obtained from online sources;
- Construction, experimentation, and refinement of model hyperparameters;
- Analysis of the results obtained and performance comparison with baselines;
- Writing of the final report, which led to its submission.

¹ An extended version of this work was published on BMC Bioinformatics 2024 (4).

3. DEVELOPMENT OF THE WORK (METHODOLOGY AND RESULTS)

The methodology, as presented in Figure 1, involves several steps: First, the dataset is normalized using the Standard Scaler. Next, the dataset is filtered using the PAM50 gene signature. The filtered dataset is then split into training, validation, and test sets to avoid overfitting. Optimal gene subset choices are assessed based on the size of the PAM50 signature. A second dataset is processed by normalizing it with the Standard Scaler and filtering it using the selected gene subsets. Subsequently, a Support Vector Machine (SVM) is trained and tested on the filtered subsets using Grid Search. Finally, the test predictions are evaluated using t-SNE, a Confusion Matrix, and the ROC-AUC curve.

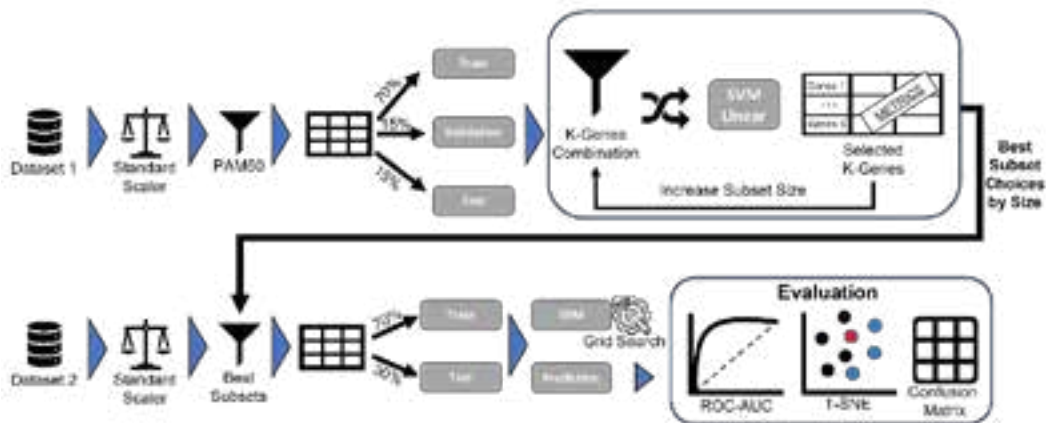


Figure 1 – Proposed approach.

This method aimed to capture fewer complex attributes in the selection phase using a linear model and trained non-linear SVM models in the evaluation phase. Results show that our three topperforming subsets (with subset sizes of 36, 35, and 34 genes) achieved AUC scores equal to or higher than the PAM50 baseline. Our method shows that we can correctly classify most of our samples, with minor misclassification between the boundaries of the classes, as shown in Figure 2.



Figure 2. t-SNE Visualization in the ACES dataset Prediction data after PAM50 filtering. The results in Table 1 present the performance of four highlighted filtering methods (PAM50, S-36, S-35, and S-34) across the four subtypes (Basal, Her2, LumA, and LumB), reporting Accuracy (Acc), Precision (Prec), Recall (Rec), F1 Score (F1), and AUC for each case

	Basal					Her2				
	Acc	Prec	Rec	F1	AUC	Acc	Prec	Rec	F1	AUC
PAM50	0.910	1.000	0.931	0.964	1.000	0.910	0.850	1.000	0.919	0.999
Subset 37	0.910	1.000	0.897	0.945	0.999	0.910	0.842	0.941	0.889	0.996
Subset 36	0.935	1.000	0.931	0.964	1.000	0.935	0.850	1.000	0.919	0.996
Subset 32	0.923	1.000	0.931	0.964	0.999	0.923	0.895	1.000	0.944	0.997

	LumA					LumB				
	Acc	Prec	Rec	F1	AUC	Acc	Prec	Rec	F1	AUC
PAM50	0.910	0.914	0.914	0.914	0.988	0.910	0.868	0.846	0.857	0.978
Subset 37	0.910	0.915	0.929	0.922	0.990	0.910	0.872	0.872	0.872	0.985
Subset 36	0.935	0.944	0.957	0.950	0.992	0.935	0.919	0.872	0.895	0.988
Subset 32	0.923	0.929	0.929	0.929	0.989	0.923	0.872	0.872	0.872	0.986

Table 1. Comparison of metrics between the state-of-the-art PAM50 and our subsets.

4. CHALLENGES AND LEARNINGS

The main challenge of this research was the curse of dimensionality, which led us to avoid deep learning approaches due to their high data and computational demands. Instead, we employed traditional machine learning techniques and feature selection methods to reduce the dimensionality of gene expression data. By using the gene subset method, we identified smaller, representative subsets of genes from the PAM50 signature, thereby simplifying and potentially improving the classification process. This approach mitigated the curse of dimensionality, enhanced computational efficiency, and maintained model performance. Through this experience, we gained valuable insights into handling high-dimensional data, the importance of feature selection, and the necessity of balancing model complexity with computational efficiency.

5. REFERENCES

1. Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M.C.U., Voduc, D., Speers, C.H., Nielsen, T.O., Gelmon, K.: Metastatic behavior of breast cancer subtypes. *Journal of clinical oncology* 28(20), 3271–3277 (2010).
2. Mendonca-Neto, R., Li, Z., Fenyő, D., Silva, C.T., Nakamura, F.G., Nakamura, E.F.: A gene selection method based on outliers for breast cancer subtype classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19(5), 2547–2559 (2021).
3. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. *Journal of Machine Learning Research* 9(Nov), 2579–2605 (2008)
4. Okimoto, L. Y., Mendonca-Neto, R., Nakamura, F. G., Nakamura, E. F., Fenyő, D., & Silva, C. T. (2024). Few-shot genes selection: subset of PAM50 genes for breast cancer subtypes classification. *BMC bioinformatics*, 25(1), 92.

Um Estudo Experimental de Técnicas de Aumentação de Dados para Reconhecimento de Entidades Nomeadas em Domínios Low-Resource

Arthur Elwing Torres¹, Altigran Soares da Silva¹

¹Universidade Federal do Amazonas, Manaus, AM, Brasil.

arthur.torres@icomp.ufam.edu.br, alti@icomp.ufam.edu.br

Resumo. Neste estudo de Reconhecimento de Entidades Nomeadas, avaliamos a eficácia de duas técnicas proeminentes de aumento de dados: Substituição de Menção e Substituição Contextual de Palavras, em dois modelos de NER amplamente utilizados, Bi-LSTM+CRF e BERT. Realizamos experimentos em quatro conjuntos de dados de domínios Low-Resource e exploramos o impacto de várias combinações de tamanhos de subconjuntos de treinamento e número de exemplos aumentados. Não apenas confirmamos que aumento de dados é particularmente benéfica para conjuntos de dados menores, mas também demonstramos que não existe um número universalmente ideal de exemplos aumentados, ou seja, os praticantes de NER devem experimentar diferentes quantidades para desenvolver seus projetos.

Abstract. In this Named Entity Recognition study, we evaluate the effectiveness of two prominent data augmentation techniques: Mention Replacement and Contextual Word Replacement, across two widely used NER models, Bi-LSTM+CRF and BERT. We conducted experiments on four low-resource domain datasets and explored the impact of various combinations of training subset sizes and numbers of augmented examples. We not only confirmed that data augmentation is particularly beneficial for smaller datasets but also demonstrated that there is no universally ideal number of augmented examples. In other words, NER practitioners should experiment with different quantities to develop their projects.

Palavras-chave: Reconhecimento de Entidades Nomeadas, Aumentação de Dados, Domínios Low-resource, Entidades Long-tail.

Nome do projeto: AmazonIA - Aplicação de Modelos de Aprendizagem de Máquina e Aprendizagem Profunda para Análise de Dados sobre Doenças Tropicais Negligenciadas na Amazônia

Links <https://github.com/arttorres0/augmented-ner-model>

1. CONTEXTO E MOTIVAÇÃO

O Reconhecimento de Entidades Nomeadas (NER - Named Entity Recognition) é uma sub-tarefa especializada do campo da Extração de Informação que se concentra na identificação e categorização de entidades do mundo real em textos [1]. É uma tarefa de aprendizado de máquina que tradicionalmente depende de aprendizado supervisionado e dados anotados. Adquirir esses dados é um desafio, especialmente em campos especializados, como os setores médico, jurídico e financeiro. Esses setores são conhecidos como domínios Low-Resource, que compreendem entidades Long-Tail, devido à escassez de dados disponíveis [2,3].

Para resolver isso, técnicas de aumento de dados estão sendo cada vez mais utilizadas para gerar instâncias de treinamento adicionais. Essa abordagem gera novas amostras de dados aplicando transformações aos dados existentes [4,5]. Apesar do crescente interesse, as pesquisas existentes não exploram suficientemente como o volume de dados aumentados influencia o desempenho dos modelos de NER. Esse é um parâmetro crucial que precisa ser determinado antes do treinamento dos modelos.

1.1 Objetivo

O objetivo neste estudo é examinar a eficácia em NER de duas técnicas específicas de aumento de dados, Substituição de Menção (MR - Mention Replacement) [6] e Substituição Contextual de Palavras (CWR - Contextual Word Replacement) [7], em quatro conjuntos de dados de domínios Low-Resource.

2. ATIVIDADES PRINCIPAIS

- Revisão bibliográfica para seleção de modelos de NER e técnicas de aumento;
- Configuração dos experimentos e seleção dos conjuntos de dados;
- Realização dos experimentos e análise dos resultados

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Em nossos experimentos, avaliamos o impacto da aumento de dados em quatro conjuntos de dados de domínios Low-Resource: BioCreative V CDR Task e i2b2-2010, do domínio biomédico, MaSciP, da ciência dos materiais, e JusBrasil, do domínio jurídico. Avaliamos diferentes tamanhos de subconjuntos (50, 150, 500 sentenças, e 25%, 50%, 75% dos conjuntos de dados originais, desde que não excedam 500 frases), além do conjunto de dados completo.

Empregamos duas técnicas populares de aumento de dados: MR e CWR. A MR aplicou a substituição de entidades por outras provenientes dos subconjuntos de treinamento, evitando fontes externas. A CWR realizou a substituição de palavras não relacionadas a entidades, utilizando BERT como ferramenta de sugestão de substituições.

Usamos as duas arquiteturas mais eficazes de NER, Bi-LSTM+CRF e BERT (aplicando SciBERT para os conjuntos dos domínios biomédico e ciência dos materiais, e BERT multilíngue para os demais). Aumentamos progressivamente os dados aumentados de 0% a 500% dos dados originais. Treinamos 10 modelos por combinação de conjuntos de dados, tamanhos de subconjuntos, arquiteturas e níveis de aumento, totalizando 784 combinações, e as avaliamos com base nas médias de F1-score.

Os experimentos mostraram que a aumento de dados com MR inicialmente melhorou a qualidade do modelo para subconjuntos menores, mas não beneficiou conjuntos de dados maiores e, às vezes, reduziu o desempenho devido ao ruído introduzido pelos dados aumentados inválidos. Em contraste, a CWR mostrou mais potencial, produzindo maior variabilidade e melhores resultados em mais modelos, embora tenha introduzido mais ruído também. Além disso, mostramos que não há uma quantidade universalmente ideal de dados aumentados, destacando a necessidade de experimentação. Notavelmente, a CWR superou a MR, e os modelos BERT consistentemente tiveram melhorias maiores com aumento de dados do que os modelos Bi-LSTM+CRF.

Além disso, das 784 combinações de modelos, simplificamos escolhendo as combinações cujas quantidades de aumento resultaram na melhor média de F1-score por subconjunto e modelo, totalizando 100 médias. Dessas 100, 10% apresentaram melhorias estatisticamente significativas em relação às médias sem aumento. Essas combinações estão listadas na Tabela 1. No geral, o estudo indica um potencial significativo para melhorar modelos de NER em domínios Low-Resource utilizando aumento de dados.

4. DESAFIOS E APRENDIZADOS

Além da revisão bibliográfica e implementação dos códigos dos experimentos, o maior desafio foi

a gestão dos experimentos, visto que foram treinados 7840 modelos, o que requereu múltiplas instâncias computacionais executando em paralelo.

Tabela 1 – Resumo dos modelos melhorados por aumento de dados

Conjunto	Tamanho do Treino	Modelo	Melhoria	Quantidade de Aumento	Técnica de Aumento
i2b2-2010	50 sentenças	BERT	25.67%	25%	CWR
i2b2-2010	150 sentenças	BERT	7.16%	50%	CWR
i2b2-2010	150 sentenças	BERT	6.70%	75%	MR
i2b2-2010	25%	Bi-LSTM + CRF	6.45%	25%	CWR
MaSciP	50 sentenças	BERT	3.68%	200%	CWR
i2b2-2010	50 sentenças	Bi-LSTM + CRF	2.92%	500%	MR
MaSciP	50 sentenças	BERT	2.32%	100%	MR
MaSciP	150 sentenças	BERT	1.50%	25%	CWR
MaSciP	25%	BERT	1.37%	100%	CWR
MaSciP	25%	BERT	1.33%	50%	MR

5. REFERÊNCIAS

- Jurafsky, D., Martin, J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 1ª ed. Prentice Hall PTR, USA (2000)
- Wang, X., Jiang, Y., Bch, N., Wang, T., Huang, Z., Huang, F., Tu, K. Automated concatenation of embeddings for structured prediction. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 2643–2660 (2021)
- Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., Li, J. Dice loss for data-imbalanced NLP tasks. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 465–476 (2020)
- Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D. Understanding data augmentation for classification: When to warp? *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 1–6 (2016)
- Shorten, C., Khoshgoftaar, T. A survey on image data augmentation for deep learning. *Journal of Big Data* 6, 1–48 (2019)
- Dai, X., Adel, H. An analysis of simple data augmentation for named entity recognition. In: *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 3861–3867 (2020)
- Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., Liu, Q. TinyBERT: Distilling BERT for natural language understanding. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 4163–4174 (2020)

Unsupervised Learning for Noise Removal in Biomedical Signals: A Case Study with Autoencoders

Waisman da Fonseca Braga

¹Programa de Pós-graduação em Ciência da Computação, UFMG, Belo Horizonte, MG

waismanfb@gmail.com

Resumo. Biomedical signals are often contaminated with noise from various sources, hindering accurate analysis and interpretation. However, a machine learning technique known as autoencoders offers a promising solution to this challenge. This study explored the application of autoencoders to remove noise from simulated biomedical signals, focusing on electrocardiograms (ECG), electroencephalograms (EEG), electromyograms (EMG) and electrooculograms (EOG). To simulate the often noisy conditions in which biomedical signals are collected, various types of noise were intentionally introduced to the signals, including powerline noise, baseline wander, electrode motion and white noise. The autoencoder model was trained on these noisy signals, learning to reconstruct the original signal in the process. Evaluation results demonstrated the model's effectiveness in reducing noise while preserving the signal features. These results highlight the potential of autoencoders to enhance the accuracy and reliability of biomedical signal analysis, ultimately leading to improved patient care.

Keywords: Autoencoders, Biomedical Signals, Machine Learning.

1. CONTEXT AND MOTIVATION

Autoencoders have demonstrated versatility in noise removal across various applications, ranging from induction motor sounds (2) to germanium detector signals (1) and musical spectrograms (3). The application of this technique to biomedical signals presents a promising field in healthcare, with the potential to enhance the accuracy of diagnoses and treatments, directly impacting quality of life.

1.1 Objective

To develop and evaluate an autoencoder model for noise reduction in simulated biomedical data, aiming to improve the quality and analysis of that data in medical and research applications.

2. KEY ACTIVITIES

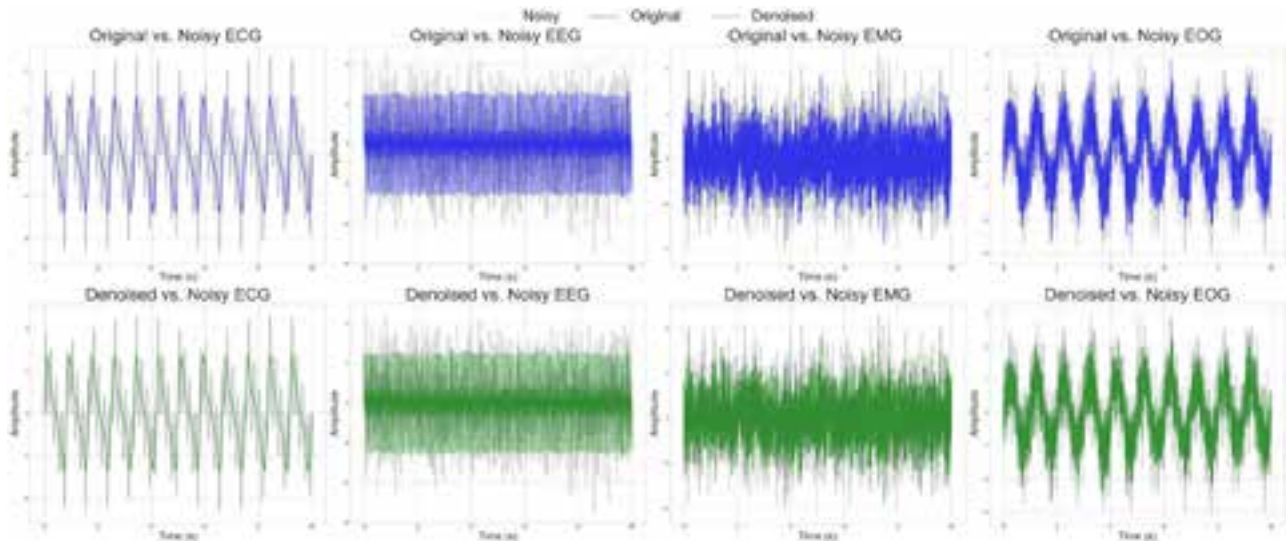
Initially, synthetic biomedical signals were generated to simulate the data. For realistic conditions, these synthetic signals were subjected to different types of noise, which were also synthetically generated. Finally, an autoencoder model was designed specifically for noise reduction and validated.

3. WORK DEVELOPMENT

Synthetic biomedical signals, including electrocardiograms (ECG), electroencephalograms (EEG), electromyograms (EMG), and electrooculograms (EOG), were generated using mathematical models. To simulate realistic conditions, these signals were subjected to synthetically generated noise, including powerline interference, baseline wander, electrode motion artifacts, and white noise. The autoencoder model was trained on these noisy signals, using a 70/30 split for training and validation. Its architecture consisted of a symmetrical design with an encoder and decoder, each composed of two dense layers and a single dropout layer to mitigate overfitting. The model was compiled with the Adam optimizer and Huber loss for robust training. The trained model was then used to denoise the noisy signals, with the results visually and quantitatively compared to the original

clean signals. Figure 1 demonstrates the effectiveness of the autoencoder through visualizations that compare the original, noisy, and denoised signals.

Figure 1 – Comparison of Original, Noisy, and Denoised Signals



Source: Created by the author

The developed model successfully filtered out various types of noise from the simulated signals, achieving up to 99% noise reduction in the examples. This suggests its potential applicability to real-world data with appropriate modifications.

4. CHALLENGES AND LEARNINGS

The development of the autoencoder model presented several challenges. Generating realistic synthetic biomedical signals and accurately simulating real-world noise required careful consideration and experimentation. Preventing overfitting during training was also a key concern, addressed through the use of regularization techniques and early stopping. Through these challenges, valuable insights were gained into the intricacies of noise removal in biomedical signals and the importance of robust model design and training strategies.

5. REFERENCES

1. Anderson MR, Basu V, Martin RD, Reed CZ, Rowe NJ, Shafiee M, et al. Performance of a convolutional autoencoder designed to remove electronic noise from p-type point contact germanium detector signals. *Eur Phys J C*. 2022 Dec 1;82(12):1084.
2. Tran T, Bader S, Lundgren J. Denoising Induction Motor Sounds Using an Autoencoder [Internet]. arXiv; 2022 [cited 2024 Jul 23]. Available from: <http://arxiv.org/abs/2208.04462>
3. Wang J, Li S, Sung Y. Deformer: Denoising Transformer for Improved Audio Music Genre Classification. *Applied Sciences*. 2023 Nov 25;13(23):12673.

Uso de Inteligência Artificial para a Prevenção de Acidentes Vasculares Encefálicos

Estevão da Silva Erthal¹, João Guilherme Bebiana de Oliveira¹,
Fernando Diniz de Azevedo Ribeiro¹, Ana Beatriz Matos Cardoso¹,
André Felipe Miranda Ferreira¹, Gabriel Maranduba Littleton Lage¹, Anrafel Fernandes Pereira^{1,2}

¹ Universidade de Vassouras, Vassouras, RJ, Brasil.

² Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

estevaoerthal@gmail.com, joaogbo2014@gmail.com, nandoribeiro02@gmail.com,
anabeatriizcardoso@gmail.com, andrefelipemed109@gmail.com,
gabriel.lage@univassouras.edu.br, anrafel.pereira@univassouras.edu.br

Resumo: A prevenção do Acidente Vascular Encefálico (AVE) por monitoramento inteligente pode representar um avanço significativo na gestão de saúde, utilizando novos métodos tecnológicos para a detecção precoce e a intervenção preventiva. Este artigo explora como a Inteligência Artificial pode auxiliar na detecção de pacientes predispostos a serem acometidos por AVE, através de uma anamnese detalhada levando em consideração dados específicos, como estilo de vida e histórico familiar. Nesse contexto, é possível identificar padrões de risco e emitir alertas a pacientes e profissionais de saúde, possibilitando intervenções rápidas e personalizadas. A abordagem proposta não apenas melhora a capacidade de resposta e personalização dos cuidados, mas também tem o potencial de reduzir a incidência e a gravidade dos AVEs, resultando em melhores desfechos clínicos e qualidade de vida para os pacientes.

Palavras-chave: Acidente Vascular Encefálico; Monitoramento Inteligente; Inteligência Artificial; Prevenção; Saúde.

1. CONTEXTO E MOTIVAÇÃO

O acidente vascular encefálico (AVE) é uma patologia grave e frequente. No Brasil, o AVE é uma das principais causas de internações e mortalidade, resultando, na maioria dos pacientes, em algum tipo de deficiência, seja parcial ou completa (ALMEIDA, 2012). Além de representar um grande problema de saúde pública, seu impacto socioeconômico vem se mostrando crescente ao longo do tempo (SAPOSNIK, 2003).

Diante desse cenário, observa-se a necessidade de propor novos métodos de diagnósticos inteligentes que possam auxiliar na detecção precoce de pessoas predispostas a serem acometidas por um AVE. Tecnologias, como Inteligência Artificial (IA) e machine learning, têm o potencial de analisar grandes volumes de dados médicos e identificar padrões sutis que indicam risco elevado de AVE (MESKÓ & GÖRÖG, 2020). Essas ferramentas podem integrar-se aos sistemas de saúde para monitorar continuamente os pacientes, permitindo intervenções antecipadas e personalizadas. A adoção dessas inovações promete não apenas melhorar os desfechos clínicos, mas também reduzir significativamente a carga econômica e social associada ao AVE.

Com o propósito de deixar mais claro o problema que o grupo vem trabalhando, um Need Statement foi preparado para este estudo. O Need Statement tem o objetivo de propor “Uma maneira de promover o monitoramento da propensão ao Acidente Vascular Encefálico (AVE) em adultos com mais de 40 anos, com o objetivo de não só prevenir o seu desenvolvimento, mas também de promover intervenções precoces e personalizadas para melhorar a qualidade de vida e reduzir os riscos associados a essa condição vascular debilitante”. Para promover o monitoramento eficaz da propensão ao AVE em adultos com mais de 40 anos, foi proposto

o desenvolvimento de um sistema de cálculo da probabilidade do paciente sofrer um AVE. Este sistema visa não apenas prevenir o desenvolvimento da doença, mas também permitir intervenções precoces e personalizadas para melhorar a qualidade de vida e reduzir os riscos associados a essa condição vascular debilitante. A proposta é que os dados sejam coletados durante a anamnese, proporcionando uma visão abrangente da saúde vascular do paciente. A partir dos dados coletados algoritmos de IA possam auxiliar nas atividades que cabem ao monitoramento inteligente.

2. ATIVIDADES PRINCIPAIS

A proposta apresentada neste trabalho é um estudo inicial, que começou a ser desenvolvido como um desafio apresentado na disciplina optativa de Inteligência Artificial Aplicada à Saúde do curso de Medicina da Univassouras no primeiro semestre de 2024. As principais atividades desenvolvidas até o momento estão relacionadas com as etapas que permeiam a metodologia Challenge Based Learning (CBL) (CBL, 2022). A CBL está estruturada em três etapas, cada uma com objetivos e conceitos específicos, que guiam os aprendizes em um processo educacional dinâmico e interativo. A primeira etapa, Engage (“Engajamento”), é focada em motivar os aprendizes a se envolverem com um desafio prático. Já a segunda etapa, Investigate (Investigação) visa criar uma fundamentação de conhecimento sólida sobre o desafio escolhido a fim de propor mais a frente alguma solução que lide com ele. Por fim, a terceira etapa da metodologia, Act (Atuação), é destinada a desenvolver de fato alguma solução proposta.

3. DESENVOLVIMENTO DO TRABALHO

Este trabalho apresenta um projeto em fase inicial de pesquisa e desenvolvimento. Dentre os métodos inicialmente utilizados estão: (i) realização de revisão de literatura; (ii) análise dos resultados obtidos; e (iii) o desenvolvimento de uma proposta de solução inicial para apoiar a Prevenção do Acidente Vascular Encefálico (AVE) por meio do monitoramento inteligente.

O desenvolvimento deste protótipo foi realizado durante a terceira etapa da CBL, chamada de Act (Atuação). Esta fase é marcada por três conceitos importantes: (i) a criação da proposta de solução para lidar com o desafio identificado; (ii) a construção do protótipo da solução; e (iii) a implementação e a avaliação da solução. A proposta é que a aplicação integre informações sobre o estilo de vida e fatores genéticos do paciente para calcular as probabilidades individuais de ocorrência de um AVE. Os principais componentes incluem: (i) a coleta e análise de dados relacionados ao estilo de vida dos pacientes, como tabagismo, etilismo, obesidade e falta de atividades físicas; (ii) consideração do histórico familiar de AVE para identificar padrões hereditários; (iii) a utilização de algoritmos para calcular as probabilidades individuais de ocorrência de um AVE; (iv) a geração de recomendações personalizadas com base nos resultados da análise; e (v) o monitoramento contínuo da propensão ao AVE ao longo do tempo.

Como próximos passos espera-se realizar uma validação preliminar da proposta por meio da realização de um focus group com especialistas. A partir dos resultados aplicar as melhorias e orientações recomendadas e iniciar efetivamente o desenvolvimento da solução, buscando algoritmos que possam apoiar esta proposta. Entre os resultados esperados incluem a melhoria na qualidade de vida dos pacientes e a redução dos riscos associados ao AVE por meio de intervenções personalizadas e monitoramento contínuo.

4. DESAFIOS E APRENDIZADOS

Os principais desafios ao desenvolver o artigo incluíram a complexidade técnica, revisão de literatura e interdisciplinaridade. Compreender e integrar tecnologias avançadas, como IA e machine learning, no contexto médico exigiu um aprofundamento significativo desses temas para

desenvolver uma proposta eficaz para o problema. A coleta e análise de uma vasta quantidade de estudos científicos sobre AVE e monitoramento inteligente demandaram tempo e esforço consideráveis, sendo crucial identificar as fontes mais relevantes e atuais para fundamentar o trabalho. Além disso, integrar conhecimentos de medicina, tecnologia da informação e estatística exige colaboração entre profissionais de diversas áreas.

Entre os principais aprendizados, destacou-se a importância da prevenção, onde a pesquisa reforçou que identificar pessoas em risco e intervir precocemente pode salvar vidas e melhorar significativamente a qualidade de vida dos pacientes. O uso de tecnologias, como a IA, mostrou-se promissora na detecção precoce e no monitoramento contínuo de pacientes em risco de AVE, com a capacidade de analisar grandes volumes de dados e identificar padrões sutis, sendo uma vantagem significativa. Por fim, a interdisciplinaridade se mostrou fundamental para desenvolver soluções inovadoras e eficazes, com a integração de perspectivas diversas enriquecendo o projeto e contribuindo para a criação de um sistema robusto.

5. REFERÊNCIAS

Almeida SRM. Análise epidemiológica do acidente vascular cerebral no Brasil. Rev Neurociência. 2012;20(4):481-2.

CBL.Apple Inc. 2024. ChallengeBasedLearning. Disponível em: <https://www.challengebasedlearning.org/framework/>. Acesso em: 08 Jul. 2024.

Meskó, Bertalan; Görög, Marton. A short guide for medical professionals in the era of artificial intelligence. NPJ digital medicine, v. 3, n. 1, p. 126, 2020.

Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(7):2064-89.

Saposnik G, Del Brutto OH. Stroke in South America: a systematic review of incidence, prevalence, and stroke subtypes. Stroke. 2003;34(9):2103-7.

Utilização de IA Generativa na Construção do Plano Diretor Estratégico (PDE) em um Hospital Universitário: Uma Experiência Exitosa

Gustavo A. Silva ¹, Leticia F. S. Machado ¹

¹ Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

gustavo.andrade2a@gmail.com, leticia.fmachado@ebserh.gov.br

Resumo: Desde 2014, os Hospitais Universitários Federais (HUFs) têm implementado Planos Diretores Estratégicos (PDEs) para padronizar a gestão e enfrentar desafios. Os hospitais da rede EBSEH seguem a metodologia estruturada pela Portaria SEI Nº 3 de 06/02/2024, envolvendo análises e contribuições de diversos grupos da comunidade local. A metodologia instituída é distribuída em três fases distintas sendo contemplado na primeira fase a participação da comunidade local com sua percepção e seus desejos para a futura estratégia empregada pelo hospital. Utilizando questionários eletrônicos, coletaram-se percepções da comunidade sobre a instituição, gerando um grande volume de dados. A inteligência artificial generativa, exemplificada pelo ChatGPT, foi aplicada para resumir as respostas dos questionários, transformando dados extensos em informações acionáveis. Os questionários, direcionados a diferentes grupos, incluíam perguntas sobre desafios, oportunidades e estratégias. O uso do ChatGPT provou ser uma solução eficaz, apesar dos desafios relacionados à capacitação dos usuários e à necessidade de revisões éticas.

O processo destacou a importância da validação contínua das informações geradas, resultando em um ganho institucional significativo, com profissionais e acadêmicos capacitados no uso da IA.

Palavras-chave: Inteligência Artificial, Gestão por Resultados, Hospital escola, Planos Diretores Estratégicos, ChatGPT, Análise de Dados.

Nome do projeto: Gestão por Resultados (extensão universitária)

1. CONTEXTO E MOTIVAÇÃO

Desde 2014, os HUFs definem sua estratégia por meio da elaboração de Planos Diretores Estratégicos (PDEs). Esses Planos possibilitaram padronizar a forma de identificar os desafios de gestão e criar estratégias para transpô-los, apresentando-se como um avanço na gestão da estratégia das unidades hospitalares.

Nesse contexto, o HUF relatado realiza a construção do seu PDE 2024-2028 percorrendo as seguintes etapas: Fase 1 - Direcionamento Estratégico: visa estabelecer uma análise situacional, envolvendo a definição de visão do HUF e identificação de macroproblemas associados aos objetivos estratégicos. Fase 2 - Planejamento de Indicadores e Projetos: busca desenvolver um portfólio que descreva os indicadores e metas locais e a definição dos projetos locais, respeitando-se, no entanto, as singularidades de cada HUF envolvido neste processo. Fase 3 – Monitoramento, Controle e Replanejamento: estabelece o processo de monitoramento e controle dos indicadores e projetos locais subsidiando a tomada de decisão para o replanejamento necessário ao longo da vigência do PDE

Para a elaboração deste documento, utiliza-se a referência do “Guia de Desdobramento da Estratégia para os Hospitais da Rede Ebserh”, instituído pela Portaria SEI Nº 3 de 06 de fevereiro de 2024. A partir dele se faz essencial na fase 1 a etapa de definição dos macroproblemas e causas raízes que conte com a contribuição da comunidade local, sendo eles: a Reitoria da UFMG, gestor SUS, trabalhadores da instituição, acadêmicos, conselhos de saúde e colegiados gestores.

Para tanto, foi elaborado formulários eletrônicos, para cada grupo, com o objetivo de colher a percepção desses atores sobre a instituição e suas demandas. O formulário toma formato de questionário contendo questões abertas. Diante de um grande volume de respondentes, houve um consequente aumento na quantidade de dados coletados. Essa realidade exigiu a adoção de uma abordagem inovadora, ainda não estabelecida na instituição, para a análise e a incorporação desses dados no PDE. Nesse contexto, a aplicação da inteligência artificial generativa, exemplificada pelo ChatGPT, emergiu como uma solução promissora para transformar grandes quantidades de dados em informações acionáveis.

1.1 Objetivo

O objetivo deste trabalho é relatar a aplicação da inteligência artificial generativa na análise de respostas de questionários utilizados no processo de reformulação do PDE em um HUF mineiro.

2. ATIVIDADES PRINCIPAIS

As atividades principais desenvolvidas neste estudo incluíram:

- Formulação dos instrumentos de coleta e as perguntas que serão abordadas
- Separação e organização dos questionários por grupo de contribuintes.
- Aplicação do ChatGPT para sumarizar as respostas dos questionários.
- Validação das informações geradas.
- Integração dos resumos gerados pela inteligência artificial após revisão ao processo de construção do PDE prediction and diagnosis (2). BNs explicitly represent causal structures, making them easily adaptable to new data and to expert opinions.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A aplicação da IA generativa ocorreu somente na fase 1 do desenvolvimento do PDE provendo os resultados para as fases seguintes. A metodologia adotada seguiu um fluxo estruturado que compreendia três principais processos: (1) coleta de dados, (2) aplicação da IA para análise e sumarização, e (3) validação dos resultados. Na fase de coleta de dados, questionários eletrônicos foram desenvolvidos utilizando o Microsoft Forms, direcionados a seis grupos de interesse, incluindo gestores, acadêmicos e conselhos de saúde. As respostas, majoritariamente qualitativas, foram processadas pela IA generativa para reduzir o volume e identificar os principais temas. Na fase de validação, os resumos gerados pela IA foram comparados com análises manuais para garantir que as informações relevantes fossem preservadas e que vieses fossem minimizados. Todo o processo foi acompanhado de capacitação contínua dos envolvidos no uso da tecnologia

O principal objetivo de utilizar o ChatGPT foi auxiliar na análise das respostas dos questionários, que geraram aproximadamente 250 contribuições qualitativas. Assim utilizou-se a IA para resumir as grandes quantidades de textos em ideias sintéticas [1]. Um aspecto crucial do processo foi a validação do prompt utilizado para gerar os resumos através do ChatGPT. Após a primeira rodada de resumos, foram realizadas revisões do prompt para ajustar a especificidade das respostas. O prompt inicial, que solicitava uma sumarização simples, foi refinado para incluir instruções mais detalhadas, como a necessidade de identificar e separar temas-chave de cada resposta e considerar o contexto das respostas para evitar generalizações indevidas. Segundo Dergaa et al [2], com o devido treinamento, é possível gerar resumos objetivos e imparciais por meio da inteligência artificial.

4. DESAFIOS E APRENDIZADOS

O uso do ChatGPT trouxe alguns desafios, como a necessidade de capacitar os profissionais para

operarem a ferramenta de forma eficaz. Houve resistência inicial ao uso da IA, principalmente devido à incerteza sobre a confiabilidade dos resumos gerados. A equipe enfrentou dificuldades para ajustar os prompts de forma que gerassem resumos consistentes e precisos. Foram necessárias várias iterações até que os resumos refletissem com precisão as respostas originais. Além disso, a necessidade de garantir a imparcialidade dos resumos foi um dos principais desafios enfrentados, especialmente considerando que as opiniões dos respondentes variavam amplamente.

Além disso, a literatura sugere que essas ferramentas podem gerar informações imprecisas e necessitam de revisões éticas quanto à sua aplicação [3]. Sendo assim é fundamental discutir os possíveis vieses introduzidos pela ferramenta. A IA, ao resumir grandes quantidades de dados, pode privilegiar determinados padrões linguísticos ou contextuais, o que pode impactar a representatividade das respostas. Houve um cuidado especial para que os profissionais realizassem uma validação ética das informações, revisando possíveis distorções geradas pela IA. Além disso, uma discussão mais aprofundada sobre os desafios éticos se faz necessária, especialmente em relação ao uso de IA em ambientes de gestão pública e hospitalar, onde a precisão dos dados pode impactar diretamente na formulação de políticas e estratégias institucionais. Foram implementados mecanismos de controle para minimizar esses vieses, todos os resultados foram validados em reuniões colegiadas.

Essa experiência resultou em ganhos para a instituição, na medida que conta com mais profissionais e acadêmicos proficientes no uso de tecnologias de IA, o que pode beneficiar futuras iniciativas de gestão e análise de dados no Hospital escola.

5. REFERÊNCIAS

1. Sampaio RC, Chagas V, Sanchez CS, Gonçalves J, Borges T, Alison MB, et al. Uma revisão de escopo assistida por inteligência artificial (IA) sobre usos emergentes de ia na pesquisa qualitativa e suas considerações éticas. *Revista Pesquisa Qualitativa* [Internet]. 2024 May 1;12(30):01-28. Available from: <https://editora.sepq.org.br/rpq/article/view/729>
2. Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*. 2023;40(2):615–22.
3. Ankita Guleria, Krishan K, Sharma V, Tanuj Kanchan. ChatGPT: ethical concerns and challenges in academics and research. *Journal of Infection in Developing Countries* [Internet]. 2023 Sep 30; 17(09):1292–9. Available from: <https://jidc.org/index.php/journal/article/view/37824352>

Realização



Apoio



Parceria



Financiadores

