

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS



Centro de Inovação em
Inteligência Artificial

ANAIS DO III SIMPÓSIO DO CI-IA SAÚDE UFMG

Cristiane dos Santos Dias
Elisa Tuler de Albergaria

Antônio Luiz Pinho Ribeiro
Adriano César Machado Pereira
Virgílio Augusto Fernandes Almeida
Wagner Meira Jr



Belo Horizonte, MG
2025



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.

Ficha catalográfica:

S612 Simpósio CI-IA Saúde da UFMG (3. : 2025 : Belo Horizonte, MG).
Anais do III Simpósio CI-IA Saúde da UFMG [recurso eletrônico] : inteligência artificial responsável na saúde. Da pesquisa à prática clínica. / organizado por Cristiane dos Santos Dias, Elisa Tuler Albergaria. - Belo Horizonte : Faculdade de Medicina da UFMG, 2025.
141f.

Requisitos do sistema: Adobe Reader

1. Inteligência Artificial. 2. Saúde. 3. Assistência à saúde. I. Dias, Cristiane dos Santos. II. Albergaria, Elisa Tuler. III. Anais do III 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

ISBN 978-65-978751-0-8

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 Comi-
tê 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:
Adriano César Machado Pereira

**Coordenadora de Educação e Difusão do Conhecimento do Comitê Exe-
cutivo - UFMG:** Elisa Tuler de Albergaria

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:
Sergio Adriano Loureiro Bersan

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

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

Letícia Santos Neto - Universidade Federal de Minas Gerais - UFMG

Fabiana Costa Peixoto - Universidade Federal de Minas Gerais - UFMG

Camila dos Reis Cunha - Universidade Federal de Minas Gerais - UFMG

Francyne Nascimento Santana - Universidade Federal de Minas Gerais - UFMG

Gilmara Viviane Silva Stolé - Universidade Federal de Minas Gerais - UFMG

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

André Chagas Lima - Universidade Federal de São João del-Rei - UFSJ

Aline Cristina Miranda Marinho Costa - Universidade Federal de Minas Gerais - UFMG

Davi Greco Resende Freitas - Universidade Federal de São João del-Rei - UFSJ

Enzo Furtado Goulart - Universidade Federal de Minas Gerais - UFMG

João Pedro Reis Moura - Universidade Federal de São João del-Rei - UFSJ

Mizael Andrade Rodrigues - Universidade Federal de Minas Gerais - UFMG

Noélida Mendes Rodrigues de Almeida - Universidade Federal de Minas Gerais - UFMG

Raphael de Campos Borges - Universidade Federal de Minas Gerais - UFMG

Comissão Científica

Coordenadora 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

Ana Paula Couto da Silva

Universidade Federal de Minas Gerais - UFMG

Antônio Pereira de Souza Júnior

Universidade Federal de São João Del Rei - UFSJ

Carla Vasconcelos Carvalho

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

Flávia Ribeiro de Oliveira

Universidade Federal de Minas Gerais - UFMG

Gabriella da Silva Reis

Universidade Federal de Minas Gerais - UFMG

Guilherme Dal Bianco

Universidade Federal da Fronteira Sul (UFFS)

Helena Duani

Universidade Federal de Minas Gerais - UFMG

Henrique Araújo Lima

Universidade Federal de Minas Gerais - UFMG

Isabela Nascimento Borges

Universidade Federal de Minas Gerais - UFMG

Isabella Viana Gomes Schettini

Universidade Federal de Minas Gerais - UFMG

João Paulo Pereira de Almeida

Universidade Federal de Minas Gerais - UFMG

Juliana Vieira Queiroz Almeida Oliveira

Universidade Federal de Minas Gerais - UFMG

Juliano de Souza Gaspar

Universidade Federal de Minas Gerais - UFMG

Leonardo Chaves Dutra da Rocha

Universidade Federal de São João Del Rei - UFSJ

Luciana Reis da Silveira Fundação

Fundação Hospitalar do Estado de Minas Gerais - FHEMIG

Luciano Lima

Universidade Federal de Minas Gerais - UFMG

Marcos André Gonçalves

Universidade Federal de Minas Gerais - UFMG

Maria do Carmo Barros

Universidade Federal de Minas Gerais - UFMG

Milena Soriano Marcolino

Universidade Federal de Minas Gerais - UFMG

Priscila Menezes Ferri Liu

Universidade Federal de Minas Gerais - UFMG

Raphael Augusto Teixeira de Aguiar

Universidade Federal de Minas Gerais - UFMG

Editorial

III 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 2025, o III Simpósio do CI-IA Saúde: IA Responsável na Saúde – Da Pesquisa à Prática Clínica foi realizado em 25 de setembro de 2025, das 8h às 17h30min, no Centro de Atividades Didáticas 3 (CAD 3), no Campus Pampulha da UFMG, em Belo Horizonte (MG). O evento contou com apresentações dos projetos desenvolvidos pelos pesquisadores do CI-IA-Saúde, apresentação de pôsteres de alunos da graduação e pós-graduação com resultados de trabalhos da aplicação da IA na área da saúde. O evento contou com ampla adesão de participantes, com 486 inscritos e 198 credenciados no dia do evento.

A abertura do simpósio foi marcada pela fala do diretor executivo do CI-IA Saúde, Prof. Virgílio Almeida, que ressaltou a importância da colaboração efetiva entre diferentes frentes para que os avanços em IA se traduzam em benefícios concretos para a população. Ele reforçou a visão multidisciplinar e multi-institucional do Centro e destacou o potencial impacto de soluções baseadas em IA para enfrentar desafios nacionais em saúde, sempre ancoradas em princípios de ética, transparência e responsabilidade.

Dentre os destaques da programação, a mesa-redonda “Pesquisa e inovação no CI-IA Saúde” reuniu diferentes áreas da UFMG para discutir os projetos e perspectivas do Centro. Foi mediada por Ana Paula Couto da Silva (Departamento de Ciência da Computação da UFMG) e contou com as falas de Deborah Carvalho Malta (Departamento de Enfermagem Materno-Infantil e Saúde Pública, Escola de Enfermagem, UFMG), Saulo Fernandes Saturnino (Faculdade de Medicina da UFMG) e Yuri Alexandre dos Santos (Centro de Pesquisa em Direito, Tecnologia e Inovação – DTIBR), evidenciando uma visão interdisciplinar de pesquisa e inovação em inteligência artificial aplicada à saúde.

A palestra “Saúde digital em debate: concepções, caminhos e propostas” trouxe o professor Naomar Monteiro de Almeida Filho (Instituto de Saúde Coletiva da UFBA) como palestrante e o professor Virgílio Almeida (Departamento de Ciência da Computação da UFMG) como mo-

derador, e discutiu diferentes visões sobre saúde digital, apresentando caminhos e propostas para o uso responsável de tecnologias digitais e inteligência artificial na organização dos sistemas de saúde.

A mesa-redonda “IA em escala: desafios da implementação na prática assistencial” discutiu como levar soluções de inteligência artificial do ambiente de pesquisa para o uso rotineiro nos serviços de saúde, enfrentando barreiras técnicas, regulatórias e organizacionais. Foi mediada por Antonio Luiz Pinho Ribeiro (Faculdade de Medicina da UFMG) e contou com as contribuições de Isabela Nascimento Borges (Faculdade de Medicina da UFMG), Robson Willian de Melo Matos (Ministério da Saúde – DATASUS) e Sergio A. L. Bersan (Unimed-BH), que compartilharam experiências práticas de implementação de IA em diferentes contextos assistenciais.

A palestra “To AI or not to AI, that is the question” discutiu de forma crítica sobre as soluções de inteligência artificial na saúde, ponderando benefícios, riscos e limites dessas tecnologias na prática clínica e na gestão. Teve como palestrante James Batchelor (University of Southampton – Clinical Informatics Research Unit) e foi moderada por Wagner Meira (Departamento de Ciência da Computação da UFMG), que conduziu o debate sobre decisões responsáveis de uso de IA em diferentes contextos de cuidado.

O evento encerrou-se com a premiação dos três melhores projetos, que abordaram temas centrais para a aplicação da inteligência artificial e da ciência de dados em saúde: “Avaliação multidimensional de redes generativas adversariais, modelos de difusão e métodos estatísticos para geração de dados sintéticos em saúde”; “Desafios para a previsão de custos hospitalares na admissão: uma análise baseada em previsões de variáveis intermediárias para internações por COVID-19”; e “Desenvolvimento de um chatbot inteligente para otimizar o processo de suporte ao usuário nos sistemas e-SUS APS”.

O eBook com os Anais do simpósio disponibiliza os 44 resumos científicos aprovados pela revisão por pares e apresentados no evento. Agradecemos aos palestrantes, autores, revisores e organizadores pelo empenho e dedicação que fizeram deste evento um sucesso. Que o III 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.

Elisa Tuler de Albergaria

Coordenadora do III Simpósio do CI-IA Saúde: IA Responsável na Saúde - Da Pesquisa à Prática Clínica

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

Sumário

| | |
|--|----|
| 1. AI-Driven in silico screening of Lapachol derivatives as potential therapeutics against Chagas Disease | 10 |
| 2. Análise e previsão de casos de dengue com Machine Learning: um enfoque municipal: resumo expandido - inovação | 13 |
| 3. Análise e visualização de dados de monitoramento glicêmico em pacientes críticos | 16 |
| 4. Aplicação da predição do encaminhamento de retinopatia diabética ao especialista em um projeto real | 19 |
| 5. Aplicação e avaliação comparativa de algoritmos de Machine Learning para o diagnóstico de câncer de mama a partir de características nucleares | 22 |
| 6. Avaliação de large language models para o reconhecimento de entidades nomeadas em prontuários médicos | 25 |
| 7. Avaliação multidimensional de redes generativas adversariais, modelos de difusão e métodos estatísticos para geração de dados sintéticos em saúde | 28 |
| 8. Avaliando os potenciais do algoritmo k-vizinhos mais próximos (KNN) na tarefa imputação em bases de dados clínicos | 31 |
| 9. Caracterização de atributos para predição de complicações graves do diabetes com base em registros administrativos de saúde | 34 |
| 10. Desafios éticos e legais da inteligência artificial na telessaúde: perspectivas sobre proteção de dados no Brasil | 37 |
| 11. Desafios para a previsão de custos hospitalares na admissão: uma análise baseada em predições de variáveis intermediárias para internações por COVID-19 | 40 |
| 12. Desenho e avaliação de taxonomia de suporte para um Chatbot Inteligente de assistência ao usuário do sistema e-SUS APS | 43 |
| 13. Desenvolvimento de um chatbot inteligente para otimizar o processo de suporte ao usuário nos sistemas e-SUS APS | 46 |
| 14. Divulgação científica sobre segurança medicamentosa: análise de formatos, qualidade e engajamento no instagram no projeto IA-Polis | 49 |
| 15. Framework bioético no design e implementação de ferramenta de IA para suporte à prescrição no SUS: experiência do projeto IA-Polis | 52 |
| 16. Gamificação com Inteligência Artificial no ensino de auditoria em saúde: uma experiência com a plataforma Rosebud | 55 |
| 17. Impacto da qualidade dos dados em indicadores perinatais: uma abordagem simulada com apoio da Inteligência Artificial | 58 |
| 18. Implementação de inteligência artificial no Sistema Único de Saúde do estado de São Paulo: redução de custos e aumento da precisão de diagnóstico em mamografias | 61 |
| 19. In silico analysis of potential anti-inflammatory agents for psoriasis | 64 |
| 20. Inovações na análise de dados sobre a hemorragia puerperal: ampliação de dados, estratégias de risco e validação com Inteligência Artificial | 67 |
| 21. Inteligência artificial como ferramenta para inclusão de pacientes oncológicos em cuidados paliativos: uma revisão de escopo | 70 |
| 22. Inteligência artificial no manejo do sangramento uterino anormal: modelos preditivos para algoritmo de tratamento personalizado | 72 |

| | |
|---|-----|
| 23. Interpretable Machine Learning for arboviruses diagnosis: a XAI approach | 75 |
| 24. Interpretação automatizada do exame de papanicolaou para triagem do câncer de colo do útero | 78 |
| 25. Letramento em inteligência artificial para a saúde: uma abordagem prática e inclusiva para idosos | 81 |
| 26. Modernização do sistema de gerenciamento de laudos anatomopatológicos no HC-UFMG: extração automatizada de dados e interoperabilidade com ferramentas analíticas | 84 |
| 27. O uso da Inteligência artificial no processo formativo do estudante de medicina: uma revisão da literatura | 87 |
| 28. Otimização inteligente de estoques em hemocentros: aplicação de algoritmos de aprendizado de máquina | 90 |
| 29. PediAI: uma ferramenta para auxiliar estudantes e pediatras no processo de raciocínio clínico e diagnóstico utilizando grandes modelos de linguagem | 93 |
| 30. Plain Language, IA e saúde digital: construindo pontes para compreender e cuidar | 96 |
| 31. Potencialidades da Inteligência Artificial na gestão do cuidado do enfermeiro | 99 |
| 32. Predição de complicações cardiovasculares em pacientes com covid-19 hospitalizados utilizando modelos supervisionados de aprendizado de máquina e técnica de inteligência artificial explicável | 102 |
| 33. Predição de mutações oncogênicas em genes reguladores associados ao câncer através de Redes Neurais de Grafos | 105 |
| 34. Predição de sobrevida de curto prazo em pacientes oncológicos na emergência utilizando aprendizado de máquina | 108 |
| 35. Predicting weekly dengue incidence in brazilian federative units | 111 |
| 36. Taxonomia de desinformação sobre vacina no WhatsApp | 114 |
| 37. Uma abordagem usando modelos generativos para apoio à decisão clínica | 117 |
| 38. Uma análise exploratória da (des)informação sobre vacinação em grupos públicos no WhatsApp no Brasil | 120 |
| 39. Uma análise qualitativa da sumarização de diálogos de pacientes usando grandes modelos de linguagem aplicados a dados em português ruidosos e informais | 124 |
| 40. Uma revisão sistemática sobre o uso de aprendizado profundo na interpretação do exame de Papanicolaou | 126 |
| 41. Uso abusivo de bebidas alcoólicas: uma abordagem baseada em inquéritos e geração de dados sintéticos | 129 |
| 42. Uso de populações sintéticas em pequenas áreas geográficas para desfechos relacionados a Doenças Crônicas Não Transmissíveis: uma revisão de escopo | 135 |
| 43. Uso de Redes Generativas Adversariais (cGANs) para geração de populações sintéticas no apoio às políticas públicas em saúde | 135 |
| 44. Utilização de RAG em grandes modelos de linguagem para aprimorar instruções de uso de medicamentos | 138 |

AI-Driven in silico screening of Lapachol derivatives as potential therapeutics against Chagas Disease

T. M. Machado^{1*}, D. E. M. de Souza¹, J T. T. Soares^{1*}, F. Guilhon-Simplicio^{1,2}

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

²Center for Innovation in Artificial Intelligence for Health (CIIA – Saúde/UFMG),
Belo Horizonte, MG, Brazil

* Scholarship of the Fundação de Amparo à Pesquisa de São Paulo of the Amazon IA project tallita.
machado@yahoo.com.br, dougemanuel40@gmail.com, joaotiagotss@gmail.com,
guilhon_simplicio@ufam.edu.br

Abstract: Chagas disease is a neglected tropical disease caused by the protozoan *Trypanosoma cruzi*, which is endemic in several countries and affects 6 to 7 million people worldwide. Given the few and obsolete therapeutic options, this study sought to design, through in silico studies, lapachol derivatives that had potential activity against Chagas disease and that presented an ideal pharmacokinetic profile for drug candidates. Thus, a protein-protein interaction network analysis was performed to determine the most promising target, which culminated in the HSP⁹⁰ enzyme, whose structure was obtained in AlphaFold and aligned through analysis in Pymol. Ten lapachol derivatives were designed and evaluated for their interaction profile with the HSP90 enzyme, where derivatives D¹, D², D³ and D⁵ obtained a complementarity profile with the enzyme in addition to a pharmacokinetic profile suitable for their application as a drug. Thus, analyses using artificial intelligence platforms allowed not only the optimization of the evaluated molecules, but also the integration of information with greater precision in the search for new therapeutic options.

Keywords: docking molecular; virtual screening; neglected diseases.

Project name: AmazonIA - Applying Machine Learning and Deep Learning Models to Analyze Data on Neglected Tropical Diseases in the Amazon.

1. CONTEXT AND MOTIVATION

Chagas Disease is part of the group of Neglected Tropical Diseases, which has been endemic to the Americas for more than nine millennia and affect more than 7 million people worldwide, mainly in Latin America, which has an annual mortality rate of 10 thousand deaths (1). Chagas disease or American trypanosomiasis is an infection caused by the protozoan *Trypanosoma cruzi*. It has an acute phase that may or may not be symptomatic, and a chronic phase, which may manifest in indeterminate (asymptomatic), cardiac, digestive or cardiodigestive forms (2).

The treatment of Chagas Disease consists of the use of nifurtimox and benznidazole, however, the effectiveness of nifurtimox and benznidazole is restricted to the initial phases of the disease, they are not suitable for the treatment of chronic cases and they can impart various collateral effects. Hence, the discovery of new drugs or drug repositioning is a top priority (3). The use of natural products represents a key source for obtaining bioactive compounds, with medicinal chemistry, particularly through strategic semi-synthetic modifications, serving as a tool in the development of new drug candidates (4). Among natural products, lapachol stands out, which is a natural 1,4 - naphthoquinone with a chemical structure belonging to the quinone class, which has antimicrobial, antioxidant, antiviral and antitumor effects, mainly due to the redox capacity of its carbonyl groups in the main chain (5).

In structure-based drug design, potential ligands for a known receptor are first screened and selected in silico before proceeding to the synthesis of the most promising compounds (6). Among the most widely used techniques in this approach is molecular docking, due to its ability to predict, with considerable accuracy, the affinity of a molecule for a biological receptor based on the conformation, type and energy

of the intermolecular interactions involved. Software tools that employ genetic algorithms, such as AutoDock, associated with artificial intelligence are increasingly used to assess the binding affinity of numerous ligands to a target receptor, making them one of the most promising options for virtual screening (7).

1. 1 Aim

To evaluate the potential inhibitory activity of lapachol and derivatives against Chagas disease enzymes in silico.

2. MAIN ACTIVITIES

- Intercession analysis and interaction network to determine the pharmacological target.
- Evaluate the ability to inhibit enzymes related to the metabolism of *Trypanosoma cruzi*. in silico;
- Evaluate the pharmacokinetic parameters of the proposed derivatives through in silico methods.

3. WORK DEVELOPMENT (METHODOLOGY AND RESULTS)

3.1 Data Mining

The molecular targets of the disease were catalogued in GeneCards®, prioritizing essential proteins, while the targets of the molecules were identified via SwissTargetPrediction®.

3.2 Development of protein-protein interaction network

The intersection of data between genes involved in Chagas Disease and lapachol targets was analyzed by Venn diagram and protein-protein interaction networks were constructed in NetworkAnalyst® identifying central nodes and priority biological pathways.

3.3. Docking molecular

Molecular docking analysis was performed using Autodock Vina® software. The evaluated target, HSP90, was obtained in (.pdb) format from the RCSB Protein Data Bank, code 3O6O, and from the Alpha-Fold Protein Structure Database, code AF-P06660-F1-v4. Water molecules and other molecules used in protein crystallization not required for catalytic activity were removed and the two proteins had their structures aligned 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 8x10x6 Å, with a grid spacing of 1. The results were visualized in AutoDock Tools® and 2D images of molecular interactions were obtained using Discovery Studio Visualizer 4.0® software.

3.4. Prediction of pharmacokinetic parameters

The prediction of pharmacokinetic parameters was carried out using AI Drug Lab platform (<https://ai-druglab.smu.edu/admet>).

3.5. Results

Initial analysis identified 648 targets associated with Chagas disease and 115 targets related to lapachol. Venn diagrams revealed 11 common targets. Protein-protein interaction network analysis of these 11 targets highlighted 4 as particularly interesting and essential for the biological processes of Chagas disease. Molecular docking was then performed with the target considered to have the greatest pathological

potential, the enzyme HSP90, which is an enzyme that is involved in multiple stages of the parasite life cycle and aids in parasite proliferation and survival (8). The enzyme obtained from the AlphaFold platform was optimized by aligning the amino acid sequence in comparison with the *Trypanosoma brucei* HSP90 enzyme, obtaining an alignment with a score of 0.80, indicating that it is aligned in the same position as the reference enzyme, which had its molecular docking method 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 molecular docking. After alignment, Ten lapachol derivatives were proposed using homology strategies, which were coded from D1 to D10 and molecular docking analysis was performed where the derivatives D1 to D7 and lapachol obtained more than 1 hydrogen bond and only D1 (-8.6 kcal/mol), D2 (-8.7 kcal/mol), D3 (-8.5 kcal/mol) and D5 (-8.9 kcal/mol) presented a higher affinity energy than the lapachol (-7.9 kcal/mol) and ligand molecule (-7.6 kcal/mol), characterizing the high affinity of the molecules for the HSP90 enzyme.

In relation to pharmacokinetic parameters, the use of artificial intelligence models stands out, which demonstrated that no molecule evaluated presented a carcinogenicity 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.

This study led to the identification of lapachol as a promising molecule for the development of molecules that can be applied in the study of inhibitors related to targets involved in the metabolism of Chagas Disease and that Artificial intelligence platforms coupled with molecular docking allow the optimization of the pharmacological evaluation.

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 an analytical model that can be applied to the research of new drugs for neglected diseases, in a robust, intuitive and secure way, which allows efficient access and manipulation of data, and is 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 optimizing its skills in bioinformatics, data science, use of database management tools and statistical analysis to carry it out.

5. REFERENCES

1. Cucunubá ZM et al. The epidemiology of Chagas disease in the Americas. *Lancet Reg Health Am.* 2024 Sep 13(37):100881.
2. Miranda CSC et al. Epidemiological Scenario of American Trypanosomiasis and Its Socioeconomic and Environmental Relations, Pará, Eastern Brazilian Amazon. *Trop. Med. Infect. Dis.* 2025 10(4):88.
3. Goldenberg S. Chagas disease treatment: a 120-year-old challenge to public health. *Mem. Inst. Oswaldo Cruz* 2022 (117).
4. Xie S et al. The latest advances with natural products in drug discovery and opportunities for the future: a 2025 update. *Expert Opin Drug Discov* 2025.
5. Almeida ZL et al. Lapachol, a natural food component, interacts with human serum albumin: Insights of its impact on the pharmacokinetics of clinically used drugs. *Int J Biol Macromol.* 2024 (282):137520.
6. Agu PC et al. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Sci Rep* 2023 (13):13398.
7. Fadahunsi AA et al. Revolutionizing drug discovery: an AI-powered transformation of molecular docking. *Med Chem Res* 2024 33:2187.
8. Jamabo M et al. In silico analysis of the HSP90 chaperone system from the African trypanosome, *Trypanosoma brucei*. *Front Mol Biosci* 2022 23(9):947078.

Análise e previsão de casos de dengue com Machine Learning: um enfoque municipal: resumo expandido - inovação

Marta Motter Magri ¹, Satiro Marcio Ignacio Junior ¹, Mercedes Morato ¹

¹*Vigilância Epidemiológica da Prefeitura Municipal de Osasco, Osasco, SP, Brasil.*

marta.ss@osasco.sp.gov.br, satiro.ss@osasco.sp.gov.br, mercedes.ve.ss@osasco.sp.gov.br

Resumo: A dengue é uma arbovirose de notificação compulsória que representa um grave problema de saúde pública no Brasil, especialmente em contextos urbanos como o município de Osasco. Diante da emergência decretada em 2025 devido ao aumento expressivo de casos, este estudo teve como objetivo identificar padrões de distribuição da doença por bairro e prever áreas de maior risco, visando subsidiar ações de prevenção e controle. Foram utilizados dados do Sistema de Informação de Agravos de Notificação (Sinan) dos anos de 2023, 2024 e início de 2025. A análise foi conduzida na linguagem Python, utilizando bibliotecas como Pandas, RapidFuzz e Scikit-learn. Após a padronização dos dados, com destaque para a correção dos nomes de bairros via fuzzy matching, foi aplicado um modelo de aprendizado de máquina do tipo Random Forest Regressor. O modelo foi treinado com 70% dos dados e testado com os 30% restantes, atingindo um R2 de 0,678 e erro absoluto médio (MAE) de 13,12 casos. Os resultados preliminares demonstram a viabilidade da aplicação de técnicas de ciência de dados na vigilância epidemiológica municipal, contribuindo para decisões mais assertivas e oportunas no enfrentamento da dengue.

Palavras-chave: dengue; Python; Machine Learning; Vigilância Epidemiológica.

1. CONTEXTO E MOTIVAÇÃO

O papel da Vigilância Epidemiológica é essencial na promoção e proteção da saúde pública, envolvendo o monitoramento de doenças e agravos à saúde, a coleta e análise de dados, e a implementação de ações preventivas e de controle ⁽¹⁾.

As arboviroses são transmitidas pelo mosquito do gênero *Aedes*, especialmente a espécie *Aedes aegypti*. Dentre elas, a dengue é a de maior relevância nas Américas, sendo considerada um dos principais problemas de saúde pública. O vírus da dengue (DENV) é um RNA-vírus pertencente ao gênero *Flavivirus*, da família *Flaviviridae*, e possui quatro sorotipos conhecidos ⁽²⁾. Trata-se de uma doença febril aguda, sistêmica e dinâmica, com ampla variedade clínica, que vai desde casos assintomáticos até formas graves com risco de óbito ⁽¹⁾. A notificação da dengue é compulsória, e os registros devem ser inseridos no Sistema de Informação de Agravos de Notificação (Sinan). A agilidade na coleta, análise e disseminação dos dados é crucial para a resposta oportuna dos órgãos de saúde, permitindo uma atuação integrada e eficaz ⁽²⁾.

Diante da situação epidemiológica, o Estado de São Paulo decretou emergência em saúde pública em 19/02/2025, e o município de Osasco adotou a mesma medida em 17/03/2025. Este trabalho foi motivado pela necessidade de direcionar as ações de controle de forma mais eficaz, em parceria com o Centro de Controle de Zoonoses.

1.1 Objetivo

Os objetivos deste trabalho são:

- Conhecer o padrão de distribuição espacial e temporal dos casos notificados de dengue em Osasco (majoritariamente autóctones — ou seja, com infecção contraída dentro da área estudada), com base nos bairros de residência;
- Prever os bairros com maior risco de concentração de casos, a fim de subsidiar ações de prevenção e controle.

2. ATIVIDADES PRINCIPAIS

As atividades da Vigilância em Saúde envolvem a alimentação e análise contínua do Sinan, com foco no acompanhamento das arboviroses, construção de indicadores epidemiológicos, avaliação de programas e apoio a estudos e pesquisas. Entre seus objetivos está a identificação precoce de áreas de maior incidência, para orientar ações integradas de prevenção, controle e organização da assistência ⁽¹⁾.

Com base na experiência prática da autora na Vigilância Epidemiológica e sua formação em andamento em Ciência de Dados pela Universidade Virtual do Estado de São Paulo, identificou-se a oportunidade de aplicar técnicas de ciência de dados na rotina de trabalho.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este estudo foi conduzido com o uso da linguagem de programação Python, amplamente adotada na análise de dados e projetos de ciência de dados devido à sua sintaxe acessível e à robustez de bibliotecas especializadas, como Pandas, NumPy, Matplotlib e Scikit-learn. A metodologia seguida baseou-se no processo *Knowledge Discovery in Databases* (KDD), composto por etapas sequenciais voltadas à descoberta de padrões úteis em grandes volumes de dados ^(3,4).

Na etapa de seleção foram utilizados dados do Sinan referentes aos anos de 2023, 2024 e ao primeiro trimestre de 2025. Considerou-se a data de início dos sintomas como referência temporal para a agregação semanal dos casos, com a inclusão de pacientes exclusivamente residentes no município de Osasco.

O pré-processamento envolveu a limpeza dos dados brutos, com a exclusão de registros com classificação final "descartado" ou "em aberto". Um dos principais desafios encontrados foi a ausência de padronização no preenchimento do campo "bairro", frequentemente sujeito a erros de digitação e abreviações variadas. Para contornar esse problema foi construída uma lista de bairros oficiais do município e aplicada a técnica de *fuzzy matching* com a biblioteca *rapidfuzz*, estabelecendo um limite de similaridade de 85% para a correspondência.

A fase de transformação dos dados compreendeu a redução da dimensionalidade do conjunto de dados, mantendo-se apenas os campos relevantes: bairro de residência, data do início dos sintomas e data da notificação. Dados pessoais foram completamente excluídos nesta etapa, em conformidade com os princípios éticos e legais de privacidade.

Para a mineração de dados foi utilizado o modelo Random Forest Regressor, escolhido por sua robustez e capacidade de lidar com dados não lineares. A base foi dividida em 70% para treinamento e 30% para teste, considerando os dados dos anos de 2023 e 2024. Com o modelo treinado foi então realizada a previsão do número de casos para o ano de 2025, distribuídos por bairro e por semana epidemiológica.

A avaliação do desempenho do modelo foi feita por meio do Erro Absoluto Médio (MAE) e do coeficiente de determinação (R²). Os resultados preliminares obtidos apontaram um R² de aproximadamente 0,678 e um MAE de cerca de 13,12 casos por bairro/semana. Esses valores indicam um bom ajuste do modelo, embora haja margem para aprimoramentos, especialmente no que se refere à inserção de variáveis ambientais e entomológicas.

4. DESAFIOS E APRENDIZADOS

A aplicação de Python e suas bibliotecas ainda representa um desafio, dada a experiência limitada da autora. No entanto, essa mesma dificuldade, inclusive, motivou o ingresso na graduação em Ciência de Dados, visando ampliar as possibilidades de aplicação prática no serviço público.

Entre os desafios do modelo destaca-se a melhoria do desempenho do MAE e R². Pretende-se incorporar o Índice de Breteau — indicador que estima a densidade do vetor com base em pesquisa larvária — como variável preditora. Também há intenção de utilizar variáveis climáticas, como temperatura e precipitação, para refinar o modelo ⁽²⁾.

5. REFERÊNCIAS

1. Centro de Vigilância Epidemiológica. Diretrizes para prevenção e controle das arboviroses urbanas no Estado de São Paulo [Internet]. São Paulo: Secretaria de Estado da Saúde de SP; [data de publicação/atualização ou [s.d.] se não houver]. Acesso em: 28 maio 2025. Disponível em: <https://portal.saude.sp.gov.br/resources/cve-centro-de-vigilancia-epidemiologica/areas-de-vigilancia/doencas-de-transmissao-por-ve-tores-e-zoonoses/doc/arboviroses/diretrizesparaaprevencaocontroledasarbovirosesurban.pdf>
2. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde e Ambiente. Departamento de Ações Estratégicas de Epidemiologia e Vigilância em Saúde e Ambiente. Guia de vigilância em saúde: volume 2. 6. Brasília: Ministério da Saúde; 2024. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/guia_vigilancia_saude_v2_6edrev.pdf
3. Fayyad UM, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. AI Mag. 1996;17(3):37–54.
4. Faceli K, Lorena AC, Almeida TA, Carvalho ACPLF. Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina. 2. São Paulo: Grupo GEN; 2021.

Análise e visualização de dados de monitoramento glicêmico em pacientes críticos

Maria Victória Fernandes Vaz¹, Camila Gabriela Mendes Santos², Maria Lúcia Bento Villela¹, Patrícia de Oliveira Salgado², Julio Cesar Soares dos Reis¹

¹*Departamento de Informática, Universidade Federal de Viçosa, Viçosa, MG, Brasil.*

²*Departamento de Medicina e Enfermagem, Universidade Federal de Viçosa, Viçosa, MG, Brasil.*

{maria.v.vaz, camila.gabriela, maria.villela, patriciasalgado, jreis}@ufv.br

Resumo: O controle glicêmico em pacientes internados em Unidades de Terapia Intensiva (UTIs) é essencial para prevenir complicações em seu estado de saúde, evitar graves desfechos e melhorar a perspectiva de prognóstico. Apesar da necessidade de monitoramento contínuo dos níveis de glicemia desses pacientes, a realização de tal tarefa pode ser complexa, dado que os profissionais de saúde enfrentam, muitas vezes, uma sobrecarga de informações clínicas, o que dificulta uma análise detalhada dos dados de cada paciente. Nesse contexto, este trabalho busca aplicar estratégias de análise de dados para trazer *insights*, históricos e informações relevantes para o monitoramento de glicemia em pacientes críticos de maneira sumarizada e visual. Para tais análises foram utilizados dados de pacientes internados em UTIs de dezembro de 2023 a janeiro de 2025 de um hospital de grande porte localizado na cidade de Belo Horizonte, em Minas Gerais. Especificamente, foram consideradas as medições de glicemia e os diagnósticos dados ao longo da internação de um paciente. Com esses dados foi possível explorar visualizações interativas que representam de maneira sintetizada o histórico de cada paciente, possibilitando uma visão ampla, ágil e completa do quadro glicêmico individual para os profissionais de saúde.

Abstract: Glycemic control in patients admitted to Intensive Care Units (ICUs) is essential to prevent complications in their health status, avoid severe outcomes, and improve prognostic perspectives. Despite the need for continuous monitoring of these patients' glycemic levels, performing such a task can be complex, given that healthcare professionals often face an overload of clinical information, which hinders a detailed analysis of each patient's data. In this regard, this work seeks to apply data analysis strategies to bring summarized and visual insights, history, and relevant information for glycemic monitoring in critically ill patients. For these analyses, data from patients hospitalized in ICUs from December 2023 to January 2025 from a large hospital located in the city of Belo Horizonte, Minas Gerais, were used. Specifically, glycemic measurements and diagnoses given throughout a patient's hospitalization were considered. With this data, it was possible to explore interactive visualizations that represent each patient's history in a summarized manner, enabling a broad, agile, and complete view of the individual glycemic picture for healthcare professionals.

Palavras-chave: glicemia; dados clínicos; análise de dados.

Nome do projeto: Predição de risco de glicemia instável em pacientes avaliados por indicadores clínicos.

1. CONTEXTO E MOTIVAÇÃO

Alterações dos níveis de glicemia, tanto para valores inferiores (hipoglicemia) quanto para superiores (hiperglicemia) aos recomendados, estão associados a diversas consequências clínicas, tais como maior vulnerabilidade a infecções, perda da capacidade cognitiva, distúrbios, disfunções, entre outros

agravantes ⁽¹⁾. Pacientes internados sob cuidados intensivos constituem-se como um grupo que está especialmente suscetível a essas variações principalmente devido ao estresse, uso de medicamentos, liberação de citocinas inflamatórias, entre outros. Essas oscilações glicêmicas estão diretamente relacionadas ao aumento da mortalidade, prolongamento do tempo de internação e maior incidência de eventos adversos ^(1,2).

Devido ao seu quadro clínico sensível, as consequências dessas variações podem ser graves, o que leva à necessidade de um monitoramento contínuo e individualizado dos níveis glicêmicos desses pacientes ⁽³⁾. Tal monitoramento envolve análises complexas e recorrentes dos dados glicêmicos e de suas inter-relações com as condições clínicas dos pacientes, demandando alto custo, tempo e esforço das equipes de saúde. Nesse sentido, a adoção de estratégias computacionais, especialmente aquelas baseadas em inteligência artificial, configura-se como uma alternativa promissora para otimizar o monitoramento glicêmico, reduzir o tempo de análise do quadro do glicêmico, trazer agilidade na tomada de decisão clínica e, conseqüentemente, contribuir para a melhoria dos desfechos assistenciais e do prognóstico do paciente.

1. 1 Objetivo

Este trabalho tem como objetivo aplicar técnicas de análise de dados para sumarizar dados clínicos em visualizações objetivas e interativas que permitam a visualização do histórico glicêmico de um paciente internado em terapia intensiva de maneira otimizada.

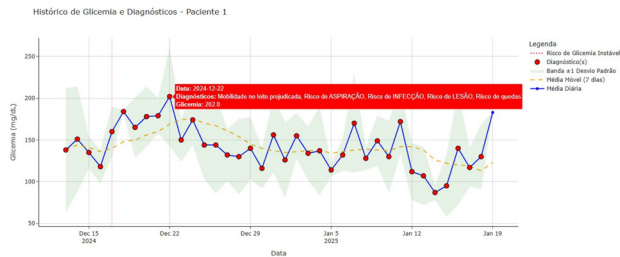
2. ATIVIDADES PRINCIPAIS

Para a realização deste trabalho, as principais atividades executadas foram: i) estudo e entendimento dos dados clínicos; ii) seleção de dados glicêmicos e de diagnósticos; iii) realização de análises estatísticas nos dados; e iv) aplicação de técnicas de visualização de dados.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

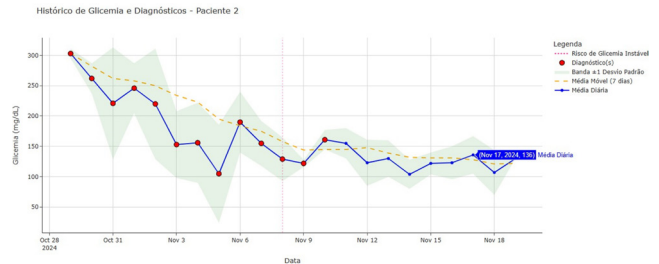
Esse projeto envolve dados de seres humanos e foi aprovado com número do CAAE: 84975724.0.0000.5153 e número do parecer: 7283503. Os dados clínicos utilizados neste trabalho são compostos por diagnósticos dos pacientes e suas datas de liberação, bem como valores das medições de glicemia e a data e hora da medição ao longo do tempo. Pensando na necessidade de representar o histórico glicêmico de maneira compacta e no objetivo de permitir a possibilidade de relacioná-lo com os diagnósticos dos pacientes, decidiu-se que uma visualização interativa seria uma maneira interessante de permitir a realização de análises cabíveis. Para possibilitar essa interatividade, foi utilizada a biblioteca *Plotly* do *Python*, mais especificamente o módulo *Graph Objects*. A visualização interativa mostra as informações do paciente por dia e possui 3 elementos principais: uma linha azul indicando o histórico de valores de glicemia do paciente (média diária), uma linha pontilhada laranja com a média móvel desses valores, uma faixa verde claro indicando a banda ± 1 desvio padrão da glicemia, uma linha tracejada vertical rosa nos dias onde o paciente recebeu diagnóstico de enfermagem de "Risco de glicemia instável", caso tenha, e por fim pontos vermelhos indicando os diagnósticos de enfermagem dados ao paciente no dia, se houverem. Passando o mouse em cima das linhas e pontos é possível visualizar os valores naquele dia. Para fins de ilustração, foram selecionados os dados de prontuários de dois pacientes (um prontuário identifica o histórico de um paciente) que foram chamados de Paciente 1 e Paciente 2. As Figuras 1 e 2 mostram as visualizações de monitoramento glicêmico para ambos os pacientes. Para acesso à visualização interativa, com as datas dos procedimentos omitidas por questões de privacidade, basta acessar o link: <https://analise-de-dados-monitoramento-glic-drab.vercel.app/>.

Figura 1 – Histórico de Glicemia e Diagnósticos do Paciente 1, mouse sobre os diagnósticos do dia 22 de dezembro de 2024.



Fonte: Autoria Própria

Figura 2 - Histórico de Glicemia e Diagnósticos do Paciente 2, mouse sobre o valor de glicemia média do dia 17 de novembro de 2024.



Fonte: Autoria Própria

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do projeto, os principais desafios enfrentados até o momento foram relacionados ao entendimento, à falta de padronização, aos valores ausentes e à seleção dos dados clínicos. No entanto, essas tarefas proporcionaram aprendizados significativos, como o domínio de bibliotecas de análise, processamento e visualização de dados, além do uso de técnicas estatísticas e da integração entre análise de dados e interpretação clínica. Com esses conhecimentos foi possível vislumbrar o potencial de aplicação de técnicas de Inteligência Artificial no monitoramento glicêmico que irão compor os próximos passos do projeto, que incluem a identificação precoce de risco de glicemia instável bem como uma avaliação do ferramental desenvolvido com potenciais usuários.

Agradecimentos. Trabalho parcialmente financiado pela FAPEMIG (Universal 2025).

5. REFERÊNCIAS

1. Gomes PM, Foss MC, Foss-Freitas MC. Controle de hiperglicemia intra-hospitalar em pacientes críticos e não-críticos. Medicina (Ribeirão Preto). 2014 Jun 30;47(2):194–200.
2. Duckworth C, Guy MJ, Kumaran A, O’Kane AA, Ayobi A, Chapman A, et al. Explainable Machine Learning for Real-Time Hypoglycemia and Hyperglycemia Prediction and Personalized Control Recommendations. Journal of Diabetes Science and Technology. 2022 Jun 13;193229682211035.
3. DeJournett J, Nekludov M, DeJournett L, Wallin M. Performance of a closed-loop glucose control system, comprising a continuous glucose monitoring system and an AI-based controller in swine during severe hypo- and hyperglycemic provocations. Journal of Clinical Monitoring and Computing. 2020 Jan 31;35(2):317–25.

Aplicação da predição do encaminhamento de retinopatia diabética ao especialista em um projeto real

Thiago Araújo¹, Bruno Morales¹, Philippe Navaux¹, Beatriz Schaan¹

¹Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.

[tsaraujo, bmmorales, navaux]@inf.ufrgs.br, bschaan@hcpa.edu.br

Resumo: A retinopatia diabética (RD) é uma das principais complicações do diabetes e pode levar à cegueira. A detecção precoce é essencial para evitar danos irreversíveis à visão, mas o acesso à avaliação oftalmológica é limitado no Sistema Único de Saúde (SUS), devido à alta demanda, escassez de especialistas e encaminhamentos muitas vezes desnecessários. Este trabalho apresenta o desenvolvimento e a avaliação de um modelo de aprendizado profundo treinado exclusivamente com imagens de pacientes brasileiros, com o objetivo de identificar automaticamente casos de RD referenciável. O modelo alcançou área sob a curva ROC (AUC) de 0,98, sensibilidade de 95,1% e especificidade de 90,9%. A aplicação foi disponibilizada via interface web desenvolvida utilizando o framework Streamlit e está em uso no Hospital de Clínicas de Porto Alegre (HCPA), como parte de um projeto voltado à triagem automatizada de pacientes com diabetes. A ferramenta realiza a predição automática a partir de imagens capturadas em ambiente clínico, oferecendo suporte imediato à decisão médica. Os resultados demonstram o potencial da inteligência artificial para otimizar o rastreamento da RD no SUS, priorizando casos graves, reduzindo filas e ampliando o acesso ao cuidado especializado.

Palavras-chave: Retinopatia diabética; inteligência artificial; rede neural convolucional; SUS.

Nome do projeto: Efeitos do rastreamento da retinopatia diabética baseado em inteligência artificial no acesso ao tratamento em tempo adequado, em pessoas com diabetes mellitus.

Links: <https://github.com/brunommorales/HCPA>

Este trabalho foi financiado pela CAPES - Código de financiamento 001 e projetos CIARS RITEs/FAPERGS e CI-IA FAPESP-MCTIC-CGI-BR.

1. INTRODUÇÃO

A retinopatia diabética (RD) é uma das principais causas de perda de visão em adultos, afetando milhões de pessoas com diabetes em todo o mundo ^[1]. Resulta da hiperglicemia crônica da diabetes, que danifica gradualmente os vasos sanguíneos da retina.

No Brasil, o SUS enfrenta desafios consideráveis para garantir o acesso rápido a consultas oftalmológicas para pacientes com RD. A alta demanda por atendimento especializado, aliada ao elevado número de encaminhamentos desnecessários, sobrecarrega o sistema resultando em longas filas de espera para avaliação com especialistas. Como consequência, o diagnóstico e o tratamento muitas vezes ocorrem de forma tardia, aumentando o risco de progressão da doença e, em casos mais graves, de perda irreversível da visão ^[2].

Este trabalho apresenta um modelo de aprendizado profundo treinado com imagens de pacientes brasileiros para identificar automaticamente casos de RD referenciável. O modelo foi carregado em uma aplicação web desenvolvida com o framework Streamlit e está em uso em um projeto no HCPA, auxiliando na triagem automatizada de pacientes com diabetes em ambiente clínico real.

2. METODOLOGIA

O estudo utilizou 15.816 imagens coloridas de retina, obtidas de 4.590 pacientes atendidos em instituições públicas brasileiras como o HCPA, Universidade Federal de São Paulo (UNIFESP) e TeleOftalmo. O estudo foi aprovado pelo comitê de ética institucional (nº 2019-0113). As imagens foram avaliadas por dois oftalmologistas e categorizadas segundo a escala ICDR (International Clinic Diabetic Retinopathy), sendo classificadas como RD não referenciável (ausente ou leve) ou referenciável (moderada a proliferativa e/ou com edema macular).

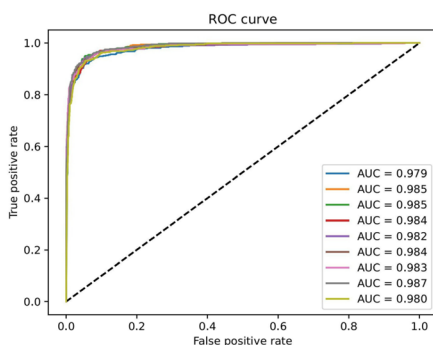
O modelo, baseado na arquitetura Inception V3 com pesos pré-treinados no ImageNet, foi adaptado para classificação binária com ativação sigmoide. Foi treinado por 200 épocas usando o otimizador Adam e função de perda binária cross-entropy, com imagens redimensionadas para 299×299 pixels. A avaliação considerou AUC, sensibilidade e especificidade, com ponto de corte de maior eficiência [3].

Para uso clínico, o modelo foi incorporado a uma aplicação web desenvolvida em Streamlit, que permite inferência sobre imagens individuais ou em lote. A ferramenta está em uso no HCPA em um projeto real de triagem, oferecendo suporte imediato à decisão médica.

3. RESULTADOS

No conjunto de dados agrupado, o modelo de aprendizado profundo alcançou uma AUC de 0,98 (IC 95%: 0,97–0,98), com sensibilidade de 95,1% (IC 95%: 93,9–96,3) e especificidade de 90,9% (IC 95%: 89,9–91,8), utilizando o ponto de corte de maior eficiência. Esses resultados demonstram a robustez da abordagem para fins de triagem. A Figura 1 apresenta as curvas ROC de nove execuções independentes do treinamento do modelo. As AUCs variaram entre 0,979 e 0,987, demonstrando consistência e estabilidade entre diferentes repetições do processo de aprendizado, o que reforça a robustez da abordagem.

Figura 1 – Curvas ROC de nove treinamentos do modelo.



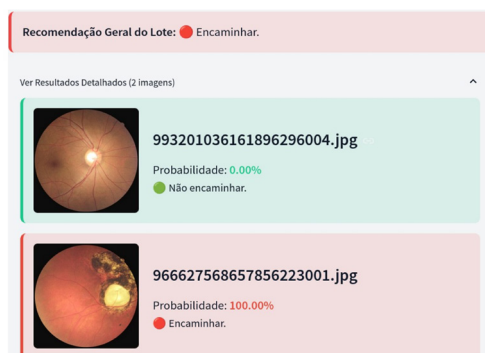
Fonte: O autor

Para aplicação prática, desenvolvemos uma interface web com Streamlit que permite carregar o modelo treinado e realizar previsões sobre imagens individuais ou em lote. Para cada imagem, a ferramenta exibe a probabilidade de RD referenciável e, com base em um limiar de 50%, classifica como “Encaminhar ao especialista” ($\geq 50\%$) ou “Não encaminhar” ($< 50\%$). Para lotes de imagens, o sistema calcula também a média aritmética, oferecendo uma visão agregada.

A aplicação está em uso no Hospital de Clínicas de Porto Alegre (HCPA), em um projeto de triagem com pacientes da atenção básica. As imagens analisadas pela ferramenta também são avaliadas por oftalmologistas. Essa validação cruzada permitirá ajustes futuros no modelo e avaliação contínua de desempenho em ambiente real.

A Figura 2 ilustra a interface em operação com duas imagens: a primeira, com 0%, que resulta em não encaminhar; a segunda, com 100% de probabilidade, foi classificada para encaminhamento. A interface foi projetada para ser intuitiva e interpretável por profissionais da saúde, mesmo sem formação em inteligência artificial.

Figura 2 – Interface da aplicação exibindo o resultado da predição para um lote de duas imagens.



Fonte: O autor

4. CONCLUSÃO

O modelo de inteligência artificial desenvolvido apresentou alto desempenho na detecção de retinopatia diabética referenciável em imagens de pacientes brasileiros. Sua disponibilização via aplicação web viabilizou o uso prático em um hospital público, permitindo triagem automatizada no HCPA. A integração entre IA e prática clínica amplia o acesso ao diagnóstico especializado e oferece base para aprimoramentos futuros e expansão do sistema.

5. REFERÊNCIAS

1. Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017;5(12):e1221–34.
2. Massuda A, Hone T, Leles FAG, De Castro MC, Atun R. The Brazilian health system at crossroads: progress, crisis and resilience. *BMJ Glob Health*. 2018;3(4):e000829.
3. Dos Reis MA et al. Advancing healthcare with artificial intelligence: diagnostic accuracy of machine learning algorithm in diagnosis of diabetic retinopathy in the Brazilian population. *Diabetol Metab Syndr*. 2024;16(1):209.

Aplicação e avaliação comparativa de algoritmos de Machine Learning para o diagnóstico de câncer de mama a partir de características nucleares

Gabriel Henrique Silva Teixeira ¹, Raphael Augusto Teixeira de Aguiar ²

¹Núcleo de Educação em Saúde Coletiva, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

²Núcleo de Educação em Saúde Coletiva, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

gabrielteixeiraufmg@gmail.com, rapha.aguiar@gmail.com

Resumo: O estudo avaliou a eficácia de seis modelos de Machine Learning (Regressão Logística, Random Forest, AdaBoost, LightGBM, CatBoost e XGBoost) do dataset Wisconsin Database Breast Cancer (WDBC) que reúne 10 variáveis características de dados clínicos obtidos a partir de imagens digitalizadas por aspiração por agulha fina (PAAF) de massa mamária. Após pré-processamento rigoroso (Winsorização, StandardScaler, SMOTE), todos os modelos tiveram desempenho excepcional. A Regressão Logística destacou-se com F1-Score de 0.976, acurácia de 98.2% e ROC AUC de 0.997. As características "piores" e "erro padrão" foram as mais preditivas. Conclui-se que Modelos de Machine Learning possuem grande potencial para sistemas de diagnóstico auxiliado por computador.

Palavras-chave: aprendizado de máquina; câncer de mama; diagnóstico auxiliado por computador.

1. CONTEXTO E MOTIVAÇÃO

A alta incidência e mortalidade do câncer de mama, aliada à subjetividade inerente aos diagnósticos histopatológicos, motivam a busca por ferramentas de apoio. A Inteligência Artificial pode oferecer uma análise quantitativa, objetiva e rápida, auxiliando os clínicos a reduzir a variabilidade e aumentar a acurácia diagnóstica.

1.1 Objetivo

O objetivo principal deste trabalho é desenvolver, treinar e avaliar comparativamente seis algoritmos de Machine Learning de classificação para distinguir tumores mamários benignos e malignos. Busca-se identificar o modelo com melhor performance preditiva e interpretar suas decisões com base nas características celulares.

2. ATIVIDADES PRINCIPAIS

Este trabalho consistiu em: carregar e explorar o dataset WDBC ⁽¹⁾; aplicar um pipeline de pré-processamento com tratamento de outliers, escalonamento e balanceamento (SMOTE); treinar seis modelos distintos; avaliar a performance com métricas (F1, ROC AUC); e interpretar as features mais relevantes do modelo vencedor.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

3.1 Metodologia

O estudo utilizou o dataset público Wisconsin Diagnostic Breast Cancer (WDBC), obtido do repositório da UCI. Este dataset contém 569 amostras, cada uma descrita por 30 features computacionais

extraídas de imagens digitalizadas de aspirado por agulha fina (FNA). As features derivam de 10 características base do núcleo celular (e.g., raio, textura, área). Para cada característica, três medidas estatísticas foram calculadas por imagem: a média (mean), representando o valor típico; o erro padrão (se), medindo a heterogeneidade; e o "pioor" valor (worst), capturando os núcleos mais anormais ⁽²⁾.

O fluxo de modelagem iniciou-se com a separação estratificada dos dados em 80% para treino e 20% para teste. Todo o pré-processamento foi ajustado exclusivamente no conjunto de treino para evitar vazamento de dados. As etapas foram: (1) tratamento de outliers via Winsorização, limitando os valores aos limites de 1.5 vezes o intervalo interquartil (IQR); (2) escalonamento das features com StandardScaler para normalizar as distribuições; (3) aplicação da técnica SMOTE para criar um conjunto de treino sinteticamente balanceado. Foram treinados e comparados seis algoritmos de classificação: Regressão Logística, Random Forest, AdaBoost, LightGBM, CatBoost e XGBoost.

3.2 Resultados

Os modelos demonstraram alta capacidade preditiva no conjunto de teste. A Regressão Logística e o AdaBoost se destacaram, ambos com um F1-Score de aproximadamente 0.976. Especificamente, a Regressão Logística alcançou uma acurácia de 98.25%, precisão de 97.62%, recall de 97.62% e uma área sob a curva ROC (ROC AUC) de 0.9974. A matriz de confusão para este modelo revelou apenas dois erros de classificação no conjunto de teste de 114 amostras: um falso positivo (um tumor benigno classificado como maligno) e um falso negativo (um tumor maligno classificado como benigno).

Os modelos baseados em Gradient Boosting (LightGBM, CatBoost) e o Random Forest também apresentaram excelente performance, com F1-Scores de 0.963. A análise de importância de features da Regressão Logística, baseada nos coeficientes do modelo, identificou texture_worst (textura "pioor"), area_se (erro padrão da área) e compactness_se (erro padrão da compactidade) como as variáveis mais influentes para a classificação.

Tabela 1 - Comparativo final dos modelos.

| Modelo | Acurácia | Precisão | Recall | F1-Score | ROC AUC |
|---------------------|----------|----------|--------|----------|---------|
| Regressão Logística | 0.982 | 0.976 | 0.976 | 0.976 | 0.997 |
| AdaBoost | 0.982 | 1 | 0.953 | 0.975 | 0.982 |
| Random Forest | 0.973 | 1 | 0.928 | 0.962 | 0.996 |
| LightGBM | 0.973 | 1 | 0.928 | 0.962 | 0.993 |
| CatBoost | 0.973 | 1 | 0.928 | 0.962 | 0.998 |
| XGBoost | 0.956 | 1 | 0.880 | 0.936 | 0.992 |

4. DESAFIOS E APRENDIZADOS

Os resultados obtidos são notáveis e indicam a alta qualidade preditiva das features extraídas do dataset WDBC. No entanto, é crucial contextualizar estes achados, reconhecendo a natureza altamente "curada" e "idealizada" do dataset. Diferente de desafios de dados clínicos do "mundo real" — que frequentemente envolvem valores faltantes, ruído significativo e desbalanceamento severo de classes — este dataset apresenta dados completos, features bem-definidas e um equilíbrio de classes favorável. Portanto, este estudo funciona como uma prova de conceito em um ambiente controlado, demonstrando o potencial máximo dos algoritmos quando alimentados com dados de alta qualidade.

O fato de um modelo linear como a Regressão Logística ter se destacado, igualando-se a um ensemble complexo como o AdaBoost e superando outros, é um achado significativo. O sucesso da Regressão Logística neste contexto não prova que o problema é inerentemente linear. Em vez disso, **sugere que um limite de decisão linear é uma aproximação altamente eficaz e eficiente para a estrutura deste**

dataset específico. A análise de importância das features reforça a validade clínica do modelo. As três variáveis mais importantes — `texture_worst`, `area_se` e `compactness_se` — correspondem diretamente a marcadores de malignidade. A `texture_worst` captura a heterogeneidade da cromatina nos núcleos mais anormais. A `area_se` e a `compactness_se` medem o pleomorfismo, ou seja, a grande variação de tamanho e a irregularidade na forma dos núcleos, respectivamente, que são características clássicas de tumores malignos. É particularmente relevante que o modelo tenha dado alta importância a features do tipo “pior” (`worst`) e “erro padrão” (`se`), validando a hipótese dos criadores do dataset de que não apenas a célula média, mas a heterogeneidade e os casos extremos são os principais indicadores de câncer.

Embora a acurácia seja alta, o único falso negativo registrado é um ponto crítico em diagnóstico oncológico. Futuras implementações poderiam investigar o ajuste do limiar de decisão do modelo para maximizar o recall, minimizando a chance de um caso maligno passar despercebido, mesmo que isso aumente o número de falsos positivos (que seriam resolvidos com exames confirmatórios).

5. REFERÊNCIAS

1. William Wolberg OM. Breast Cancer Wisconsin (Diagnostic) [Internet]. UCI Machine Learning Repository; 1993 [citado 16 de junho de 2025]. Disponível em: <https://archive.ics.uci.edu/dataset/17>
2. Street WN, Wolberg WH, Mangasarian OL. Nuclear feature extraction for breast tumor diagnosis. Electronic imaging [Internet]. 1993 [citado 16 de junho de 2025]; Disponível em: <https://www.semanticscholar.org/paper/Nuclear-feature-extraction-for-breast-tumor-Street-Wolberg/53f0fbb425bc14468eb3bf96b2e1d41ba8087f36>

Avaliação de large language models para o reconhecimento de entidades nomeadas em prontuários médicos

Cauã G. Marvila (aluno)¹, Cristiano da S. Colombo (orientador)¹

¹Instituto Federal do Espírito Santo, Cachoeiro de Itapemirim, ES, Brasil.

caua.privg@gmail.com, cristiano.colombo@gmail.com

Resumo: Este trabalho apresenta um estudo inicial sobre o uso de Large Language Models (LLMs) para a tarefa de *Named Entity Recognition* (NER) em prontuários médicos em português brasileiro. O projeto encontra-se em estágio preliminar e busca mapear o potencial e os desafios da aplicação de LLMs na extração automática de informações clínicas, com foco na prática responsável e ética. Foram realizados testes exploratórios com modelos como HUMADEx e Medialbertina, utilizando prontuários anonimizados e anotados manualmente em categorias como Doença, Medicamento e Procedimento. Os resultados iniciais mostraram limitações no desempenho (F1-Score inferior a 0,40), mas forneceram aprendizados relevantes para ajustes metodológicos e planejamento de *fine-tuning*, além de destacar a necessidade de ampliação do conjunto de dados. Este estudo contribui para o debate sobre o uso seguro e eficaz da IA na saúde, apontando caminhos para o desenvolvimento de soluções alinhadas às necessidades clínicas e à ética em saúde digital.

Abstract: This work presents an initial study on the use of language models (LLMs) for the task of Named Entity Recognition (NER) in medical records written in Brazilian Portuguese. The project is in a preliminary stage and seeks to map the potential and challenges of applying LLMs for the automatic extraction of clinical information, with a focus on responsible and ethical practices. Exploratory tests were conducted with models such as HUMADEx and Medialbertina, using anonymized medical records manually annotated with categories such as Disease, Medication, and Procedure. The initial results showed performance limitations (F1-Score below 0.40) but provided valuable insights for methodological adjustments and fine-tuning planning, as well as highlighting the need to expand the dataset. This study contributes to the discussion on the safe and effective use of AI in healthcare, pointing out pathways for the development of solutions aligned with clinical needs and ethics in digital health.

Palavras-chave: Inteligência Artificial na Saúde; LLM; NER; prontuário médico.

Nome do projeto: Plataforma Inteligente para Análise de Prontuários e Condições de Saúde.

1. CONTEXTO E MOTIVAÇÃO

A Inteligência Artificial (IA), especialmente os *Large Language Models* (LLMs), vem se consolidando ⁽¹⁾ como ferramenta promissora para apoiar atividades na área da saúde, como a análise de prontuários médicos e o reconhecimento de entidades clínicas. Embora sistemas de *Named Entity Recognition* (NER) adaptados ao domínio biomédico sejam essenciais ⁽²⁾ para captar terminologias e relações próprias da prática médica, há escassez de modelos treinados para o português brasileiro, o que dificulta o avanço de soluções realmente precisas e responsáveis para uso clínico. Este projeto, ainda em fase inicial, busca explorar e mapear os desafios e possibilidades do uso de LLMs em tarefas de NER aplicadas a prontuários médicos em português, contribuindo para o debate sobre o desenvolvimento ético e seguro de tecnologias de IA na saúde ⁽⁵⁾.

1. 1 Objetivo

Este trabalho tem como objetivo principal realizar uma avaliação exploratória do desempenho de modelos de linguagem (LLMs) na tarefa de NER em prontuários médicos em português do Brasil. A partir dessa avaliação inicial, busca-se identificar os modelos com maior potencial para posterior especialização por meio de *fine-tuning*, com foco em aplicações clínicas. Durante o projeto, foi realizada a seleção de modelos disponíveis em repositórios públicos (como *Hugging Face*), pela definição de prontuários médicos a serem utilizados nos testes, pela anotação manual das entidades, pela execução e análise dos resultados dos modelos, além da obtenção de resultados a partir de métricas como precisão, recall e F1-score. Essa experiência permite o aprofundamento em conhecimentos de *Natural Language Processing* (NLP), técnicas de avaliação de modelos e boas práticas na organização de experimentos científicos.

2. ATIVIDADES PRINCIPAIS

O experimento inicial para a avaliação de LLMs em aplicações de saúde seguiu uma metodologia estruturada, que incluiu as seguintes etapas principais:

- Seleção dos modelos de LLMs de uso geral;
- Escolha de LLMs treinados em Língua Portuguesa;
- Escolha dos prontuários;
- Anotação manual de entidades em prontuários;
- Execução dos modelos no Google Colab;
- Comparação preliminar dos resultados;
- Cálculos das métricas.

O desenvolvimento do trabalho iniciou-se com a seleção de modelos de LLMs de uso geral, priorizando aqueles com boa reputação em tarefas de PLN. Em seguida, foram escolhidos modelos especificamente treinados para o português, com ênfase naqueles voltados ao domínio médico. Também foram selecionados prontuários médicos reais, representativos de diferentes estruturas textuais e níveis de complexidade. Para garantir a qualidade da avaliação, realizou-se a anotação manual das entidades nomeadas presentes nesses textos, como doenças, medicamentos, procedimentos, sintomas e profissionais de saúde, criando-se um padrão ouro confiável. Após isso, os modelos foram executados no ambiente Google Colab, onde processaram os mesmos prontuários. Os resultados foram então comparados preliminarmente com as anotações manuais, e, por fim, foram calculadas as métricas de precisão, recall e F1-score, permitindo a análise inicial do desempenho dos modelos e a identificação dos mais promissores para futuras aplicações em saúde digital.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

O desenvolvimento do trabalho seguiu uma abordagem experimental com foco na avaliação preliminar de modelos de linguagem para a tarefa de NER em prontuários médicos escritos em Língua Portuguesa para o Brasil. A primeira etapa consistiu na seleção de LLMs de uso geral e de modelos pré-treinados no domínio médico, com ênfase em versões desenvolvidas especificamente para o português. Os modelos foram obtidos principalmente da plataforma Hugging Face⁽³⁾, incluindo: HUMADEx / portuguese_medical_ner, BioBERTpt-clin, BioBERTpt-all, portugueseNLP/medialbertina_pt-pt_1.5b_NER e CardioBERTpt. Foram escolhidos dois prontuários médicos representativos, retirados do dataset SemClinBr, sendo um mais curto e outro com texto mais extenso. A opção por dois prontuários clínicos com perfis contrastantes (um curto e outro mais extenso) visou explorar como os modelos se comportam frente a diferentes níveis de complexidade textual. Embora a amostra reduzida limite a generalização dos resultados, essa estratégia permitiu obter subsídios iniciais sobre os desafios e potencialidades do uso de LLMs em NER clínico. Realizou-se então a anotação manual das entidades nomeadas presentes nos textos, categorizando-as em classes como: doença, sintoma, tratamento, medicamento, procedimento, tempo, anatomia, profissional de saúde e estado clínico. A execução dos modelos foi feita utilizando o ambiente Google Colab,

e os resultados produzidos por cada modelo foram comparados com as anotações manuais. Para avaliar o desempenho foram calculadas as métricas clássicas de NLP⁽⁴⁾: precisão, recall e F1-score, que indicam, respectivamente, a taxa de acerto entre as entidades previstas, a cobertura sobre as entidades reais e o equilíbrio entre ambas. Os modelos HUMADEx e Medialbertina apresentaram os melhores resultados entre os avaliados até o momento, conforme apresentado na Tabela 1. Os resultados obtidos reforçam a necessidade de ajustes e etapas futuras de fine-tuning. Para trabalhos futuros pretende-se aprofundar na avaliação e parametrização de LLMs, ampliar o conjunto de prontuários e realizar ajustes nos modelos para melhor adaptação ao contexto clínico nacional.

Tabela 1 - Resultados obtidos nos experimentos

| Modelo | Precisão | Recall | F1-SCORE |
|--|-------------|-------------|-------------|
| HUMADEx/portuguese_medical_ner | 0,36 | 0,42 | 0,39 |
| portugueseNLP/medialbertina_pt-pt_1.5b_NER | 0,42 | 0,33 | 0,37 |

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do estudo foram identificados desafios significativos que reforçam a complexidade da aplicação de LLMs em tarefas de NER clínico em português brasileiro. Um dos principais obstáculos foi a dificuldade para encontrar modelos de linguagem treinados especificamente para o português, especialmente no domínio da saúde. Além disso, muitos dos modelos disponíveis careciam de documentação detalhada, o que dificultou a correta aplicação e ajuste dos *pipelines*. Modelos de uso geral, não especializados no domínio clínico, apresentaram saídas inconsistentes, com baixa precisão no reconhecimento das entidades desejadas. Apesar dos desafios, o projeto proporcionou aprendizados valiosos, incluindo o aprofundamento em técnicas de anotação manual e validação de dados (padrão-ouro), o domínio de ferramentas como Google Colab para experimentação em PLN e o fortalecimento da compreensão sobre as limitações atuais dos LLMs no contexto clínico nacional. Esses aprendizados fundamentam os próximos passos do estudo, como o planejamento de *fine-tuning* e a ampliação do conjunto de dados para melhorar a acurácia das soluções propostas.

5. REFERÊNCIAS

1. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019; 380(14):1347-1358. doi:10.1056/NEJMra1814259.
2. Paganote B, et al. SemClinBr: A Brazilian Clinical Corpus annotated for NER. J Biomed Inform. 2023;134:104300. doi:10.1016/j.jbi.2022.104300.
3. Wolf T, et al. Transformers: State-of-the-art natural language processing. EMNLP. 2020;38-45. doi:10.18653/v1/2020.emnlp-demos.6.
4. Jurafsky D, Martin JH. Speech and Language Processing. 3rd ed. draft. Stanford University; 2020.
5. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-399. doi:10.1038/s42256-019-0088-2.

Avaliação multidimensional de redes generativas adversariais, modelos de difusão e métodos estatísticos para geração de dados sintéticos em saúde

Oluwatoyin Joy Omole, Celso França, Samuel N. Alves,
Regina Bernal, Deborah Malta, Marcos André Gonçalves, Jussara M. Almeida

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

joyomole@dcc.ufmg.br

Resumo: Este trabalho apresenta uma análise comparativa de três abordagens para geração de dados sintéticos em saúde: um método estatístico (Gaussian Copula), uma rede adversarial generativa (CTGAN) e um modelo simplificado de difusão neural (SimpleTableDiffusion). Utilizando uma amostra de dados do inquérito Vigitel (Belo Horizonte, 2006-2018) como estudo de caso, esses modelos foram avaliados em três dimensões principais: fidelidade estatística, utilidade preditiva e proteção da privacidade. A técnica cGAN demonstrou o desempenho mais consistente em todas as métricas, enquanto a cópula gaussiana demonstrou força na modelagem de dependências condicionais e o modelo baseado em difusão apresentou resultados competitivos com sobrecarga computacional reduzida. Os resultados mostram os compromissos de desempenho de cada modelo e apontam para direções promissoras para abordagens híbridas.

Palavras-chave: Modelos Generativos; Dados Tabulares Sintéticos; Análise Comparativa Multidimensional.

Abstract: This work presents a comparative analysis of three approaches for generating synthetic health data: a statistical method (Gaussian Copula), a conditioned generative adversarial network (CTGAN), and a simplified neural diffusion model (SimpleTableDiffusion). Using a sample of the Vigitel dataset (BH, 2006–2018) as a case study, these models were assessed across three key dimensions: statistical fidelity, predictive utility, and privacy protection. cGAN demonstrated the most consistent performance across all metrics. At the same time, the Gaussian Copula showed strength in modelling conditional dependencies, and the diffusion-based model delivered competitive results with reduced computational overhead. The findings show the trade-offs of each modelling paradigm and point toward promising directions for hybrid approaches.

Keywords: Generative Models; Synthetic Tabular Data; Multi-dimensional comparative analysis.

Nome do projeto: Uso de Algoritmos de machine learning e de Inteligência Artificial para predição em pequenas áreas de indicadores de saúde relacionados às doenças crônicas não transmissíveis.

1. CONTEXTO E MOTIVAÇÃO

O acesso a dados reais de saúde é frequentemente limitado por questões éticas e regulatórias, tornando os dados sintéticos uma alternativa valiosa para, por exemplo, a avaliação de políticas de gestão de saúde pública em cenários sintéticos mas realistas, assim como para estimativas robustas de desfechos de interesse, especialmente em cenários com grandes limitações de dados reais. Este projeto investiga o desempenho de diferentes paradigmas de geração de dados sintéticos aplicados a conjuntos de dados tabulares de saúde. O objetivo é avaliar sua adequação para pesquisas em saúde pública no mundo real, onde a sensibilidade e a escassez de dados se cruzam.

1.1 Objetivo

O projeto teve como objetivo comparar abordagens estatísticas e baseadas em aprendizado de máquina para geração de dados sintéticos na área da saúde considerando três dimensões complementares: fidelidade estatística, utilidade preditiva e preservação da privacidade dos dados reais. Nossa contribuição inclui a implementação dos modelos e a avaliação do desempenho das três técnicas em uma amostra de dados extraídos do inquérito Vigitel (Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico) [2] restrita à cidade de Belo Horizonte no período de 2006-2018 (Vigitel-BH), considerando as três dimensões indicadas.

2. ATIVIDADES PRINCIPAIS

As principais atividades envolveram a implementação dos modelos Cópula Gaussiana [5], Rede Adversarial Generativa Tabular Condicional (CTGAN) [1] e SimpleTableDiffusion [6]. Um modelo de Cópula Gaussiana de dependências entre variáveis contínuas; CTGAN eficaz o processo de treinamento ao manipular distribuições multimodais e capturar dependências entre variáveis contínuas e categóricas de forma eficaz. SimpleTableDiffusion é uma versão simplificada do modelo Difusão Tabular de Truda [6], com menor complexidade e ajustes para dados de saúde. Os métodos foram avaliados no conjunto Vigitel-BH, cuja diversidade e sensibilidade o tornam ideal para testar fidelidade estatística e preservação de privacidade.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Os modelos foram avaliados com validação cruzada de cinco folds, considerando três dimensões: fidelidade, utilidade e privacidade. A fidelidade mede o quanto os dados sintéticos preservam propriedades estatísticas dos dados reais, usando métricas como os complementos de KS e TV [6], similaridade de correlação entre atributos numéricos e de tabelas de contingência para variáveis categóricas. A utilidade foi avaliada por meio da tarefa de classificar indivíduos fumantes ou não, utilizando três algoritmos, Logistic Regression (LR), Random Forest (RF) e XGBoost (XGB), e as métricas Micro e Macro-F1. Esses métodos estão descritos em [7] e [8]. A dimensão de privacidade quantifica o risco de reidentificação, com base na inferência de atributos sensíveis.

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, sob o parecer nº 6.538.883. Os resultados médios dos folds, com intervalo de confiança de 95%, estão apresentados na Tabela 1.

- **Fidelidade:** CTGAN teve o melhor desempenho em métricas de distribuição marginal (complementos de KS e TV), mostrando eficácia na modelagem de dados categóricos diversos e desbalanceados via geração condicional. Gaussian Copula destacou-se na modelagem de dependências entre variáveis, com as melhores pontuações em similaridade de correlação, mas mostrou fragilidade em interações categóricas. O SimpleTableDiffusion obteve resultados equilibrados em todas as métricas, oferecendo um bom compromisso entre complexidade e expressividade.

- **Utilidade:** O CTGAN obteve os resultados mais próximos dos dados reais no Micro-F1 e manteve desempenho estável entre os classificados. A Cópula Gaussiana apresentou desempenho inferior especialmente na Regressão Logística, enquanto o SimpleTableDiffusion apresentou desempenho competitivo em todos os modelos, com destaque para RF e LR, mesmo com uma arquitetura mais simples reforçando sua aplicabilidade em contextos com restrições computacionais.

- **Privacidade:** foi avaliada com uma Pontuação de Proteção de Exposição baseada no risco de inferência de atributos. O CTGAN obteve a maior pontuação média, seguido pelo Gaussian Copula e pelo SimpleTableDiffusion. Apesar de não contar com mecanismos formais de privacidade, o modelo de difusão teve pontuações competitivas, indicando que sua natureza estocástica pode oferecer benefícios inerentes à privacidade.

Tabela 1 - Comparação dos métodos (valores entre 0 (pior) e 100 (melhor)).

| Métrica de avaliação | | CTGAN | | Gaussian Copula | | SimpleTableDiffusion | |
|----------------------|------------------------------|-----------|-----------|-----------------|-----------|----------------------|-----------|
| Fidelidade | Complemento KS | 89.9(1.3) | | 78.0(6.0) | | 78.3(1.1) | |
| | Complemento TV | 97.9(0.2) | | 71.3(2.0) | | 81.0(1.0) | |
| | Correlação Similaridade | 95.2(0.2) | | 99.0(0.3) | | 97.2(0.4) | |
| | Similaridade de Contingência | 91.7(1.3) | | 54.8(1.5) | | 65.1(1.8) | |
| Utilidade | | Real | Sintético | Real | Sintético | Real | Sintético |
| | Micro-F1 (LR) | 87.7(0.0) | 87.7(0.0) | 87.7(0.1) | 84.8(5.2) | 87.7(0.1) | 87.7(0.0) |
| | Macro-F1 (LR) | 47.7(1.8) | 46.7(0.0) | 47.7(1.8) | 51.8(7.3) | 47.1(1.8) | 46.8(0.1) |
| | Micro-F1 (RF) | 87.5(0.2) | 87.6(0.1) | 87.5(0.2) | 87.7(0.0) | 85.7(0.2) | 87.0(0.0) |
| | Macro-F1 (RF) | 50.9(1.3) | 46.8(0.1) | 50.9(1.3) | 46.9(0.3) | 50.9(1.3) | 46.8(0.0) |
| | Micro-F1 (XGB) | 87.7(0.1) | 87.7(0.0) | 87.7(0.1) | 87.5(0.4) | 87.7(0.1) | 85.1(4.4) |
| | Macro-F1 (XGB) | 47.8(1.9) | 46.7(0.0) | 47.8(1.9) | 47.7(2.8) | 47.8(1.9) | 49.7(3.4) |
| Privacidade | Pontuação de privacidade | 88.6(0.6) | | 86.6(1.4) | | 83.6(0.1) | |

De forma geral, os resultados evidenciam que não há um modelo que domine todas as dimensões simultaneamente. O CTGAN se destacou pela consistência entre fidelidade e privacidade, sendo especialmente eficaz em dados categóricos. O método Gaussian Copula mostrou-se robusto na preservação de dependências estatísticas, enquanto o modelo de difusão ofereceu uma alternativa promissora com desempenho competitivo e menor complexidade computacional. Esses achados reforçam o potencial de abordagens híbridas e a importância de escolhas metodológicas alinhadas aos objetivos específicos de cada aplicação em saúde pública.

4. DESAFIOS E APRENDIZADOS

O principal desafio foi a implementação do modelo baseado em difusão. Diferentemente dos modelos estatísticos e GAN, disponíveis na biblioteca SDV^[4], a construção de um modelo personalizado exigiu aderência aos princípios de difusão e adaptação aos desafios dos dados tabulares. Equilibrar eficiência computacional e qualidade de geração demandou experimentação iterativa. Nesse processo, aprofundamos a compreensão da dinâmica do modelo, solucionamos instabilidades de treinamento e ganhamos insights sobre os compromissos arquiteturais em contextos de dados de saúde.

Agradecimentos. Esse trabalho conta com o apoio do CI-IA-Saúde, CNPq, FAPESP, FAPEMIG, Unimed-BH e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).

5. REFERÊNCIAS

As referências foram condensadas no link a seguir:

<https://docs.google.com/document/d/1PaG8FUxhBgrYPwMgKDmeADxkYhH485Ep6JQQ1NZiDNXU/edit?usp=sharing>

Avaliando os potenciais do algoritmo k-vizinhos mais próximos (KNN) na tarefa imputação em bases de dados clínicos

Izadora Monken Ganem¹, Guilherme D. Bianco³, Luciano Lima¹, José Carlos Serufo Filho¹, Milena Soriano Marcolino¹, Leonardo Rocha², Marcos André Gonçalves¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG. ²Universidade Federal de São João del-Rei, São João del-Rei, MG. ³Universidade Federal da Fronteira Sul

izadoraganem@gmail.com, magoncalv@gmail.com

Resumo: A integridade dos dados é crucial para o desempenho de modelos de aprendizado de máquina em saúde, sendo a ausência de valores um dos principais desafios. Este estudo compara cinco métodos de imputação de dados em um conjunto clínico com dados de mais de 16.000 pacientes com diagnóstico de covid-19. Os resultados mostram que o MissForest apresenta maior acurácia, embora com alto custo computacional. Como alternativa, propõe-se um imputador baseado no algoritmo k-vizinhos mais próximos (KNN), customizado para o contexto analisado. Apesar da menor acurácia, destaca-se pela eficiência computacional. A análise indica que características como homogeneidade dos dados e entropia local afetam seu desempenho. Em cenários com baixa variabilidade, o método mostrou-se promissor, sugerindo sua aplicabilidade em abordagens híbridas, equilibrando robustez e eficiência na tarefa de imputação.

Abstract: Data integrity is critical for the performance of machine learning models in healthcare, with missing values posing a major challenge. This study compares five data imputation methods using a clinical dataset of over 16,000 COVID-19 patients. Results show that MissForest achieves the highest accuracy but at a high computational cost. As an alternative, we propose a KNN-based imputer tailored to the dataset. While less accurate, it offers superior computational efficiency. Analysis reveals that data homogeneity and local entropy influence its performance. In low-variability settings, the method proves promising, supporting its use in hybrid approaches that balance robustness and efficiency in missing data imputation.

Palavras-chave: Imputação de dados; Inteligência Artificial; COVID-19; aprendizado de máquina.

Nome do projeto: Predição de desfechos clínicos e econômicos por meio de representações semânticas multi-modais de pacientes resilientes a drifts temporais.

1. CONTEXTO E MOTIVAÇÃO

Soluções baseadas em aprendizado de máquina (AM) têm sido cada vez mais aplicadas na saúde devido à sua acurácia e escalabilidade ⁽¹⁾. A eficácia desses modelos depende da qualidade dos dados, frequentemente comprometida por valores ausentes causados por falhas na integração, erros de coleta ou omissões. Essas lacunas, comuns em bases clínicas, podem introduzir vieses e comprometer a acurácia preditiva. Para mitigar esse problema, técnicas de imputação de dados, estatísticas ou baseadas em AM têm sido amplamente estudadas ⁽²⁾, especialmente na saúde, onde imputações imprecisas podem prejudicar o apoio de decisões médicas pelos modelos.

1.1 Objetivos

Este trabalho tem três objetivos. O primeiro é realizar uma comparação experimental entre diferentes métodos de imputação, com o intuito de identificar aqueles com melhor desempenho. O segundo é avaliar a eficácia do algoritmo K-nearest neighbors (KNN) como imputador, partindo da hipótese de que

a utilização de pacientes com quadros clínicos similares ao caso-alvo tende a gerar bons resultados. Por fim, o terceiro objetivo é analisar os fatores que influenciam o desempenho do KNN, visando identificar os contextos em que o método se mostra mais eficaz.

2. METODOLOGIA E RESULTADOS

Trata-se de estudo observacional retrospectivo, que utilizou base de dados com 16 mil pacientes com covid-19 confirmada, contendo 62 variáveis clínicas que descrevem o quadro de saúde dos pacientes ⁽³⁾. Avaliamos a qualidade da imputação de cinco métodos (KNN, MissForest ⁽⁵⁾, *Multivariate Imputation by Chained Equations* (MICE)⁽⁶⁾, redes neurais adversárias (GAIN) e média dos valores das demais instâncias) sobre a variável relação saturação periférica de oxigênio por fração inspirada de oxigênio (*SatFio2*), escolhida por ter uma forte relação com o desfecho clínico de óbito ⁽³⁾. Para simular a ausência de dados, removeu-se de maneira artificial 50% dos dados presentes nessa variável utilizando o método *Missing Completely at Random* (MCAR) ⁽³⁾, que insere valores faltantes de forma aleatória, sem viés associado a outras variáveis. Para garantir uma avaliação robusta, utilizou-se a estratégia de validação cruzada entre os dados ausentes dividindo-os em 10 subconjuntos disjuntos de teste, utilizando os demais para o treinamento dos modelos. O desempenho de cada método foi avaliado por meio do erro absoluto médio (MAE) e do erro quadrático médio (MSE) em relação aos valores originais, além do tempo de execução (segundos), apresentados na Tabela 1. O valor médio da variável *SatFio2* é de 369.83 e a mediana é de 428.7; ambos servem de referência para interpretar os erros.

Tabela 1 - Efetividade e Custo Computacional (tempo) dos Métodos de Imputação.

| Métricas | Média | KNN | KNN++ | MissForest | MICE | GAIN |
|--------------|-------|-------|-------|------------|-------|-------|
| MAE | 95.74 | 94.47 | 87.32 | 52.17 | 64.88 | 94.51 |
| MSE | 19061 | 25731 | 20101 | 11281 | 15742 | 18013 |
| Tempo (seg.) | 0.014 | 0.16 | 0.082 | 228.95 | 5.15 | 9.39 |

Os resultados indicam que os métodos MissForest (MAE = 52.17) e MICE (MAE = 64.88) se destacam pela efetividade, porém apresentam alto custo computacional, enquanto a média (MAE= 95.74) e GAIN (MAE = 94.51) demonstram desempenho inferior. O KNN (MAE = 94.47), embora com baixa efetividade, se diferencia pelo baixo custo de processamento. Exploramos otimizações do KNN levando em conta fatores que influenciam seu desempenho: as métricas de distância, o tamanho da vizinhança e a quantidade de variáveis no conjunto de treino. Segundo avaliação preliminar, propomos o KNN++ – uma versão otimizada do KNN –, configurado com distância de Manhattan, k=3 e 10 variáveis mais relacionadas ao *SatFio2*, selecionadas por permutação de importância: VM (ventilação mecânica), *po2fio2* (pressão parcial de oxigênio/fração inspirada de oxigênio), FR (frequência respiratória), óbito, HospitalGDP, plaquetas, FC (frequência cardíaca), bicarbonato, PH (ph sanguíneo), PCR (proteína c reativa). Como mostrado na Tabela 1, o KNN++ melhora tanto na efetividade quanto no tempo de execução em relação ao KNN original.

Aprofundando a avaliação do KNN++, analisamos seu desempenho considerando a estrutura das vizinhanças. Após validação cruzada e imputação, representamos cada paciente em um espaço bidimensional dividido em quatro quadrantes pela média de dois fatores: distância média aos vizinhos mais próximos (eixo X) e entropia da diferença de vizinhança (eixo Y). A Tabela 2 mostra que o KNN++ apresenta desempenho competitivo em cenários de alta homogeneidade (quadrantes com alta similaridade e baixa entropia). Em regiões mais heterogêneas seu desempenho é inferior em comparação aos demais métodos. O quadrante com melhor desempenho do KNN++ (baixa entropia e distância) concentra 42% dos pacientes, enfatizando a relevância prática em contextos específicos.

Tabela 2 - Média dos erros por quadrante com significância estatística entre os erros do KNN em relação aos demais. Os símbolos † e * reportam perda e empate estatístico, respectivamente, do KNN++ em relação aos demais.

| Quadrante | KNN++ | MissForest | MICE | Porcentagem da Coorte |
|-------------------|--------------|------------|-------|-----------------------|
| Direita Inferior | 79.78 (↓,*) | 65.61 | 79.03 | 9% |
| Direita Superior | 146.40 (↓,↓) | 54.94 | 79.32 | 28% |
| Esquerda Inferior | 59.97 (↓,↓) | 47.25 | 52.59 | 42% |
| Esquerda Superior | 63.14 (↓,*) | 50.44 | 60.62 | 21% |

3. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do projeto, diversos desafios foram enfrentados, tanto no tratamento dos dados quanto na avaliação dos métodos de imputação. Uma das principais dificuldades encontradas foi a alta heterogeneidade de dados, que ilustra o cenário instável da pandemia de covid-19. Além disso, a parametrização do KNN++ foi particularmente desafiadora, exigindo diversas experimentações e validações para chegar a um resultado positivo que se destacasse. Concluímos que, em cenários nos quais a vizinhança é mais homogênea — com alta similaridade média e baixa entropia — o desempenho do KNN++ pode ser competitivo em relação aos métodos estado da arte, com custo computacional significativamente menor (aproximadamente 70 e 2.850 vezes menor que o tempo do MICE e MissForest, respectivamente). Esses resultados indicam a importância de considerar a estrutura local da vizinhança dos pontos ao se utilizar o KNN e a possibilidade do desenvolvimento de algoritmos híbridos para realizar a imputação.

Agradecimentos. Ministério da Saúde, Secretaria de Atenção Primária à Saúde (TED 31/2022, Processo 25000.058753/2022-59). Esse trabalho conta com o apoio do CNPq, FAPEMIG, FAPESP, Unimed-BH, CIIA-Saúde e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação - INCT-TILD-IAR.

4. REFERÊNCIAS

1. Paiva BBM, 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;13(1):3463.
2. Liu, M., Li, S., Yuan, H., Ong, Y., Xie, F., Saffari, S. E., Shang, Y., Volovici, V., Chakraborty, B., et al. (2023). Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques. *Artificial intelligence in medicine*, 142:102587.
3. Lana, F. C. B., Marinho, C. C., de Paiva, B. B. M., Valle, L. R., do Nascimento, G. F., da Rocha, L. C. D., Carneiro, M., Batista, J. d. L., Anschau, F., Paraíso, P. G., Bartolazzi, F., Cimini, C. C. R., Schwarzbald, A. V., Rios, D. R. A., Gonçalves, M. A., and Marcolino, M. S. (2025). Unraveling relevant cross-waves pattern drifts in patient-hospital risk factors among hospitalized covid-19 patients using explainable machine learning methods. *BMC Infectious Diseases*, 25(1):537
4. Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
5. Shadbahr, T., Roberts, M., Stanczuk, J., Gilbey, J., Teare, P., Dittmer, S., Thorpe, M., Torné, R. V., Sala, E., Lío, P., Patel, M., Preller, J., Rudd, J. H. F., Mirtti, T., Rannikko, A. S., Aston, J. A. D., Tang, J., and Schönlieb, C.-B. (2023). The impact of imputation quality on machine learning classifiers for datasets with missing values. *Communications Medicine*, 3(1):139.
6. Jarrett, D., Cebere, B., Liu, T., Curth, A., and van der Schaar, M. (2022). Hyperimpute: Generalized iterative imputation with automatic model selection.

Caracterização de atributos para predição de complicações graves do diabetes com base em registros administrativos de saúde

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

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

Resumo: A implantação de um programa efetivo de Saúde Preventiva exige o tratamento e a análise de dados para antecipar eventos de saúde monitorados. Sistemas computacionais, aliados a métodos estatísticos e aprendizado de máquina, facilitam a análise de grandes volumes de informações e a construção de modelos preditivos. Este estudo tem como objetivo apresentar a estratégia adotada na construção da variável resposta para um modelo preditivo voltado à identificação de indivíduos com diabetes com alta propensão a eventos graves como complicações retinianas, renais, cardiovasculares, entre outras. Para isso, foram utilizadas informações de exames e procedimentos extraídos de registros administrativos codificados pela TUSS, referentes a beneficiários diabéticos de uma grande operadora de saúde da região Centro-Sul do Brasil. A proposta visa apoiar a prevenção e a gestão clínica, melhorando a qualidade de vida dos pacientes e a eficiência dos serviços de saúde.

Palavras-chave: Diabetes mellitus; modelo preditivo; aprendizado de máquina; eventos graves; saúde suplementar; registros administrativos.

Abstract: The implementation of an effective Preventive Health program requires data processing and analysis to anticipate monitored health events. Computational systems, combined with statistical methods and machine learning, facilitate the analysis of large volumes of information and the development of predictive models. This study aims to present the strategy adopted in constructing the response variable for a predictive model designed to identify individuals with diabetes who are at high risk of experiencing severe events, such as retinal, renal, cardiovascular complications, among others. To this end, data on tests and procedures were extracted from administrative records coded according to the TUSS standard, referring to diabetic beneficiaries of a large health insurance provider in the South-Central region of Brazil. The proposal seeks to support prevention and clinical management, improving patients' quality of life and the efficiency of healthcare services.

Keywords: Diabetes mellitus; predictive model; machine learning; severe health events; supplementary health system; administrative health data.

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

1. CONTEXTO E MOTIVAÇÃO

O diabetes mellitus é uma doença crônica que, quando não adequadamente controlada, pode levar a uma série de eventos graves, tanto de natureza aguda quanto crônica. Essas complicações representam riscos significativos à saúde, podendo resultar em incapacidades permanentes ou até mesmo em óbito⁽¹⁾. De acordo com os dados da Pesquisa Nacional de Saúde (PNS) 2019, conduzida pelo IBGE em parceria com o Ministério da Saúde, 7,7% da população brasileira com 18 anos ou mais relataram ter recebido diagnóstico médico de diabetes. Esse percentual corresponde a aproximadamente 12,3 milhões de pessoas, representando um aumento expressivo em relação aos 6,2% registrados em 2013⁽²⁾. Diante desse cenário,

a Agência Nacional de Saúde Suplementar (ANS) tem incentivado as operadoras de planos de saúde a implementarem programas de promoção da saúde e prevenção de doenças crônicas, com o objetivo de melhorar a qualidade de vida dos beneficiários⁽³⁾. A ênfase em ações preventivas se justifica pelo fato de que a maioria das doenças crônicas, incluindo o diabetes tipo 2, é passível de prevenção e controle eficaz, especialmente por meio do acompanhamento adequado e mudanças no estilo de vida.

1. 1 Objetivo

Este estudo tem como objetivo apresentar a estratégia adotada na construção da variável resposta para um modelo preditivo voltado à identificação de indivíduos com diabetes com alta propensão a eventos graves decorrentes do mau controle da doença, como complicações retinianas, renais, cardiovasculares, entre outras.

2. ATIVIDADES PRINCIPAIS

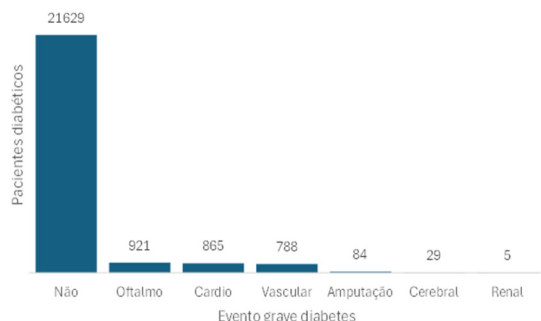
A construção das variáveis preditoras e da variável resposta do modelo foi realizada com base exclusivamente nos registros administrativos disponíveis na base de dados de uma grande operadora de saúde da região Centro-Sul do Brasil. Esses registros incluem informações sobre a realização de exames e procedimentos, com respectivas datas e códigos TUSS (Terminologia Unificada da Saúde Suplementar), mas não possuem os resultados clínicos ou laboratoriais dos exames.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A variável resposta do modelo, denominada evento grave de diabetes, foi construída com base nos procedimentos realizados por pacientes com diagnóstico de diabetes no período de 2015 a 2024. Para isso, foram utilizados dados administrativos de atendimentos de saúde, codificados segundo a TUSS. Os dados foram inicialmente agrupados por código TUSS, permitindo a identificação e o agrupamento de todos os procedimentos realizados por cada paciente ao longo do período analisado. Em seguida, com base na descrição e na finalidade clínica de cada procedimento, um médico especialista realizou a classificação dos códigos TUSS em categorias que representam eventos graves associados ao diabetes, conforme as principais complicações conhecidas da doença. Os eventos foram categorizados nos seguintes grupos: Oftalmológicos (ex.: retinopatia com introflexão escleral); Renais (ex.: angioplastia renal); Amputações (ex.: amputação ao nível do pé); Cardiovasculares (ex.: revascularização do miocárdio); Vasculares periféricos (ex.: angioplastia); Cerebrais (ex.: tratamento cirúrgico da isquemia cerebral). Dos 24.321 pacientes analisados, a maioria (21.624) não apresentou evento grave no período observado. Entre os que apresentaram eventos graves, 921 foram classificados como oftalmológico, 865 como cardiovascular, 788 como vascular, 84 como amputação, 29 como cerebral e 5 pacientes como renal (Figura 1). Para o modelo preditivo foi considerado a presença de pelo menos um procedimento classificado em qualquer uma dessas categorias como indicativo de um evento grave relacionado ao diabetes, e a variável resposta foi codificada como binária (1 = ocorrência de evento grave; 0 = ausência). A Figura 2 ilustra a distribuição de uma variável preditora – tempo médio entre exames de hemoglobina glicada – entre pacientes com e sem evento grave. Observa-se uma concentração de pacientes com evento grave em intervalos inferiores a 500 dias, indicando um acompanhamento laboratorial mais frequente neste grupo. Por outro lado, pacientes sem evento grave tendem a apresentar intervalos médios superiores a 500 dias entre os exames. Essa diferença pode refletir maior atenção clínica nos casos com complicações ou uma resposta relativa à progressão da doença.

Figura 1 – Distribuição dos pacientes segundo evento grave de diabetes.

Figura 2 - Distribuição dos pacientes segundo tempo médio de exame de hemoglobina glicada por tipo de evento grave.



Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da UFMG, sob o parecer nº 6.538.883.

4. DESAFIOS E APRENDIZADOS

O principal desafio deste estudo foi a construção da variável resposta, limitada pela falta de dados clínicos nos registros administrativos. Muitos laboratórios da rede credenciada realizam os exames, mas não compartilham os laudos com a operadora, restringindo o acesso a desfechos clínicos objetivos, como níveis de glicemia. Diante da ausência de indicadores clínicos diretos, adotou-se uma abordagem alternativa baseada na classificação dos procedimentos realizados, conforme codificação TUSS. Esse processo exigiu rigor metodológico e interpretação médica especializada para representar adequadamente a gravidade dos eventos relacionados ao diabetes.

Como aprendizado relevante, destaca-se a importância da atuação de uma equipe multidisciplinar, composta por médicos, estatísticos e profissionais da saúde, cuja colaboração foi essencial para a validação clínica dos dados, a construção das variáveis e a interpretação adequada dos resultados. Essa integração entre conhecimento técnico e clínico se mostrou fundamental para superar as limitações dos dados e garantir a robustez do modelo proposto.

Agradecimentos. Esse trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unimed-BH e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).

5. REFERÊNCIAS

1. World Health Organization. (2021). Diabetes. <https://www.who.int/news-room/fact-sheets/detail/diabetes>
2. BRASIL. Instituto Brasileiro de Geografia e Estatística – IBGE. Pesquisa Nacional de Saúde 2019: percepção do estado de saúde, estilos de vida, doenças crônicas e saúde bucal. Rio de Janeiro: IBGE, 2020. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101764.pdf>. Acesso em: 17 maio 2025.
3. BRASIL. Agência Nacional de Saúde Suplementar – ANS. Programa de Promoção da Saúde e Prevenção de Riscos e Doenças. Brasília: ANS, 2022. Disponível em: <https://www.gov.br/ans/pt-br/assuntos/operadoras/compromissos-e-interacoes-com-a-ans-1/programas-ans-1/promoprev>. Acesso em: 17 maio 2025.

Desafios éticos e legais da inteligência artificial na telessaúde: perspectivas sobre proteção de dados no Brasil

Juliana Almeida de Moraes ¹, Cesar Nobre Pereira ²

^{1,2} Programa de Pós-Graduação em Ciências da Saúde – Infectologia e Medicina Tropical, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

ju.almeidademoraes@gmail.com, c.nobrepereira@gmail.com

Resumo: A inteligência artificial (IA) tem se destacado como uma tecnologia promissora na área da saúde, especialmente por sua capacidade de processar grandes volumes de dados e otimizar processos clínicos. No contexto da telessaúde, seu uso pode contribuir significativamente para ampliar o acesso, sobretudo em regiões com carência de profissionais especializados e de difícil acesso. No entanto, sua aplicação envolve desafios éticos e legais, em especial no que se refere à proteção de dados sensíveis, conforme as diretrizes da Lei Geral de Proteção de Dados (LGPD). Este estudo realizou uma revisão narrativa da literatura com o objetivo de analisar como a IA pode ser integrada à telessaúde de forma segura e em conformidade com a legislação brasileira. Foram selecionados quatro artigos, a partir de um total de 15 trabalhos identificados, que abordam aspectos técnicos, éticos e regulatórios. Os resultados apontam que, embora a IA represente uma ferramenta promissora na área da saúde, sua adoção exige atenção aos riscos relacionados à confidencialidade e ao uso ético das informações dos pacientes. O estudo reforça a importância de diretrizes técnicas e jurídicas claras para orientar a incorporação responsável da IA no campo da telessaúde.

Palavras-chave: inteligência artificial; telessaúde; proteção de dados; LGPD.

Nome do projeto: Projeto de mestrado: Pós-Graduação em Ciências da Saúde, Infectologia e Medicina Tropical.

1. CONTEXTO E MOTIVAÇÃO

A inteligência artificial (IA) vem se consolidando como uma ferramenta estratégica na área da saúde, apresentando aplicações que consistem desde o apoio diagnóstico até a gestão de fluxos assistenciais mais complexos. A sua capacidade de processar grandes volumes de dados e identificar padrões relevantes permite avanços importantes de eficiência e precisão, sobretudo em contextos de alta demanda e limitações de recursos. No entanto, a incorporação da IA à prática clínica demanda uma análise criteriosa dos impactos éticos, técnicos e regulatórios provenientes de sua adoção. Nesse cenário, a telessaúde surge como uma alternativa para ampliar o acesso aos serviços de saúde, sobretudo em regiões de difícil acesso e com carência de profissionais técnicos especializados.

Os procedimentos médicos digitais e a circulação intensa de informações sensíveis exigem estruturas robustas de segurança da informação e de proteção de dados. Se devidamente utilizada, a IA pode desempenhar um papel fundamental na integração, validação e uso inteligente desses dados. Contudo, o seu uso requer a observância aos dispositivos legais vigentes, sobretudo da Lei Geral de Proteção de Dados (LGPD). Diante disso, torna-se essencial investigar de que maneira a IA pode ser utilizada de forma segura, ética e em conformidade com a legislação brasileira, contribuindo para a qualificação das práticas em telessaúde sem comprometer os direitos fundamentais à privacidade e à confidencialidade dos dados dos pacientes.

1.1 Objetivo

Analisar como a Inteligência Artificial pode contribuir para a integração segura de dados em telessaúde conforme as normas de proteção de dados brasileira.

2. ATIVIDADES PRINCIPAIS

As atividades desenvolvidas neste estudo consistiram em:

- Levantamento das bases de dados e seleção dos artigos, considerando o objetivo da proposta;
- Leitura crítica dos estudos elegíveis;
- Elaboração da revisão narrativa com abordagem da temática multidisciplinar de forma a contribuir para a produção científica sobre o tema;
- Organização dos pontos relevantes dos artigos selecionados em forma de tabela, destacando as contribuições da IA ética e segura na telessaúde.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Foi realizada uma revisão narrativa da literatura com o foco da contribuição da IA na telessaúde considerando os aspectos pertinentes à LGPD. A busca foi realizada nas seguintes bases de dados: Scielo, PubMed e Plataforma Sucupira, utilizando os respectivos termos de busca: ("inteligência artificial") AND ("saúde digital" OR "telessaúde") AND ("proteção de dados" OR "LGPD") AND ("ética" OR "bioética"). Foram identificados 15 artigos por meio dos critérios de inclusão, disponíveis em texto completo e relacionados à temática proposta. Foi realizada uma triagem dos artigos após leitura dos seus respectivos resumos sendo selecionados quatro artigos publicados nos últimos cinco anos que atendiam aos critérios de inclusão previamente definidos. Posteriormente, foi realizada uma análise qualitativa dos estudos, os quais foram compilados e estratificados e categorizados em uma tabela a fim de elencar os principais pontos e resultados dos estudos selecionados.

3.2 Resultados

Os principais achados segundo os itens encontrados nos estudos estão dispostos na Tabela 1:

| AUTORES | OBJETIVO DO ESTUDO | METODOLOGIA | PRINCIPAIS RESULTADOS |
|--------------------------|--|--|---|
| Yadav et al. (2023) | Analisar os desafios e oportunidades da aplicação da IA na saúde em relação à privacidade de dados | Revisão Sistemática da literatura com base em estudos empíricos e teóricos. | A IA pode melhorar a eficiência e personalização do cuidado, mas impõe riscos à privacidade se não houver regulamentação clara. |
| Nascimento et al. (2024) | Analisar os desafios éticos e legais da IA na saúde digital em conformidade com a LGPD | Estudo teórico reflexivo com base na legislação brasileira e na literatura especializada | Verificou a importância de diretrizes éticas, consentimento informado e transparência nos sistemas de IA. |
| Santana et al. (2023) | Analisar o uso da IA na saúde pública | Revisão sistemática | Dificuldade de controle, privacidade e governança de dados em saúde no SUS. |
| Brito C, Lopes L (2024) | Analisar a aplicação da IA no atendimento médico | Revisão Sistemática | O papel principal consiste em ajudar os médicos a melhorar sua eficiência e precisão. |

4. DESAFIOS E APRENDIZADOS

Durante a elaboração do trabalho, um dos principais desafios foi localizar estudos que articulassem, de forma multidisciplinar, a temática da IA, telessaúde e proteção de dados no contexto brasileiro. Muitos estudos abordam apenas um desses eixos temáticos, o que exigiu uma maior dedicação na articulação desses temas de forma multidisciplinar. Outro desafio consistiu na interpretação crítica dos aspectos técnicos da IA sob a ótica das exigências legais da LGPD, considerando que essas áreas nem sempre estabelecem uma conexão direta. O processo proporcionou um aprendizado multidisciplinar relevante, destacando a importância da avaliação ética e normativa de tecnologias digitais aplicadas à saúde. Ademais, o desenvolvimento da revisão narrativa contribuiu para o aprimoramento da capacidade de síntese e para a identificação de tendências e lacunas que podem orientar e sugerir futuras pesquisas na área.

5. REFERÊNCIAS

1. Yadav N, Pandey S, Gupta A, Dudani P, Gupta S, Rangarajan K. Data Privacy in Healthcare: In the Era of Artificial Intelligence. *Indian Dermatology Online Journal*. dezembro de 2023;14(6):788.
2. Nascimento SM, Paiva TMGD, Kasuga MPM, Silva TDAF, Crozara CMG, Byk J, et al. Inteligência artificial e suas implicações éticas e legais: revisão integrativa. *Rev Bioét*. 2024;32:e3729PT.
3. Santana GA de, Vieira EL de C, Duarte Z, Mello RC. Inteligência Artificial na saúde pública do Brasil: princípios éticos aplicados à privacidade e construção de algoritmos. *Revista Fontes Documentais*. 2023;6(Ed. Especial):73–5.
4. Brito CMS, Lopes LEC. Aplicação e impacto da inteligência artificial no atendimento médico: uma revisão sistemática da literatura. *ResearchGate* [Internet]. 20 de dezembro de 2024 [citado 30 de maio de 2025]; Disponível em: https://www.researchgate.net/publication/387168576_Aplicacao_e_impacto_da_inteligencia_artificial_no_atendimento_medico_uma_revisao_sistemica_da_literatura

Desafios para a previsão de custos hospitalares na admissão: uma análise baseada em predições de variáveis intermediárias para internações por COVID-19

Ricardo Bertoglio Cardoso¹, Guilherme Fonseca do Nascimento², Leonardo Chaves Dutra da Rocha³, Milena Soriano Marcolino², Ana Paula Beck da Silva Etges⁴, Carisi Anne Polanczyk¹, Jussara M. Almeida², Marcos André Gonçalves²

¹Universidade Federal do Rio Grande do Sul. ²Universidade Federal de Minas Gerais.

³Universidade Federal de São João Del-Rei. ⁴Hospital de Clínicas de Porto Alegre.

autor correspondente: ricardobcardoso@gmail.com

Resumo: A predição de custos hospitalares na admissão hospitalar é uma tarefa importante para a gestão eficiente de recursos em saúde. No entanto, trata-se de uma tarefa desafiadora pela escassez natural de informações do paciente, sendo ainda mais crítica em contextos de alta incerteza clínica, como durante a pandemia de COVID-19. Neste trabalho, investigamos o quanto é possível melhorar a predição de custos nesses cenários a partir de predições intermediárias de três variáveis sabidamente relacionadas a custos ⁽³⁾: admissão em UTI, tempo de permanência e desfecho clínico. Para isso, propomos e avaliamos três fluxos preditivos multiestágio (Pipeline) baseados no XGBoost que realizam essas predições intermediárias, combinando-as para estimar o custo final, tendo nossos resultados indicando grande potencial das abordagens propostas.

Abstract: Predicting hospital costs at the time of hospital admission is an important task for the efficient management of healthcare resources. However, it is a challenging task due to the natural scarcity of patient information and is even more critical in contexts of high clinical uncertainty, such as during the COVID-19 pandemic. In this work, we investigate how much it is possible to improve cost prediction in these scenarios based on intermediate predictions of three variables known to be related to costs ⁽³⁾: ICU admission, length of stay, and clinical outcome. To this end, we propose and evaluate three multistage predictive flows (Pipeline) based on XGBoost that perform these intermediate predictions, combining them to estimate the final cost. Our results indicate significant potential for the proposed approaches.

Palavras-chave: custos em saúde; inteligência artificial; COVID-19; aprendizado de máquina.

Nome do projeto: Predição de desfechos clínicos e econômicos por meio de representações semânticas multi-modais de pacientes resilientes a drifts temporais.

1. CONTEXTO E MOTIVAÇÃO

Predição de custos em saúde é um componente essencial para a alocação adequada de recursos e o suporte a modelos de cuidado baseados em valor. Previsões antecipadas e precisas melhoram a eficiência da gestão desses recursos. No entanto, essa tarefa torna-se particularmente desafiadora no momento da admissão hospitalar, quando há informações limitadas sobre o diagnóstico e o plano terapêutico. Fatores clínicos dinâmicos como infecções hospitalares e resposta individual ao tratamento, e variáveis externas aos protocolos assistenciais podem impactar significativamente o uso de recursos e os custos associados ⁽¹⁾. Um exemplo que ilustra bem esse cenário foi a pandemia de COVID-19, marcada por interrupções nas operações hospitalares, falhas nas cadeias de suprimentos, medidas emergenciais de controle de infecção e aumento abrupto na demanda por atendimento ⁽²⁾. A escassez de conhecimento sobre a doença, a imprevisibilidade da evolução clínica e a ausência de protocolos consolidados dificultaram ainda mais a predição precisa dos custos nesse cenário crítico.

1.1 Objetivos

Avaliar o quanto é possível melhorar a qualidade da predição de custos hospitalares de pacientes no momento de admissão a partir de predições intermediárias de três variáveis sabidamente relacionadas a custos ⁽³⁾: admissão em UTI, tempo de permanência e desfecho clínico.

2. METODOLOGIA E RESULTADOS

Este estudo de coorte retrospectivo e multicêntrico analisou dados clínicos e de custos hospitalares de 717 pacientes com COVID-19, internados entre março e agosto de 2020 em cinco hospitais das regiões Nordeste e Sul do Brasil. Os dados foram coletados no âmbito do projeto do IATSCNPq e os custos estimados pelo método Time-Driven Activity-Based Costing (TDABC), apresentados em dólares internacionais (Int\$), com paridade de poder de compra de 2021. A pesquisa foi aprovada pelo comitê de ética sob o CAAE: 30525920.7.1001.0008.

Três fluxos de previsão multiestágio - Pipelines - baseados em XGBoost (4) foram propostos para estimar os custos hospitalares, abordando tarefas de classificação (categorias de custos: baixo [Int\$-58-Int\$982], médio [Int\$983-Int\$3.218] e alto [Int\$3.219-Int\$55.641]) e regressão (valor contínuo). As pipelines incorporaram a previsão de variáveis intermediárias como admissão em UTI (UTI=1, UTI=0), tempo de permanência (≤ 3 dias, 4-14, 15-30, ≥ 31) e desfecho clínico (alta ou óbito). A **Pipeline 1** empregou uma abordagem sequencial e estratificada, onde as previsões de admissão em UTI, desfecho clínico e tempo de permanência foram usadas como entradas para um modelo final de previsão de custos. A **Pipeline 2**, uma versão simplificada da Pipeline 1, utilizou apenas a previsão de admissão em UTI para estratificar os pacientes, treinando modelos de custo separados para cada grupo sem incorporar previsões para tempo de permanência ou desfecho clínico. A **Pipeline 3** representa o desempenho dos modelos, o qual foi avaliado usando o Macro F1 para a tarefa de classificação e o Mean Absolute Percentage Error (MAPE) para a tarefa de regressão. Cenários com "oráculos" (substituição das previsões intermediárias por valores perfeitamente precisos) foram utilizados para avaliar o quanto um conhecimento perfeito dessas variáveis pode melhorar na predição de custo final. A partição dos dados foi realizada utilizando a estratégia hold-out (50% treinamento, 50% teste), repetida dez vezes. Os resultados foram relatados como valores médios com intervalos de confiança de 95%. Nas comparações estatísticas utilizamos o teste t, com correção de Bonferroni.

Tabela 1 - Análise comparativa de modelos de previsão de custos. O desempenho das Pipelines foi comparado ao modelo tradicional. Diferenças estatísticas são destacadas em verde.

| Tarefa (todos os hospitais) | Tradicional | Pipeline 1 | | Pipeline 2 | | Pipeline 3 | |
|-----------------------------|-------------|------------|------------|------------|------------|------------|------------|
| | | - | Oráculo | - | Oráculo | - | Oráculo |
| Classificação | 50.5(1.3) | 49.1(1.2) | 72.2(1.2) | 49.8(1.4) | 59.6(1.3) | 47.8(2.9) | 73.9(1.4) |
| Regressão | 2.61(0.24) | 1.91(0.15) | 0.80(0.05) | 2.29(0.14) | 1.60(0.14) | 1.80(0.16) | 0.67(0.02) |

Na Tabela 1 apresenta-se uma comparação das três pipelines propostas (e as respectivas versões utilizando oráculos) com a estratégia que realiza a predição de custos diretamente (Tradicional). Os resultados mostram que as Pipelines 1 e 3 foram capazes de melhorar a qualidade das regressões do modelo tradicional, porém nenhuma superou a estratégia tradicional na tarefa de classificação. Ao utilizarem o oráculo, todas as Pipelines apresentaram resultados superiores ao modelo tradicional, indicando que há um grande potencial de se utilizar predições de variáveis intermediárias. Ao mesmo tempo, a tarefa de prever essas variáveis intermediárias é bastante difícil. O destaque fica por conta do Pipeline 3 com o melhor resultado de regressão e o maior potencial (oráculo) tanto na predição quanto na regressão.

Tabela 2 - Análise isolada das variáveis-chave do Pipeline 3. Diferenças estatísticas do modelo com seu próprio oráculo são destacadas em verde.

| Tarefa (todos os hospitais) | Tradicional | Apenas Admissão UTI | | Apenas Tempo de Permanência | | Apenas Desfecho Clínico | |
|--------------------------------|-------------|---------------------|------------|-----------------------------|------------|-------------------------|------------|
| | | - | Oráculo | - | Oráculo | - | Oráculo |
| Classificação | 50.5(1.3) | 36.6(0.9) | 47.8(0.7) | 38.3(1.6) | 62.8(1.0) | 28.6(1.1) | 34.1(1.1) |
| Regressão | 2.61(0.24) | 2.30(0.10) | 1.68(0.11) | 2.31(0.17) | 1.00(0.04) | 3.26(0.14) | 3.33(0.18) |

Focando na Pipeline 3 de maior potencial, avaliamos a qualidade final da predição de custos considerando apenas uma das variáveis individualmente, seja explorando a saída da previsão intermediária como entrada para o modelo de custo, seja substituindo esta previsão por um oráculo, para cada variável (Tabela 2). Conforme esperado, nenhuma das variáveis isoladamente é capaz de realizar predições melhores que a estratégia tradicional. No entanto, observando os resultados relacionados ao uso de oráculo em cada uma delas, temos que para todos há um grande potencial para melhorias, ou seja, melhorar a capacidade preditiva dos modelos de cada uma das variáveis tende a melhorar significativamente a qualidade da predição de custos final. O destaque é o grande potencial que melhorar a predição para a variável “Tempo de Permanência” tem na predição dos custos finais.

3. DESAFIOS E APRENDIZADOS

Os resultados demonstram que o uso de pipelines com predições intermediárias das variáveis do modelo pode aprimorar a estimativa de custos hospitalares. Análises com oráculos, simulando predições perfeitas para essas variáveis intermediárias, evidenciaram a relevância dessas variáveis e seu impacto no desempenho dos modelos. O desenvolvimento de modelos robustos e precisos para predição de custos contribui diretamente para a otimização de recursos, o planejamento financeiro e a melhoria na prestação de serviços de saúde.

Agradecimentos. Esse trabalho conta com o apoio do CNPq, FAPEMIG, FAPESP, Unimed-BH, CIIA-Saúde e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação - INCT-TILD-IAR.

4. REFERÊNCIAS

1. Veselovská L. Supply chain disruptions in the context of early stages of the global COVID-19 outbreak. *Problems and Perspectives in Management*. 2020;18(2):490-500.
2. Xu Z, Elomri A, Kerbache L, El Omri A. Impacts of COVID-19 on global supply chains: Facts and perspectives. *IEEE engineering management review*. 2020;48(3):153-66.
3. Zhou J, Wu R, Williams C, Emberson J, Reith C, Keech A, et al. Prediction Models for Individual-Level Healthcare Costs Associated with Cardiovascular Events in the UK. *PharmacoEconomics*. 2023;41(5):547-59. doi: 10.1007/s40273-022-01219-6.
4. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*. 2021;54(3):1937-67. doi: 10.1007/s10462-020-09896-5.

Desenho e avaliação de taxonomia de suporte para um Chatbot Inteligente de assistência ao usuário do sistema e-SUS APS

Raiane A. Asevedo (aluna)¹, Isaias J. R. de Oliveira (técnico)¹, Aline Amorim (técnica)¹, Francy Webster Pereira (técnico)², Gestefane Rabbi (aluno)³, Marco A. Costa (técnico)¹, Claudio M. V. de Andrade (aluno)³, Julia P. de Viterbo (aluna)³, Adriana S. Pagano (orientadora)⁴, Marcos André Gonçalves (orientador)³, Zilma S. N. Reis (orientadora)¹

¹ Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil. ² Ministério da Saúde, SAPS, Brasília, DF, Brasil. ³ Departamento de Ciência da Computação da UFMG, Belo Horizonte, MG, Brasil.

⁴ Faculdade de Letras da UFMG, Belo Horizonte, MG, Brasil.

raiaasevedo@ufmg.br, isaias@medicina.ufmg.br, aline@medicina.ufmg.br, webster.pereira@saude.gov.br, gestefane@dcc.ufmg.br, massiscosta@gmail.com, claudio.valiense@dcc.ufmg.br, juliapaes@dcc.ufmg.br, apagano@ufmg.br, mgoncalv@dcc.ufmg.br, zilma@ufmg.br.

Resumo: Este estudo apresenta o desenho e avaliação de uma taxonomia de suporte para um Chatbot inteligente voltado aos usuários do sistema de prontuário eletrônico e aplicativos e-SUS APS. Demandas de usuários foram classificadas visando uma taxonomia de base para rotulagem automática com aprendizado de máquina supervisionado, possibilitando um mapeamento preciso de novas demandas e sua vinculação com fontes de proveniência validada — manuais técnicos, apostilas do curso Educa e-SUS e documentação dos desenvolvedores. O sistema de recuperação de informação utilizando a técnica de RAG permite responder automaticamente a consultas rotineiras, enquanto direciona casos mais complexos para analistas humanos. A taxonomia constitui uma contribuição relevante para a Saúde Digital, oferecendo uma estrutura escalável e alinhada às constantes mudanças do e-SUS APS. A abordagem potencializa a busca por fontes corretas de informação, aprimorando a qualidade do suporte técnico aos usuários do e-SUS APS.

Abstract: This study introduces the design and evaluation of taxonomy to support an intelligent chatbot tailored to users of the electronic health record system and e-SUS Primary Health Care (e-SUS APS) applications. User queries were classified to build a foundational taxonomy for the automatic categorization of recurring queries using supervised machine learning. This classification allows for accurate mapping of new queries and links them to validated sources such as technical manuals, Educa-eSUS training materials, and developer documentation. By leveraging Retrieval-Augmented Generation (RAG), the system delivers automated answers to routine queries and redirects more complex issues to human analysts. The taxonomy offers a valuable contribution to Digital Health, providing a scalable solution that adapts to the ongoing evolution of e-SUS APS. The approach enhances access to accurate information and strengthens the quality of user support.

Palavras-chave: decision support techniques; generative artificial intelligence; health information systems; primary health care.

Nome do projeto: Desenvolvimento de um Chatbot inteligente para otimizar o processo de suporte ao usuário do sistema nacional e-SUS APS - Ministério da Saúde

Links: educaesusaps.medicina.ufmg.br; <http://150.164.2.44/chatbot-esus>

1. CONTEXTO E MOTIVAÇÃO

O sistema e-SUS APS é composto pelo Prontuário Eletrônico do Cidadão e aplicativos móveis, adotados nacionalmente na Atenção Primária à Saúde. Os municípios brasileiros, contando com sua capacidade operacional, operam os *softwares* oferecidos sem custos pelo Ministério da Saúde em cenários variados, com diferentes níveis de infraestrutura e maturidade digital. Nesse contexto, o suporte técnico aos usuários é essencial. O modelo de suporte vigente é operado por humanos através de central nacional online com respostas manuais a demandas de usuários (tickets). Diante do aumento da demanda e das constantes atualizações do sistema, torna-se necessário adotar soluções automatizadas e escaláveis para responder automaticamente a consultas rotineiras e direcionar casos mais complexos para analistas humanos. Chatbots inteligentes, com o uso de modelos de linguagem de larga escala (LLMs), apoiados por *Retrieval Augmented Generation* (RAG) ^[1], representam uma alternativa promissora para agilizar o suporte ^[2]. Para implementar essa abordagem foi criada uma proposta de taxonomia que organize e classifique as demandas dos usuários ^[3], permitindo a recuperação eficiente de informações e a geração de respostas mais precisas e contextualizadas. Essa taxonomia é base de um sistema de classificação automática de textos ^[4] por aprendizado de máquina que direciona as demandas para as fontes mais adequadas para aplicação de RAG.

1. 1 Objetivo

Descrever o desenho e avaliação de uma taxonomia de suporte para um Chatbot inteligente, com classificação automatizada de demandas e recuperação de informações de fontes validadas, visando otimizar o atendimento a usuários do e-SUS APS.

2. ATIVIDADES PRINCIPAIS

O desenho da taxonomia envolveu a avaliação humana de uma base de demandas de usuários do suporte técnico do sistema e-SUS APS para a elaboração de categorias, considerando a complexidade, a mutabilidade do sistema e as necessidades dos usuários. Foram estabelecidos vínculos entre as categorias e fontes validadas de informação (manuais, cursos, conteúdos de suporte e FAQs do desenvolvedor). A nomenclatura e hierarquia da taxonomia foi alinhada à base de conhecimento e ao perfil do público do chatbot. A categorização manual subsidiou o treinamento de modelos de linguagem (LLM) para classificação automática em tempo real.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Foi definido um conjunto de rótulos alinhados às funcionalidades do sistema e-SUS APS, às suas atualizações frequentes, aos perfis profissionais dos usuários que acionaram o suporte e às fontes de informação disponíveis para a técnica RAG. Uma amostra de 811 demandas do suporte do e-SUS APS, dados sem informações sensíveis disponibilizados em acordo de cooperação técnica TED 31/2022 entre a UFMG e a Secretaria de Atenção Primária à Saúde (SAPS), foi analisada e rotulada manualmente. Ajustes dos rótulos, como a criação de subcategorias, foram realizados para aprimorar a clareza e usabilidade, facilitando sua aplicação no Chatbot. A relação final dos rótulos, suas subcategorias e a distribuição das demandas está apresentada no Quadro 1. Todos os rótulos e subcategorias foram relacionadas às fontes utilizadas: manuais técnicos (disponíveis no GitHub do projeto), e-Books de cursos da plataforma Educa e-SUS APS (organizados por trilhas temáticas), documentos do Laboratório Bridge, documentos do Ministério da Saúde e Governo Federal (como notas técnicas e resoluções). Essa vinculação sistemática permitiu o treinamento da LLM para classificação e busca automática em tempo real das fontes mais relevantes, otimizando a recuperação de respostas, gerando um texto sumarizado e direcionando os usuários às fontes mais relevantes conforme a demanda (ex.: questões sobre instalação do sistema são associadas ao Manual do Prontuário Eletrônico e cursos específicos, enquanto dúvidas sobre aplicativos acionam respostas baseadas no manual próprio de cada um). Casos complexos, sem correspondência ou com respostas inadequadas, são redirecionados para analistas humanos.

Quadro 1 - Distribuição de uma amostra de 811 tickets na taxonomia proposta para direcionamento das demandas dos usuários do sistema e-SUS APS.

| RÓTULO | SUBCATEGORIA | N (%) | TOTAL N (%) |
|---|------------------------------|-------------|--------------------|
| Instalação e Implantação do e-SUS APS | Instalação e Implantação | 54 (6,7%) | 91 (11,3%) |
| | Configuração e Administração | 37 (4,6%) | |
| Manutenção do Software e Versões | Manutenção e Versões | 22 (2,7%) | 155 (19,1%) |
| | Erros do Sistema | 117 (14,4%) | |
| | Atualizações | 16 (2,0%) | |
| Cadastro de Pessoas (Cidadão) | - | - | 43 (5,3%) |
| Processo de Atendimento | - | - | 107 (13,2%) |
| Vacina | - | - | 33 (4,1%) |
| Demandas de ACS e ACE | - | - | 66 (8,1%) |
| Aplicativos e-SUS | e-SUS Território | 26 (3,2%) | 36 (4,4%) |
| | e-SUS Atividade Coletiva | 0 (0%) | |
| | Gestão e-SUS APS | 7 (0,9%) | |
| | e-SUS Vacinação | 2 (0,2%) | |
| | e-SUS Atenção Domiciliar | 1 (0,1%) | |
| Cadastro de Profissional | - | - | 31 (3,8%) |
| Preenchimento da Coleta de Dados Simplificada | - | - | 33 (4,1%) |
| Gestão de Recursos, Serviços ou Relatórios | - | - | 57 (7,0%) |
| Demais Sistemas Relacionados à APS | e-Gestor e SISAB | 71 (8,8%) | 128 (15,8%) |
| | Outros Sistemas | 57 (7,0%) | |
| Não Classificável | - | - | 31 (3,8%) |

4. DESAFIOS E APRENDIZADOS

Um desafio central foi lidar com a diversidade das demandas e falta de clareza dos textos enviados pelos usuários ao sistema de suporte. O processo demandou refinamento contínuo da taxonomia para assegurar clareza e aplicabilidade, equilibrando precisão técnica e usabilidade para diferentes perfis de usuários. A experiência adquirida nesta pesquisa poderá subsidiar a construção de outras taxonomias similares voltadas a sistemas conversacionais na área da Saúde baseados em Inteligência Artificial.

Financiamento e Agradecimentos. Ministério da Saúde, Secretaria de Atenção Primária à Saúde (TED 31 /2022, Processo 25000.058753/2022-59). Esse trabalho conta com o apoio do CNPq, FAPEMIG e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação - INCT-TILD-IAR.

5. REFERÊNCIAS

1. Celso França, Gestefane Rabbi, Thiago Salles, Washington Cunha, Leonardo Rocha, and Marcos André Gonçalves. Optimizing Tail-Head Trade-off for Extreme Multi-Label Text Classification (XMTTC) with RAG-Labels and a Dynamic Two-Stage Retrieval and Fusion Pipeline. Aceito para publicação na ACM SIGIR (SIGIR '25). (2025).
2. Authors, Clark M, Bailey S. Chatbots in Health Care: Connecting Patients to Information: Emerging Health Technologies [Internet]. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2024 [citado: 20 de maio de 2025]. (CADTH Horizon Scans). Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK602381/>
3. Shachak A, Montgomery C, Dow R, Barnsley J, Tu K, Jadad AR, et al. End-user support for primary care electronic medical records: a qualitative case study of users' needs, expectations and realities. Health Syst (Basingstoke). 1o de novembro de 2013;2(3):198–212.
4. de Andrade, C. M., Belem, F. M., Cunha, W., França, C., Viegas, F., Rocha, L., & Gonçalves, M. A. (2023). On the class separability of contextual embeddings representations—or “the classifier does not matter when the (text) representation is so good!”. Inf. Proc. & Management, 60(4), 103336.

Desenvolvimento de um chatbot inteligente para otimizar o processo de suporte ao usuário nos sistemas e-SUS APS

Gestefane Rabbi (aluno)¹, Claudio Andrade (aluno)¹, Júlia Paes (aluna)¹, Isaias R. Oliveira (técnico)², Marco A. Costa (técnico)², Aline Amorim (técnico)², Raiane Asevedo (aluna)², Adriana Pagano (orientadora)⁴, Vania Priamo (técnica)³, Zilma S. N. Reis (orientadora)¹, Marcos André Gonçalves (orientador)¹

¹ Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil. ² Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil. ³ Ministério da Saúde, SAPS, Brasília, DF, Brasil. ⁴ Faculdade de Letras da UFMG, Belo Horizonte, MG, Brasil.

gestefane@dcc.ufmg.br, claudiovaliense@gmail.com, juliapaes@dcc.ufmg.br, isaias.ramos@gmail.com, massiscosta@gmail.com, aline@medicina.ufmg.br, raiasevedo@ufmg.br, apagano@ufmg.br, vania.priamo@saude.gov.br, zilma@ufmg.br, mgoncalv@dcc.ufmg.br

Resumo: Este projeto desenvolve um chatbot inteligente para aprimorar o suporte ao sistema e-SUS APS, atualmente realizado por atendentes humanos do Ministério da Saúde. Usa a arquitetura RAG (Retrieval-Augmented Generation), que combina recuperação semântica e modelos de linguagem de grande escala (LLM) para gerar respostas precisas. A base inclui manuais, materiais de treinamento e demandas categorizadas. Desafios enfrentados incluem desequilíbrio nas categorias e escassez de dados rotulados. Os resultados mostram potencial para reduzir o tempo de resposta e melhorar o suporte, embora perguntas complexas e o contexto entre interações ainda apresentem dificuldades.

Abstract: This project develops an intelligent chatbot to improve support for the e-SUS APS system, currently handled by human agents from Ministério da Saúde. It uses the RAG architecture (Retrieval-Augmented Generation), combining semantic retrieval with large-scale language models (LLMs) to generate accurate responses. The knowledge base includes manuals, training materials, and categorized user requests. Challenges such as category imbalance and limited labeled data remain, but initial results show potential to reduce response time and enhance support, despite ongoing difficulties with complex questions and maintaining context.

Palavras-chave: Chatbot; Retrieval-Augmented Generation; Large Language Model.

Nome do projeto: Desenvolvimento de um chatbot inteligente para otimizar o processo de suporte ao usuário do sistema nacional e-SUS APS - Ministério da Saúde

1. CONTEXTO E MOTIVAÇÃO

Os sistemas que compõem o ecossistema e-SUS são amplamente utilizados no Brasil, o que gera uma elevada demanda por suporte técnico. Responder rapidamente a esse grande volume de chamados é desafiador, envolvendo fatores críticos como a preparação técnica contínua dos atendentes, atualização dos manuais e fontes de informação, além da disponibilidade suficiente de pessoal. Para garantir atendimento rápido e eficaz em larga escala, desenvolvemos uma solução com um chatbot inteligente, capaz de responder às dúvidas mais frequentes utilizando como base de conhecimento demandas de atendimento categorizadas, materiais didáticos e manuais dos sistemas, desenvolvidos pela equipe de suporte ao usuário do e-SUS APS.

1.1 Objetivo

Este estudo propõe um chatbot inteligente que acessa uma base de conhecimento com manuais, registros de tickets e material de treinamento dos sistemas e-SUS, interagindo em linguagem natural. O objetivo é que o *chatbot* compreenda e gere respostas de alta qualidade, semelhante às de atendentes humanos qualificados. Para isso, utiliza a arquitetura RAG (*Retrieval-Augmented Generation*), que combina recuperação semântica e modelos de linguagem de grande escala (LLM). Seus componentes incluem um sistema de recuperação e um modelo de linguagem, integrados para fornecer respostas precisas e fundamentadas em documentos selecionados.

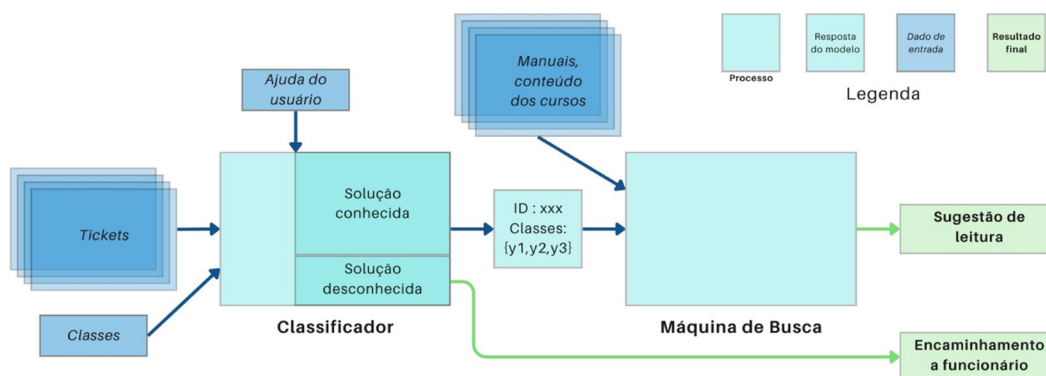
2. ATIVIDADES PRINCIPAIS

Desenvolver um *chatbot* inteligente para otimizar o processo de suporte ao usuário do sistema nacional e-SUS APS - Ministério da Saúde.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A Figura 1 ilustra o fluxograma do chatbot E-SUS APS. O modelo de resposta ao usuário fundamenta-se na arquitetura RAG ⁽²⁾, composta por duas etapas: (i) a preparação da base de conhecimento para recuperação semântica, e (ii) o fluxo de interação, que integra recuperação eficiente e geração de linguagem natural para respostas automatizadas de qualidade.

Figura 1 - Fluxograma chatbot E-SUS APS.



Fase 1 - Preparação da Base de Conhecimento: A preparação da base de conhecimento converte documentos do e-SUS em PDFs e conteúdos HTML, em representações vetoriais para recuperação semântica. O processo envolve: (i) extração do conteúdo, convertendo diferentes formatos em texto estruturado; (ii) segmentação em unidades menores (*chunks*); (iii) vetorização, que gera *embeddings* que capturam semântica e contexto; e (iv) armazenamento, onde os vetores são indexados para consultas rápidas por similaridade. Essas etapas garantem buscas eficientes e relevantes para a geração de respostas.

Fase 2 - Fluxo de Interação com o Usuário: Após a preparação da base de conhecimento, o sistema passa a interagir com outros componentes, processando perguntas dos usuários e fornecendo respostas.

Classificação: Com um conjunto de 1131 registros de demanda de atendimento categorizados em classes de tópico por humanos, foi aplicada validação cruzada com *5-folds*, dividindo os dados em treinamento e teste. Um modelo de regressão logística classifica a dúvida do usuário em 11 categorias, auxiliando o sistema RAG na formulação da resposta. A eficácia do modelo chega a 66.9% de acurácia, enquanto para a métrica macro-f1, que considera igual importância para todas as categorias, é alcançado 64.1%. Outra

forma de avaliar o modelo é considerando as 3 primeiras (*top-3*) categorias com maior probabilidade do modelo (escolhemos 3 para evitar sobrecarga de avaliação do usuário); dessa forma em 91% das vezes a categoria correta é apresentada dentre as *top-3* categorias. Desta forma, criamos um sistema que a partir da dúvida do usuário, o chatbot retorna as *top-3* categorias e o usuário confirma umas das 3 categorias da sua dúvida, que em seguida é apresentado a resposta a partir do modelo baseado em RAG, caso nenhuma das 3 categorias esteja relacionado a sua dúvida, o usuário é direcionado para um atendente humano.

Geração da resposta: O sistema recebe a pergunta, cuja classificação delimita o espaço de busca. A questão é convertida em vetor e comparada ao índice pelo algoritmo KNN, que identifica os *k* fragmentos mais semanticamente similares. Os *top-k*, selecionados por similaridade, formam o contexto enviado ao modelo de linguagem (LLM). Este gera a resposta, focada no material, prevenindo alucinações e assegurando aderência ao conteúdo oficial do e-SUS. A resposta é devolvida ao sistema, garantindo rastreabilidade e transparência.

Tabela 1 - Desempenho médio de acurácia e macro-f1 do modelo de predição de classificação das categorias do ticket, seguido do intervalo de confiança com 95%

| Método | Acurácia | Macro-F1 |
|------------------|--------------|--------------|
| Top-1 Categoria | 66.9 +/- 2.8 | 64.1 +/- 5.0 |
| Top-3 Categorias | 91.4 +/- 2.2 | 87.9 +/- 2.7 |

Como parte da validação, testes com perguntas reais de operadores foram realizados coletando erros e comentários. A avaliação ocorreu em duas etapas: inicialmente, com apenas material didático eletrônico, e depois, com ampliação da base. Na versão inicial, enfrentamos limitações como falta de contexto, respostas incompletas e dificuldades com perguntas complexas, além de recomendações desalinhadas. Na versão ajustada, que incluiu materiais didáticos, manuais e registros de demanda, houve melhora significativa na objetividade, alinhamento com os procedimentos oficiais e eliminação de ruídos, além de maior cobertura de tópicos essenciais. Contudo, desafios como manter o contexto e interpretar consultas técnicas complexas ainda persistem, embora os avanços tenham aumentado a aderência e a clareza das respostas.

4. DESAFIOS E APRENDIZADOS

Enfrentamos o desafio de suportar usuários com um chatbot inteligente para dúvidas sobre o sistema de saúde E-SUS APS. Utilizamos registros de demanda (textos, descrições e categorias) e materiais didáticos. Um obstáculo é a distribuição desigual entre categorias, com algumas predominantes e outras mais críticas, raras e com poucos exemplos rotulados. O volume limitado de dados rotulados, comum em sistemas de saúde pública em países em desenvolvimento, dificulta a modelagem, pois a rotulagem depende da disponibilidade de especialistas.

Financiamento e Agradecimentos. Ministério da Saúde, Secretaria de Atenção Primária à Saúde (TED 31/2022, Processo 25000.058753/2022-59). Esse trabalho conta com o apoio do CNPq, FAPEMIG e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação - INCT-TILD-IAR.

5. REFERÊNCIAS

1. Hu, J., Ruder, S., Siddhant, A., Neubig, G., Firat, O., & Johnson, M. XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation., Proceedings of the 37th International Conference on Machine Learning, pages 4411–4421. PMLR. (2020).
2. Celso França, Gestefane Rabbi, Marcos André Gonçalves et al. Optimizing Tail-Head Trade-off for Extreme Multi-Label Text Classification (XMTC) with RAG-Labels and a Dynamic Two-Stage Retrieval and Fusion Pipeline. Aceito para publicação na ACM SIGIR (SIGIR '25). (2025).

Divulgação científica sobre segurança medicamentosa: análise de formatos, qualidade e engajamento no instagram no projeto IA-Polis

Marina Cavalcanti (aluno)¹, Plácido Lima (aluno)¹, Sara Cristina Leone Lemos (aluno)¹, Zilma Nogueira Reis (orientador)¹

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

placidolima@ufmg.br, marinagcr@ufmg.br, saracll2021@ufmg.br, zilma@ufmg.br

Resumo: Erros no uso de medicamentos ainda representam uma das principais causas evitáveis de danos à saúde no mundo. A compreensão limitada das orientações contidas nas prescrições contribui significativamente para esse problema, especialmente entre populações com baixo letramento em saúde. O projeto IA-Polis visa promover a divulgação científica sobre o uso correto de medicamentos a partir de uma solução baseada em inteligência artificial (IA) desenvolvida por instituições parceiras no Brasil, Portugal e Moçambique para gerar instruções personalizadas em receitas digitais. Por meio da plataforma Instagram, buscamos democratizar o conhecimento sobre o uso seguro de medicamentos, traduzindo a linguagem técnica das prescrições para um formato acessível, visual e culturalmente sensível. Este trabalho apresenta as estratégias utilizadas na criação de conteúdos digitais e os primeiros resultados sobre o engajamento nas redes sociais, que servirá de canal para ampla validação dos produtos do projeto.

Palavras-chave: medicamentos; inteligência artificial; mídias sociais; letramento em saúde; redes sociais; adesão à terapêutica.

Abstract: Mistakes in the use of medications still represent one of the main avoidable causes of health harm worldwide. The limited understanding of the instructions on the prescriptions contributes significantly to this problem, especially among populations with low health literacy. The IA-Polis project aims to promote science communication related to the appropriate use of medications through an AI-based solution developed by partner institutions from Brazil, Portugal and Mozambique, which will generate personalized instructions in digital prescriptions. Through the Instagram platform, we seek to democratize the knowledge about the safe use of medications, translating the technical language of prescriptions into an accessible, visual, and culturally sensitive format. This work presents the strategies used in the creation of digital content and the initial results regarding social media engagement, which will serve as a channel for the broad validation of the project's outcomes.

Keywords: medications; artificial intelligence; social media; health literacy.

Nome do projeto: IA-Polis Links: <https://www.instagram.com /apolisufmg/>

1. CONTEXTO E MOTIVAÇÃO

A não adesão ao tratamento e o uso incorreto de medicamentos são problemas de saúde pública no Brasil e no mundo, representando um dos maiores custos evitáveis para os sistemas de saúde ⁽¹⁾. Com a ascensão do uso da inteligência artificial (IA) na saúde, surge a possibilidade de utilizar essa tecnologia, treinada em fontes baseadas em ciência, para prover explicações claras, corretas e personalizadas sobre como utilizar um medicamento prescrito por um profissional de saúde. Além disso, ferramentas tecnológicas demandam validação antes de serem utilizadas e precisam ser amplamente reconhecidas e avaliadas pelo seu público-alvo. Nesse sentido, estratégias de divulgação e de interação direta com o público em mídias sociais e com linguagem acessível, pode contribuir para difundir as boas práticas para uso correto de medicamentos. Ao mesmo tempo, o ambiente das mídias propicia interação com a sociedade. O perfil no Instagram, denominado “@iapolisufmg”, realiza o trabalho de democratizar o acesso às boas práticas para o uso correto de medicamentos, abordando o uso da IA para apoiá-las. O conhecimento a respeito

do engajamento do público à mídia social permitirá validar os produtos de pesquisa diretamente com a sociedade.

O uso incorreto de medicamentos, muitas vezes resultante da má interpretação das prescrições, está entre os principais fatores que comprometem a eficácia terapêutica e a segurança do paciente⁽²⁾. No contexto do prontuário eletrônico do SUS para Atenção Primária à Saúde, o e-SUS APS, a funcionalidade que permite a prescrição de medicamentos é realizada com padrões internacionais para a especificação dos elementos de uma receita: identificação do usuário, escolha do princípio ativo, concentração, via de administração e posologia são padronizados. Mesmo assim, a linguagem técnica da receita gerada no sistema pode ser um empecilho para compreensão dos usuários. O perfil do projeto IA-Polis no Instagram surge como uma iniciativa de comunicação científica, com foco na divulgação acessível das recomendações de uso de medicamentos, a partir de tecnologias de IA treinadas com dados clínicos validados. O conteúdo das postagens é divulgado em mídias sociais com linguagem simplificada e apoio visual.

Objetivo: o presente estudo descreve o processo de produção das postagens, suas formas de apresentação e alcance desde seu lançamento.

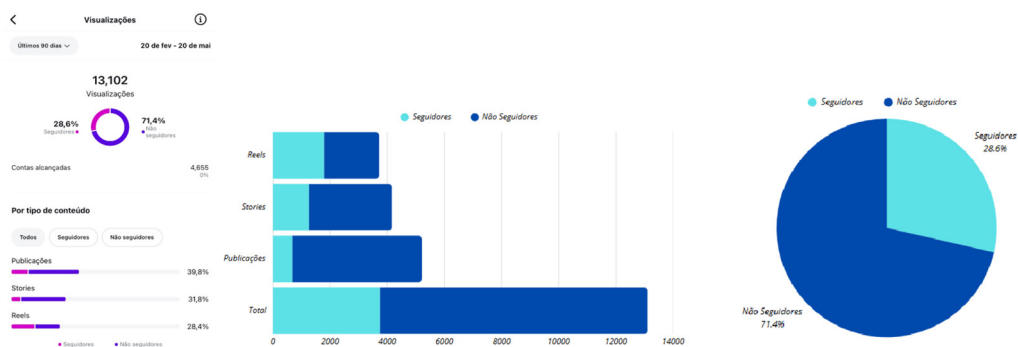
2. ATIVIDADES PRINCIPAIS

A equipe de iniciação científica é formada por acadêmicos do Curso de Medicina da UFMG, sob orientação dos pesquisadores do projeto de pesquisa: Development and Evaluation of an Intelligent System for Generating Guidelines for Prescribing Medication that is Safe, Accessible and Adapted to Different Cultural Contexts, acessível em <https://gcgh.grandchallenges.org/grant/development-and-evaluation-intelligent-system-generating-guidelines-prescribing-medication>. Três formatos foram utilizados para as postagens no Instagram: reels (vídeos curtos e dinâmicos), stories (conteúdos interativos e temporários) e carrosséis (sequências de imagens explicativas). As fontes de informação científica para produção do conteúdo das postagens são diversificadas em diretrizes de entidades reconhecidas, como a Organização Mundial de Saúde e o Ministério da Saúde, em normas e recomendações emitidas pela ANVISA e na literatura científica atual sobre segurança do uso de medicamentos. Cada publicação aborda aspectos diferentes do uso correto de medicamentos — como horários, conservação, vias de administração, interações medicamentosas e adesão terapêutica⁽²⁾. Além disso, são explorados conteúdos que incluem atualizações e notícias recentes sobre fármacos, para informar a população sobre temas de relevância no momento. Também são promovidas discussões acerca da importância do uso da inteligência artificial no âmbito da saúde, especialmente no que se refere ao apoio à prescrição segura e eficaz de medicamentos. As postagens são, então, elaboradas pelo corpo discente do projeto e validadas por docentes pesquisadores da equipe. Além da produção de conteúdo, foi realizada a análise do desempenho das publicações, com métricas de alcance, curtidas, compartilhamentos e salvamentos.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Utilizou-se uma abordagem qualitativa e quantitativa para avaliação das publicações. A linguagem das fontes de informação científica foi adaptada a partir de técnicas de simplificação textual e princípios de comunicação em saúde⁽³⁾. Os materiais foram editados com *softwares* livres como Canva® e CapCut®. Para a avaliação do desempenho das publicações, utilizou-se o painel de *insights* disponibilizado pelo Instagram, que fornece métricas sobre o alcance da conta. Nos últimos 90 dias ele alcançou 4.655 contas, totalizando 13.102 visualizações em todos os formatos de publicação. Dentre essas visualizações, 39,8% ocorreram em publicações de **Feed**, 31,8% em **Stories** e 28,4% em **Reels** (Figura 1). Ainda, entre os resultados preliminares, destaca-se o crescimento de 280% no número de seguidores em três meses, com média de alcance de 1.200 contas por postagem. Os reels apresentaram maior engajamento, especialmente aqueles que envolvem simulações de situações cotidianas ou que desmistificam erros comuns na administração de medicamentos.

Figura 1 - Painel de insights do perfil @iapolisufmg no Instagram, referente ao período de 20 de fevereiro a 20 de maio de 2025.



Fonte: Captura de tela dos insights do perfil @iapolisufmg no Instagram.

4. DESAFIOS E APRENDIZADOS

Um dos principais desafios foi transformar conteúdos técnicos e densos em publicações acessíveis sem perder o rigor científico. Também se evidenciaram barreiras relacionadas à construção de uma identidade visual para o perfil @iapolisufmg que refletisse seriedade, mas que, ao mesmo tempo, fosse atrativa e compreensível para públicos diversos. Ademais, identificou-se que, nos últimos 90 dias, 71,4% das visualizações foram realizadas por não-seguidores, o que aponta para dificuldade de converter interações eventuais em seguidores contínuos. Esse aspecto representa um desafio importante, uma vez que a formação de tal base sólida e engajada de seguidores é crucial para ampliar o alcance e potencializar os impactos do projeto. Por fim, o aprendizado mais significativo foi reconhecer a potência das redes sociais como ferramentas de educação em saúde, e a importância de utilizar a IA não apenas como instrumento técnico, mas como facilitadora de inclusão, compreensão e cuidado.

Financiamento. Ministério da Saúde DECIT/CNPq (00758/2024-5) e em parceria com a Gates Foundation [Grant Number INV-009289]. O estudo teve apoio do Ministério da Saúde/SAPS na forma de bolsas de iniciação científica (TED 31/2022).

5. REFERÊNCIAS

1. World Health Organization. Medication Without Harm – WHO’s Third Global Patient Safety Challenge. Geneva: WHO; 2017.
2. Ministério da Saúde (Brasil). Protocolo de Segurança na Prescrição, Uso e Administração de Medicamentos. Brasília: MS; 2013.
3. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int. 2000;15(3):259–67.

Framework bioético no design e implementação de ferramenta de IA para suporte à prescrição no SUS: experiência do projeto IA-Polis

Gabriella Reis¹, Cristiane Dias², Eura Lage³, Isaias Oliveira⁴, Gláucia Varella⁵, Ricardo Correa⁶, Adalberto Santos⁷, Igor Oliveira⁸, Erico Mineiro⁹, Marco Costa¹⁰, Davi Jesus¹¹, Antônio Souza Júnior¹², Igor Gomes¹³, Juliana Oliveira¹⁴, Flávia Oliveira¹⁵, Adriana Pagano¹⁶, Elisa Tuler¹⁷, Leonardo Rocha¹⁸, Daniel Costa¹⁹, Zilma Reis²⁰

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

autor correspondente: gabriellareis@gmail.com

Resumo: Este estudo apresenta os cuidados com os princípios da IA responsável, adotados no desenvolvimento de um sistema inteligente para gerar instruções personalizadas para uso de medicamentos. O projeto interdisciplinar utilizou metodologia de trabalho colaborativo, com reuniões regulares para desenvolvimento progressivo da solução. A análise comparativa do desempenho de *Large Language Models* (LLMs) levou em conta requisitos bioéticos como: equidade e justiça, transparência e explicabilidade, autonomia e responsabilidade, beneficência e não maleficência, além de privacidade de dados. A experiência proporcionou um aprendizado sobre integração responsável de IA na saúde pública, destacando a importância de considerações éticas no desenvolvimento tecnológico para populações vulneráveis.

Abstract: This study presents the responsible AI principles adopted in the development of an intelligent system designed to generate personalized medication usage instructions. The interdisciplinary project employed a collaborative work methodology, with regular meetings to ensure the progressive development of the solution. The comparative analysis of Large Language Model (LLM) performance considered bioethical requirements, including fairness and justice, transparency and explainability, autonomy and accountability, beneficence and non-maleficence, as well as data privacy. The experience provided valuable insights into the responsible integration of AI in public health, highlighting the importance of ethical considerations in technological development for vulnerable populations.

Palavras-chave: bioética; Inteligência Artificial; IA responsável.

Nome do projeto: Desenvolvimento e Avaliação de um Sistema Inteligente para Gerar Diretrizes para a Prescrição de Medicamentos que seja Seguro, Acessível e Adaptado a Diferentes Contextos Culturais.

1. CONTEXTO E MOTIVAÇÃO

A Inteligência Artificial (IA) tem transformado a forma como lidamos com a saúde por meio de sistemas capazes de simular comportamentos inteligentes, tais como: aprendizado, raciocínio, adaptação, e interação ^(1,2,3). No entanto, seu uso na saúde exige considerações bioéticas para assegurar acesso equitativo e respeitar a autonomia dos pacientes. A sofisticação da IA, com processos cada vez mais autônomos, traz novos desafios éticos ⁽¹⁾, como privacidade, segurança, estigmatização, vieses, responsabilidade na implementação e necessidade de formação ética e educação continuada ⁽⁴⁾. Embora as diretrizes técnicas estejam em expansão, há poucos relatos sobre a incorporação prática da bioética no desenvolvimento de IA. O projeto IA-Polis busca preencher essa lacuna, desenvolvendo um sistema que gera orientações medicamentosas seguras, acessíveis e culturalmente adaptadas para a atenção primária do SUS (e-SUS APS), de forma ética e responsável.

1.1 Objetivo

Compartilhar um *framework* bioético para o desenvolvimento de IA responsável, aplicada à saúde pública e baseado na experiência de desenvolvimento do projeto IA-Polis. A proposta integra princípios bioéticos (i.e. autonomia, beneficência, não maleficência, justiça e equidade) à arquitetura de um LLM, para tornar as prescrições mais compreensíveis, especialmente para populações vulneráveis, sempre com validação realizada por profissionais de saúde.

2. ATIVIDADES PRINCIPAIS

O projeto IA-Polis desenvolveu um sistema de IA para personalizar orientações medicamentosas na atenção primária do SUS (e-SUS APS). As atividades foram realizadas por uma equipe multidisciplinar (computação, medicina, fisioterapia, linguística, farmácia e design), com reuniões semanais durante um ano. Foi feita uma análise comparativa de LLMs, considerando segurança de dados, custo, adaptabilidade ao português e capacidade computacional. As atividades se organizaram em dois eixos: técnico-tecnológico e ético-normativo. Desenvolvimento de prompts e simulações clínicas: Foram elaboradas versões progressivamente refinadas de prompts, testadas em ambiente de simulação com 180 casos fictícios baseados em dados do e-SUS APS. Após a análise de mais de 30 LLMs, o modelo Llama 3.1 foi selecionado por apresentar melhor equilíbrio entre custo, eficiência, segurança, ocorrência de erros e vieses. Sua característica de código aberto permite maior controle e personalização do sistema. Foram incorporadas bulas da ANVISA contemplando a implementação de RAG (*Retrieval Augmented Generation*), a fim de aprimorar a capacidade da IA de fundamentar suas orientações em fontes oficiais e confiáveis, aumentando a precisão e a segurança das informações fornecidas. Validação técnica e ética de especialistas: A validação humana ocorreu continuamente em todas as fases do IA-Polis. As orientações de prescrição de medicamentos comumente usados no e-SUS APS foram geradas por IA e foram avaliadas por 5 médicas que integram a equipe IA-Polis. Posteriormente, foram validadas por médicos, cirurgiões-dentistas e enfermeiros voluntários, com base em questionários com critérios de adequação, confiabilidade e compreensibilidade. Essa etapa foi estruturada em ciclos iterativos de desenvolvimento, com feedback contínuo da equipe, respeitando os princípios da beneficência, da não maleficência, além da autonomia e responsabilidade profissional. A supervisão humana é considerada essencial neste sistema, planejado para que o prescritor monitore, suspenda ou modifique as entregas. Implementação de mecanismos de transparência e explicabilidade: Foram concebidas interfaces em camadas para promover a transparência, visando garantir que as informações geradas fossem compreensíveis e confiáveis para profissionais de saúde e pacientes, independente do seu letramento digital. Para isso, foram usadas orientações sobre as boas práticas de prescrição e foram testadas várias versões de orientações a fim de promover uma linguagem clara, personalizada, acessível e adaptada culturalmente. O uso de pictogramas foi incluído, com o objetivo de ampliar a acessibilidade das orientações geradas. O objetivo foi evitar tecnicismos excessivos ou simplificações redutoras, promovendo acessibilidade e a autonomia do paciente. Proteção de dados e conformidade ética: O projeto seguiu as regras de proteção de dados pessoais sensíveis, conforme previsto na LGPD e nas diretrizes da OMS (5). Não foram utilizados dados reais de pacientes; todos os cenários foram simulados. A abordagem respeitou a privacidade, incorporando salvaguardas éticas desde a concepção da ferramenta. Testes de confiabilidade: Realizou-se análise detalhada dos erros cometidos pelos modelos, categorizando-os em sete tipos, além de testes de variabilidade intra e interobservador para garantir consistência e reprodutibilidade.

Infraestrutura de produto e disseminação científica: O projeto resultou na construção de uma API para integração com o PEC e-SUS APS, em um ambiente de simulação virtual, em instrumentos de avaliação padronizados e em um banco de dados de casos clínicos, além de contribuições científicas (artigos, apresentações em congresso etc.).

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Estudo de caso qualitativo e descritivo, fundamentado na análise de documentos internos do projeto IA-Polis, incluindo protocolos de desenvolvimento, registros de reuniões da equipe e reflexões interdisciplinares realizadas ao longo do processo. A abordagem metodológica adotou um referencial dedutivo orientado por quatro princípios bioéticos: equidade/justiça; autonomia, beneficência e não maleficência, além de implicações de ética e segurança relacionadas à transparência, explicabilidade de sistema de IA e proteção de dados pessoais.

O projeto foi aprovado pelo Comitê de Ética em Pesquisa da UFMG (CAAE 78883924.7.0000.5149) e contou com a participação voluntária de profissionais de saúde na etapa de validação. Todos os avaliadores assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), em conformidade com as diretrizes éticas aplicáveis à pesquisa com seres humanos. O desenvolvimento do IA-Polis foi conduzido de forma interativa e colaborativa, com participação ativa de equipe multidisciplinar composta por pesquisadores, profissionais da saúde e especialistas em tecnologia. Desde março de 2024, foram realizadas 31 reuniões científicas semanais registradas, visando aprimoramentos contínuos e integração de perspectivas clínicas, técnicas e éticas. O projeto contou com financiamento do CNPq-Inova Fiocruz e da Fundação Gates, e teve como objetivo central contribuir para a incorporação ética e responsável da IA no sistema público de saúde. A ênfase recaiu na promoção da supervisão e do controle humano sobre as recomendações geradas, priorizando a segurança dos pacientes, a confiabilidade da informação e a preservação da dignidade nas interações mediadas por tecnologia.

4. DESAFIOS E APRENDIZADOS

Diante de uma equipe multidisciplinar, a mediação entre linguagem técnica e a clareza na comunicação constituiu o principal desafio e também um aprendizado central. Garantir que as instruções geradas por IA fossem compreensíveis para pacientes em diferentes níveis de letramento exigiu atenção aos riscos do simplismo excessivo, evitando práticas paternalistas e assegurando o respeito à autonomia e a promoção de equidade.

Agradecimentos. O estudo foi realizado com recursos do CNPq-Inova Fiocruz, Fundação Gates e FAPEMIG.

5. REFERÊNCIAS

1. Souza-Zinader JP. A Estratégia da Saúde Digital para o Brasil. *Journal of Health Informatics* [Internet]. 30 de novembro de 2020 [citado 16 de abril de 2025];12(4). Disponível em: <https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/792>
2. Adegbesan A, Akingbola A, Ojo O, Jessica OU, Alao UH, Shagaya U, et al. Ethical Challenges in the Integration of Artificial Intelligence in Palliative Care. *Journal of Medicine, Surgery, and Public Health*. 1º de dezembro de 2024;4:100158.
3. Ransbotham S, Khodabandeh S, Kiron D, Candelon F, Chu M, LaFountain B. Expanding AI's Impact With Organizational Learning. *MIT SMR* [Internet]. 2º de outubro de 2020 [citado 16 de junho de 2025]; Disponível em: <https://sloanreview.mit.edu/projects/expanding-ais-impact-with-organizational-learning/>
4. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. 1º de junho de 2020;58:82–115.
5. Regulatory considerations on artificial intelligence for health [Internet]. [citado 16 de junho de 2025]. Disponível em: <https://www.who.int/publications/i/item/9789240078871>.

Gamificação com Inteligência Artificial no ensino de auditoria em saúde: uma experiência com a plataforma Rosebud

Priscila G. Maia¹, Sônia Maria Nunes Viana², Fabiana Maria Kakehasi³

¹Mestranda em Saúde Pública - UFMG, MG, Brasil.

²Professora do Departamento de Gestão em Saúde - EEUFMG, MG, Brasil

³Professora associada do Departamento de Pediatria UFMG, MG, Brasil

priscilagmaiabr@gmail.com, soniamnv²@gmail.com, Fabiana.Kakehasi@ebserh.gov.br

Resumo: Este relato descreve a experiência de implementação de uma atividade gamificada com inteligência artificial (IA) na disciplina de Auditoria em Saúde do curso de Gestão de Serviços de Saúde da UFMG. Utilizando a plataforma Rosebud AI, desenvolvemos um jogo onde os estudantes interagiram com um personagem virtual (Elara) para conduzir uma auditoria simulada em ambiente hospitalar fictício. A atividade, alinhada às diretrizes da ANVISA^[1], permitiu aos alunos vivenciarem na prática a identificação de não conformidades, com foco em três situações pré-definidas: falhas na documentação de higienização, ausência de grades de proteção em leitos e controle inadequado de estoques^[2]. Os resultados apontaram elevado engajamento dos estudantes, aprimoramento de habilidades analíticas e de interpretação normativa, embora tenham sido relatadas limitações técnicas e necessidade de elementos visuais mais imersivos. Como perspectivas futuras, pretende-se submeter a proposta a um comitê de ética para validação em turmas subsequentes e investir em recursos gráficos adicionais, favorecendo a replicabilidade da iniciativa.

Palavras-chave: gamificação; Inteligência Artificial; educação em saúde; auditoria em saúde; simulação virtual.

Links: <https://play.rosebud.ai/games/51d371a1-4410-46a5-a7f6-3ba89212f091>

1. CONTEXTO E MOTIVAÇÃO

O ensino superior enfrenta desafios contínuos para integrar tecnologias emergentes que tornem a aprendizagem mais engajadora e eficaz^[5]. A gamificação, combinada com IA, surge como uma estratégia promissora para aumentar a motivação e personalizar a aprendizagem^[6]. No contexto da Auditoria em Saúde, disciplina que exige habilidades analíticas e aplicação prática de normas, a gamificação com IA oferece um ambiente seguro para simular situações reais, como identificação de não conformidades em hospitais^[1]. Este estudo foi motivado pela necessidade de inovar metodologias pedagógicas, alinhando-se às demandas do século XXI por competências como pensamento crítico e resolução de problemas^[7].

1.1 Objetivo

A disciplina de Auditoria em Saúde demanda abordagens que integrem conhecimento teórico-normativo com habilidades práticas de investigação^[4]. A gamificação com IA surgiu como estratégia para:

- Tornar o aprendizado mais engajante através de elementos lúdicos^[5];
- Simular situações reais sem riscos^[6];
- Oferecer feedback imediato e personalizado via IA^[7].

2. ATIVIDADES PRINCIPAIS

A atividade central consistiu no desenvolvimento de um jogo interativo na plataforma Rosebud AI, onde os estudantes assumiram o papel de auditores em um hospital fictício. O jogo foi estruturado em

torno de três não conformidades pré-definidas, baseadas em normas da ANVISA^[1,8]. Os alunos interagiram com a personagem virtual Elara, que mediou a simulação por meio de diálogos em linguagem natural. A atividade incluiu:

- Exploração de cenários hospitalares simulados.
- Identificação de falhas em processos e documentação.
- Preenchimento de um formulário reflexivo para avaliação da experiência^[2].

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

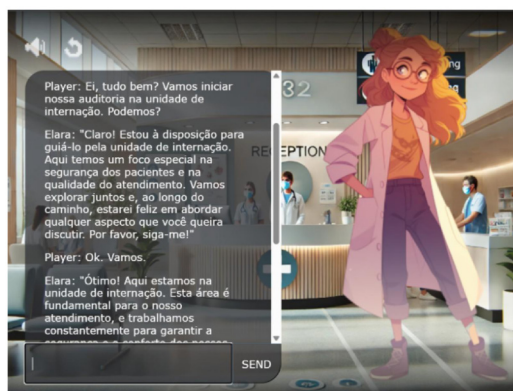
A metodologia envolveu a criação da personagem Elara, configurada para responder contextualmente às ações dos alunos, simulando desafios reais de auditoria^[9]. A plataforma permitiu interações flexíveis, adaptando-se a diferentes estratégias investigativas. Os resultados mostraram:

- Engajamento elevado: Os alunos relataram maior motivação devido à abordagem lúdica e à interação com a IA^[2].
- Aprimoramento de habilidades: Melhoria na capacidade de interpretação normativa e investigação crítica^[10].
- Limitações: Dificuldades técnicas para alguns alunos e a necessidade de elementos visuais mais imersivos^[3].

Figura 1 - Tela inicial do jogo de auditoria desenvolvido na plataforma Rosebud AI.



Figura 2 - Exemplo de interação com a personagem Elara



Fonte: Próprios autores - Rosebud AI

4. DESAFIOS E APRENDIZADOS

A implementação destacou desafios como:

- Heterogeneidade discente: Necessidade de adaptar a ferramenta para diferentes perfis de aprendizagem ^[5].
- Questões éticas: Privacidade de dados e transparência algorítmica ^[11].
- Capacitação docente: Formação contínua para integrar tecnologias de forma pedagógica ^[12].
- Os aprendizados reforçam a importância de um design inclusivo, alinhamento com objetivos pedagógicos e investimento em infraestrutura e formação ^[4].

5. CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

A iniciativa mostrou-se promissora e alinhada às tendências educacionais contemporâneas. Para superar as limitações, sugere-se a criação de um guia rápido ou tutorial para familiarizar os usuários com a plataforma e a exploração de recursos visuais mais detalhados em versões futuras do jogo. Como perspectivas futuras, planeja-se submeter a proposta a um comitê de ética, viabilizando avaliações com estudantes e docentes em turmas futuras, bem como explorar recursos visuais mais imersivos. A continuidade do projeto poderá explorar cenários de auditoria mais complexos e a integração com outras ferramentas de simulação.

6. REFERÊNCIAS

1. Agência Nacional de Vigilância Sanitária (ANVISA). Manual de Princípios Diretrizes e Regras de Auditoria do SUS. Brasília: ANVISA; 2017.
2. Fernandes CS, Lima A, Melo FLNB. Tecnologia e Gamificação Aplicada à Saúde. ADITGAMES; 2024.
3. Zhonggen Y. A Meta-Analysis of Use of Serious Games in Education. Int J Comput Games Technol. 2019;2019:4797035.
4. Subhash S, Cudney EA. Gamified learning in higher education: A systematic review. Comput Human Behav. 2018;87:192-206.
5. Daghestani LF, et al. Adapting gamified learning systems using educational data mining techniques. Comput Appl Eng Educ. 2020;28(3):568-89.
6. World Economic Forum. The Future of Jobs Report 2020. Geneva: WEF; 2020.
7. Melo FLNB. Gamificação e aversão ao risco: um estudo com estudantes de administração. Rev Adm Empresas. 2024;64(2):1-15.
8. ANVISA. Relatório da avaliação nacional das práticas de segurança do paciente. Brasília: ANVISA; 2023.
9. Rosebud AI. Plataforma de desenvolvimento de personagens virtuais. 2025. Disponível em: <https://play.rosebud.ai>.
10. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation. Am Psychol. 2000;55(1):68-78.
11. Maleki M, Alizadeh M. A Review of AI-based Gamification in Medical Education. J Med Educ Dev. 2024;15(1):1-12.

Impacto da qualidade dos dados em indicadores perinatais: uma abordagem simulada com apoio da Inteligência Artificial

Jannine dos Santos Nascimento¹, Luciana Vinhal dos Santos Ferreira¹, Zilma Silveira Nogueira Reis¹

¹Programa de Pós-Graduação em Saúde da Criança e do Adolescente. Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

ninne.nascimento@hotmail.com, lucianavinhaldotsantosferreira@gmail.com,
zilma.medicina@gmail.com

Resumo: A qualidade de dados em registros de saúde é determinante para se obter indicadores de saúde confiáveis e relevantes para o planejamento de políticas públicas. Este estudo avaliou o impacto da completude de dados sobre indicadores perinatais, por meio de um experimento simulado que comparou cenários de alta e baixa qualidade de dados, gerados por Large Language Model (LLM). Os resultados demonstram que dados incompletos podem gerar distorções significativas nos indicadores calculados, com potenciais consequências na sua utilização para tomada de decisão. A análise sugere que a incorporação de abordagens de inteligência artificial para avaliar a qualidade dos dados pode contribuir na gestão da informação sobre a saúde populacional.

Abstract: Data quality in health records is crucial for obtaining reliable health indicators that are essential for planning public policies. This study evaluated the impact of data completeness on perinatal indicators through a simulated experiment that compared high- and low-quality data scenarios, generated using a Large Language Model (LLM). The results show that incomplete data can lead to significant distortions in the calculated indicators, potentially affecting their use in decision-making. The analysis suggests that incorporating artificial intelligence approaches to assess data quality can support the management of population health information.

Palavras-chave: qualidade dos dados; saúde digital; registros de saúde.

1. CONTEXTO E MOTIVAÇÃO

A confiabilidade dos registros de saúde é fundamental para a vigilância epidemiológica, o planejamento de políticas públicas e o suporte à decisão clínica ⁽¹⁻²⁾. No entanto, a baixa qualidade dos dados – refletida por variáveis incompletas, inconsistentes ou desatualizadas – compromete a validade de indicadores e gera vieses nas análises ⁽³⁾. Em contextos materno-infantis, onde decisões são sensíveis e urgentes, essa limitação pode ocultar desigualdades ou desinformar gestores ⁽⁴⁾.

1.1 Objetivo

Este estudo avalia o impacto da perda de qualidade nos registros sobre a acurácia de indicadores perinatais, comparando cenários simulados de dados completos e incompletos.

2. ATIVIDADES PRINCIPAIS

A principal atividade desenvolvida consistiu na aplicação de técnicas de Inteligência Artificial (IA) à área da saúde, com foco na análise de dados relacionados à qualidade da informação em registros neonatais.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este estudo utilizou uma simulação de base de dados com registros hipotéticos, elaborados exclusivamente para fins de análise metodológica. Por isso, não foi necessária submissão ao Comitê de Ética em Pesquisa, conforme a Resolução CNS nº 510/2016. A proposta com inteligência artificial segue as diretrizes da Lei Geral de Proteção de Dados (LGPD), assegurando anonimato, segurança e transparência no uso de informações.

Foi realizado um estudo experimental em ambiente de simulação, utilizando-se o Chat GPT4.0 para gerar uma base sintética de dados, com 1.000 registros hipotéticos de nascimentos. No cenário de alta qualidade todas as variáveis estavam completas e consistentes. No cenário de baixa qualidade aplicou-se uma taxa de 30% de dados ausentes/inconsistentes aleatoriamente, simulando perdas típicas observadas em bancos de dados reais. A ausência de dados foi aplicada de forma aleatória e independente para cada variável, resultando em pequenas variações no número de registros válidos por indicador. O prompt utilizado foi:

<https://drive.google.com/file/d/1RQx4XeRgOaLzmEiliQhMISypdi5bSNC8/view>

As variáveis obtidas por geração sintética foram aquelas necessárias para calcular indicadores de qualidade utilizados em saúde perinatal: (1) cobertura de pré-natal, cujo indicador de qualidade esperado é de ≥ 6 consultas. (2) gestante adolescente, cujo indicador é idade < 20 anos. (3) taxa de prematuridade, cujo indicador é a razão entre a idade gestacional < 37 semanas e o número de nascimentos. (4) baixo peso ao nascer, cujo indicador é a razão entre crianças com peso de nascimento < 2500 g e o número de nascimentos. (5) Tipo de parto, cujo indicador é a razão do número de partos cesariana e o número total de partos. (6) Apgar no 5º minuto de vida < 7 . (7) Internação em UTI neonatal.

Os indicadores foram calculados conforme definições estabelecidas por consenso internacional. As diferenças observadas entre os indicadores gerados pela base de dados completa e a base de dados com qualidade comprometida foram analisadas quanto aos seus valores. Também foram realizadas inferências livres sobre o seu impacto potencial de indicadores gerados com dados sem qualidade na prática clínica e no planejamento de políticas públicas em saúde. Os resultados foram apresentados na **Tabela 1**.

Tabela 1 - Indicadores de saúde materno-infantil nos cenários de baixa e alta qualidade dos dados gerados por LLM (n = 1.000).

| Indicador | Baixa Qualidade | Alta Qualidade |
|---|------------------|-------------------|
| Cobertura de pré-natal adequado (≥ 6 consultas) | 59,1% (n=700) | 70,3% (n=1000) |
| Gestantes adolescentes (< 20 anos) | 17,2% (n=698) | 20,1% (n=1000) |
| Prematuridade (< 37 semanas) | 7,1% (n=703) | 10,3% (n=1000) |
| Baixo peso ao nascer (< 2500 g) | 12,7% (n=696) | 15,3% (n=1000) |
| Cesarianas | 39,1% (n=695) | 39,0% (n=1000) |
| Apgar < 7 no 5º minuto | 11,9% (n=700) | 16,0% (n=1000) |
| Internação em UTI neonatal | 7,4% (n=701) | 9,8% (n=1000) |

Fonte: Modelo de aprendizado de máquina.

A análise comparativa dos indicadores calculados com os dados sintéticos indicam que a perda de qualidade subestimou quase todos os indicadores. Em cenários reais, hipotetizamos que a distorção

observada pode comprometer a priorização de ações, alocação de leitos, dimensionamento de equipes e distribuição de insumos.

Em relação ao indicador de cobertura de pré-natal, as falhas de preenchimento podem ocultar deficiências na assistência pré-natal e comprometer o monitoramento da qualidade da atenção à gestante. Enquanto que a aparente redução de gestantes adolescentes, por exemplo, pode desviar políticas de proteção social desse grupo. Da mesma forma, a subnotificação de prematuridade e baixo peso tem potencial para comprometer estratégias de prevenção e cuidados intensivos neonatais.

Em contrapartida, a taxa de cesarianas mostrou menor sensibilidade à perda de qualidade. Essa estabilidade sugere que variáveis padronizadas e obrigatórias, muitas vezes exigidas para faturamento hospitalar, tendem a apresentar maior completude e consistência nos registros.

A subestimação do Apgar <7 no 5º minuto pode ocultar a real gravidade das complicações neonatais, o que impacta diretamente a avaliação da qualidade da assistência ao parto e nascimento. Já a subnotificação das internações em UTI neonatal pode afetar a alocação de recursos e a regulação de leitos, dada a alta complexidade e custo desse cuidado.

4. DESAFIOS E APRENDIZADOS

Acreditamos que estratégias para avaliar sistematicamente a qualidade dos dados de saúde, essenciais para o cálculo de indicadores de saúde, podem ser úteis em sistemas de informação em saúde. Ferramentas de IA, como modelos preditivos e algoritmos de auditoria automatizada, podem aumentar a precisão dos registros e facilitar o reconhecimento precoce de intervenções necessárias em saúde pública.

5. REFERÊNCIAS

1. Harkener S, Stausberg J, Hagel C, Siddiqui R. Towards a Core Set of Indicators for Data Quality of Registries. *Stud Health Technol Inform*. 2019 Sep 3;267:39-45. doi: 10.3233/SHTI190803.
2. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013 Jan 1;20(1):144-51. doi: 10.1136/amiajnl-2011-000681.
3. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc*. 2002 Nov-Dec;9(6):600-11. doi: 10.1197/jamia.m1087.
4. Makhoul N. Revisão de indicadores e métricas de qualidade de dados e sugestões de indicadores e métricas para monitoramento da saúde estrutural. *ABEN*. 2022;3(17). doi: 10.1186/s43251-022-00068-9.

Implementação de inteligência artificial no Sistema Único de Saúde do estado de São Paulo: redução de custos e aumento da precisão de diagnóstico em mamografias

Maria Angélica Dedoné Gonçalves ¹, Orientadora: Prof.^a Dr^a Soraya Regina Gasparetto Lunardi ²

¹ Universidade Estadual Paulista Júlio De Mesquita Filho, CAMPUS ARARAQUARA - FCLAR, SP, BRASIL.

² Universidade Estadual Paulista Júlio De Mesquita Filho, CAMPUS ARARAQUARA - FCLAR, SP, BRASIL.

ma.gonçalves@unesp.br, orientadora: soraya.gasporetto@unesp.br.

Resumo: Este projeto tem como objetivo identificar os impactos da inteligência artificial no campo da saúde, sobretudo no que tange a redução de custos e precisão de diagnósticos de mamografia no Estado de São Paulo. Para isso, são realizados movimentos metodológicos, que consistem em: revisão sistemática e análise estatística. Entre os resultados esperados estão: redução no tempo de análise, redução de falsos positivos e negativos, além de uma economia anual em exames desnecessários.

Palavras-chave: Inteligência Artificial; mamografias; SUS-SP; redução de custos; saúde pública.

1. INTRODUÇÃO E SÍNTESE BIBLIOGRÁFICA

A Inteligência Artificial (IA) tem se consolidado como uma tecnologia de impacto crescente, especialmente na área da saúde. Definida como a capacidade de máquinas executarem tarefas que exigem inteligência humana (Russell; Norvig, 2021), sua aplicação vem acompanhada de desafios éticos e legais, como a proteção de dados sensíveis (Obermeyer; Emanuel, 2016).

Entre as principais abordagens, destacam-se o *machine learning* e o *deep learning*, cujas técnicas permitem a análise autônoma de grandes volumes de dados complexos, sendo amplamente utilizadas em exames por imagem. Evidências empíricas, como as de Ardila et al. (2019), demonstraram que algoritmos baseados em *deep learning* podem prever o risco de câncer de pulmão por meio da análise de tomografias computadorizadas, superando a acurácia de radiologistas experientes, reforçando o potencial transformador da IA na medicina, ao mesmo tempo em que se exige regulamentação específica, sobretudo em sistemas públicos como o SUS.

A IA tem se destacado como tecnologia inovadora na saúde, com aplicações que vão de diagnósticos precisos a terapias personalizadas. Em imagem médica, algoritmos detectam anomalias em tempo real, otimizando decisões clínicas.

Na oncologia, sistemas como o *Watson for Oncology* sugerem tratamentos baseados em evidências, reduzindo subjetividades. No rastreamento do câncer de mama, algoritmos superam radiologistas em precisão, com redução de 5,7% nos falsos positivos e 9,4% nos falsos negativos (Mckinney et al., 2020). O NHS britânico obteve ganhos de custo e eficiência com IA em mamografias (Esteve et al., 2017). No Brasil, hospitais como Sírio-Libanês e A.C. Camargo já aplicam IA em diagnósticos, validando sua efetividade, desempenham um papel crucial no diagnóstico e tratamento do câncer de mama.

A análise dos centros de referência é fundamental para entender como a I.A. pode ser integrada ao SUS, ampliando a precisão diagnóstica e reduzindo a sobrecarga dos radiologistas. Contudo, sua adoção enfrenta desafios estruturais, como escassez de especialistas, filas prolongadas, desigualdades regionais e infraestrutura tecnológica defasada (Dourado, 2023; OCDE, 2023). Diante disso, este estudo busca suprir lacunas na literatura ao avaliar a viabilidade da IA em mamografias nas 15 regiões administrativas do SUS-SP, propondo diretrizes regulatórias e estimando seus impactos clínicos e econômicos.

2. METODOLOGIA

A metodologia consiste em: **1. Revisão Sistemática:** será realizada em bases como PubMed e SciELO, com recorte dos últimos 10 anos, utilizando os termos: “IA em mamografias”, “custo-efetividade” e “implementação no SUS”. O objetivo é reunir evidências atualizadas sobre a aplicação da IA no rastreamento mamográfico, além de prevenir a obsolescência da revisão bibliográfica, priorizando-se a inclusão de fontes recentes e atualizadas; **2. Coleta de Dados Secundários:** serão extraídos dados do DATASUS (SIASIH), INCA e Hospital Sírio-Libanês sobre número de mamografias, estágios do câncer e custos (quimioterapia, cirurgias, biópsias). A amostra será composta por um município de cada região administrativa do Estado de São Paulo, selecionados com base em IDH, população e localização; **3. Análise Estatística e Modelagem:** serão aplicadas análises descritivas e inferenciais, seguidas de modelagem estatística em R. Serão projetadas reduções de custos e aumento na precisão diagnóstica com o uso da IA, além de análise de sensibilidade com base em dados reais; **4. Validação com Evidências Internacionais:** os resultados serão comparados com estudos internacionais, como McKinney et al. (2020), para validação dos cenários simulados; **5. Proposta de Diretrizes para o SUS-SP:** serão elaboradas diretrizes preliminares para implementação da IA no SUS, abordando infraestrutura, capacitação profissional, segurança de dados e regulamentação. Assim, a abordagem se certifica como, quantitativa e exploratória, com dados anonimizados (Resolução CNS 510/2016).

3. RESULTADOS PARCIAIS

Com base na literatura científica e em dados públicos do SUS-SP, espera-se que a implementação da Inteligência Artificial (IA) na triagem de mamografias traga benefícios clínicos e econômicos significativos. As projeções são: **1) Redução de falsos positivos:** cerca de **5,7%**, evitando exames e biópsias desnecessárias. **Impacto estimado:** Em aproximadamente 1 milhão de mamografias realizadas anualmente no SUS-SP, isso representa até **57.000 falsos positivos evitados**; **2) Redução de falsos negativos:** cerca de **9,4%**, permitindo a identificação precoce de casos inicialmente não detectados. **Impacto clínico:** maior probabilidade de tratamento precoce, o que pode aumentar as chances de cura em até **30%** (INCA, 2023). **3) Melhoria da acurácia diagnóstica geral:** incremento de **10 a 12%** na sensibilidade e especificidade, aproximando o desempenho da IA ao de especialistas. **4) Redução do tempo de análise dos exames:** a automação pode diminuir em até **30 a 40% o tempo de leitura dos laudos**, possibilitando a realocação de profissionais para casos complexos. E, **economia anual estimada:** considerando o custo médio de **R\$ 60,00 por mamografia** (SUS-SP), a redução de exames e procedimentos desnecessários pode gerar uma economia entre **R\$ 10 e 15 milhões por ano**.

4. REFERÊNCIAS

1. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25:954-61.
2. Brasil. Ministério da Saúde. Departamento de Informática do SUS (DATASUS). Sistema de Informação Ambulatorial do SUS (SIA/SUS). [Internet]. [cited 2025 May 30]. Available from: <http://www2.datasus.gov.br>
3. Dourado DA. A regulação da inteligência artificial na saúde no Brasil: desafios e perspectivas. *Cad Saúde Publica*. 2023;39(2):e00190222. doi:10.1590/0102-311XEN190222.
4. Instituto Nacional de Câncer (INCA). Câncer de mama: estimativas 2023. Rio de Janeiro: INCA; 2023 [cited 2025 May 30]. Available from: <https://www.inca.gov.br>
5. Obermeyer Z, Emanuel EJ. Predicting the future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med*. 2016;375(13):1216-9.
6. Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Health at a Glance: Latin

America and the Caribbean 2023. Paris: OECD Publishing; 2023 [cited 2025 May 30]. Available from: <https://www.oecd.org/health>

7. Russell S, Norvig P. Inteligência artificial. 3rd ed. São Paulo: Pearson; 2021.

8. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-8. doi:10.1038/nature21056.

9. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577:89-94. doi:10.1038/s41586-019-1799-6.

In silico analysis of potential anti-inflammatory agents for psoriasis

D. E. M de Souza¹, T.M. Machado^{1*}, J.T.T. Soares^{1*}, F. Guilhon-Simplicio¹

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

*Scholarship of Fundação de Amparo à Pesquisa de São Paulo

dougemanuel⁴⁰@gmail.com, tallita.machado@yahoo.com.br, joaothiagotss@gmail.com, guilhon_simplicio@ufam.edu.br

Abstract: Psoriasis is a chronic autoimmune skin disease with no effective and specific therapies for this pathology. In this sense, this in silico study investigated Amazonian compounds for new treatments. Molecular targets of psoriasis were identified and cross-referenced with potential targets of the molecules, revealing 18 common targets. Interaction network analysis highlighted six priority targets, with the Chymase 1 gene, essential in psoriasis inflammation, selected for study. Molecular docking was performed in AutoDock[®] using the 4KPO structure, validated by RMSD <2Å, showing that dilapiol had the highest affinity of -7.7 kcal/mol. Three derivatives were designed to optimize their pharmacological properties. All surpassed dilapiol in binding energy: DEMS002: -9.5 kcal/mol; DEMS001: -8.2 kcal/mol; DEMS003: -7.7 kcal/mol, and showed favorable interactions in the active site. Pharmacokinetic and toxicological evaluation confirmed that the derivatives have superior profiles to dilapiol, with DEMS003 standing out for the best Caco-2 intestinal permeability: -5.15 log cm/s, 71% absorption, blood-brain barrier permeability, and favorable metabolism as a substrate of CYP2D6, CYP3A4, and CYP2C9. It was concluded that dilapiol is the most promising compound and its derivatives, especially DEMS003, are optimized candidates for future studies against psoriasis.

Keywords: molecular docking; dilapiol; derivatives; bioinformatics.

Project Name: AmazonIA - Application of Machine Learning and Deep Learning Models for Data Analysis on Neglected Tropical Diseases in the Amazon.

1. CONTEXT AND MOTIVATION

Psoriasis is a chronic inflammatory skin disease, typically manifesting as red, scaly, and itchy lesions that can appear on various parts of the body ⁽¹⁾. Although its exact cause remains unknown, psoriasis is recognized as an autoimmune condition in which the immune system mistakenly attacks healthy skin cells ⁽²⁾. Genetic and environmental factors, such as stress and infections, also contribute to its development. Given the continuous need for new therapies, this study focused on exploring Amazonian compounds, including lupeol, amyryn, spilanthol, dilapiol, and lapachol, recognized for their anti-inflammatory and antioxidant activities. Using computational modeling and databases previously developed by the research group, we performed computational screening to identify the most promising compound for psoriasis.

1.1 Objective

Investigate the anti-psoriasis potential of six molecules through database screening and, based on the most promising compound, design and compare at least three derivatives to optimize their pharmacological activity using artificial intelligence and bioinformatics.

2. MAIN ACTIVITIES

- Identification of Molecular Targets and Intersection and Interaction Network Analysis;
- Target Selection and Molecular Docking;
- Derivative Design and Evaluation;
- Pharmacokinetic and Toxicological Assessment (ADME-Tox);
- Results Analysis and Selection.

3. WORK DEVELOPMENT (METHODOLOGY AND RESULTS)

3.1 Methodology

3.1.1 Data mining

Molecular targets of the disease were cataloged in GeneCards®, prioritizing essential proteins, while molecular targets of the compounds were identified via SwissTargetPrediction®.

3.1.2 Protein-protein interaction network development

The intersection was analyzed using a Venn diagram, and protein-protein interaction (PPI) networks were constructed in NetworkAnalyst®, identifying hub nodes and priority biological pathways.

3.1.3 Docking Molecular

Molecular docking was performed using AutoDock® with the 4KP0 structure (Protein Data Bank), validated by redocking with RMSD < 2 Å. Subsequently, molecular docking was conducted to identify the most promising ligand for the target. With grid box parameters set as: center_x = 25.277; center_y = -9.181; center_z = -12.02 e size_x = 10; size_y = 8 e size_z = 10.

3.1.4 Design and development of new derivatives

Three derivatives were developed using the Maestro® platform to optimize pharmacological activity. Parameters followed the molecule's pharmacophore group—the functional group responsible for biological activity. Modifications were made exclusively at the C2 and C7 carbons, based on literature data indicating these are non-essential sites.

3.1.5 Pharmacokinetic and toxicological parameters

Pharmacokinetic and toxicological parameters were evaluated using the artificial intelligence platforms AldrugLab and PROTOX.

3.2 Results

The initial analysis identified 10,584 targets associated with psoriasis and the studied molecules. The Venn diagram revealed 18 common targets. Protein-protein interaction network analysis of these 18 targets highlighted six as particularly interesting and essential for psoriasis' biological processes. Molecular docking was then performed with the target considered to have the highest pathological potential: the chymase 1 enzyme essential in inflammatory processes, tissue remodeling, and immune response, characteristic of psoriasis. Other targets will be investigated in future studies. Molecular docking showed dilapiol had the highest affinity (-7.7 kcal/mol), outperforming amyryn (-3.3 kcal/mol), lupeol (-5.9 kcal/mol), spilanthol (-5.3 kcal/mol), lapachol (-7.1 kcal/mol), and the co-crystallized ligand (-7.5 kcal/mol).

Based on this result, three dilapiol derivatives the DEMS001, DEMS002, and DEMS003 were designed for pharmacological optimization. Molecular docking with these derivatives revealed all surpassed the original compound in binding energy: DEMS002 (-9.5 kcal/mol), DEMS001 (-8.2 kcal/mol), and DEMS003

(-7.7 kcal/mol). Furthermore, the derivatives exhibited more favorable interactions with key amino acid residues in the active site GLY216, LYS192, ALA190, GLY193, SER195, including hydrogen bonds, and replicated interactions observed with the crystallographic ligand (3), suggesting strong affinity and essential features for enzyme inhibitory activity.

Pharmacokinetic and toxicological parameters were evaluated using AldrugLab and PROTOX platforms. All derivatives met druglikeness and pharmacokinetic criteria, showing more favorable toxicity profiles than dilapiol. DEMS002 stood out in solubility (Log K = 0.14 log mol/L; moderate aqueous solubility = -3.88 log mol/L). DEMS003 achieved the best predicted Caco-2 permeability (- 5.15 log cm/s). For human intestinal absorption, DEMS001 showed 30%, while DEMS002 and DEMS003 reached 70% and 71%, respectively. DEMS001 and DEMS002 were substrates of CYP2D6 and CYP3A4, while DEMS003 was a substrate of CYP2D6, CYP3A4, and CYP2C9, indicating potentially favorable metabolism with no predicted problematic inhibition or drug interactions. DEMS003 also showed the best predicted blood-brain barrier permeability and overall distribution. Regarding toxicity, DEMS001 was the safest, with no alerts for mutagenicity (Ames), carcinogenicity (rats/mice), or hepatotoxicity.

4. CHALLENGES AND LEARNINGS

The complexity of psoriasis required managing extensive databases (10,584 initial targets) and precise identification of relevant molecular targets among thousands of possibilities. Selecting a single priority target from six promising biological candidates involved difficult decisions, leaving other potential targets for future studies. Molecular docking validation through redocking demanded meticulous execution to avoid errors, while designing dilapiol derivatives faced the challenge of balancing molecular affinity with pharmacokinetic properties, a common dilemma in drug development. However, this work demonstrated that Amazonian natural compounds, particularly dilapiol, are viable foundations for developing anti-psoriatic agents, validating the computational screening strategy. Molecular modification revealed that structural changes to dilapiol could simultaneously enhance target affinity and optimize pharmaceutical properties, overcoming limitations of the original compound. The experience further highlighted that an ideal candidate (like DEMS003) is not necessarily the one with highest affinity, but rather one combining efficacy, safety, and favorable ADMET characteristics. Finally, it confirmed that the *in silico* approach efficiently reduces costs and time in early drug discovery phases.

5. REFERENCES

1. Guo, J., et al. "Signaling pathways and targeted therapies for psoriasis." *Signal transduction and targeted therapy* 8.1 (2023):437.
2. Radu, A., Tit, D.M., Endres, L.M., et al. "Naturally derived bioactive compounds as precision modulators of immune and inflammatory mechanisms in psoriatic conditions." *Inflammopharmacol* 33 (2025):527-549.
3. Matsumoto, Y., Kakuda, S., Koizumi, M., Mizuno, T., Muroga, Y., Kawamura, T., et al. "Crystal structure of a complex of human chymase with its benzimidazole derived inhibitor." *J. Synchrotron Rad.* 20 (2013):914-918.

Inovações na análise de dados sobre a hemorragia puerperal: ampliação de dados, estratégias de risco e validação com Inteligência Artificial

Desenvolvimento Tecnológico

Flávia Ribeiro de Oliveira¹, Ana Paula Couto da Silva², Zilma Silveira Nogueira Reis¹

¹Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil

²Departamento de Ciência da Computação da UFMG, Belo Horizonte, MG, Brasil

flaviagob@ufmg.br, anacouto@dcc.ufmg.br, zilmareis@ufmg.br

Resumo: A hemorragia puerperal (HPP) representa uma das principais causas de morbimortalidade materna no Brasil. Apesar da existência de protocolos assistenciais, a identificação precoce do risco de HPP ainda apresenta lacunas, especialmente em contextos com recursos limitados. Este projeto tem como objetivo validar modelos de estratificação de risco com o uso de técnicas de inteligência artificial (IA), aplicadas a bases de dados clínicas provenientes de dois centros públicos de referência. A proposta contempla a ampliação da base de dados retrospectiva (2014 a 2024), análise de variáveis clínicas relevantes e aplicação de algoritmos de aprendizado de máquina para predição de risco hemorrágico. A iniciativa integra diferentes áreas do conhecimento, como ginecologia, obstetrícia, ciência de dados e saúde digital. O trabalho está vinculado ao Programa de Pós-Doutorado em Saúde da Mulher da Faculdade de Medicina da UFMG, e insere-se em uma proposta maior de desenvolvimento tecnológico com foco na melhoria da vigilância e do cuidado materno. Espera-se contribuir para a personalização da assistência obstétrica baseada em evidências.

Palavras-chave: hemorragia puerperal; Inteligência Artificial; aprendizado de máquina; risco materno; saúde digital.

Nome do projeto: Inovações na Análise da Hemorragia Puerperal: Ampliação de Dados, Estratégias de Risco e Validação com Inteligência Artificial.

1. CONTEXTO E MOTIVAÇÃO

A hemorragia puerperal (HPP) é uma das principais causas evitáveis de morte materna no Brasil e no mundo. Sua identificação precoce e manejo adequado são essenciais para a redução da morbimortalidade materna, especialmente em contextos de alta demanda e recursos limitados ⁽¹⁾. No entanto, os métodos tradicionais de estratificação de risco ainda são falhos, baseando-se em fatores isolados ou modelos clínicos pouco sensíveis.

Com o avanço da ciência de dados e da inteligência artificial (IA), surgem oportunidades inéditas para desenvolver ferramentas preditivas mais eficazes e capazes de integrar múltiplas variáveis clínicas e assistenciais ⁽²⁾. A aplicação de algoritmos de aprendizado de máquina à base de dados de mulheres atendidas em maternidades públicas, pode ampliar a capacidade de previsão do risco de HPP e apoiar a tomada de decisão em tempo real ⁽³⁾. Esta abordagem inovadora responde à necessidade de personalização da assistência e aprimoramento dos sistemas de vigilância materna, especialmente no Sistema Único de Saúde (SUS) ^(4 e 5).

1.1 Objetivo

Desenvolver e validar modelos preditivos de risco para hemorragia puerperal por meio de técnicas

de aprendizado de máquina, utilizando dados clínicos retrospectivos e prospectivos de mulheres atendidas em maternidades públicas, com o objetivo de subsidiar estratégias mais eficazes de prevenção e manejo no contexto da saúde pública brasileira.

2. ATIVIDADES PRINCIPAIS

As principais atividades deste projeto estão organizadas em etapas sequenciais e integradas, com previsão de execução durante o ano de 2025. São elas:

- 2.1. Revisão da literatura sobre predição HPP, abordagens de IA aplicadas à saúde materna e diretrizes nacionais e internacionais relacionadas à vigilância obstétrica.
- 2.2. Ampliação da base de dados clínica com inclusão de registros retrospectivos de partos ocorridos entre 2014 e 2024 nos centros participantes (HC-UFMG e MOV- FHEMIG), assegurando o cumprimento de critérios éticos e de qualidade dos dados.
- 2.3. Curadoria, limpeza e padronização dos dados, com definição das variáveis relevantes, tratamento de dados faltantes e transformação dos dados para análise por algoritmos de machine learning.
- 2.4. Seleção e implementação de algoritmos de AM, incluindo técnicas supervisionadas para construção de modelos preditivos de risco de hemorragia puerperal.
- 2.5. Validação dos modelos com aplicação de métricas estatísticas de desempenho (sensibilidade, especificidade, acurácia, AUC), análise de importância de variáveis e avaliação externa com dados independentes.
- 2.6. Discussão dos resultados e elaboração de produtos de divulgação científica, como artigos, apresentações e proposta de protótipo de ferramenta de apoio à decisão clínica.

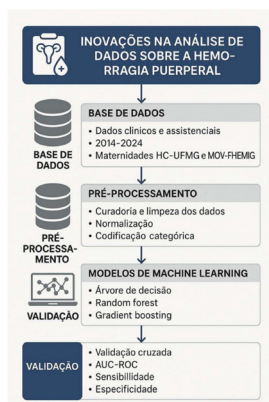
3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este estudo está vinculado ao Programa de Pós-Doutorado em Saúde da Mulher da Faculdade de Medicina da UFMG e ao projeto "Inovações na Análise da Hemorragia Puerperal: Ampliação de Dados, Estratégias de Risco e Validação com Inteligência Artificial", aprovado pelos Comitês de Ética em Pesquisa da UFMG (CAAE: 55087421.0.0000.5149) e da FHEMIG (CAAE: 55087421.0.3001.5119). O projeto foi contemplado com financiamento pelo Edital Ampliado CI-IA Saúde – FINEP/MCTI/SESAI/SCTIE/MS, o que fortalece sua execução e viabilidade institucional. Trata-se de um estudo observacional retrospectivo, com previsão de etapas prospectivas para validação externa. A base de dados incluirá informações clínicas e assistenciais de mulheres que pariram nas maternidades HC-UFMG e MOV-FHEMIG entre os anos de 2014 e 2024. As variáveis contemplam dados sociodemográficos, antecedentes obstétricos, fatores de risco clínicos e evolução da assistência ao parto, com foco específico na ocorrência de hemorragia puerperal (HPP).

Atualmente, o projeto encontra-se na etapa de coleta e organização dos dados junto à FHEMIG, com a curadoria sendo conduzida conforme os parâmetros éticos e técnicos estabelecidos. Após essa etapa, será realizado o pré-processamento dos dados (incluindo normalização e codificação categórica), seguido da seleção e aplicação de algoritmos supervisionados de aprendizado de máquina, como árvore de decisão, *random forest* e *gradient boosting*. A avaliação dos modelos preditivos ocorrerá por meio de validação cruzada, com análise de métricas como sensibilidade, especificidade, acurácia e área sob a curva ROC (AUC-ROC).

Até o presente momento, não há resultados disponíveis. No entanto, a metodologia já estruturada, o suporte interinstitucional e o financiamento vigente garantem bases sólidas para o desenvolvimento das próximas etapas do projeto.

Figura 1 - Fluxograma do projeto com etapas de desenvolvimento e aplicação da inteligência artificial na predição da hemorragia puerperal.



Fonte: Elaboração própria com auxílio do ChatGPT (2025).

4. DESAFIOS E APRENDIZADOS

Durante a execução deste projeto, foram identificados diversos desafios relacionados à natureza interdisciplinar da proposta e à sua implementação em instituições públicas de saúde. Um dos primeiros desafios enfrentados foi a obtenção das aprovações éticas e institucionais, especialmente no contexto da expansão do estudo para múltiplos centros. O alinhamento entre os cronogramas dos comitês de ética, núcleos de pesquisa e setores de tecnologia das instituições participantes demandou persistência e articulação contínua. Outro desafio importante tem sido a formação e a manutenção de uma equipe multidisciplinar qualificada, incluindo profissionais da saúde, estatísticos e cientistas de dados, o que exige planejamento de recursos humanos e financeiros. A qualidade dos dados também se apresenta como um ponto crítico, uma vez que a coleta em sistemas distintos requer padronização, curadoria rigorosa e estratégias de tratamento de inconsistências.

O acesso seguro às bases de dados clínicas, respeitando a LGPD e os protocolos institucionais, envolve etapas técnicas e legais complexas. Finalmente, a elaboração dos modelos preditivos exige decisões metodológicas cuidadosas para lidar com desbalanceamento de classes, variáveis correlacionadas e validação em contextos distintos. Apesar dos desafios, o projeto tem proporcionado aprendizados significativos, como a construção de soluções viáveis para análise de grandes volumes de dados clínicos reais, o fortalecimento da colaboração interinstitucional e a aplicação prática de modelos de inteligência artificial voltados à melhoria da saúde materna.

5. REFERÊNCIAS

1. Say L, Chou D, Gemmill A, Tunçalp Ö, Moller AB, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. *Lancet Glob Health*. 2014;2(6):e323-33. doi: 10.1016/S2214-109X(14)70227-X. Epub 2014 May 5. PMID: 25103301.
2. Knight M, Callaghan WM, Berg C, Alexander S, Bouvier-Colle MH, Ford JB, et al. Trends in postpartum hemorrhage in high-resource countries: a review and recommendations from the International Postpartum Hemorrhage Collaborative Group. *BMC Pregnancy Childbirth*. 2009;9:55. doi: 10.1186/1471-2393-9-55. PMID: 19943928; PMCID: PMC2790440.
3. Ministério da Saúde (BR). *Boletim Epidemiológico: Mortalidade materna no Brasil, 2009–2020 [Internet]*. Brasília (DF): Ministério da Saúde; 2022 [citado 2025 set 3]. Vol. 53, nº 20:19-29. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2022/boletim-epidemiologico-vol-53-no20/view>
4. Pinheiro YT, Silva RA. Artificial intelligence in gynecology and obstetrics: from the enthusiasm of use in practice to the challenges of implementation. *Rev Bras Ginecol Obstet*. 2024;46:e-rbgo41. doi: 10.61622/rbgo/2024rbgo41.
5. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317-8. doi: 10.1001/jama.2017.18391. PMID: 29532063.

Inteligência artificial como ferramenta para inclusão de pacientes oncológicos em cuidados paliativos: uma revisão de escopo

A. Santos (aluna)¹, J. Vega Castañeda (aluno)¹, H. Rocha (colaborador)², I. Quispe Mendoza (colaboradora)³, G. Simino (coorientadora)³, M. Cherchiglia (orientadora)²

¹Programa de Pós Graduação em Saúde Pública, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG

²Departamento de Medicina Preventiva e Social, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG

³Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, MG

pssaline@gmail.com, alejandrovcast@gmail.com, rochahugoandre@gmail.com, isabelyovana@gmail.com, gsimino@gmail.com, mcherchiglia@gmail.com

Resumo: Introdução: Os cuidados paliativos (CP), buscam melhorar a qualidade de vida de pacientes com doenças graves e de seus familiares, por meio do controle de sintomas e atenção a aspectos físicos, psicossociais e espirituais. Na oncologia, os CP são fundamentais para alinhar o cuidado aos valores do paciente e evitar intervenções desnecessárias. A Inteligência Artificial (IA), por sua capacidade de analisar grandes volumes de dados e apoiar decisões clínicas, tem se destacado como ferramenta promissora nesse contexto. **Objetivo:** Mapear as tecnologias de IA utilizadas para inclusão de pacientes oncológicos em CP, identificando os modelos aplicados, desfechos e impactos na qualidade de vida. **Metodologia:** Trata-se de uma revisão de escopo baseada nas diretrizes do Instituto Joanna Briggs e do PRISMA-ScR. O protocolo foi registrado no OSF e as buscas foram realizadas em nove bases de dados e na literatura cinzenta, sem restrição de idioma ou período. **Resultados:** Foram identificados 833 estudos até abril de 2025. Atualmente a triagem encontra-se em andamento. **Conclusão:** O trabalho desenvolvido visa oferecer uma visão abrangente e atual da literatura, apoiando estratégias que favoreçam o encaminhamento oportuno aos CP. Dessa forma, o estudo poderá subsidiar políticas públicas e favorecer o acesso equitativo à saúde.

Palavras-chave: inteligência artificial; cuidados paliativos; oncologia; revisão de escopo; saúde pública.

Nome do projeto: Integração Inteligente: Estratégias de Cuidados Paliativos Orientadas por IA para Pacientes com Câncer.

1. CONTEXTO E MOTIVAÇÃO

Segundo dados da Worldwide Hospice Palliative Care Alliance (WHPCA) no ano de 2020, mundialmente, estima-se que mais de 56 milhões de pessoas apresentem demandas direcionadas aos cuidados paliativos (CP). Cerca de 25 milhões destas encontram-se em fase final de vida, possuindo idades médias acima dos 50 anos e residindo, em sua maior parte, em países de baixa ou média renda.⁽¹⁻³⁾

Estudos que envolvem a aplicação de inteligência artificial (IA) em cuidados paliativos têm se tornado cada vez mais relevantes, especialmente considerando as demandas e a necessidade de cuidados mais personalizados e eficazes. Desta forma, tendo em vista o grande potencial de aplicação da IA na saúde, sobretudo para benefício de pacientes oncológicos e ampliação do acesso aos CP, emerge a necessidade de realizar um mapeamento dos estudos que utilizam a IA com a finalidade de direcionar a inclusão de pacientes oncológicos em CP e ainda identificar quais intervenções esses pacientes estavam sendo submetidos e desfechos após a inclusão em CP.

2. ATIVIDADES PRINCIPAIS

As principais atividades deste trabalho envolveram a elaboração e registro do protocolo na plataforma Open Science Framework (OSF), seguido pela construção de uma estratégia de busca abrangente, em colaboração com um bibliotecário especializado. As buscas foram realizadas em nove bases de dados científicas (PubMed, Cochrane, Lilacs, Scopus, Embase, Livivo, Web of Science, IEEE Xplore e ACM Digital Library) e em fontes de literatura cinzenta (Google Scholar e ProQuest), sem restrições de idioma ou data. Os resultados foram organizados e gerenciados na plataforma Rayyan, onde dois revisores realizaram a triagem por título e resumo de forma independente. Em casos de discordância, um terceiro revisor foi acionado para tomada de decisão.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Inicialmente foram identificados 2.029 estudos provenientes de 11 bases de dados. Após a remoção de 1.196 estudos duplicados, 833 artigos foram submetidos à triagem por título e resumo. Atualmente, o estudo encontra-se na etapa de leitura completa dos artigos selecionados, com extração de dados estruturada para mapeamento dos modelos de IA aplicados, desfechos clínicos e impactos na qualidade de vida dos pacientes oncológicos em cuidados paliativos.

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento desta revisão, houve desafios em relação ao grande volume inicial de estudos identificados, o que exigiu um processo detalhado de triagem para a remoção de duplicatas e a análise de títulos e resumos. A etapa de elegibilidade demandou uma definição clara dos critérios para exclusão, o que evidenciou a necessidade de maior alinhamento entre os revisores. Em contrapartida, esse processo permitiu importantes aprendizados, como o aprimoramento da estratégia de busca e o refinamento dos critérios de seleção. Esses elementos contribuíram para fortalecer a consistência e a transparência do estudo, além de ressaltar a importância do rigor científico na condução da revisão.

5. REFERÊNCIAS

1. World Health Organization. National cancer control programmes: policies and managerial guidelines. 2. ed. Geneva: World Health Organization; 2002. 180 p.
2. Global Atlas of Palliative Care. 2o ed. Geneva: World Health Organization (WHO) e Worldwide Palliative Care Alliance (WPCA); 2020.
3. Why Palliative Care is an Essential Function of Primary Health Care. Geneva; 2018. Report No.: WHO/HIS/SDS/2018.39.
4. Papachristou N, Kotronoulas G, Dikaios N, Allison SJ, Eleftherochorinou H, Rai T, et al. Digital Transformation of Cancer Care in the Era of Big Data, Artificial Intelligence and Data-Driven Interventions: Navigating the Field. *Semin Oncol Nurs.* junho de 2023;39(3):151433.
5. Charalambous A, Dodlek N. Big Data, Machine Learning, and Artificial Intelligence to Advance Cancer Care: Opportunities and Challenges. *Semin Oncol Nurs.* junho de 2023;39(3):151429.

Inteligência artificial no manejo do sangramento uterino anormal: modelos preditivos para algoritmo de tratamento personalizado

Tipo: Desenvolvimento Tecnológico

Lívia Leni de Oliveira do Nascimento¹, Lucas Emanuel Cunha Melo¹, Zilma Silveira Nogueira Reis¹, Flávia Ribeiro de Oliveira¹

¹Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil

livialeninasc@gmail.com, lucasecm@ufmg.br, zilmareis@ufmg.br, flaviagob@ufmg.br

Resumo: O Sangramento Uterino Anormal (SUA) compromete significativamente a qualidade de vida das mulheres e, frequentemente, exige procedimentos diagnósticos complexos e terapias padronizadas. Este projeto propõe o uso de modelos de Aprendizado de Máquina (AM) para aprimorar o diagnóstico e desenvolver um algoritmo de tratamento personalizado, abordando uma lacuna relevante na prática ginecológica. Trata-se de um estudo observacional retrospectivo com dados de pacientes diagnosticadas com SUA, atendidas na Maternidade Odete Valadares – FHEMIG entre 2019 e 2023. Serão utilizados algoritmos supervisionados, como redes neurais e árvores de decisão, com o objetivo de construir modelos preditivos robustos. A amostragem será por conveniência, considerando a complexidade e a disponibilidade dos dados. Solicitou-se a dispensa do Termo de Consentimento Livre e Esclarecido (TCLE), conforme normativas éticas para estudos retrospectivos. A confidencialidade e o sigilo das informações serão rigorosamente garantidos. A pesquisa foi aprovada pelos Comitês de Ética da UFMG e FHEMIG. Com base nos achados, será proposto um algoritmo clínico com potencial de aplicação prática, promovendo avanços no cuidado à saúde da mulher. Este estudo contribui para a inovação na atenção ginecológica, com impacto positivo para pacientes, profissionais de saúde e a pesquisa científica.

Palavras-chave: Sangramento Uterino Anormal; inteligência artificial; aprendizado de máquina; tratamento personalizado; saúde digital.

1. CONTEXTO E MOTIVAÇÃO

O Sangramento Uterino Anormal (SUA) é uma condição clínica comum, que acomete entre 10% e 30% das mulheres em idade reprodutiva e também pode ocorrer no período pós-menopausa. Trata-se de uma das principais causas de demanda por atendimento ginecológico⁽¹⁾. A Federação Internacional de Ginecologia e Obstetrícia (FIGO) propôs a classificação PALM-COEIN para a etiologia do SUA, que agrupa as causas em categorias estruturais (Pólipo, Adenomiose, Leiomioma, Malignidade e Hiperplasia) e não estruturais (Coagulopatia, Ovulatória, Endometrial, Iatrogênica e Não classificada)⁽²⁾. No entanto, na prática clínica, muitas pacientes ainda recebem tratamentos padronizados, com pouca consideração por suas características individuais, o que pode comprometer a efetividade terapêutica e a satisfação com o cuidado recebido^(3,4).

A crescente disponibilidade de bases de dados clínicos e os avanços na inteligência artificial, especialmente em modelos de Aprendizado de Máquina (AM), oferecem novas possibilidades para a análise de grandes volumes de dados e identificação de padrões que podem orientar decisões clínicas de forma mais personalizada e precisa^(3,4,5).

Diante desse cenário, este projeto busca inovar ao aplicar modelos de AM na predição de padrões clínicos de SUA e na elaboração de um algoritmo de decisão terapêutica personalizada. A proposta se insere na convergência entre ciência de dados e saúde da mulher, promovendo uma abordagem mais eficiente, individualizada e baseada em evidências para o manejo do SUA.

1.1 Objetivo

Desenvolver um modelo de aprendizado de máquina capaz de classificar com precisão os diferentes tipos de Sangramento Uterino Anormal (SUA) segundo os critérios PALM-COEIN, utilizando dados clínicos, laboratoriais e de imagem de pacientes atendidas no ambulatório especializado da Faculdade de Medicina da UFMG, com vistas à personalização do diagnóstico e à otimização da conduta terapêutica.

2. ATIVIDADES PRINCIPAIS

A elaboração e submissão do projeto ao Comitê de Ética da UFMG, seguida de emenda para inclusão da MOV-FHEMIG como centro coparticipativo, bem como a busca por financiamento por meio de editais de fomento.

Foi realizada uma revisão de escopo da literatura sobre o uso da inteligência artificial (IA) no manejo do Sangramento Uterino Anormal (SUA), com o intuito de mapear abordagens existentes e identificar lacunas de conhecimento.

A base de dados retrospectiva está sendo organizada com informações clínicas, laboratoriais e de imagem de mulheres com diagnóstico de SUA. Os casos foram classificados segundo os critérios PALM-COEIN e validados por especialistas.

O pré-processamento dos dados incluiu a normalização, tratamento de dados ausentes e codificação categórica. Em seguida, modelos preditivos estão sendo desenvolvidos e validados com algoritmos de aprendizado de máquina (*Decision Tree, Random Forest, SVM e XGBoost*), avaliados por métricas como acurácia, sensibilidade, especificidade, F1-score e AUC-ROC.

A análise comparativa entre os modelos visa identificar aqueles com maior potencial de aplicação clínica. Os resultados estão sendo organizados para divulgação em artigos científicos e congressos especializados.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este estudo integra o projeto de doutorado em Ciências Aplicadas à Cirurgia e Oftalmologia da Faculdade de Medicina da UFMG, intitulado "Aplicação de Modelagem de Aprendizado de Máquina na Predição de Padrões de Sangramento Uterino Anormal e na Elaboração de Algoritmo de Tratamento Personalizado". Aprovado pelos Comitês de Ética em Pesquisa da UFMG (CAAE: 82310024.0.0000.5149) e da FHEMIG (CAAE: 82310024.0.3001.5119), o projeto segue um delineamento observacional retrospectivo, com validação futura dos modelos preditivos.

A base de dados é composta por informações clínicas, laboratoriais e de imagem de pacientes com diagnóstico de Sangramento Uterino Anormal (SUA) atendidas no ambulatório da Faculdade de Medicina da UFMG (centro proponente) e na Maternidade Odete Valadares – FHEMIG (centro coparticipante). As variáveis incluem características sociodemográficas, hormonais, reprodutivas, morfológicas e evolução clínica.

A metodologia prevê curadoria e pré-processamento dos dados, classificação das pacientes segundo os critérios PALM-COEIN, e aplicação de algoritmos de aprendizado de máquina supervisionado (*Decision Tree, Random Forest, SVM e XGBoost*). A performance dos modelos será avaliada com métricas como acurácia, sensibilidade, especificidade, F1-score e AUC-ROC.

A revisão de escopo encontra-se em fase de submissão, e o projeto está atualmente na etapa de coleta de dados no centro coparticipante, além de seguir em busca de financiamento complementar. Até o momento, os esforços estão voltados à estruturação da base de dados e à preparação para as etapas subsequentes de análise com inteligência artificial, que visam explorar seu potencial no suporte à decisão clínica personalizada para o manejo do Sangramento Uterino Anormal (SUA).

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do projeto, foram enfrentados desafios relevantes que exigiram estratégias de superação e proporcionaram aprendizados significativos. Entre os principais obstáculos, destacam-se:

- A tramitação ética e regulatória, incluindo submissão de projeto e emendas nos sistemas CEP/ CONEP, especialmente para inclusão de centro coparticipante;
- A organização da base de dados clínicas, com necessidade de padronização, curadoria manual e tratamento de informações incompletas;
- A interface entre conhecimentos da área da saúde e da ciência de dados, exigindo uma sólida articulação interdisciplinar;
- A busca ativa por financiamento, com elaboração de propostas adequadas aos editais disponíveis;
- O domínio das etapas de pré-processamento e modelagem com algoritmos de aprendizado de máquina, como Decision Tree, SVM e XGBoost.

Esses desafios impulsionaram o crescimento acadêmico e científico da doutoranda, promovendo o desenvolvimento de competências essenciais para a atuação em pesquisa translacional. A vivência direta com revisão sistemática, metodologia científica, submissão a órgãos reguladores, análise de dados clínicos e aplicação de inteligência artificial tem ampliado sua autonomia, pensamento crítico e capacidade de inovar na área de Ginecologia e Obstetrícia.

5. REFERÊNCIAS

1. Federação Brasileira das Associações de Ginecologia e Obstetrícia (FEBRASGO). Sangramento uterino anormal. São Paulo: FEBRASGO; 2017.
2. Munro MG, Critchley HOD, Broder MS, Fraser IS. FIGO classification system (PALM-COEIN) for causes of abnormal uterine bleeding in nongravid women of reproductive age. *Int J Gynaecol Obstet.* 2011;113(1):3-13.
3. Schork NJ. Artificial Intelligence and Personalized Medicine. *Cancer Treat Res.* 2019;178:265-83.
4. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ.* 2019;7:e7702.
5. Choi Y, Park J, Kim H, Lee YJ, Lee Y, Choi YS, et al. Artificial intelligence models predicting abnormal uterine bleeding after COVID-19 vaccination. *Sci Rep.* 2025;15:7081. doi:10.1038/s41598-025-91882-4.

Interpretable Machine Learning for arboviruses diagnosis: a XAI approach

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

Abstract: Machine Learning (ML) models play a crucial role in optimizing healthcare resources by reducing diagnostic errors and preventing financial waste. In this work, we propose an ML classification pipeline to address two binary classification problems: distinguishing between chikungunya and dengue and classifying cases as arboviruses or non-arboviruses. First, we applied the Chi-squared test to reduce the dimensionality of tabular data extracted from the Recife open data portal and the SINAN database. Then, we employed a novel Monte Carlo-inspired approach to optimize the hyperparameters of three shallow ML models. To ensure the robustness of our approach, we performed k-fold cross-validation to assess the effectiveness of the best hyperparameters found in the previous step. Finally, we evaluated five classification metrics: Accuracy, Recall, Precision, F1-Score, and Area Under the ROC Curve. To enhance model explainability, we used SHAP values to analyze the best-performing model, which was the Gradient Boosting Classifier, for both classification tasks. Our results outperform previous studies that used the same dataset.

Keywords: Dengue and Chikungunya; Machine Learning; Feature Selection; SHAP Explainability; Monte Carlo Cross-Validation.

Project Name: Cover for blind review.

1. CONTEXT AND MOTIVATION

Arboviruses are ecologically defined viruses capable of replicating in both arthropods and vertebrates, primarily transmitted through bites from mosquitoes such as *Aedes aegypti* and *Aedes albopictus*. Dengue and chikungunya are common arboviral infections in tropical regions, presenting similar early symptoms like fever, rash, and muscle and joint pain, which complicates clinical diagnosis. Dengue is often linked to thrombocytopenia, leukopenia, and increased hematocrit, while chikungunya is marked by arthritis and prolonged joint pain. Although most dengue cases are mild, some may progress to severe, potentially life-threatening conditions if not properly managed. In this context, Machine Learning (ML) models can enhance diagnostic accuracy, assist clinical decision-making, and reduce unnecessary treatments, ultimately improving patient outcomes and reducing healthcare costs.

1.1 Objective

We propose an ML classification pipeline to tackle two tasks using the dataset described in ⁽⁴⁾: (i) determine whether a case is arboviral—defined here as either chikungunya or dengue (A)—or not (NA, including inconclusive or unrelated cases); and (ii) distinguish between chikungunya and dengue (CHIK × DEN). We use SHAP values ⁽¹⁾ to interpret model outputs and identify the most relevant features.

2. MAIN ACTIVITIES

The main activities included:

- Feature dimensionality reduction via Chi-squared test;
- Splitting the dataset into two binary tasks (CHIK × DEN and A × NA) for targeted classification;

- Training shallow models Gradient Boosting Classifier (GBC), Random Forest Classifier (RFC), and Support Vector Machine (SVM) to mitigate overfitting on limited data;
- Tuning model hyperparameters using Monte Carlo-inspired Cross-Validation (MCICV) and validating via k-fold Cross-Validation;
- Assessing model performance using metrics such as AUC, Precision, Recall, and F1-score;
- Interpreting model decisions through SHAP to identify key predictive features;
- Comparing performance against prior studies using the same dataset to benchmark results.

3. DEVELOPMENT OF THE WORK (METHODOLOGY AND RESULTS)

We developed a classification pipeline based on 14 features selected via Sequential Feature Algorithm (SFA) and refined by domain experts, presented in ⁽⁴⁾. These include 13 binary variables and one integer (DIAS), standardized using Z-score. The dataset was split into two tasks: Chikungunya vs. Dengue (CHIK × DEN) and Arboviruses vs. Non-Arboviruses (A × NA). Pearson’s Chi-squared test was used for feature selection, reducing dimensionality to 10 relevant variables: ‘DIAS’, ‘ARTRALGIA’, ‘FEBRE’, ‘MIALGIA’, ‘CEFALEIA’, ‘DOR_RETRO’, ‘EXANTEMA’, ‘NAUSEA’,

‘ARTRITE’, and ‘DOR_COSTAS’. Given the limited sample size (up to 17,172 instances), we focused on shallow models—GBC, RFC, and SVM—to prevent overfitting. Hyperparameters were tuned using Monte Carlo-inspired Cross-Validation (MCICV), followed by 3-fold validation. Due to computational constraints, SVM was evaluated with only three MCICV iterations, while GBC and RFC used 100. SHAP values ⁽¹⁾ provided model interpretability, revealing the most influential features in each task.

3.1 CHIK x DEN task

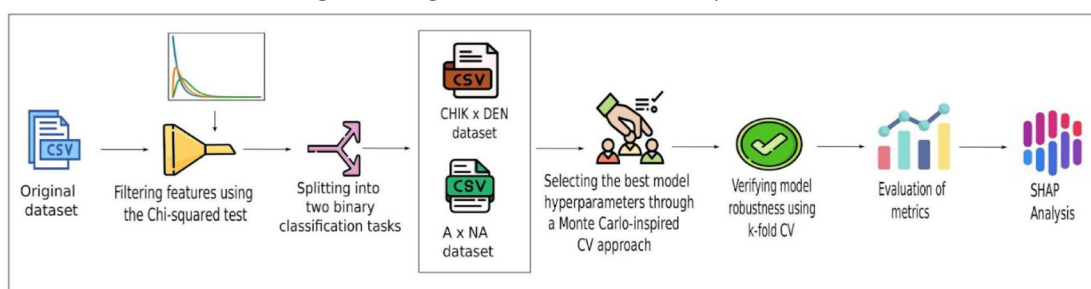
GBC achieved the best performance, with ACC = 0.77 and AUC = 0.83, outperforming RFC and SVM. It also showed higher Recall (0.74) and Precision (0.79), confirming strong class discrimination and generalization. SHAP analysis highlighted arthralgia as the most influential feature—though counterintuitively associated with dengue in this dataset. Eye pain presence favored chikungunya classification, while headache, symptom duration (DIAS), and arthritis also contributed. Fever had limited impact, reflecting its ubiquity in both diseases.

3.2 A x NA task

Despite class imbalance (A: 11,448; NA: 5,724), the models showed similar overall performance, differing mainly in the Precision–Recall balance. RFC made fewer NA-to-A misclassifications, but GBC and SVM had fewer false negatives and achieved the highest Recall (0.85), making them preferable in healthcare contexts. GBC stood out with the highest Precision and F1-Score (0.81), outperforming XGBoost ⁽³⁾ and matching DeepInsight CNN ⁽²⁾, despite using only ten features.

Feature importance analysis shows that the absence of arthralgia, rash, arthritis, nausea, and eye pain favors NA classification, while longer symptom duration also indicates non-arboviral cases.

Figura 1 - A general overview of the experiment.



4. CHALLENGES AND LEARNINGS

This study proposed a ML pipeline to classify arbovirus vs. non-arbovirus (A × NA) and chikungunya vs. dengue (CHIK × DEN) cases. Dimensionality was reduced from 14 to 10 features using the Chi-squared test, preserving model performance. Hyperparameters were tuned via Monte Carlo-inspired Cross Validation (MCICV), followed by K-fold validation. SHAP values enhanced interpretability by highlighting influential features. The GBC model achieved state-of-the-art performance on the A × NA task (F1-score of 0.81 with 10 features), surpassing prior studies ⁽²⁾ and ⁽³⁾. CHIK × DEN classification yielded an AUC of 0.83, showing strong class separability. Compared to previous studies—specifically works ⁽²⁾ and ⁽⁴⁾—which used 14 or 27 features, this approach achieved comparable or superior performance using fewer input variables. SHAP explanations occasionally diverged from clinical expectations, likely due to non-linear effects and feature interactions ⁽¹⁾. Future work should test generalization on other datasets, investigate new dimensionality reduction methods (e.g., PCA, Autoencoders), and refine data curation to minimize undersampling bias ⁽⁴⁾.

5. REFERENCES

1. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS); 2017 Dec; Long Beach, CA, USA. Curran Associates Inc. p. 4765–74. Available from: <https://dl.acm.org/doi/10.5555/3295222.3295230>
2. Neto LM, da Silva Neto SR, Endo PT. A comparative analysis of converters of tabular data into image for the classification of arboviruses using convolutional neural networks. PLoS One. 2023;18:e0295598.
3. Neto SRDS, Tabosa T, Neto LM, Teixeira IV, Sadok S, Sampaio VDS, Endo PT. Binary models for arboviruses classification using machine learning: A benchmarking evaluation. In: Proceedings of the 56th Hawaii International Conference on System Sciences; 2023. p. 2820–9.
4. Oliveira TT, da Silva Neto SR, Teixeira IV, de Oliveira SBA, de Almeida Rodrigues MG, Sampaio VS, Endo PT. A comparative study of machine learning techniques for multi-class classification of arboviral diseases. Front Trop Dis. 2022;2.

Interpretação automatizada do exame de papanicolaou para triagem do câncer de colo do útero

Mayara K. M. Gomides¹, Leonardo A. Ferreira¹, Mateus A. da Silva¹, Bruna L. M. dos Santos¹, Cauan C. Marotta¹, Giulia Z. de Castro¹, Lucas E. Alves¹, Lucas M. Herinques¹, Franciele A. Barbosa², Marcelo A. Pascoal Xavier¹, Marilene Vale de C. Monteiro¹, Frederico G. Guimarães¹, Walmir M. Caminhas¹

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

²Universidade Estadual de Montes Claros, Montes Claros, MG, Brasil.

{mgomides, leaufferreira, silvaamaralmateus, blms, cauanmarotta, giuliaz, lucaseliasalves, lucas-
mont, pascoal, marilenevale, fredericoguimaraes, caminhas}@ufmg.br,
francielealvesb¹⁰@gmail.com

Resumo: O câncer do colo do útero representa um relevante problema de saúde pública, com alta incidência e mortalidade, especialmente em países em desenvolvimento. A análise automatizada de imagens citológicas por meio de algoritmos de deep learning tem se mostrado promissora na detecção e classificação de lesões cervicais precursoras deste câncer. No entanto, a maioria dos modelos tem uma capacidade de generalização limitada devido à variabilidade entre equipamentos e protocolos, configurando o problema conhecido como mudança de domínio (domain shift). Diante desses desafios, o presente projeto propõe o uso de estruturas de aprendizado federado com privacidade diferencial como solução viável para corrigir o problema da limitação de generalização sem comprometer a privacidade dos dados. Conduzido por uma equipe interdisciplinar de estudantes e orientadores, o estudo apresenta como resultados parciais a realização de experimentos com modelos aplicados em bases de dados cruzadas, visando avaliar a robustez e a capacidade de generalização dos modelos em contextos diversos e realistas.

Palavras-chave: deep learning; aprendizado federado; câncer cervical.

Nome do projeto: Interpretação Automatizada do Exame de Papanicolaou para Triagem do Câncer de Colo do Útero.

1. CONTEXTO E MOTIVAÇÃO

O câncer do colo do útero constitui um importante problema de saúde pública. No Brasil, estimam-se cerca de 17.010 novos casos anuais entre 2023 e 2025, com uma taxa de mortalidade de 4,51 por 100 mil mulheres em 2021 ⁽¹⁾. Nesse cenário, técnicas de inteligência artificial, em especial algoritmos de Deep Learning (DL), têm se destacado na análise automatizada de imagens citológicas, com resultados promissores na extração de características clínicas relevantes ⁽²⁾.

Contudo, tais modelos geralmente são treinados com dados de uma única instituição, o que limita sua capacidade de generalização diante da variabilidade entre equipamentos e métodos de coloração utilizados por diferentes centros. Isso acarreta o problema de mudança de domínio (*domain shift*), que compromete o desempenho em dados externos ⁽³⁾. Uma solução viável é o uso de estruturas de aprendizado federado com privacidade diferencial, que possibilitam a análise colaborativa de imagens médicas sem compartilhamento direto dos dados, assegurando a confidencialidade e ampliando a diversidade dos conjuntos utilizados no treinamento dos modelos ⁽⁴⁾.

2. ATIVIDADES PRINCIPAIS

O projeto é desenvolvido por uma equipe interdisciplinar formada por estudantes de graduação, pós-graduação e pesquisadores das áreas de engenharia, ciência da computação e medicina.

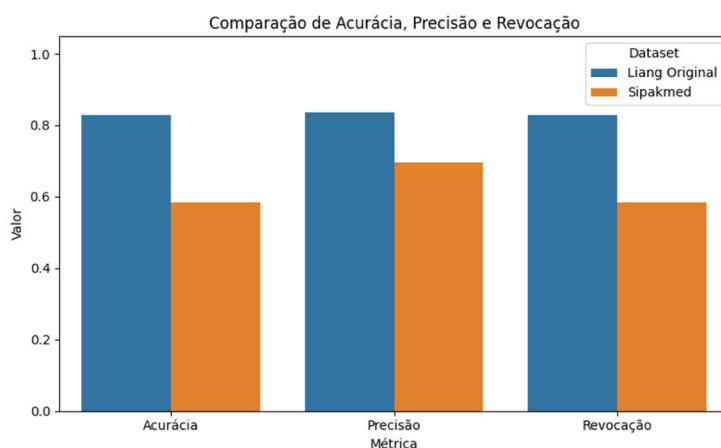
As atividades são organizadas em duas reuniões semanais: uma técnica, voltada à discussão de desafios computacionais, ajustes na arquitetura dos modelos e implementação de soluções; e outra, com toda a equipe, dedicada à apresentação de resultados, avaliação crítica e planejamento das tarefas subsequentes.

Entre as principais atividades, destacam-se: (i) revisões sistemáticas sobre aprendizado profundo aplicado à análise de imagens médicas; (ii) estudo de modelos automatizados para detecção e classificação celular; e (iii) atualização contínua sobre aspectos clínicos e epidemiológicos do câncer cervical. Essa estrutura de trabalho assegura que o desenvolvimento da ferramenta esteja alinhado às evidências científicas atuais e às demandas clínicas reais.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Atualmente, está sendo conduzida uma investigação sobre a perda de desempenho decorrente da limitada capacidade de generalização dos modelos, atribuída ao fenômeno de overfitting. Para essa análise, foram realizados experimentos nos quais modelos previamente treinados foram avaliados em conjuntos de dados distintos daqueles utilizados em sua etapa de treinamento. Especificamente, o modelo Comparison Detector proposto por Liang et al. (5) foi aplicado ao conjunto de dados SIPaKMeD (6), enquanto o modelo desenvolvido por Plissiti et al. (6) foi avaliado no conjunto de dados originalmente utilizado por Liang et al. (5). Os resultados obtidos (Figura 1) evidenciam uma degradação no desempenho dos modelos ao serem expostos a dados externos, indicando uma limitação em sua capacidade de generalização.

Figura 1 - Comparação das métricas de desempenho do modelo proposto por Liang et al. (5). Em azul, os resultados obtidos quando o modelo foi treinado e testado com o conjunto de dados original descrito pelos autores. Em laranja, o desempenho do mesmo modelo ao ser aplicado no conjunto de dados SIPaKMeD (6), o que evidencia o impacto da mudança de domínio.



4. DESAFIOS E APRENDIZADOS

Como desafio, destacamos a necessidade de uma infraestrutura computacional apropriada, estratégias eficientes de processamento de imagem e o envolvimento interdisciplinar entre engenheiros, cientistas de dados e especialistas em patologia e ciências médicas. Além dos entraves técnicos, barreiras administrativas e de recursos humanos também impactaram o andamento do projeto. Houve atraso na implementação das bolsas pela agência financiadora, o que comprometeu a alocação tempestiva de pesquisadores nas tarefas de desenvolvimento. Paralelamente, identificou-se dificuldade na captação de

bolsistas com o perfil técnico e acadêmico adequado.

Como aprendizado relevante ressaltamos a constatação de que, mesmo com modelos avançados, como o Comparison Detector ⁽⁵⁾, a capacidade de generalização ainda representa uma limitação substancial. A superação desse desafio está exigindo estudos aprofundados sobre estratégias de generalização e implementação de técnicas de aprendizado federado, com base em abordagens recentes da literatura ^(3,4).

5. REFERÊNCIAS

1. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Dados e números sobre câncer do colo do útero – 2023. Rio de Janeiro: INCA; 2023. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/dados_e_numeros_col_o_22marco2023.pdf. Acesso em: 21/05/2025.
2. Jiang H, Zhou Y, Lin Y, Chan RCK, Liu J, Chen H. Deep learning for computational cytology: A survey. *Medical Image Analysis*. 2023 Feb 1;84:102691.
3. Pan Q, Yang B, Chen D, Wang M. Enhancing Model Generalization of Cervical Fluid-Based Cell Detection through Causal Feature Extraction: A Novel Method. Em: *Proceedings of the 15th Asian Conference on Machine Learning* [Internet]. PMLR; 2024 [citado 22 de maio de 2025]. p. 1055–70. Disponível em: <https://proceedings.mlr.press/v222/pan24a.html>.
4. Adnan M, Kalra S, Cresswell JC, Taylor GW, Tizhoosh HR. Federated learning and differential privacy for medical image analysis. *Sci Rep*. 2022 Feb 4;12(1):1953.
5. Liang Y, Tang Z, Yan M, Chen J, Liu Q, Xiang Y. Comparison-Based Convolutional Neural Networks for Cervical Cell/Clumps Detection in the Limited Data Scenario [Internet]. arXiv; 2018 [citado 22 de maio de 2025]. Disponível em: <http://arxiv.org/abs/1810.05952>.
6. Plissiti M.E., Dimitrakopoulos P, Sfikas G, Nikou C, Krikoni O, Charchanti A. SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images. In: *IEEE International Conference on Image Processing (ICIP)*; 2018; Athens, Greece. IEEE; 2018.

Letramento em inteligência artificial para a saúde: uma abordagem prática e inclusiva para idosos

Davi Greco^{1,2}, João Pedro Reis^{1,2}, Andre Lima^{1,2}, Francyne Nascimento¹, Cristiane Dias¹, Leonardo Rocha², Elisa Tuler (orientadora)^{1,2}

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

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

davigreco21@gmail.com

Resumo: Com o crescimento do uso de tecnologias digitais na área da saúde e o envelhecimento da população, torna-se fundamental promover o letramento digital de pessoas idosas. Este trabalho apresenta o desenvolvimento de um curso introdutório sobre Inteligência Artificial (IA) aplicada à saúde, voltado especificamente para esse público. A iniciativa é coordenada pela Diretoria de Educação e Difusão do Conhecimento do CI-IA Saúde, que atua na promoção da educação em saúde digital. O curso está sendo elaborado com foco em uma linguagem simples, utilizando o WhatsApp como principal canal de interação, devido à sua ampla familiaridade entre pessoas idosas. Os conteúdos abordam conceitos básicos de IA, seu uso na saúde e no cotidiano, por meio de cards informativos, vídeos, perguntas interativas e atividades lúdicas. O objetivo é promover a inclusão digital, fortalecer a autonomia e o conhecimento, validando a proposta com os próprios participantes, garantindo que ela atenda às suas necessidades reais.

Palavras-chave: letramento digital; educação em saúde; tecnologia interativa.

Nome do projeto: Letramento em Inteligência Artificial para a Saúde.

1. CONTEXTO E MOTIVAÇÃO

A Diretoria de Educação e Difusão do Conhecimento do CI-IA Saúde tem se destacado na promoção da educação em saúde digital, utilizando mídias sociais, cursos e eventos para disseminar informações e capacitar profissionais e a comunidade em geral. Com o objetivo de ampliar o alcance dessas iniciativas, surge a necessidade de expandir os esforços para diversos públicos-alvo, incluindo a população idosa.

O envelhecimento da população é uma realidade crescente em muitas sociedades, incluindo o Brasil. Segundo dados do Censo Demográfico de 2022¹, o país conta com 22,2 milhões de pessoas com 65 anos ou mais, representando um aumento de 57% em relação a 2010. Esse fenômeno impõe desafios significativos para a saúde pública, tornando essencial a adoção de tecnologias que promovam o bem-estar e a qualidade de vida dos idosos.

A inteligência artificial (IA) tem se mostrado uma ferramenta promissora nesse contexto. Soluções baseadas em IA, como assistentes virtuais, dispositivos de monitoramento de saúde e sistemas de alerta para quedas, por exemplo, têm demonstrado eficácia na promoção da autonomia, segurança e saúde mental na população com 60 anos ou mais ^(1,2,4,5). Para que essas tecnologias alcancem seu potencial máximo, é fundamental que os idosos adquiram competências digitais que lhes permitam utilizar essas ferramentas de forma eficaz.

1.1 Objetivo

Este trabalho propõe a criação de um curso de letramento em IA para a saúde, direcionado aos idosos, com o intuito de capacitar esse público no uso de tecnologias digitais aplicadas à saúde. Para tornar o curso mais acessível e familiar ao público-alvo, a interação será realizada por meio do WhatsApp

— aplicativo de mensagens muito utilizado por pessoas 60+ no Brasil. Sua interface intuitiva, ampla adoção e facilidade de uso tornam essa plataforma ideal para reduzir barreiras tecnológicas.

2. METODOLOGIA

A metodologia adotada para o desenvolvimento do curso de letramento em IA está sendo estruturada em duas frentes complementares: ⁽¹⁾ elaboração do conteúdo pedagógico e ⁽²⁾ desenvolvimento da solução tecnológica de entrega e interação.

2.1 Elaboração do Conteúdo

A construção do conteúdo seguiu os princípios da educação para adultos ⁽³⁾, incluindo a autoimagem do aprendiz, a experiência prévia, a prontidão para aprender, a orientação para a aprendizagem e a motivação interna. Além disso, foram considerados os conceitos de alfabetização digital em saúde, levando em conta as necessidades cognitivas, comunicativas e culturais dos idosos. Nessa frente, estamos realizando as seguintes etapas: levantamento de temas-chave, definição dos objetivos de aprendizagem, elaboração de roteiros com linguagem acessível e adequada ao público-alvo, desenvolvimento de materiais gráficos e, por fim, será feita uma validação com os usuários. Os roteiros, em particular, têm papel central na estruturação do conteúdo, funcionando como guias pedagógicos que orientam a construção das mensagens, a escolha dos recursos visuais e o ritmo da apresentação, garantindo coerência e eficácia na transmissão da informação.

2.2 Solução Tecnológica e Interface de Entrega

O projeto utiliza quatro tecnologias principais — *Python*², *NextJS*³, *Moodle* e *Firebase*⁴⁴ — escolhidas estrategicamente para atender às necessidades técnicas e sociais da plataforma.

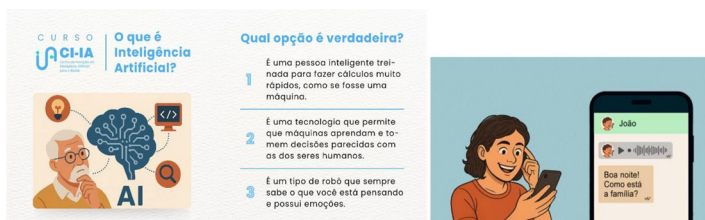
O *backend* está sendo desenvolvido em *Python*, devido à sua simplicidade e robustez, permitindo o gerenciamento dos dados dos usuários, a comunicação via API com o WhatsApp e o controle do progresso nos cursos. Ele será responsável por interpretar as mensagens recebidas e fornecer respostas personalizadas. A interface gráfica voltada à equipe pedagógica será construída com *NextJS*, uma ferramenta em *Python* que permite criar aplicações web de forma rápida e intuitiva. Com isso, a equipe de conteúdo poderá criar e editar cursos, incluir perguntas com botões interativos e acompanhar relatórios de forma acessível, sem a necessidade de conhecimento técnico. Por fim, o armazenamento das informações será feito com o *Firebase*, um banco de dados em nuvem do Google que permite acesso em tempo real aos dados e facilita o desenvolvimento de aplicações escaláveis.

3. O CURSO

Como apresentado na metodologia, o curso está sendo desenvolvido com linguagem simples e acolhedora, buscando apresentar os conceitos básicos de IA de forma acessível e prática. O conteúdo está em processo de elaboração, contemplando temas como definições no contexto da Inteligência Artificial, suas aplicações no cuidado diário com a saúde, prevenção de doenças e apoio ao diagnóstico médico. Para enriquecer a experiência de aprendizado, estão sendo criados cards informativos com conceitos-chave (Figura 1), “jogos” interativos, questões de verdadeiro e falso e vídeos explicativos (Figura 1), todos pensados para estimular o engajamento e facilitar a assimilação dos conteúdos. A interação ocorre por meio de interações no WhatsApp, que conduz o aprendizado com explicações curtas, perguntas com botões de resposta e recursos visuais. O objetivo é despertar o interesse, promover o letramento digital e mostrar como a tecnologia pode ser uma aliada no cuidado com a saúde e na promoção do bem-estar.

¹ <https://www.ibge.gov.br/estatisticas/sociais/saude/22827-censo-demografico-2022.html>

Figura 1 - Cards que apresentam conceitos e Print do vídeo “Reconhecimento de Voz e Acessibilidade com IA”



4. DESAFIOS E RESULTADOS ESPERADOS

Como apresentamos, este projeto busca oferecer um curso acessível e interativo que capacite idosos no letramento em Inteligência Artificial aplicada à saúde. Pretendemos, por meio de uma abordagem simples e acolhedora, facilitar a inclusão digital e a autonomia desse público. A validação do curso será realizada diretamente com os próprios idosos, garantindo que o conteúdo e a plataforma atendam às suas necessidades e promovam um aprendizado efetivo e significativo.

5. REFERÊNCIAS



² <https://docs.python.org/3/>

³ <https://nextjs.org/>

⁴ <https://firebase.google.com/docs/firestore?hl=pt-br>

⁵ <https://sora.chatgpt.com/>

Modernização do sistema de gerenciamento de laudos anatomopatológicos no HC-UFMG: extração automatizada de dados e interoperabilidade com ferramentas analíticas

Victor Huggo Bartholomeu Gardino (aluno)¹, Eduardo Felipe Venuti Coelho (aluno)¹, Carlos Henrique Rolemberg da Rocha (aluno)¹, Adriano Lages dos Santos (orientador)², Cristiano Xavier Lima (Orientador)¹, Paula Vieira Teixeira Vidigal¹ (orientador)

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

²Instituto Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

bartholomeuvictor@gmail.com, paula.vidigal@gmail.com, adrianocomp@gmail.com

Resumo: Em conformidade com a Lei nº 13.787/2018 e com a Lei Geral de Proteção de Dados (LGPD), instituições de saúde devem garantir o armazenamento de prontuários médicos e exames por longos períodos. No Hospital das Clínicas da Universidade Federal de Minas Gerais, o sistema de gerenciamento de laudos anatomopatológicos de biópsias tornou-se obsoleto, dificultando sua operação e manutenção. Este projeto tem como objetivo desenvolver um novo sistema com usabilidade aprimorada e interoperabilidade com ferramentas de análise de dados. As etapas principais incluem categorização da estrutura de dados dos laudos, extração automatizada das informações com uso de expressões regulares, validação progressiva de dados por especialistas e organização dos dados em formato tabular. O sistema conta com interface gráfica amigável, geração de logs e relatórios. A iniciativa visa não apenas cumprir exigências legais, mas também ampliar o aproveitamento da base de dados para fins administrativos e de saúde pública. Atualmente em fase de validação, o projeto apresenta resultados promissores.

Abstract: In compliance with Brazilian legislation, healthcare institutions must ensure accessible long-term storage of patient medical records and exams. At Hospital das Clínicas da Universidade Federal de Minas Gerais, the current anatomic pathology biopsy management system has become obsolete, compromising its operability and maintainability. This project aims to develop a new system with improved usability and interoperability with data analysis tools. The key stages of execution include data categorization from biopsy reports data structure, automatically extracting information using regular expressions, progressive data validation by professionals, and organizing the data in a tabular format. The system features a user-friendly graphical interface, log generation and reporting capabilities. This initiative seeks not only to comply with legal requirements but also to enhance the use of the database for administrative and public health purposes. Currently in validation phase, the project shows promising results.

Palavras-chave: biópsia; planejamento hospitalar; laudo médico; gerenciamento de dados; expressões regulares.

Nome do projeto: Recuperação e digitalização do arquivo de laudos de biópsias do serviço de anatomia patológica e necrópsia do HC-UFMG/Ebserh.

Links: <https://github.com/victor-bartho/gsheets-biopsy>

1. CONTEXTO E MOTIVAÇÃO

De acordo com a Lei nº 13.787/2018, conhecida como Lei do Prontuário Digital, hospitais possuem o dever de armazenar e permitir acesso a prontuários e exames de cada paciente por um período mínimo de 20 anos, que pode ser estendido a *ad infinitum* no caso de judicialização.

Essa conduta tem como finalidade a segurança, continuidade e manutenção do cuidado do paciente. Princípios esses que são ratificados tanto pela Resolução nº 1.821/2007 do Conselho Federal de Medicina quanto pela Lei Geral de Proteção de Dados Pessoais (LGPD - Lei nº 13.709/2018), que ditam requisitos sobre segurança, integridade e acessibilidade de dados.

Nesse contexto, é imperativo que os serviços de saúde mantenham sua infraestrutura de tecnologia da informação atualizada para obedecer a esse requisito ético e legal. Diante disso, surge a necessidade de modernização do sistema de gerenciamento de biópsias do Hospital das Clínicas da Universidade Federal de Minas Gerais (HC-UFMG). Tal necessidade decorre da obsolescência iminente do sistema atual, cuja interface baseada em *Command Line Interface* (CLI) torna a usabilidade hostil e conta com poucos profissionais no hospital capazes de operá-lo. Além disso, a situação é agravada pelo fato de o único profissional do hospital, com perícia no uso do programa, encontrar-se aposentado, e de a empresa proprietária ter encerrado o suporte, impossibilitando atualizações para aumentar a vida útil do recurso. Assim, a transição de sistema de gerenciamento de biópsias torna-se imperativa.

Outro ponto relevante nessa temática é a possibilidade de maior aproveitamento da base de dados de biópsias para *insights* tanto administrativos quanto de saúde pública a partir da análise de dados. Para a concretização disso, é importante a integração com a infraestrutura tecnológica existente, a ver ferramentas de *office*, *business intelligence* e análise estatística. No entanto, o sistema legado, presente desde a década de 80, não oferece interoperabilidade com sistemas de gerenciamento de dados atuais, nem sequer permite a extração em lote de dados para análise em ferramentas externas. Logo, uma transição de sistemas não só resolve a vulnerabilidade jurídica do serviço como também potencialmente beneficia o sistema de saúde a partir de conhecimentos gerados a partir de uma base de dados de um serviço referência de alto volume de dados.

1.1 Objetivo

Elaboração de um sistema de gerenciamento de laudos de biópsias interoperável com ferramentas de análise de dados e com usabilidade amigável.

2. ATIVIDADES PRINCIPAIS

Tendo o objetivo final de elaboração de um novo sistema de gerenciamento de laudos de biópsias em mente, a equipe realizou um estudo piloto com as biópsias hepáticas do hospital, com expansão pós-validação planejada. Essa execução possui dois principais desafios: a migração de dados pré-existentes e o armazenamento adequado de novos laudos de biópsias hepáticas. Assim, as principais atividades do projeto consistem em sistematização e categorização das informações dos laudos, extração das informações e armazenamento em formato tabular, compatível com sistemas de base de dados relacionais, e validação contínua do sistema.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A fim de viabilizar a migração e sistematização dos dados de biópsias hepáticas do HC-UFMG, foi adotada uma abordagem dividida em três grandes frentes: (1) estudo e categorização da estrutura dos laudos existentes, (2) desenvolvimento de um sistema automatizado de extração e estruturação das informações (3) validação progressiva dos dados pela equipe técnica e médica do serviço.

Inicialmente, foi realizado um mapeamento detalhado de laudos retroativos, visando identificar os campos relevantes e sua organização lógica, processo que envolveu reuniões com médicos patologistas e profissionais da área técnica. Em seguida, partiu-se para o desenvolvimento de um sistema de extração dessas informações que tivesse como saída as informações de um arquivo de laudo de biópsia hepática, organizadas e estruturadas em tabela, que passou a alimentar uma base de dados.

Com relação aos aspectos técnicos, a funcionalidade central é baseada no script de extração de dados, capaz de realizar a migração da base de dados pré-existente, enquanto também realiza a inclusão

de novas biópsias. A partir da leitura do arquivo PDF do laudo, é feita a extração de campos específicos (como nome, material da biópsia, número do pedido, entre outros) a partir de expressões regulares (regex). Os dados podem ser formatados conforme demanda para servirem de entrada para qualquer base de dados. A linguagem escolhida para o desenvolvimento foi a Python, visando a legibilidade e facilidade de manutenção por profissionais primariamente da área da saúde. Além disso, tendo em mente que o programa, até o momento atual, não tem previsão de um grande número de usuários simultâneos, escalabilidade e performance, não foi uma preocupação, sendo possível aproveitar o melhor que a linguagem e seu ecossistema tem a oferecer sem sofrer com suas limitações. Por fim, foi escolhido o paradigma orientado a objetos que favorece a modularidade e possibilidade de extensão futura para outros tipos de laudos. O programa conta com funcionalidades adicionais como: interface visual guiada, geração de logs de operação e relatórios de arquivos que requerem revisão manual. Por fim, no que diz respeito a base de dados, inicialmente foi utilizada a estrutura de arquivos "xlsx" em nuvem, via consumo de API (Application Programming Interface) do Google Cloud, mas está prevista nas próximas etapas do projeto a adoção de um banco de dados relacional localmente hospedado.

Atualmente, 7 anos de dados retroativos já foram migrados, com dados atuais também sendo registrados sob o novo sistema, que atua em simultaneidade com o legado na fase de validação. Os próximos passos pretendidos pelo time de planejamento e desenvolvimento do sistema são: melhoria progressiva do módulo extrator de dados, com *feedback* dos funcionários usuários (etapa atual do projeto), adoção de um banco de dados relacional para maior robustez, introdução de grandes modelos de linguagem para incrementação de tags adicionais a partir de interpretação de trechos dos laudos e extensão do módulo a outros tipos de biópsias, totalizando aproximadamente 200.000 unidades ou 38GB de documentos retroativos.

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do projeto, um dos principais desafios foi a tradução mútua entre os dois universos técnicos: a medicina e a computação. Ao entender o fluxo de trabalho do hospital, foi notável que, embora o objetivo fosse manter a funcionalidade do sistema de gerenciamento de documentos do hospital, as formas de enxergar o problema, visualizar soluções e vocabulário técnico das equipes colidiram. O verdadeiro obstáculo era a dificuldade de comunicação entre os profissionais de saúde, focados em raciocínio clínico e fluxos assistenciais, e os profissionais técnicos, focados em infraestrutura computacional. Dessa forma, a presença de um colaborador com vivências em ambos os campos de atuação foi essencial para direcionar e conciliar as capacidades técnicas das duas áreas. Ademais, é importante ressaltar a reduzida equipe de desenvolvimento, o que impôs a necessidade de priorização de etapas essenciais, embora limitasse a velocidade de implementação de novas funcionalidades. Por fim, além de dar oportunidade de aprimorar entendimentos sobre arquitetura e usabilidade de software em um contexto de uso real, o projeto deixou claro que a comunicação interprofissional eficiente entre as áreas assistenciais e tecnológicas é crucial, sendo fator limitante para o desenvolvimento de soluções de tecnologia em saúde, independentemente de capacidade técnica e disponibilidade de tecnologias avançadas.

5. REFERÊNCIAS

1. Link para referências: [Referências](#).

QR Code para referências:



O uso da Inteligência artificial no processo formativo do estudante de medicina: uma revisão da literatura

Enzo Furtado Goulart (aluno)¹, André Felipe de Andrade Santos (aluno)¹, Paula Pereira de Souza (aluno)², Gabriel Silva Soares², Luiza Haikal de Paula² (aluno), Elisa Tuler de Albergaria (professora)¹, Francyne Nascimento Santana¹, Cristiane dos Santos Dias (orientador)¹

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

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

enzo-goulart@ufmg.br, andresants@ufmg.br, paulaps@ufmg.br, gss617@ufmg.br,
luizahaikal@ufmg.br, etuler@gmail.com, francyne@ufmg.br, cristianedias@ufmg.br

Resumo: A Inteligência Artificial (IA) tem sido progressivamente incorporada a Educação Médica. Este estudo explora as principais aplicações da IA no ensino médico, delineando suas vantagens e desvantagens em comparação com metodologias tradicionais. Adicionalmente, são analisadas as percepções dos estudantes em relação ao uso da IA, bem como as limitações e os desafios inerentes a sua integração eficaz no ambiente educacional médico.

Abstract: Artificial Intelligence (AI) has been progressively integrated into Medical Education. This study explores the main applications of AI in medical teaching, outlining its advantages and disadvantages compared to traditional methodologies. Additionally, students' perceptions regarding the use of AI are analyzed, as well as the limitations and challenges inherent in its effective integration into the medical educational environment.

Palavras-chave: Inteligência Artificial; educação médica; ensino médico.

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

Links: <https://ciia-saude.dcc.ufmg.br/>

1. INTRODUÇÃO GERAL

A inteligência artificial (IA) tem transformado tanto a prática clínica quanto a formação médica, por meio de simuladores inteligentes, assistentes virtuais e plataformas adaptativas, que promovem aprendizagem ativa, personalizada e com feedback contínuo ^(2,3).

Além de auxiliar no diagnóstico e prever desfechos clínicos, a IA também apoia estudantes na revisão de conteúdos, na geração de casos clínicos interativos e na simulação de atendimentos com linguagem natural. Esse cenário traz desafios, como a capacitação docente, a atualização curricular, a garantia de infraestrutura digital e a preparação dos alunos para o uso crítico da tecnologia. Questões éticas, como o viés algorítmico e a responsabilidade na tomada de decisão assistida por IA, também merecem reflexões profundas ⁽¹⁾.

2. PANORAMA DAS APLICAÇÕES DA IA NO ENSINO MÉDICO

A incorporação da inteligência artificial (IA) no ensino médico é um tema recorrente na literatura. Estudos indicam que estudantes reconhecem seu potencial como apoio à aprendizagem e à prática clínica, e para transformar áreas como oncologia, radiologia e patologia, embora ainda apresentem familiaridade

limitada com seus recursos ^(3,4,5).

Dentre as principais aplicações, destacam-se os sistemas adaptativos de aprendizagem, tutores inteligentes e simuladores cirúrgicos, além de assistentes baseados em linguagem natural para simulação de interações clínicas ^(6,7).

Esses recursos oferecem *feedback* imediato e permitem treinos personalizados e repetitivos, sendo proveitosos nos currículos e na futura prática da medicina.

3. PERCEPÇÕES DOS ESTUDANTES SOBRE O USO DA IA

É imprescindível avaliar a percepção dos estudantes de medicina sobre o uso da IA no contexto acadêmico. O objetivo é compreender as demandas desse grupo e propiciar um cuidado médico completo aos pacientes.

Dados recentes revelam que 91,5% dos estudantes de medicina consideram o treinamento em IA útil para a prática médica. Contudo, 91,2% dos participantes relataram que suas instituições de ensino não oferecem essa formação ou que desconhecem sua existência ⁽⁴⁾. Em outro estudo, com 4.492 alunos entrevistados, 92,4% não receberam treinamento formal, e 87,1% exibiram um baixo nível de conhecimento ⁽⁵⁾. Ambos os achados evidenciam uma lacuna no ensino médico sobre IA, apesar da percepção positiva dos estudantes quanto à sua importância ⁽⁵⁾. Além disso, a maioria destaca a necessidade de uma abordagem multidisciplinar para esse aprendizado ⁽²⁾.

Em suma, os estudos indicam que, na percepção dos acadêmicos de medicina, o ensino de IA deve focar em suas aplicações clínicas e em métodos inovadores de aprendizado multidisciplinar. Para isso, é necessário um currículo básico e holístico que aborde tanto os aspectos técnicos quanto as aplicações clínicas da IA ⁽³⁾.

4. ESTUDOS COMPARATIVOS: IA x ENSINO TRADICIONAL

Estudos comparativos entre métodos tradicionais e abordagens baseadas em inteligência artificial no ensino médico apresentam resultados variados. Em habilidades técnicas, como cirurgia, simuladores com IA podem mostrar eficácia semelhante à instrução presencial, com a vantagem adicional da escalabilidade ⁽⁷⁾.

No entanto, há limitações claras no uso de pacientes virtuais para simulação clínica. Apesar de auxiliarem no treinamento de habilidades diagnósticas, esses modelos carecem de expressividade e empatia — componentes essenciais da prática médica ⁽⁶⁾. Foram observadas melhoras pontuais no desempenho acadêmico entre estudantes que utilizaram IA como ferramenta auxiliar, embora os efeitos positivos estejam restritos a contextos específicos e ainda careçam de validação ampla ⁽⁸⁾.

Adicionalmente, identificaram-se dificuldades na aceitação da IA por parte dos estudantes, que frequentemente expressam dúvidas sobre a precisão das respostas e a profundidade das interações simuladas ^(3,5).

Diante disso, a IA deve ser vista como um complemento ao ensino tradicional, cuja eficácia depende de sua aplicação crítica, contextualizada e integrada a abordagens centradas na experiência humana.

5. LIMITAÇÕES, DESAFIOS E OPORTUNIDADES

A integração da inteligência artificial (IA) na educação médica ainda enfrenta limitações, como a escassez de treinamento específico sobre o uso da IA durante a formação dos estudantes de medicina ⁽²⁾. Isso leva muitos a buscarem fontes não acadêmicas, como mídias populares, o que pode resultar em uma compreensão superficial do real papel dessas tecnologias ⁽⁴⁾.

Entre os principais desafios da IA na educação médica está a incapacidade de replicar competências humanas, como empatia, julgamento clínico e comunicação ⁽⁶⁾. Há também uma limitada colaboração entre medicina e ciência da computação, além da dificuldade em criar materiais didáticos acessíveis. A implementação da IA tende a se concentrar em contextos de alta renda e com elevada alfabetização di-

gital ^(3,5).

Por outro lado, a IA traz impactos positivos, como a personalização do ensino, com feedback imediato e individualizado, além de práticas ilimitadas em ambientes virtuais a custos reduzidos, como demonstram simuladores cirúrgicos baseados em IA, a exemplo do Virtual Operative Assistant (VOA) ⁽⁷⁾. Além disso, a IA oferece uma oportunidade para o desenvolvimento de currículos mais interdisciplinares, promovendo uma prática médica mais alinhada às demandas contemporâneas ⁽¹⁾.

Portanto, uma integração eficaz da IA requer investimentos em capacitação docente, infraestrutura tecnológica e parcerias interdisciplinares entre instituições de ensino, empresas de tecnologia e sistemas de saúde, assim como uma reflexão ética crítica ^(2,5).

5. CONCLUSÃO

A aplicação da inteligência artificial no ensino médico é promissora, oferecendo recursos inovadores, embora limitada na replicação de competências humanas. O uso da IA entre estudantes cresce e é visto positivamente, mas os currículos ainda negligenciam seus conteúdos e implicações. Portanto, sua integração eficaz requer capacitação docente e atualização curricular, garantindo uma formação tecnológica alinhada à centralidade no cuidado ao paciente.

6. REFERÊNCIAS



Otimização inteligente de estoques em hemocentros: aplicação de algoritmos de aprendizado de máquina

T. O. Lopes ^{1 2 3}

¹Universidade de Brasília (UnB), Brasília, DF, Brasil.

²Fundação Hemocentro de Brasília (FHB), Brasília, DF, Brasil.

³Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), Brasília, DF, Brasil.

thalissonolopes@gmail.com

Resumo: Este trabalho propõe a aplicação de técnicas de machine learning para otimização do estoque de sangue em um hemocentro público brasileiro. Utilizando dados históricos de 2015 a 2020, foram implementados e comparados três modelos de regressão (Ridge, CART e Random Forest) para prever a demanda de bolsas sanguíneas. O Random Forest obteve o melhor desempenho com MSE de 23.66, demonstrando potencial para reduzir desperdícios e escassez no ciclo do sangue. A metodologia incluiu pré-processamento de dados, validação cruzada e busca em grade para ajuste de hiperparâmetros. Os resultados sugerem que abordagens baseadas em aprendizado de máquina podem auxiliar na gestão estratégica de estoques sanguíneos, especialmente em contextos de demanda incerta e perecibilidade do produto.

Palavras-chave: machine learning; gestão de estoques; bancos de sangue; saúde digital.

Nome do projeto: Inteligência Artificial Aplicada a Sistemas de Saúde.

1. CONTEXTO E MOTIVAÇÃO

A gestão de estoques de sangue representa um desafio complexo para os sistemas de saúde, combinando características de demanda incerta, perecibilidade do produto (35-42 dias para concentrado de hemácias) e necessidade de disponibilidade contínua ⁽¹⁾. No Brasil, apenas 1,8% da população doa sangue regularmente, enquanto o desperdício atinge 20% das unidades coletadas ⁽²⁾. Essa dualidade entre escassez e desperdício motivou a aplicação de técnicas de machine learning para otimização do ciclo do sangue. Estudos recentes demonstram o potencial de modelos preditivos na cadeia de suprimentos de saúde ⁽³⁾, porém poucos trabalhos abordam especificamente a realidade dos hemocentros brasileiros. Este trabalho se alinha à Estratégia de Saúde Digital para o Brasil ⁽⁴⁾ ao propor soluções baseadas em dados para problemas críticos do sistema de saúde.

1.1 Objetivo

O objetivo deste estudo é analisar e comparar modelos preditivos para gestão de estoques de sangue, utilizando dados do Hemocentro de Brasília, verificando o desempenho dos respectivos modelos por meio do Mean Squared Error (MSE).

2. ATIVIDADES PRINCIPAIS

As atividades desenvolvidas incluíram:

1. Pré-processamento de dados: Limpeza de dados inconsistentes, normalização de variáveis, gerando um histórico de 24.000 registros (2015-2020).
2. Engenharia de *features*: Seleção das variáveis preditoras (tipo de bolsa sanguínea) e da variável alvo (uso em transfusão).

3. Implementação dos modelos *Ridge Regression*, *CART* e *Random Forest*, realizando validação cruzada (10-*fold* com 30 repetições) e busca em grade para otimização de hiperparâmetros.

4. Avaliação de desempenho, comparação sistemática dos modelos através de métricas de erro e análise de importância de variáveis.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A metodologia adotada seguiu os passos de coleta de dados, pré-processamento dos dados, modelagem, validação dos resultados e testes com os dados, tendo seu fluxo representado na Figura 1

Figura 1 - Cards que apresentam conceitos e Print do vídeo “Reconhecimento de Voz e Acessibilidade com IA”



Aplicando o fluxo citado na Figura 1, temos um resumo dos experimentos trabalhados a partir da base de dados gerada no momento de limpeza dos dados, podendo ser visualizados na Tabela 1, a seguir.

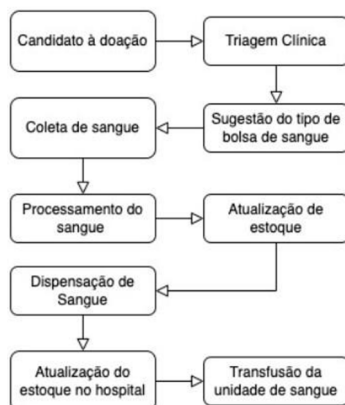
Tabela 1 - Resultados dos experimentos de acordo com os modelos.

| Modelo | MSE | Parâmetros |
|-------------------------|-------|--------------------------------------|
| <i>Ridge Regression</i> | 26.73 | Melhor configuração com 10 iterações |
| CART | 24.17 | Profundidade máxima de 5 |
| <i>Random Forest</i> | 23.66 | 200 árvores com profundidade de 15 |

É fundamental destacar que a implementação do algoritmo *Random Forest* demonstrou uma performance superior significativa, apresentando uma melhoria de 12% em relação ao modelo *Ridge Regression*, evidenciando a eficácia desta abordagem de *ensemble learning* no contexto analisado. Esta diferença percentual representa um avanço considerável na capacidade preditiva do sistema, sugerindo que a natureza não-linear dos dados é melhor capturada pela estrutura baseada em árvores de decisão do *Random Forest*.

Para contextualizar adequadamente a complexidade e a abrangência do cenário operacional em que estes modelos estão inseridos, a Figura 2 apresenta uma representação visual detalhada do *workflow* completo do processo, abrangendo desde as etapas iniciais de coleta e processamento dos dados até o momento crítico da transfusão sanguínea. Esta, permite identificar claramente os pontos estratégicos da aplicação dos modelos preditivos ao longo de toda a cadeia operacional.

Tabela 1 - Resultados dos experimentos de acordo com os modelos.



4. DESAFIOS E APRENDIZADOS

Como principais desafios e aprendizados: Trabalhar com base de dados inconsistente, por ser um sistema antigo, gerando dados incompletos e a necessidade de desenvolver técnicas de imputação para registros com variáveis faltantes. Além disso, a pandemia de COVID-19 causou impacto no dataset gerado, visto que, a exclusão de 2020 do estudo foi necessária, por representar um padrão atípico de demanda. Os aprendizados/desafios técnicos foram interessantes, visto o aprendizado da manipulação do Python para este caso, e, o entendimento da validação cruzada e a otimização dos hiperparâmetros. E, como um grande desafio, a integração com os profissionais de saúde para melhor compreensão dos requisitos clínicos para que os dados fizessem sentido ao dia a dia da instituição.

5. REFERÊNCIAS

1. Hamdan B, Diabat A. A two-stage multi-echelon stochastic blood supply chain problem. *Comput Oper Res.* 2019;101:130-43.
2. Ministério da Saúde. Caderno de informação: sangue e hemoderivados. Brasília: MS; 2021.
3. Abbasi B, et al. Predicting solutions of large-scale optimization problems via machine learning. *Comput Oper Res.* 2020;119:104941.
4. Souza-Zinader JP. A Estratégia da Saúde Digital para o Brasil. *J Health Inform.* 2020;12(4).

PediAI: uma ferramenta para auxiliar estudantes e pediatras no processo de raciocínio clínico e diagnóstico utilizando grandes modelos de linguagem

Mizael A. Rodrigues¹, Noélida Mendes Rodrigues de Almeida¹, Aline Cristina Miranda Marinho Costa¹, Victor Huggo Bartholomeu Gardino¹, Adriano Lages dos Santos², Cristiane Dias¹, Eduardo A. Oliveira²

¹Faculdade de Medicina UFMG, Belo Horizonte, MG, Brasil.

²Instituto Federal de Minas Gerais, Brasil.

mizaelrodrigues20014@gmail.com, noelidamendes@gmail.com, aline12021995@gmail.com, bartholomeuvictor@gmail.com, adrianocomp@gmail.com, profacristianedias@gmail.com, eduolive812@gmail.com

Resumo: A sepse pediátrica é uma emergência de saúde pública, sendo responsável por milhões de mortes por ano em todo o mundo. Um dos grandes desafios no enfrentamento dessa condição é a complexidade do diagnóstico, o que pode comprometer a tomada de decisão clínica. Nesse contexto, a inteligência artificial (IA) desponta como uma ferramenta inovadora e promissora, especialmente por meio do uso dos Grandes Modelos de Linguagem (LLMs), como o ChatGPT. Essas tecnologias têm se mostrado capazes de oferecer suporte relevante aos profissionais de saúde, promovendo maior acurácia diagnóstica, redução de erros médicos e ampliação do acesso ao conhecimento. Diante disso, o presente estudo teve como objetivo principal o desenvolvimento de um sistema de suporte à decisão clínica voltado à Pediatria, baseado em LLMs, com ênfase no diagnóstico rápido e preciso da sepse, visando melhorar os desfechos clínicos e aumentar a sobrevivência dos pacientes pediátricos.

Palavras-chave: inteligência artificial; medicina; pediatria; clínica médica.

Nome do projeto: PediAI: Uma ferramenta para auxiliar estudantes e pediatras no processo de raciocínio clínico e diagnóstico utilizando grandes modelos de linguagem.

1. CONTEXTO E MOTIVAÇÃO

A sepse pediátrica é um problema de saúde pública que causa cerca de 3,3 milhões de mortes anualmente pelo mundo ^[1]. O Escore de Sepse de Phoenix ^[2], criado no Congresso da Society of Critical Care Medicine de 2024, é o mais recente escore relacionado ao diagnóstico de sepse pediátrica e seu uso deve ser usado mundialmente. Esse score baseia-se na avaliação dos sistemas cardiovascular, respiratório, neurológico e de coagulação. Apesar dos avanços fornecidos pelo novo escore, com maior valor preditivo positivo comparado aos critérios anteriores, ainda apresenta limitações. Segundo nota técnica do Instituto Latino Americano da Sepse – ILAS, publicada em maio de 2024 ^[3], o Escore de Sepse de Phoenix não foi criado como ferramenta de triagem precoce, assim, o reconhecimento clínico rápido e preciso da sepse ainda necessita de ferramentas que aumentem a precisão e agilidade do diagnóstico clínico de sepse.

Assim, o diagnóstico médico de sepse é um processo importante para o cuidado à saúde, exigindo expertise, conhecimento médico aprofundado e a capacidade de integrar diversas informações do paciente. Apesar da experiência e do conhecimento dos médicos, o processo de diagnóstico desta doença pode ser desafiador e sujeito a erros, assim como em casos de outras doenças raras ou complexas ^[4,5]. O julgamento clínico é particularmente importante quando exigem discrepâncias entre os achados clínicos e resultados de exames diagnósticos. O diagnóstico precoce é importante, pois sepse é uma emergência médica e seu tratamento deve ser priorizado ^[6]. Os pacientes devem receber antibioticoterapia adequada

o mais rápido possível. Culturas de sangue, bem como outras culturas de locais sob suspeita de infecção, devem ser colhidas na tentativa de detectar o agente causador da doença.

Nesse contexto, a inteligência artificial (IA) surge como uma ferramenta promissora para auxiliar os profissionais de saúde na tomada de decisões clínicas, especialmente no campo do diagnóstico médico. Os grandes modelos de linguagem (LLMs), como Llama, ChatGPT e Gemini, demonstram potencial significativo para processar e analisar grandes volumes de dados médicos, incluindo textos, imagens e registros eletrônicos de saúde, e identificar padrões que podem auxiliar no diagnóstico de doenças [6, 7].

Em estudo publicado em abril de 2024 na revista JAMA Internal Medicine [8], O ChatGPT demonstrou maior acurácia para realizar diagnósticos em consultas em comparação com diagnósticos realizados por clínicos. Foram comparadas as justificativas de diagnósticos feitos por médicos assistentes, estudantes e pelo ChatGPT e todos os diagnósticos realizados pelo LLM obtiveram nota máxima, ou seja, pontuação 10 pela ferramenta r-IDEA [9], enquanto médicos assistentes e residentes obtiveram notas 9 e 8, respectivamente.

Esses estudos demonstram o potencial dos LLMs para auxiliar no diagnóstico médico, contribuindo para a redução de erros diagnósticos. Além disso, são diversos os benefícios de um sistema que utiliza LLMs para realizar diagnóstico médico, principalmente em localidades remotas que contam com equipes de saúde reduzidas. Um sistema desta natureza poderia auxiliar médicos com uma segunda opinião e em caso de encaminhamento para hospitais em grandes centros os pacientes seriam encaminhados já com um relatório diagnóstico maior acurácia. Destaca-se o potencial de sistemas como este como ferramentas valiosas no aprendizado e treinamento de estudantes de medicina de graduação e residentes médicos.

1.1 Objetivo

Desenvolver um sistema de suporte à decisão para diagnóstico médico em Pediatria que utilize Grandes Modelos de Linguagem (LLMs) para auxiliar médicos e outros profissionais de saúde no processo de diagnóstico de sepse, com o objetivo de reduzir o tempo de diagnóstico, aumentar a precisão do diagnóstico e melhorar a chance de sobrevivência dos pacientes.

2. ATIVIDADES PRINCIPAIS

Até o presente momento de desenvolvimento deste trabalho criamos um sistema web com uma interface para que os profissionais de saúde possam inserir dados clínicos e de exames dos pacientes em atendimento. Após a inserção de informações, o sistema através de um LLM (Llama 3) irá apresentar para o profissional uma taxa de classificação de risco que está dividida em quatro níveis: sem risco, baixo, médio risco e alto risco. Além disso, é exibido para o profissional de saúde a explicação para o diagnóstico gerado pela ferramenta de IA. Não há nesse primeiro momento a inserção de texto em linguagem natural. Isso irá ocorrer na segunda etapa do projeto.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Fase de coleta e pré-processamento de dados clínicos: Dados da Sociedade brasileira de pediatria e também da associação americana de cardiologia. Além disso, será incorporado também o UpToDate de sepse pediátrica (<https://www.uptodate.com/contents/sepsis-in-children-definitions-epidemiology-clinical-manifestations-and-diagnosis?csi=2586a422-dfd6-43cd-8c93-c63c403a43ae&source=contentShare>) contendo um sumário com as melhores diretrizes clínicas para médicos para o tópico de sepse pediátrica. O UpToDate é uma base de informações médicas, baseada em evidências e revisada por pares. O material é escrito por mais de 4400 médicos, atuando como autores, editores e revisores. Também iremos incorporar os checklists do protocolo de sepse pediátrica e o fluxograma de tratamento para o referido quadro clínico retirados do ILAS (Instituto Latino Americano de SEPSE) <https://ilas.org.br/materiais-pediatria-e-neo/>. Também podemos adicionar ao treinamento do modelo, casos clínicos anonimizados positivos e negativos criados por um dos membros do projeto.

Definição de protocolo de diagnóstico e conduta: Para protocolo principal de diagnóstico e conduta para treinamento dos LLMs utilizamos o escore de Phoenix para sepse pediátrica que descreve a forma de diagnóstico e conduta para casos do tipo.

Criação de um sistema web como interface para o modelo de linguagem: Nesta etapa criamos uma Sistema web onde o profissional de saúde pode inserir dados clínicos e de exames do paciente e o modelo de linguagem Llama 3 sem nenhum ajuste fornece diagnóstico de sepse de acordo com os dados fornecidos além de explicabilidade sobre o diagnóstico. A Figura 1 mostra a interface explicável da ferramenta desenvolvida utilizando IA para diagnóstico de sepse pediátrica.

Figura 1 - Interface do sistema web de diagnóstico de sepse pediátrica utilizando grande modelo de linguagem.



4. DESAFIOS E APRENDIZADOS

O desenvolvimento de um projeto deste tipo traz aprendizados valiosos para a equipe de iniciação científica do projeto. Desde a fase de consulta a bibliografia até o treinamento e adaptação dos modelos para o diagnóstico de sepse e integração de um modelo com um sistema web traz desafios e novos conhecimentos para a equipe, uma vez que estamos lidando com conhecimentos recentes e com muitas oportunidades de pesquisa multidisciplinar entre computação e medicina.

5. REFERÊNCIAS



Plain Language, IA e saúde digital: construindo pontes para compreender e cuidar

Lucas E. C. Melo¹ (aluno), Marina G. C. Ribeiro¹, Julia G. Bernardo¹, Eric M. F. Martins¹, Juliana V. Q. A. Oliveira², Prof.^a Adriana S. Pagano³, Prof.^a Cristiane S. Dias⁴, Prof. Érico F. Mineiro⁵, Prof.^a Eura M. Lage⁶, Prof.^a Zilma S. N. Reis⁶, Prof. Isaias J. R. Oliveira⁷, Prof. Marco A. A. Costa⁷, Prof.^a Flávia R. de Oliveira⁶ (orientador).

¹ Estudantes de Graduação, Curso de Medicina, Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil. ² Estudante de Pós-graduação em Ciências Aplicadas à Cirurgia e à Oftalmologia (Doutorado), UFMG, Belo Horizonte, MG, Brasil. ³ Diretoria, Faculdade de Letras, UFMG, Belo Horizonte, MG, Brasil. ⁴ Departamento de Pediatria, Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil. ⁵ Departamento de Tecnologia da Arquitetura e Urbanismo, Escola de Arquitetura da UFMG, Belo Horizonte, MG, Brasil. ⁶ Departamento de Ginecologia e Obstetrícia, Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil. ⁷ Centro de Informática, Faculdade de Medicina da UFMG, Belo Horizonte, MG, Brasil.

lucasecm@ufmg.br, marinagcr@ufmg.br, jgb06@ufmg.br, ericmatheus@ufmg.br, julianaoliveiramd@cir.dout.ufmg.br, apagano@ufmg.br, cristianedias@ufmg.br, ericomineiro@ufmg.br, erico.acad@gmail.com, euramartinslage@gmail.com, zilma@ufmg.br, isaias@medicina.ufmg.br, marcocosta@ufmg.br, flaviagob@ufmg.br.

Resumo: O presente trabalho descreve o projeto de extensão universitária “Saúde Digital e Letramento em Saúde: Fortalecendo a Compreensão e o Acesso à Informação na Atenção Primária e Além”, vinculado à Iniciação Científica e ao Ensino Curricular da disciplina Iniciação à Atenção Primária à Saúde II da Faculdade de Medicina da UFMG. A proposta busca promover a equidade na comunicação em saúde por meio da integração entre Plain Language, tecnologias digitais e Inteligência Artificial (IA). Por meio de uma metodologia extensionista participativa, a atuação dos(as) estudantes têm incluído desde a adaptação de conteúdos técnicos para linguagem simples até a criação e execução de atividades interativas com o público em feiras de extensão e ações comunitárias. Entre as estratégias desenvolvidas, destacam-se quiz premiado, jogos com pictogramas, caixas de mensagens educativas, painel de impacto, gravação de depoimentos e divulgação científica por meio do perfil @iapolis no Instagram, promovendo o engajamento digital do público. Os resultados preliminares apontam para engajamento por parte do público, interesse nas dinâmicas propostas e reflexões críticas iniciais por parte dos(as) discentes quanto aos desafios da comunicação em saúde. A experiência evidencia o potencial transformador da articulação entre extensão, ensino e pesquisa na formação médica e na promoção do letramento em saúde no SUS.

Palavras-chave: letramento em saúde; inteligência artificial; linguagem simples; extensão universitária; saúde digital.

Nome do projeto: Saúde Digital e Letramento em Saúde: Fortalecendo a Compreensão e o Acesso à Informação na Atenção Primária e Além (orientador: Flávia Ribeiro de Oliveira).

Links: <https://www.instagram.com/iapolisufmg?igsh=MXhvdDh3ZTJYanY2Zg==>

1. CONTEXTO E MOTIVAÇÃO

A promoção da equidade em saúde exige que a informação seja compreensível, acessível e adaptada às diversas realidades socioculturais da população. No Brasil, índices de letramento em saúde seguem baixos, com implicações na adesão a tratamentos, autocuidado e segurança do paciente ^(1,2). Nesse contexto, a saúde digital e a IA podem oferecer soluções para facilitar a comunicação entre profissionais e usuários do sistema de saúde. No entanto, o uso dessas tecnologias precisa estar ancorado em práticas éticas, seguras e culturalmente adequadas ^(3,4). O projeto surgiu da necessidade de aproximar estudantes de medicina da prática da comunicação acessível, por meio da linguagem simples (Plain Language), da escuta qualificada e do uso responsável da IA em atividades reais de extensão universitária.

Além disso, a iniciativa está articulada ao projeto de pesquisa aprovado pelo Comitê de Ética em Pesquisa da UFMG, intitulado “Desenvolvimento e avaliação de um sistema inteligente de geração de orientações para a prescrição de medicamentos segura, acessível e adaptada a distintos contextos culturais” (CAAE: 78883924.7.0000.5149).

2. ATIVIDADES PRINCIPAIS

A atuação dos(as) estudantes de graduação e pós-graduação contempla:

- Participação do planejamento, da execução e da apresentação de pôster na II Feira de Extensão da Faculdade de Medicina da UFMG, promovendo ações interativas voltadas ao público leigo com foco em saúde digital, letramento em saúde e uso ético de tecnologias.
- Produção de materiais educativos com linguagem simples, pictogramas e recursos digitais, buscando tornar a informação em saúde mais acessível e compreensível.
- Aplicação de instrumentos para avaliação do letramento em saúde (como Test of Functional Health Literacy in Adults – Short Version e Health Literacy Survey - European Questionnaire), incorporando ferramentas digitais e sistematizando dados para aprimorar as estratégias de comunicação.
- Desenvolvimento de conteúdos lúdicos e interativos – como quizzes digitais, jogos com pictogramas e painéis de interação – para estimular a autonomia dos usuários na interpretação de informações em saúde.
- Contribuição para a implementação de ações em escolas e espaços comunitários, promovendo o uso de tecnologias educacionais e inteligência artificial de forma contextualizada e responsiva às realidades locais.
- Participação de análises preliminares da compreensão de mensagens de saúde, incluindo observações qualitativas sobre o uso de soluções tecnológicas na mediação da informação.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

O projeto adota uma abordagem integrada entre ensino, pesquisa e extensão, com foco na comunicação acessível em saúde por meio de linguagem simples, recursos digitais e inteligência artificial (IA). As ações incluem capacitação de estudantes, produção de conteúdos educativos com apoio de IA generativa e aplicação de estratégias interativas em atividades comunitárias e escolares.

Durante a II Feira de Extensão da Faculdade de Medicina da UFMG, os(as) estudantes participaram da execução de jogos, quizzes, painéis e uso de instrumentos de avaliação do letramento em saúde. Por se tratar de uma fase inicial do projeto, os resultados ainda são preliminares, mas indicam boa aceitação do público e relevância das abordagens utilizadas para o fortalecimento do letramento em saúde no SUS.

4. DESAFIOS E APRENDIZADOS

A fase inicial do projeto apresentou desafios como a adaptação de conteúdos técnicos à linguagem simples e a seleção de recursos visuais compreensíveis para diferentes públicos. A utilização ética e res-

responsável da inteligência artificial também exigiu discussões orientadas e acompanhamento constante.

Entre os principais aprendizados dos(as) estudantes, destacam-se o desenvolvimento de habilidades em comunicação em saúde, o reconhecimento das barreiras de letramento enfrentadas pela população e a importância de estratégias educativas que valorizem a escuta e a participação ativa dos usuários. A experiência em atividades práticas com o público contribui para a formação crítica e humanizada, reforçando o papel transformador da extensão universitária.

5. REFERÊNCIAS

1. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health*. 2012;12:80.
2. Moraes KL, Brasil VV, Ribeiro FA. Letramento em saúde e segurança do paciente: revisão integrativa. *Rev Enferm UFPE online*. 2021;15:e245862.
3. World Health Organization. WHO guideline: recommendations on digital interventions for health system strengthening. Geneva: WHO; 2019.
4. Morley J, Luciano F. How to design a governable digital health ecosystem. *The Lancet Digital Health*. 2021;3(7):e425–e427.

Potencialidades da Inteligência Artificial na gestão do cuidado do enfermeiro

Carolina da Silva Caram¹, Lívia Metzker Glória Alves De Deus², João Pedro Tôrres Costa², Isabela Silva Câncio Velloso³; Fernanda Martins Azzi de Carvalho⁴, Kethlin Maia Mariano⁴

¹ Orientadora, Universidade Federal de Minas Gerais, Escola de Enfermagem, Departamento de Enfermagem Aplicada, Belo Horizonte, MG, Brasil.

² Discente de Enfermagem, Universidade Federal de Minas Gerais, Escola de Enfermagem, Belo Horizonte, MG, Brasil.

³ Docente, Universidade Federal de Minas Gerais, Escola de Enfermagem, Departamento de Enfermagem Aplicada, MG, Brasil

⁴ Mestrandas do Programa de Pós-graduação em Enfermagem, Universidade Federal de Minas Gerais, Escola de Enfermagem, Belo Horizonte, MG, Brasil.

caram.carol@gmail.com, liviametzker@outlook.com, pedrojoaocosta2012@hotmail.com, isacancio@gmail.com, fernanda.azzi.carvalho@gmail.com, kethlinmaiam@gmail.com.

Resumo: O avanço das tecnologias digitais, especialmente da Inteligência Artificial (IA), tem provocado transformações na prática da enfermagem, em especial na gestão do cuidado. Este estudo investigou, de forma exploratória e descritiva, ferramentas de IA com potencial para otimizar a gestão do cuidado em enfermagem, promovendo melhoria contínua, segurança do paciente e protagonismo do enfermeiro gestor. A partir da análise documental e temática de conteúdo, foram selecionadas 58 ferramentas específicas na área, organizadas em três categorias principais: planejamento estratégico; qualidade e melhoria contínua; e cuidado ao paciente e gestão de pessoas. Os resultados indicam que essas tecnologias automatizam atividades burocráticas, agilizam a organização de dados assistenciais e possibilitam atuação estratégica do enfermeiro. Contudo, a incorporação da IA enfrenta desafios relacionados à integração dos sistemas, capacitação profissional e aspectos éticos. Destaca-se a importância da alfabetização digital e do desenvolvimento de ambientes de discussão para a construção de práticas seguras, éticas e centradas no cuidado humano, ampliando a eficiência e a qualidade na gestão do cuidado do enfermeiro.

Palavras-chave: Inteligência Artificial; enfermagem; gestão da assistência de enfermagem.

Abstract: The advancement of digital technologies, especially Artificial Intelligence (AI), has transformed nursing practice, particularly in care management. This exploratory and descriptive study investigated AI tools with potential to optimize nursing care management, promoting continuous improvement, patient safety, and nurse leadership. Through documentary analysis and thematic content analysis, 58 specific tools were selected and categorized into three main groups: strategic planning, quality and continuous improvement, and patient care and team management. The results indicate that these technologies automate bureaucratic tasks, streamline the organization of care data, and enable a more strategic role for nurses. However, the integration of AI faces challenges related to system interoperability, professional training, and ethical considerations. Emphasizing digital literacy and the creation of forums for discussion are crucial to developing safe, ethical, and human-centered practices, enhancing efficiency and quality in nursing care management.

Keywords: Artificial Intelligence; nursing; nursing care management.

Nome do projeto: Configuração da Prática de Gestão do Cuidado em Diferentes Cenários de Atuação

do Enfermeiro.

1. CONTEXTO E MOTIVAÇÃO

O avanço das tecnologias digitais e a crescente incorporação da Inteligência Artificial (IA) nos sistemas de saúde têm provocado transformações na prática da enfermagem. A IA é um conjunto de tecnologias capazes de simular funções cognitivas humanas — como análise, aprendizado e tomada de decisão — apresentando-se como ferramenta promissora na reorganização dos processos de trabalho em saúde ⁽¹⁾. Na enfermagem, tais transformações são particularmente relevantes diante da complexidade das atividades e da multifuncionalidade do enfermeiro, cuja prática engloba assistência, gestão, educação, pesquisa e ação política ⁽²⁾. Nesse contexto, a gestão do cuidado emerge como eixo estratégico ao articular planejamento, organização e avaliação das ações de enfermagem com foco na qualidade assistencial ⁽³⁾. Ao desempenhar essa função, o enfermeiro organiza o trabalho e os recursos humanos, objetivando viabilizar condições adequadas tanto para a assistência ao paciente quanto para o desempenho da equipe ⁽⁴⁾. O processo de gestão do cuidado envolve planejamento de recursos, adoção de dispositivos e ferramentas, avaliação contínua da qualidade e reformulação de planos de ação diante de insuficiências ⁽⁵⁾. Considerando a prática da gestão do enfermeiro, este profissional não apenas provê dados para os sistemas inteligentes, mas também interpreta a informação gerada pela IA, posicionando-se como gestor analítico capaz de integrar a tecnologia e o cuidado ⁽¹⁾. Estudos indicam que, longe de substituir o profissional da enfermagem, a IA contribui para ampliar sua eficiência ao reduzir a carga de tarefas burocráticas, ao otimizar o tempo disponível, ao fortalecer o processo de trabalho, ao aperfeiçoar o cuidado prestado à beira-leito e ao elevar a qualidade da assistência. Esses benefícios requerem capacitação contínua para incorporação adequada das tecnologias. Este estudo, inserido na Saúde Digital, investiga ferramentas de IA que otimizam a gestão do cuidado, promovendo qualidade, segurança e protagonismo do enfermeiro.

1.1 Objetivo

Analisar ferramentas de inteligência artificial com potencial de aplicação na gestão do cuidado em enfermagem.

2. ATIVIDADES PRINCIPAIS

As atividades principais atreladas a este projeto de pesquisa foram: Formulação do problema de pesquisa e objetivos; Levantamento e seleção de dados na base “There’s an AI for that”; Aplicação de critérios de inclusão e exclusão para ferramentas voltadas à enfermagem; Organização das ferramentas em planilhas com informações detalhadas; Classificação das ferramentas em áreas de atuação e formas de aplicação; Categorização temática com base na análise de conteúdo; Produção de fluxograma (PRISMA-ScR) e gráficos descritivos e Redação do trabalho e análise crítica fundamentada na literatura.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Esta pesquisa utilizou a abordagem qualitativa voltada à investigação documental, exploratória e descritiva. Foi possível, de forma ampliada e precisa, identificar, analisar e compreender conceitos e funcionalidades acerca da IA relacionada à enfermagem. Os dados foram coletados em janeiro de 2024, a partir do repositório “There’s an AI for that” (ferramenta gratuita e de livre acesso), criada em dezembro de 2022 por um desenvolvedor independente, que rapidamente se tornou conhecida entre os usuários de internet, alcançando mais de 100 mil visitas já na primeira semana de lançamento, sendo possível identificar e selecionar, diante dos documentos, as informações relacionadas à IA no contexto da Enfermagem. Foram aplicados os seguintes critérios de inclusão: a ferramenta deveria estar relacionada à gestão do cuidado de enfermagem e possuir ao menos uma IA listada dentro da task. A análise dos dados seguiu os preceitos da Análise Temática de Conteúdo (ATC), conforme os polos cronológicos descritos por Bar-

din(5). Iniciou-se com a leitura e organização das funcionalidades das ferramentas de IA, transcritas no Excel. Em seguida, as informações foram codificadas e categorizadas manualmente de acordo com sua aplicabilidade na gestão do cuidado em enfermagem. Por fim, os dados foram interpretados à luz da literatura científica. Este estudo utilizou dados provenientes de artigos e documentos disponibilizados de maneira pública e gratuita em meios de comunicação e bancos de dados, dispensando a aprovação do Comitê de Ética em Pesquisa. Identificou-se na base de dados "There's an AI for that" 10.700 IAs, após filtrar para a profissão "Nurse Manager", foram encontradas 900 e, após aplicar os critérios de inclusão a amostra final foi composta por 58 ferramentas. As 58 ferramentas foram organizadas em áreas de atuação e possíveis aplicações na gestão do cuidado em enfermagem. Essas tecnologias foram agrupadas em três grandes categorias: 1 - Planejamento Estratégico, que envolve IAs voltadas para a organização de escalas, gestão de tempo, cronogramas e planos de cuidado; 2 - Qualidade e Melhoria Contínua, com ferramentas voltadas à avaliação de indicadores, construção de relatórios e análise de dados; e 3 - Cuidado ao Paciente e Gestão de Pessoas, com ênfase na comunicação, gestão de equipe, prontuários inteligentes e suporte à tomada de decisão clínica. Os resultados preliminares evidenciam que, ao automatizar atividades burocráticas e organizar dados assistenciais de forma mais ágil e estruturada, essas ferramentas contribuem para atuação estratégica do enfermeiro na gestão do cuidado. Além disso, demonstram potencial para apoiar processos como auditoria interna, educação permanente, coordenação de equipes e monitoramento de metas assistenciais, reforçando a importância da alfabetização digital como competência essencial para o enfermeiro contemporâneo.

4. DESAFIOS E APRENDIZADOS

A aplicação da IA na saúde tem sido reconhecida por seu potencial em qualificar decisões e otimizar processos. Contudo, sua implementação na enfermagem enfrenta obstáculos como a fragmentação dos sistemas, a escassez de capacitação técnica e a falta de regulamentações éticas. Frente a esses desafios é importante discutir a ética, a segurança e a efetividade no uso da IA valorizando o protagonismo e a autonomia do enfermeiro. O desenvolvimento de soluções inovadoras deve integrar a tecnologia à prática, promovendo cuidado resolutivo, humano e baseado em evidências. Investir em formação crítica e tecnológica da equipe é fundamental para a construção de práticas sustentáveis e centradas no paciente.

5. REFERÊNCIAS

1. Robert N. How artificial intelligence is changing nursing. Nurs Manag (Springhouse) [Internet]. Set 2019 [citado 27 maio 2025];50(9):30-9. Disponível em: <https://doi.org/10.1097/01.numa.0000578988.56622.21>.
2. Sanna MC. Os processos de trabalho em Enfermagem. Rev Bras Enferm [Internet]. Abr 2007 [citado 27 maio 2025];60(2):221-4. Disponível em: <https://doi.org/10.1590/s0034-71672007000200018>
3. Mororó DD, Enders BC, Lira AL, Silva CM, Menezes RM. Análise conceitual da gestão do cuidado em enfermagem no âmbito hospitalar. Acta Paul Enferm [Internet]. Maio 2017 [citado 30 maio 2025];30(3):323-32. Disponível em: <https://doi.org/10.1590/1982-0194201700043>.
4. McGrow K. Artificial intelligence. Nursing [Internet]. Set 2019 [citado 30 maio 2025];49(9):46-9. Disponível em: <https://doi.org/10.1097/01.nurse.0000577716.57052.8d>.
5. Bardin L. Análise de conteúdo. São Paulo: Edições 70; 2011.

Predição de complicações cardiovasculares em pacientes com covid-19 hospitalizados utilizando modelos supervisionados de aprendizado de máquina e técnica de inteligência artificial explicável.

Isabella Viana Gomes Schettini¹, Guilherme Fonseca do Nascimento¹, Leonardo Chaves Dutra da Rocha², Fernanda Cristina Barbosa Lana¹, Luciano Romero Soares de Lima¹ Marcos André Gonçalves¹, Milena Soriano Marcolino¹

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

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

{silvaivg, guilhermefonseca8426, lcrocha,fernandacblana, magoncalv, milenamarc}
@gmail.com, luciano@dcc.ufmg.br

Resumo: Este estudo de coorte retrospectivo multicêntrico incluiu 10.700 pacientes adultos hospitalizados com covid-19 entre março/2021 e agosto/2022, em 25 hospitais. O desfecho composto abrangeu arritmia, insuficiência cardíaca aguda, infarto agudo do miocárdio, miocardite e pericardite. Dois modelos foram desenvolvidos com o algoritmo Light Gradient-Boosting Machine: modelo 1 com 59 variáveis (demográficas, clínicas, laboratoriais e socioeconômicas) e modelo 2 com 52 variáveis (demográficas, clínicas, laboratoriais e socioeconômicas). Ambos apresentaram área sob a curva Receiver Operating Characteristic moderada (0,752 e 0,760) e alta acurácia (94,6% e 94,5%). No entanto, o desequilíbrio entre as classes (5,3% com eventos) comprometeu a predição de desfechos, resultando em baixos valores de Macro-F1 (51,2% e 50,7%) e F1 (5,2% e 4,2%) para a classe minoritária. A análise Shapley Additive Explanation indicou idade, ureia, plaquetas e razão entre saturação de oxigênio e fração de oxigênio inspirado como preditores relevantes. Apesar da alta acurácia, os modelos tiveram baixa sensibilidade para eventos cardiovasculares, reforçando a necessidade de lidar com o desequilíbrio entre classes além de adotar métricas de avaliação não-enviesadas que considerem a efetividade em ambas as classes consideradas (desfecho e não-desfecho).

Abstract: This multicenter retrospective cohort study included 10,700 adult patients hospitalized with covid-19 between March 2021 and August 2022 across 25 hospitals. The composite outcome included arrhythmia, acute heart failure, acute myocardial infarction, myocarditis, and pericarditis. Two predictive models were developed using the Light Gradient Boosting Machine (LightGBM) algorithm: Model 1 incorporated 59 variables (demographic, clinical, laboratory, and socioeconomic), and Model 2 included 52 variables (demographic, clinical, and laboratory). Both models demonstrated moderate discrimination, as indicated by the area under the receiver operating characteristic curve (AUC) of 0.752 and 0.760, respectively, and high accuracy (94.6% and 94.5%, respectively). However, class imbalance (5.3% with cardiovascular events) hindered the prediction of outcomes, resulting in low macro-F1 scores (51.2% and 50.7%) and F1 scores for the minority class (5.2% and 4.2%). Shapley Additive Explanation (SHAP) analysis identified age, urea, platelet count, and the ratio of oxygen saturation to inspired oxygen fraction as relevant predictors. Despite the high accuracy, the models had low sensitivity for cardiovascular events, underscoring the need to address the class imbalance in addition to adopting unbiased evaluation metrics that consider the effectiveness in both classes (outcome and non-outcome).

Palavras-chave: Inteligência artificial; covid-19; doenças cardiovasculares; modelos preditivos.

Nome do projeto: Predição de desfechos clínicos e econômicos por meio de representações semânticas multi-modais de pacientes resilientes a drifts temporais.

1. CONTEXTO E MOTIVAÇÃO

A pandemia de covid-19 impactou intensamente os sistemas de saúde globais, não apenas pelas complicações respiratórias, mas também pelos efeitos significativos no sistema cardiovascular, com maior risco de complicações como lesão miocárdica, arritmias e eventos tromboembólicos, especialmente em pessoas com doenças cardiovasculares pré-existentes. Em regiões como a América Latina e o Caribe, esse cenário foi agravado pela alta prevalência de comorbidades e pelas desigualdades no acesso aos serviços de saúde^[1-2].

Compreender a interação entre fatores clínicos e sociais no risco de complicações cardiovasculares torna-se fundamental. Investigações que explorem essas associações podem contribuir para o desenvolvimento de estratégias de identificação precoce e intervenções mais eficazes, capazes de reduzir a carga cardiovascular imposta pela covid-19.

1.1 Objetivo

Desenvolver modelos de inteligência artificial (IA) explicável para identificar preditores de eventos cardiovasculares em pacientes hospitalizados com covid-19.

2. METODOLOGIA E RESULTADOS

Trata-se de um estudo de coorte retrospectivo, envolvendo 10.700 pacientes adultos hospitalizados com diagnóstico laboratorial confirmado de covid-19 [3], internados entre março/2021 e agosto/2022 em 25 hospitais brasileiros. O estudo foi aprovado pela Comissão Nacional de Ética em Pesquisa (CAAE: 30350820.5.1001.0008). O desfecho composto considerou a ocorrência de arritmia, insuficiência cardíaca aguda, infarto agudo do miocárdio, miocardite e pericardite. A base de dados apresenta um desequilíbrio significativo entre as classes, com 10.129 indivíduos (94,7%) sem eventos cardiovasculares e apenas 571 (5,3%) com ao menos um evento.

Foram desenvolvidos dois modelos preditivos com o algoritmo *Light Gradient Boosting Machine* (LightGBM). O modelo 1 incorporou 59 variáveis, incluindo informações demográficas, clínicas, laboratoriais e socioeconômicas. O modelo 2 utilizou as mesmas variáveis, com exceção das socioeconômicas, totalizando 52 variáveis. O objetivo da construção dos modelos foi avaliar a contribuição das variáveis socioeconômicas na capacidade preditiva de eventos cardiovasculares. O desempenho foi avaliado por meio de um processo de validação cruzada de 5 partes, considerando as métricas de acurácia, área sob a curva *Receiver Operating Characteristic* (AUROC), Macro-F1, bem como precisão, sensibilidade e F1-score por classe, acompanhadas de seus respectivos intervalos de confiança. Para verificar a existência de diferenças estatisticamente significativas entre os desempenhos dos modelos 1 e 2, foi aplicado o teste t de Student.

Na análise dos resultados, ambos os modelos apresentaram valores moderados de AUROC: 0,752 (0,027) para o modelo 1 e 0,760 (0,026) para o modelo 2, além de elevadas acurácias, 94,6% (0,1%) e 94,5% (0,1%), respectivamente. Contudo, o desbalanceamento entre as classes comprometeu a capacidade de predição do desfecho composto, como evidenciado pelos baixos valores de Macro-F1: 51,2% (1,4%) para o Modelo 1 e 50,7% (1,5%) para o Modelo 2. Para a classe majoritária (ausência de evento cardiovascular), ambos os modelos demonstraram desempenho excelente, com F1-score de 97,2% (0,1%). Em contraste, o desempenho para a classe minoritária (presença de evento cardiovascular) foi significativamente inferior, com F1-scores de 5,2% (2,7%) e 4,2% (2,9%) para os modelos 1 e 2, respectivamente, evidenciando a limitação dos modelos em identificar corretamente esses casos. Adicionalmente, não foram observadas diferenças estatisticamente significativas em Micro-F1 ($p = 0,3821$) e Macro-F1 ($p = 0,4531$) entre os modelos. Esses achados reforçam a importância da escolha criteriosa das métricas de avaliação em estudos preditivos na área da saúde, especialmente em contextos marcados por forte desbalanceamento entre as classes, como é comum em bases de dados clínicas. Por fim, observa-se que, preliminarmente, a inclusão de variáveis socioeconômicas não contribuiu para a melhoria da performance preditiva dos mo-

delos. Para corroborar esses resultados, foi realizada uma análise de importância das variáveis com base nos valores de *Shapley Additive Explanations* (SHAP) [4], extraídos do modelo LightGBM. Os principais preditores identificados para o desfecho composto foram idade, ureia, contagem de plaquetas e a razão entre saturação de oxigênio e fração de oxigênio inspirado ($\text{SatO}_2/\text{FiO}_2$).

3. DESAFIOS E APRENDIZADOS

O principal desafio enfrentado neste estudo foi o acentuado desbalanceamento entre as classes, decorrente da baixa proporção de pacientes que apresentaram complicações cardiovasculares. Essa assimetria dificultou a capacidade dos modelos em identificar corretamente os casos positivos, afetando, sobretudo, sua sensibilidade. Como principais aprendizados, destaca-se o potencial dos modelos de aprendizado de máquina como ferramentas de apoio à estratificação de risco clínico, bem como a relevância de métodos de interpretabilidade, como o *Shapley Additive Explanations* (SHAP), que contribuem para a transparência e a aplicabilidade dos modelos em contextos assistenciais.

Ainda que as variáveis socioeconômicas apresentem reconhecida importância conceitual, a semelhança de desempenho entre os modelos avaliados sugere que sua exclusão pode ser viável, especialmente em aplicações clínicas que demandam maior simplicidade e operacionalização.

Apesar do desempenho preditivo global satisfatório, ambos os modelos apresentaram sensibilidade limitada para a detecção de eventos cardiovasculares em pacientes hospitalizados com covid-19. Embora os valores de AUROC tenham indicado capacidade discriminativa aceitável, não refletiram plenamente as limitações dos modelos na identificação dos casos positivos, conforme evidenciado pelos baixos valores de F1-score para a classe minoritária. A análise de importância das variáveis indicou que idade, níveis de ureia, contagem de plaquetas e a razão entre a saturação de oxigênio e a fração de oxigênio inspirado ($\text{SatO}_2/\text{FiO}_2$) foram os preditores mais relevantes para o desfecho, reforçando a relevância clínica desses marcadores na avaliação do risco cardiovascular em pacientes com covid-19.

Agradecimentos. Ministério da Saúde, Secretaria de Atenção Primária à Saúde (TED 31/2022, Processo 25000.058753/2022-59), Instituto de Avaliação de Tecnologias para saúde, CNPq, FAPEMIG e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação - INCT-TILD-IAR. FAPESP, Unimed-BH, CIIA-Saúde.

4. REFERÊNCIAS

1. Krishna BA, Metaxaki M, Sithole N, Landín P, Martín P, Salinas-Bostrán A. Cardiovascular disease and covid-19: A systematic review. *Int J Cardiol Heart Vasc.* 2024; 54; 1-7.
2. Herrera CA, Kerr AC, Eberwein JD, Bedregal P, Kringos D, Klazinga N. Navigating the divide: inequalities in household experiences of healthcare disruption in Latin American and the Caribbean countries amidst COVID-19. *Int J Equity Health.* 2024;23(1).
4. World Health Organization. Diagnostic testing for SARS-CoV-2. Interim guidance. Set 2020. <https://www.icao.int/safety/CAPSCA/COVID19Docs/WHO-2019-nCoV-laboratory-2020.6-eng.pdf>
5. Chen H, Janizek JD, Lundberg S, Lee SI. True to the Model or True to the Data? 2020 Jun 29 [cited 2025 Mar 17]; Available from: <http://arxiv.org/abs/2006.16234>

Predição de mutações oncogênicas em genes reguladores associados ao câncer através de Redes Neurais de Grafos

Lucas Moraes dos Santos^{1,2}, Tatiane Senna Bialves^{1,2}, Luana Luiza Bastos^{1,2}, Diego Mariano^{1,2}, Adenilson Arcanjo de Moura Junior^{1,2}, Sheila Cruz Araújo^{1,2}, Leonardo Silva^{1,2}, Márcia Bigão^{1,2}, Milenna Pirovani^{1,2}, Ana Luisa Araújo Bastos^{1,2}, Raquel Cardoso de Melo-Minardi^{1,2}

¹Departamento de Ciência da Computação, UFMG, Belo Horizonte, MG, Brasil.

²Programa de Pós-Graduação em Bioinformática, UFMG, Belo Horizonte, MG, Brasil.

{lucas.santos, raquelcm}@dcc.ufmg.br

Resumo: A compreensão funcional de regiões proteicas intrinsecamente desordenadas (IDRs) e de seu papel em patologias como o câncer permanece um desafio, apesar dos recentes avanços em bioinformática estrutural. Nesse contexto, hipotetizamos que motivos estruturais específicos nessas regiões, como hélices de poliprolina tipo II, modulam a tolerância a mutações associadas à carcinogênese. Neste estudo, desenvolvemos um modelo preditivo baseado em aprendizado profundo para classificar variantes missense como potencialmente condutoras ou não condutoras de câncer. A partir do COSMIC, um banco de dados de mutações somáticas em câncer humano, desenvolvemos um conjunto de dados balanceados com 205 mutações condutoras e um número equitativo de variantes não condutoras em 45 genes reguladores. Após a predição das estruturas tridimensionais com AlphaFold2 (pLDDT médio de 81,74), cada proteína foi representada como um grafo que integra informações sequenciais e espaciais, permitindo a classificação por meio de uma Rede Neural de Grafos (GNN). Em um conjunto de teste independente, o modelo demonstrou desempenho robusto, com acurácia de 79,1%, precisão de 0,80 e F1-score de 0,79.

Palavras-chave: câncer; aprendizado de máquina; regiões intrinsecamente desordenadas.

Nome do projeto: Modelos multimodais na busca e compreensão de padrões de sequência-estrutura-função relacionados ao desenvolvimento do câncer.

1. CONTEXTO E MOTIVAÇÃO

Avanços recentes em bioinformática estrutural, impulsionados pela rápida expansão de dados ômicos (1), têm elucidado os mecanismos de enovelamento proteico (2-4). Nesse contexto, o sequenciamento genômico de nova geração tem sido fundamental para aprofundar a compreensão do câncer. Contudo, ainda existem lacunas no entendimento dessas macromoléculas, sobretudo em suas regiões intrinsecamente desordenadas (IDRs) (5).

Apesar de representarem cerca de 30% dos proteomas e desempenharem papéis funcionais relevantes, as IDRs não adotam uma conformação tridimensional (3D) única; em vez disso, existem como um conjunto de conformações dinâmicas (6). Estudos recentes demonstram que as hélices de poliprolina tipo II são componentes estruturais ubíquos nas IDRs. A abundância de prolina nessas regiões favorece os ângulos torsionais característicos dessa estrutura (7). Essas hélices têm sido associadas a mutações missense em genes reguladores, algumas das quais, conhecidas como condutoras, estão envolvidas no desenvolvimento do câncer (8).

Neste sentido, a literatura sugere que padrões de sequência e propriedades estruturais dessas regiões modulam sua elevada flexibilidade conformacional e, conseqüentemente, sua maior tolerância ao acúmulo de mutações (9,10). Essas alterações, ao comprometerem a função e a estabilidade proteica, estariam implicadas no processo de carcinogênese. Assim, hipotetizamos que a aplicação de estratégias integradas baseadas em inteligência artificial (IA) emerge como uma abordagem promissora para a identificação e caracterização de mutações clinicamente relevantes em diversas doenças, incluindo o câncer.

De acordo com Green *et al.* (11), a impraticabilidade de testar experimentalmente todas as variantes

genéticas em múltiplos contextos biológicos torna o aprendizado de máquina uma alternativa célere e essencial à triagem empírica. A utilização dessas abordagens na análise de elementos estruturais em IDRs, como as hélices de poliprolina do tipo II, pode fornecer um esquema inicial para a compreensão mecanicista subjacente ao desenvolvimento de tumores⁽¹²⁾. Esse entendimento é, por sua vez, fundamental para o desenho racional de novas terapias, como fármacos que visam a estabilização estrutural ou atuam como inibidores moleculares.

1.1 Objetivo

Desenvolvimento de um modelo baseado em aprendizado profundo para prever a oncogenicidade de variantes missense em genes reguladores, com ênfase em como motivos nas regiões intrinsecamente desordenadas se associam a mutações oncogênicas.

2. METODOLOGIA E RESULTADOS PRELIMINARES

Para o desenvolvimento de um repositório curado e abrangente sobre o espectro de variantes no contexto oncológico, utilizou-se o Catálogo de Mutações Somáticas no Câncer (COSMIC) (13). Inicialmente, procedeu-se à identificação de variantes genéticas condutoras em múltiplos tipos de câncer, a partir do conjunto de dados do Censo de Mutações do Câncer (CMC) do COSMIC (v100), alinhado ao genoma de referência GRCh37. Com base nesses dados, foram conduzidas análises sobre as frequências e distribuições de mutações condutoras e não condutoras, considerando genes classificados nos diferentes níveis do Censo de Genes do Câncer (CGC) e a significância das mutações conforme os padrões da *Human Genome Variation Society (HGVS)*.

A análise dos dados revelou um aumento de 13,17% nas mutações condutoras e de 15,56% nos genes de Nível 1 do CGC, em comparação aos achados de Meirson et al. (8). Dentre os genes mais frequentemente mutados, destacaram-se PIK3CA, NRAS e CTNNB1, com 20, 17 e 15 mutações condutoras, respectivamente. Adicionalmente, constatou-se um esperado e significativo desbalanceamento de classes entre as instâncias condutoras (232) e não condutoras (73.761). A fim de mitigar a disparidade de classes, aplicou-se uma subamostragem estratificada ao conjunto de não-condutoras para obter um conjunto de dados balanceado (1:1), no qual a representatividade das instâncias manteve-se proporcional à frequência de mutações por gene.

O conjunto de dados inicial compreendeu 232 mutações condutoras do tipo missense, representando 73% de todos os tumores sequenciados e catalogados no COSMIC. Essas mutações foram mapeadas para as sequências canônicas em formato FASTA, obtidas do UniProt (14). Após a exclusão de 11,63% das instâncias – devido a falhas de mapeamento ou posições de mutação fora do intervalo descrito no banco de dados – o conjunto de dados para modelagem passou a ser composto por 205 mutações condutoras, distribuídas em 45 genes reguladores.

As estruturas 3D para cada variante foram previstas utilizando o AlphaFold2 (v2.3.0)(3) e a qualidade das modelagens foi avaliada através da métrica pLDDT (predicted Local Distance Difference Test). Os modelos 3D de melhor classificação demonstraram uma alta confiança preditiva global, com um pLDDT médio de 81,74. Destes, 74 apresentaram um nível de confiança comparável ao de estruturas resolvidas experimentalmente por difração de raios-X ou outras técnicas (pLDDT > 90,00) e outros 89 exibiram boa confiança (70,00 < pLDDT ≤ 90,00).

Cada proteína foi representada como um grafo, no qual os nós (resíduos de aminoácidos) contêm características definidas por vetores de incorporação do modelo de linguagem de proteínas ESM-2 (2), os quais codificam um conjunto de propriedades evolutivas, de conservação e físico-químicas locais. As relações espaciais e interações de longo alcance, constituem as arestas cujas geometrias foram inferidas a partir das coordenadas 3D previstas pelo AlphaFold2. Essa representação, baseada na multimodalidade de informações sequenciais e espaciais, serve como entrada para uma Rede Neural de Grafos (GNN), treinada para processar a topologia e as características dos grafos, visando classificar as mutações como condutoras ou não condutoras. Para a otimização dos hiperparâmetros¹, utilizou-se a técnica de validação cruzada k-fold (k = 5), resultando em uma acurácia média de 80,4% ± 11,5%. A variância observada indica

uma sensibilidade do modelo às diferentes partições dos dados, possivelmente decorrente da heterogeneidade intrínseca dos grafos, como variações no número de nós e na densidade das arestas (i.e., o número de interações químicas). Contudo, na avaliação subsequente realizada em um conjunto de teste independente (*hold-out*), o modelo selecionado na validação demonstrou um desempenho robusto, com acurácia de 79,1%, precisão de 0,80, recall de 0,79 e F1-score de 0,79 (loss: 0,377). Esses resultados demonstram a eficácia da abordagem e corroboram a importância da conformação estrutural das IDRs como fator relevante na predição de mutações oncogênicas.

3. DESAFIOS E APRENDIZADOS

Este trabalho abordou desafios metodológicos relevantes, notadamente a heterogeneidade inerente aos grafos — em termos de dimensionalidade (número de resíduos de aminoácidos / nós) e topologia (densidade de interações / arestas) — e o significativo desbalanceamento de classes entre mutações condutoras e não condutoras. Em contrapartida, demonstrou-se grande eficácia na integração de informações de modalidades distintas (i.e., sequenciais e espaciais).

4. REFERÊNCIAS

1. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform.* 2017;18(5):851–69.
2. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* 17 de março de 2023;379(6637):1123–30.
3. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* agosto de 2021;596(7873):583–9.
4. Chronowska M, Stam MJ, Woolfson DN, Di Costanzo LF, Wood CW. The Protein Design Archive (PDA): insights from 40 years of protein design. *Nat Biotechnol.* 2025;43(5):669–71.
5. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 7 de janeiro de 2022;50(D1):D439–44.
6. Lotthammer JM, Ginell GM, Griffith D, Emenecker RJ, Holehouse AS. Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat Methods.* março de 2024;21(3):465–76.
7. Stapley BJ, Creamer TP. A survey of left-handed polyproline II helices. *Protein Sci.* 1999;8(3):587–95.
8. Meirson T, Bomze D, Schueler-Furman O, Stemmer SM, Markel G. Systemic structural analysis of alterations reveals a common structural basis of driver mutations in cancer. *NAR Cancer.* 1 de março de 2023;5(1):zac040.
9. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* janeiro de 2015;16(1):18–29.
10. Chillón-Pino D, Badonyi M, Semple CA, Marsh JA. Protein structural context of cancer mutations reveals molecular mechanisms and candidate driver genes. *Cell Rep.* 26 de novembro de 2024;43(11):114905.
11. Green ED, Gunter C, Biesecker LG, Di Francesco V, Easter CL, Feingold EA, et al. Strategic vision for improving human health at The Forefront of Genomics. *Nature.* outubro de 2020;586(7831):683–92.
12. Meirson T, Bomze D, Schueler-Furman O, Stemmer SM, Markel G. Systemic structural analysis of alterations reveals a common structural basis of driver mutations in cancer. *NAR Cancer.* 1 de março de 2023;5(1):zac040.
13. Sondka Z, Dhir NB, Carvalho-Silva D, Jupe S, Madhumita, McLaren K, et al. COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res.* 5 de janeiro de 2024;52(D1):D1210–7.
14. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Adesina A, et al. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* 2025;(D1):D609–17.

¹ A implementação do modelo foi desenvolvida em Python, utilizando a biblioteca PyTorch (versão 2.5.1 com suporte a CUDA 12.6). O treinamento foi conduzido em uma GPU NVIDIA A100 com 80GB de VRAM.

Predição de sobrevida de curto prazo em pacientes oncológicos na emergência utilizando aprendizado de máquina

Sebastian N. Geiger¹, Gregório Victor Rodrigues², Maria Dele Pilar Esteves Diz¹,
Fauze Maluf-Filho²

¹ICESP, São Paulo, SP, Brasil.

²Faculdade de Medicina da Universidade de Minas Gerais, Minas Gerais, MG, Brasil.

contato@patients2python.com.br, sebastian.geiger@hc.fm.usp.br

Resumo: A identificação precoce de pacientes com câncer com sobrevida inferior a seis meses durante visitas à emergência pode facilitar a transição oportuna para os cuidados paliativos (CP). Neste trabalho, aplicamos um modelo de aprendizado de máquina baseado em registros clínicos eletrônicos, utilizando dados de 51.311 visitas ao pronto-socorro no Hospital X (2018–2019). O modelo Gradient Boost Machine (GBM) apresentou desempenho superior (AUC de 0,973 no treinamento e 0,823 na validação cruzada). As principais variáveis preditoras incluíram o escore ECOG e o diagnóstico oncológico. O modelo demonstrou 97,57% de especificidade e 80,36% de sensibilidade na detecção de pacientes com sobrevida de curto prazo. Uma análise de interpretabilidade do modelo (SHAP) revelou a direção e magnitude do impacto de cada variável no desfecho. Essa abordagem possibilita um suporte à decisão clínica na emergência, otimizando recursos clínicos e promovendo intervenções precoces de CP.

Abstract: Early identification of cancer patients with survival times of less than six months during emergency department visits can facilitate timely transition to palliative care (PC). In this work, we applied a machine learning model based on electronic clinical records, using data from 51,311 emergency department visits at Hospital X (2018–2019). The Gradient Boost Machine (GBM) model demonstrated superior performance (AUC of 0.973 in training and 0.823 in cross-validation). The main predictor variables included ECOG score and oncologic diagnosis. The model demonstrated 97.57% specificity and 80.36% sensitivity in detecting patients with short-term survival. A model interpretability analysis (SHAP) revealed the direction and magnitude of each variable's impact on the outcome. This approach enables clinical decision support in the emergency department, optimizing clinical resources and promoting early interventions in PC.

Palavras-chave: aprendizado de máquina; cuidados paliativos; emergência; câncer; sobrevida; predição.

Nome do projeto: [não se aplica].

Links (opcional): <https://app-predicaopaliativo.patients2python.com.br/>

1. CONTEXTO E MOTIVAÇÃO

A transição para cuidados paliativos é frequentemente adiada por estimativas excessivamente otimistas de sobrevida [1–3]. Visitas ao pronto-socorro representam oportunidades chave para esse encaminhamento, especialmente em pacientes com sobrevida inferior a seis meses [4–6]. No entanto, a triagem sistemática de pacientes elegíveis no contexto da emergência é desafiadora. As ferramentas existentes de predição de sobrevida, como questionários, frequentemente apresentam baixa precisão [7] e demandam tempo [8], o que as torna impraticáveis para uso rotineiro sem sobrecarregar a equipe clínica. Modelos de aprendizado de máquina baseados em prontuários eletrônicos têm se mostrado eficazes na predição de mortalidade em pacientes oncológicos [9,10], mas poucos utilizam dados disponíveis em tempo real na admissão hospitalar [11].

1.1 Objetivo

Este estudo teve como objetivo aplicar e validar modelos de aprendizado de máquina para prever a sobrevida de curto prazo (<6 meses) em pacientes oncológicos na admissão da emergência, utilizando dados disponíveis em tempo real de prontuários eletrônicos de um centro oncológico terciário. O foco foi no desenvolvimento de uma ferramenta confiável e clinicamente aplicável, cuja confiabilidade foi avaliada por meio de

validação cruzada rigorosa e métricas de desempenho balanceadas, com o intuito de aprimorar a tomada de decisão clínica e melhorar a qualidade dos cuidados de fim de vida para pacientes com câncer.

2. ATIVIDADES PRINCIPAIS

Todas as etapas do projeto foram realizadas pelos autores do estudo, sob orientação e supervisão da equipe responsável: definição da coorte, extração e pré-processamento dos dados dos prontuários eletrônicos, desenvolvimento dos modelos usando a biblioteca H2O AutoML, validação cruzada e análise dos resultados. Também foi implementada uma ferramenta online para uso clínico do modelo. Os dados utilizados estavam anonimizados e o projeto foi aprovado pelo comitê de ética do Hospital das Clínicas da FMUSP.

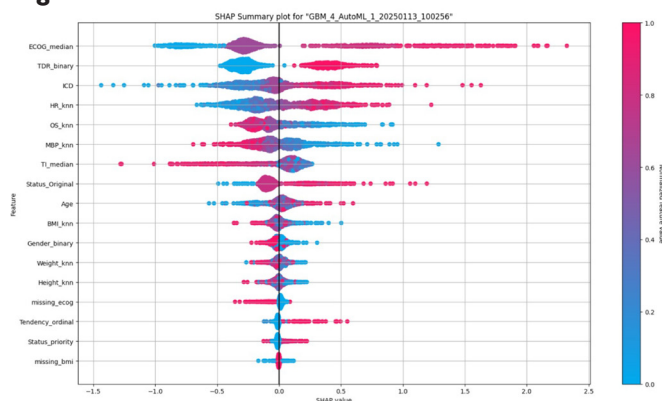
3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Foram analisadas 51.311 visitas à emergência realizadas por 17.189 pacientes no Hospital X entre 2018 e 2019. Os dados clínicos e demográficos foram extraídos dos prontuários eletrônicos no momento da admissão.

3.1 Variáveis e Pré-processamento

As variáveis de entrada utilizadas no modelo incluíram dados demográficos como idade e sexo; sinais vitais incluindo frequência cardíaca (HR - batimentos por minuto), Pressão Arterial Média (MBP - mmHg), frequência respiratória, saturação de oxigênio e temperatura; escalas clínicas como o Escore de Performance ECOG; diagnóstico e histórico incluindo tipo de câncer primário e histórico recente de atendimentos na emergência; e condições agudas como classificação de risco da triagem. O pré-processamento dos dados incluiu a imputação de valores missing utilizando a moda para variáveis categóricas e a mediana para variáveis numéricas, com base na distribuição global dos dados. As variáveis categóricas foram codificadas utilizando encoding, enquanto as variáveis numéricas contínuas foram padronizadas para ter média 0 e desvio padrão 1. O desfecho foi a sobrevida pós-visita, classificada como curto (<6 meses) ou longo prazo.

Figura 1 - SHAP values das Variáveis Normalizadas no Modelo GBM.



Fonte: Elaborado pelos autores com base nos resultados do modelo GBM.

3.2 Treinamento do Modelo e Métricas de Avaliação

Foram treinados cinco modelos de aprendizado de máquina utilizando a biblioteca H2O AutoML, com validação cruzada de 5 folds. O modelo GBM obteve o melhor desempenho. Para avaliar o modelo, utilizamos as seguintes métricas, formalmente definidas: Sensibilidade (Recall), entendida como a proporção de pacientes com sobrevida inferior a 6 meses que foram corretamente identificados pelo modelo; Especificidade, definida como a proporção de pacientes com sobrevida superior a 6 meses que foram corretamente classificados como tal; AUC-ROC (Área Sob a Curva Característica de Operação do Receptor), que mede a capacidade do modelo de discriminar entre as duas classes em todos os limiares de decisão possíveis; e Escore F0.5, uma média harmônica entre precisão e recall que atribui maior peso à precisão (reduzindo falsos positivos). Esta última métrica foi escolhida para balancear a necessidade de identificar pacientes elegíveis (sensibilidade) com

a imperativa segurança clínica de evitar a interrupção precoce de tratamentos em pacientes com sobrevida potencialmente longa (especificidade/precisão).

3.3. Resultados

O modelo GBM obteve um AUC de 0,973 no conjunto de treinamento e 0,823 na validação cruzada, demonstrando boa capacidade de generalização. Desenvolveu-se uma aplicação web em *Streamlit* para simulação de uso, ainda em fase experimental.

3.4. Limitações

Este estudo apresenta algumas limitações: a utilização de Dados de Centro Único, provenientes de um único centro terciário, o que pode limitar a generalização dos resultados para outras populações ou contextos hospitalares; o Viés Retrospectivo inerente à natureza retrospectiva do estudo e sujeito a vieses inerentes aos dados de prontuário eletrônico; as Variáveis Não Incluídas, pois variáveis potencialmente prognósticas importantes, como resultados de exames laboratoriais ou tratamentos sistêmicos em curso, nem sempre estavam consistentemente disponíveis no momento da admissão e não foram incluídas no modelo; e a necessidade de Validação Externa, pois o modelo necessita de validação externa em um conjunto de dados independente para confirmar sua robustez e acurácia antes da implantação clínica em larga escala.

4. DESAFIOS E APRENDIZADOS

Os principais desafios envolveram a limpeza, padronização e imputação de dados clínicos heterogêneos extraídos dos prontuários eletrônicos. Outro ponto crítico foi a escolha de uma métrica de avaliação que refletisse as implicações clínicas do modelo. Optou-se pelo uso do escore F0.5, uma média harmônica entre precisão e sensibilidade que atribui maior peso à precisão. Essa escolha foi motivada pela necessidade de reduzir ao máximo os falsos positivos — isto é, evitar que pacientes com sobrevida potencialmente longa sejam classificados incorretamente como casos de curto prazo, o que poderia levar à interrupção precoce de tratamentos com benefício terapêutico. Ao mesmo tempo, o uso do F0.5 ainda considera a sensibilidade, garantindo que o modelo mantenha capacidade de identificar pacientes elegíveis para cuidados paliativos. Essa decisão refletiu um equilíbrio entre segurança clínica e efetividade na triagem.

5. REFERÊNCIAS

1. Patra L, Ghoshal A, Damani A, Salins N. Cancer palliative care referral: patients' and family caregivers' perspectives – a systematic review. *BMJ Support Palliat Care*. 2024;14(e1):e143–e149.
2. Mulville AK, Widick NN, Makani NS. Timely referral to hospice care for oncology patients: a retrospective review. *Am J Hosp Palliat Care*. 2019;36(6):466–471.
3. Hawley P. Barriers to access to palliative care. *Palliat Care*. 2017;10:1178224216688887.
4. Wang DH, Heidt R. Emergency department admission triggers for palliative consultation may decrease length of stay and costs. *J Palliat Med*. 2021;24(4):554–560.
5. Wu FM, Newman JM, Lasher A, Brody AA. Effects of initiating palliative care consultation in the emergency department on inpatient length of stay. *J Palliat Med*. 2013;16(11):1362–1367.
6. Wilson JG, et al. End-of-life care, palliative care consultation, and palliative care referral in the emergency department: a systematic review. *J Pain Symptom Manage*. 2020;59(2):372–383.e1.
7. Zhuang Q, Zhang AY, Cong RSTY, et al. Towards proactive palliative care in oncology: developing an explainable EHR-based machine learning model for mortality risk prediction. *BMC Palliat Care*. 2024;23(1):124.
8. Parikh RB, Manz C, Chivers C, et al. Machine learning approaches to predict 6-month mortality among patients with cancer. *JAMA Netw Open*. 2019;2(10):e1915997.
9. Elfiky AA, Pany MJ, Parikh RB, et al. Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw Open*. 2018;1(3):e180926.
10. Lu SC, Xu C, Nguyen CH, et al. Machine learning–based short-term mortality prediction models for patients with cancer using electronic health record data: systematic review and critical appraisal. *JMIR Med Inform*. 2022;10(3):e33182.
11. Elfiky AA, Pany MJ, Parikh RB, et al. Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw Open*. 2018;1(3):e180926.

Predicting weekly dengue incidence in brazilian federative units

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

Abstract: Effective epidemic management requires proactive measures, making the accurate prediction of weekly dengue cases in specific regions essential for prevention and control strategies. In this study, we evaluated the effectiveness of classical statistical techniques and machine learning (ML) methods in forecasting the number of weekly dengue cases across 27 Brazilian federative units. For each unit, we explored a multivariate one-step prediction strategy, applying wavelet filtering to the features to enhance signal decomposition and improve predictive performance. We investigated the univariate LightGBM model trained on data from 26 cities (cross-learning) and validated it individually in each federative unit using the univariate leave-one-out (LOO) technique with one-step predictions. Additionally, we demonstrated the model's generalizability by training LightGBM on data from all Brazilian federative units and validating it in a different geographic location, San Juan. The LightGBM LOO model exhibited superior generalization compared to other shallow, deep, and foundation models, including TimeGPT-1 and MOIRAI.

Keywords: dengue; time series forecasting; machine learning; wavelets; cross-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

Dengue cases in the Americas have surged dramatically over the past four decades, with Brazil reporting over 7.8 million cases in the first 23 weeks of 2024 alone—a 230% increase compared to the same period in 2023 (5). Despite the availability of a vaccine, anticipating dengue trends through time series forecasting remains essential for efficient public health planning (1). Recent advances include deep learning models based on encoder-decoder architectures with self-attention mechanisms, such as Transformers, which operate under the universal forecasting paradigm and leverage zero-shot learning without task-specific fine-tuning. Pre-trained foundation models like TimeGPT-1 and MOIRAI are promising in this context, but they often ignore covariates and lack hyperparameter tuning. Traditional ML models like Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) have been shown to outperform deep learning methods in many forecasting tasks, yet few studies compare them directly against foundation models for dengue prediction in Brazil. Furthermore, cross-learning—training on multiple related datasets to improve generalization—can be a powerful strategy in this setting, while time series-specific validation techniques, such as leave-one-out (LOO) cross-validation across Federative Units, are needed to properly assess model performance.

1.1 Objective

We propose and evaluate a machine learning regression pipeline to forecast weekly dengue cases across Brazil's 27 federative units (FUs), comparing a wide range of models—including classical statistical methods, shallow regressors, deep learning architectures, and foundation models. The statistical models include Random Walk (RW), Simple Exponential Holt (SEH), Simple Exponential Smoothing (SES), and Holt-Winters (HW), ARIMA, and SARIMA; deep learning models involve CNNs, LSTMs, and their hybrids; foundation models comprise TimeGPT-1 and MOIRAI; and shallow regressors cover Linear Regression (LR), Least Ab-

solute Shrinkage and Selection Operator (LASSO), Ridge Regression (Ridge), Support Vector Regressor (SVR), Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), LightGBM, and XGBoost. We also explore dataset reformatting strategies (2) and multiple validation schemes, such as time-dependent cross-validation (3), univariate model combinations, and cross-learning (6).

2. MAIN ACTIVITIES

The main activities included:

- Preprocessing epidemiological and meteorological time series from online sources;
- Applying wavelet filtering to smooth and denoise relevant features;
- Reformulating datasets using one-step and multi-step learning strategies;
- Implementing and evaluating diverse forecasting models (statistical, ML, and foundation models);
- Tuning LightGBM hyperparameters via randomized cross-validation;
- Testing cross-learning and leave-one-out strategies across 27 Brazilian federative units;
- Analyzing model performance using MAE, RMSE, Mean Absolute Scaled Error (MASE), Root Mean Squared Scaled Error (RMSSE), and BIAS;
- Evaluating generalization on external data from San Juan, Puerto Rico.

3. DEVELOPMENT OF THE WORK (METHODOLOGY AND RESULTS)

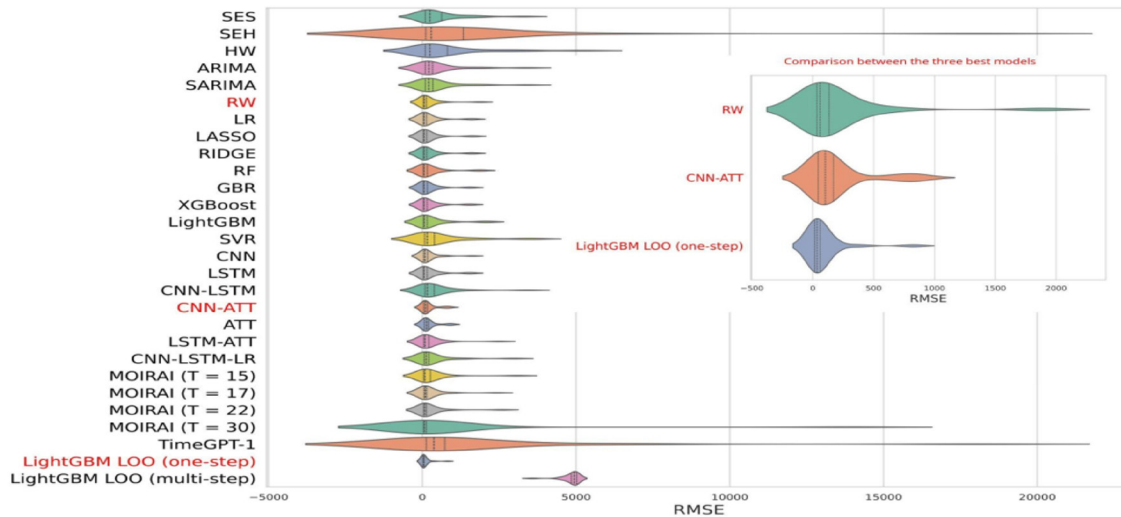
3.1 InfoDengue Results

Each Brazilian federative unit from the InfoDengue (4) database was modeled individually using multivariate one-step approaches, applying wavelet filtering to enhance feature quality. We reserved the last 100 instances for validation. In parallel, we adopted a cross-learning strategy using only the weekly dengue case counts (univariate), concatenating data from all 27 units into supervised learning matrices. Two validation strategies were applied: randomized K-fold (with $K = 15$ and $R = 50$ rounds) and leave-one-out (LOO). LightGBM hyperparameters were optimized using grid search, with the best configuration set to window size $T = 30$, learning rate = 0.01, number of leaves = 127, maximum depth = 10, and regularization hyperparameters $L1 = 9$, and $L2 = 7$. This configuration was fixed across experiments. LightGBM LOO (one-step) outperformed all models in RMSE across Brazilian cities, showing low error variance and no significant bias. Multi-step versions suffered from accumulated error. Deep learning models exhibited higher BIAS and were often outperformed by simpler methods. Classical models such as SEH performed poorly due to their linear assumptions, while the random walk model remained a strong baseline.

3.2 San Juan Generalization

To assess generalization, we applied the best-performing model—LightGBM LOO (one-step)—to the San Juan dataset. It was evaluated on the full series and on the last 187 weeks. Although it outperformed all univariate models from prior work, it did not surpass the multivariate models developed specifically for San Juan in this study, such as LR (with data filtered using the sym20 wavelet) and CNN-LSTM-LR (with data filtered using the sym20 wavelet). This performance gap is likely due to the presence of auxiliary features in the San Juan dataset, absent in InfoDengue.

Figure 1 - Numerical distributions of the RMSE performance metrics across all evaluated models.



4. CHALLENGES AND LEARNINGS

Our study highlighted several important insights. First, cross-learning proved effective for improving forecast accuracy across different regions, including all 27 Brazilian federative units and external data from San Juan, Puerto Rico. Second, the univariate LightGBM model consistently outperformed deep learning and pre-trained foundation models such as TimeGPT-1 and MOIRAI, even without wavelet preprocessing. Finally, our pipeline showed strong generalization capacity, achieving performance comparable to models tailored specifically for San Juan. Key challenges included limited hyperparameter tuning, small validation samples for external regions, and high feature dimensionality in multivariate setups. These findings suggest future directions involving dimensionality reduction, seasonality modeling, and integration of exogenous variables.

5. REFERENCES

1. Anumanthan G, Sahay B, Mergia A. Current dengue virus vaccine developments and future directions. *Viruses*. 2025;17:212.
2. Benidis K, Rangapuram SS, Flunkert V, Wang Y, Maddix D, Bohlke C, et al. Deep learning for time series forecasting: tutorial and literature survey. *ACM Comput Surv*. 2022;55(6):1–36.
3. Bergmeir C, Hyndman RJ, Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput Stat Data Anal*. 2018;120:70–83.
4. Codeco C, Coelho F, Cruz O, Oliveira S, Castro T, Bastos L. Infodengue: a nowcasting system for the surveillance of arboviruses in Brazil. *Rev Epidemiol Sante Publique*. 2018;66(Suppl):S386.
5. Organização Pan-Americana da Saúde. Atualização Epidemiológica: Aumento de casos de dengue na Região das Américas. June 18, 2024. Washington, D.C.: PAHO/WHO; 2024. Available from: <https://www3.paho.org/data/index.php/es/temas/indicadores-dengue.html>
6. Semenoglou AA, Spiliotis E, Makridakis S, Assimakopoulos V. Investigating the accuracy of cross-learning time series forecasting methods. *Int J Forecast*. 2021;37(3):1072–84.

Taxonomia de desinformação sobre vacina no WhatsApp

Thales Henrique Silva¹, Jonatas Henrique dos Santos¹, Matheus Gontijo Guimaraes¹, Glaucio S. Souza^{1,3}, Philipe Melo², Julio C. S. Reis², Marco Antônio G. Rodrigues¹, Marcos André Gonçalves¹, Fabrício Benevenuto¹, Cristiano X. Lima¹

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

²Universidade Federal de Viçosa (UFV), MG, Brasil.

³Universidade Federal de Juiz de Fora, MG, Brasil.

thaleshenrique@ufmg.br, jonatashds@ufmg.br, guimaraesmatheus22@gmail.com, glaucio.gss@gmail.com, philipe.freitas@ufv.br, jreis@ufv.br, magro.mg@terra.com.br, mgoncalv@dcc.ufmg.br, fabricio@dcc.ufmg.br, cxlima@ufmg.br

Resumo: Este trabalho apresenta uma taxonomia proposta para classificação de diferentes tipos de desinformação médica, com foco específico em conteúdos sobre vacinação. A taxonomia será utilizada para categorizar automaticamente, por meio de técnicas de aprendizado de máquina supervisionadas, fenômenos como distorções científicas, teorias da conspiração, alarmismo e uso indevido de autoridade. Fundamentada na expansão da literatura existente, a iniciativa busca apoiar ações de monitoramento, combate e educação contra a desinformação em saúde no Brasil.

Abstract: This work presents a taxonomy for classifying different types of medical misinformation, with a specific focus on content related to vaccination. The taxonomy will be used to automatically categorize phenomena such as scientific distortions, conspiracy theories, alarmism, and misuse of authority through supervised machine learning techniques. Based on an expansion of the existing literature, this initiative aims to support monitoring, mitigation, and educational efforts to combat health misinformation in Brazil.

Palavras-chave: WhatsApp; vacinação; desinformação em saúde; covid-19; taxonomia.

Nome do projeto: Contribuições da Inteligência Artificial no Estudo da Desinformação Médica em Mídias Sociais; Cristiano Xavier Lima.

1. CONTEXTO E MOTIVAÇÃO

A desinformação em saúde constitui um desafio crescente, especialmente em contextos de crise sanitária, como a pandemia de COVID-19. A disseminação de informações falsas, enganosas ou distorcidas compromete a confiança da população em práticas médicas, como a vacinação, e enfraquece iniciativas de saúde pública. Este projeto tem como objetivo gerar conhecimento inovador que subsidie o desenvolvimento de estratégias para enfrentar a desinformação médica nas redes sociais, visando prevenir ou mitigar seus impactos negativos na saúde humana. Por meio de técnicas de Processamento de Linguagem Natural apoiadas por Inteligência Artificial, busca-se mapear padrões discursivos, identificar tópicos recorrentes e propor uma taxonomia dos principais tipos de desinformação médica. Tal taxonomia constitui um passo essencial para o desenvolvimento de abordagens automáticas e escaláveis de detecção e classificação de conteúdos desinformativos, contribuindo para estratégias mais eficazes de monitoramento, mitigação e promoção da saúde pública.

2. ATIVIDADES PRINCIPAIS

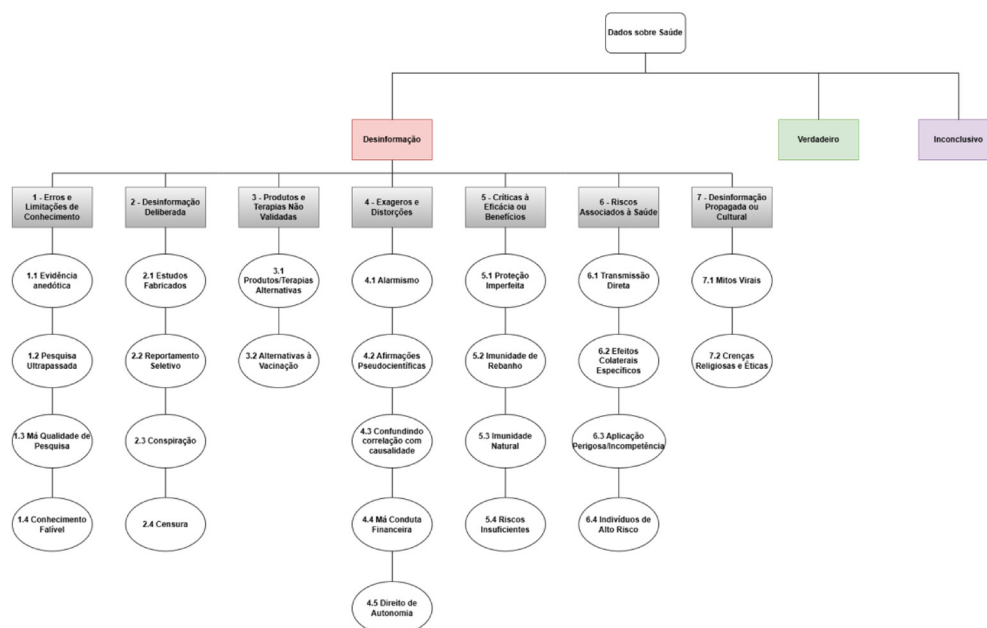
Realizou-se uma revisão da literatura sobre desinformação em saúde e o tema vacina se mostrou relevante ⁽¹⁾. Com base em trabalhos anteriores ⁽²⁾ ⁽³⁾, propôs-se uma taxonomia voltada à classificação de

desinformação sobre vacinas, a partir de dados coletados durante a pandemia de COVID-19 (2020–2022) em grupos públicos com viés político no WhatsApp. Por fim, foi conduzida uma validação preliminar da proposta com base nesses dados.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Este trabalho propõe uma taxonomia autoral para a classificação da desinformação em saúde, baseada em mensagens coletadas de redes sociais. A estrutura hierárquica da taxonomia permite, em um primeiro nível, classificar mensagens como “Desinformação”, “Verdadeiro” ou “Inconclusivo”. As mensagens rotuladas como “Desinformação” são então subdivididas em categorias mais específicas, como distorção de evidências, uso indevido de autoridade, alarmismo, teorias conspiratórias, dentre outras (Figura 1). Para validação preliminar, foram identificadas manualmente mensagens representativas de cada subcategoria, evidenciando a presença e diversidade desses discursos. Exemplos ilustrativos (Figura 2), escolhidos por palavras-chave e verificação manual, confirmam a presença de cada tipo na base e destacam a importância de classificações que captem as nuances da desinformação médica. O processo de rotulagem consiste numa seleção aleatória de mensagens após aplicar um filtro de palavras-chave (excluindo-se mensagens irrelevantes para o contexto de vacinas). Três especialistas em saúde analisam e rotulam as mensagens de forma independente. Em seguida, calculam-se métricas de concordância e ocorre uma discussão conjunta para alinhar decisões e esclarecer dúvidas. Já foram classificadas cerca de 1000 mensagens, e novas rodadas de rotulagem estão previstas para refinar o processo.

Figure 1 - Taxonomia para desinformação em grupos de saúde.



4. DESAFIOS E APRENDIZADOS

Um dos principais desafios foi equilibrar a simplicidade de uma taxonomia genérica com a precisão de uma classificação detalhada. A primeira facilita aplicação e interpretação, mas tende a não capturar nuances dos discursos. Já a segunda permite uma análise mais rigorosa, porém exige maior esforço de anotação. Essa tensão proporcionou aprendizados relevantes sobre o potencial de cada abordagem. Nas próximas etapas, o projeto buscará refinar a taxonomia, conciliando a objetividade da classificação simplificada com a profundidade da tipologia descritiva, visando a criação de ferramentas para o monitoramento da desinformação em saúde.

Figure 2 - Exemplos de mensagens.

| Categoria | Subcategoria | Exemplos Mensagens |
|---|--|---|
| 1 - Erros e Limitações de Conhecimento | 1.1 Evidência anedótica | "...Minha filha tomou todas as vacinas desde que nasceu e mesmo assim pegou catapora, caxumba e coqueluche..." |
| | 1.2 Pesquisa Ultrapassada | "*Trump foi devidamente avisado por esta Doutora...e nada fez para Salvar e Livrar a Humanidade da Escravidão e da Morte...* PLANDEMIC - LEGENDADO EM PORTUGUES..." |
| | 1.3 Má Qualidade de Pesquisa | "...quem quiser que tome a vacina, eu tbm tomaria se fosse eficaz. sou totalmente favorável a saúde, a medicação da cura, se fosse testada com tempo de estudo e aprovada em seu tempo..." |
| | 1.5 Conhecimento Falível | "...Falam que a ômicron é tão somente uma maneira de denominar os inúmeros efeitos colaterais das próprias vacinas. Nunca saberemos..." |
| 2 - Desinformação Deliberada | 2.1 Estudos Fabricados | "Remédio secreto é pior que cloroquina para Covid-19, diz estudo chinês Empresas trabalham na criação da vacina contra coronavirus..." |
| | 2.2 Reportamento Seletivo | "...Coronavac tem eficácia de 42% acima dos 70 anos. Acima de 80 é ainda menor..." |
| | 2.3 Conspiração | "...Mano tecnologia de localização em vacinas juntos com 5G, pesquiem..." |
| | 2.4 Censura | "...um bocado de pesquisador e de médico sérios foram calados e perseguidos. Parece que está tudo vindo a tona..." |
| 3 - Produtos e Terapias Não Validadas | 3.1 Produtos/Terapias Alternativas | "...PREFERIRAM VENDER VACINAS SEM COMPROVAÇÃO CIENTÍFICA.... AGORA ELES VÃO ADOTAR A HIDROXICLOROQUINA E A IVERMECTINA..." |
| | 3.2 Alternativas a Vacinação | "Na falta da vacina, não seria recomendável o uso do tratamento precoce, como uma forma de mitigar as mortes pelo Covid19? Ou o senhor se enquadra no grupo científico-genocida?" |
| 4 - Exageros e Distorções | 4.1 Alarmismo | "O que você acha que acontecerá dentro do seu corpo quando as partículas de grafeno existentes nas vacinas covid-888 entrarem em contato com a frequência 5G?" |
| | 4.2 Afirmações Pseudocientíficas | "Estudo: Máscaras faciais contribuíram para altas taxas de mortalidade por COVID-19. Um relatório bombástico compartilhado pela Dra. Simone Gold..." |
| | 4.3 Confundindo correlação com causalidade | "AS MORTES TRIPLICARAM DEPOIS DA VACINAÇÃO. O QUE ESTARÁ ACONTECENDO? SE AS VACINAS ERAM PRA CURAR E IMUNIZAR AS PESSOAS CONTRA O CORONAVIRUS..." |
| | 4.4 Má Conduta Financeira | "*A pressa para vacinar é meramente financeira, para poder liberar o CARNAVAL, mentindo criminosamente para o povo brasileiro..." |
| | 4.5 Direito de Autonomia | "Direito de escolha!!!! O tratamento médico e planejado pelo médico e pelo seu paciente..." |
| 5 - Críticas à Eficácia ou Benefícios | 5.1 Proteção Imperfeita | "Primeiro tocam o terror - Impõe a vacina em massa - Depois afirmam que a vacina não protege contra infecções do vírus..." |
| | 5.2 Imunidade de Rebanho | "...DEVO SER VACINADO? Não. Você não precisa. Mais cedo ou mais tarde todos nós teremos contato com o vírus (93% são assintomáticos). A imunidade acontecerá naturalmente, a chamada imunidade de rebanho..." |
| | 5.3 Imunidade Natural | "...o Projeto Veritas obteve as confissões, via câmera escondida, de vários cientistas da Pfizer admitindo que a imunidade natural é melhor que a vacina..." |
| | 5.4 Riscos Insuficientes | "...Aliás, muitos tem imunidade natural genética, outros já contaminados apresentam apenas efeitos como uma 'gripezinha' mesmo..." |
| 6 - Erros e Limitações de Conhecimento | 6.1 Transmissão Direta | "...DEPOIS aprovam a MORTE pela VACINAÇÃO OBRIGATÓRIA, após DISTO o povo ficam doentes vai pro hospital..." |
| | 6.2 Efeitos Colaterais Específicos | "...Minha Sobrinha transplantada está com COVID19... e, no hospital viu uma senhora que teve trombose consequência do uso prolongado destes medicamentos..." |
| | 6.3 Aplicação Perigosa/Incompetência | "...CRIANÇA É VÍTIMA DE APLICAÇÃO ERRADA DA VACINA PARA COVID E MÃE DENUNCIA..." |
| | 6.4 Indivíduos de Alto Risco | "Tenho certeza que vacinas em crianças é para tornarem estéreis e reduzir a população" |
| 7 - Desinformação Propagada ou Cultural | 7.1 Mitos Virais | "*Vacina magnetizada? Microchips na injeção?..." |
| | 7.2 Crenças Religiosas e Éticas | "Temos que confiar só mente em deus vacina não vai livrar ninguém não" |

Agradecimentos. Esse trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unimed-BH e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).

5. REFERÊNCIAS

1. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. Journal of medical Internet research. 2021 Jan 20;23(1):e17187.
2. Stureborg R, Nichols J, Dhingra B, Yang J, Orenstein W, Bednarczyk RA, Vasudevan L. Development and validation of VaxConcerns: A taxonomy of vaccine concerns and misinformation with Crowdsourced-Viability. Vaccine. 2024;42(10):2672-2679
3. Ismail H, Hussein N, Elabyad R, Abdelhalim S, Elhadeif M. Aspect-based classification of vaccine misinformation: a spatiotemporal analysis using Twitter chatter. BMC Public Health. 2023 Jun 21;23(1):1193.

Uma abordagem usando modelos generativos para apoio à decisão clínica

Bruno B. de Paiva¹, Guilherme Dal Bianco³, Caio Jorge¹, Milena Marcolino¹, Leandro Marcolino⁴,
Mariangela Leal Cherchiglia⁴, Hugo André da Rocha⁴, Leonardo Rocha²,
Marcos André Gonçalves¹

¹Universidade Federal de Minas Gerais, Belo Horizonte, MG. ²Universidade Federal de São João
Del-Rei, São João del-Rei, MG. ³Universidade Federal da Fronteira Sul,
⁴Lancaster University

autor correspondente: brunobarbosa.mpaiva@gmail.com

Resumo: Abordagens recentes de inteligência artificial, como modelos generativos, destacam-se por identificar padrões complexos em grandes volumes de dados. Ao analisar sequências de eventos clínicos, esses modelos mostram potencial para prever ocorrências futuras, como o agravamento de doenças. Este trabalho propõe uma abordagem baseada em um modelo generativo para gerar decisões personalizadas e apoiar a predição de eventos médicos. O modelo é treinado com dados clínicos, em que cada evento é uma unidade discreta posicionada por deltas temporais, permitindo capturar relações temporais. Resultados preliminares mostram melhora de 6,2% na predição de óbito em relação às linhas de base.

Abstract: Recent AI approaches, particularly generative models, excel at identifying complex patterns in large clinical datasets. By analyzing event sequences, these models can predict future outcomes, such as disease progression. This work proposes a generative model to support personalized decision-making in predicting medical events. Trained on clinical data with temporally ordered discrete events, the model captures time-based relationships. Preliminary results show a 6.2% improvement in mortality prediction over baseline methods.

Palavras-chave: Inteligência artificial; COVID-19; modelos preditivos.

Nome do projeto: Predição de desfechos clínicos e econômicos por meio de representações semânticas multimodais de pacientes resilientes a drifts temporais.

1. CONTEXTO E MOTIVAÇÃO

A rápida expansão dos dados de saúde traz oportunidades para uma medicina mais precisa, mas também desafios devido ao volume e à complexidade das informações ^(1,2). A pandemia de COVID-19 destacou a necessidade de otimizar recursos e priorizar intervenções, reforçando a importância de abordagens preditivas. Avanços em aprendizado de máquina, especialmente com redes neurais profundas, têm permitido modelar padrões complexos em grandes conjuntos clínicos ⁽³⁾. Métodos como o Pat2Vec ⁽⁴⁾ e adaptações do BERT ao domínio médico ⁽⁵⁾ ilustram esse progresso. No entanto, tais trabalhos não exploram uma modelagem explícita da dimensão temporal dos eventos clínicos — essencial para prever desfechos com precisão, já que variações temporais (drifts) podem distorcer os modelos ^(2,3).

2. ATIVIDADES PRINCIPAIS

Descrevemos o funcionamento do Modelo Generativo Proposto (MGP) para aprendizagem de representações de pacientes e modelagem de eventos clínicos. Baseado na arquitetura GPT-2, o MGP é treinado em duas fases: (i) predição do próximo token em sequências de eventos de saúde; e (ii) ajuste fino para refinar o espaço de incorporações. Introduzimos uma adaptação que substitui incorporações posicionais aprendidas por codificações de delta de tempo, refletindo a estrutura temporal dos registros.

Por exemplo, em exames mensais ao longo de um ano, os deltas em dias a partir do primeiro evento seriam: 0 (primeiro exame), 30 (segundo) e 360 (último).

Pré-treinamento: Modelamos o histórico médico de um paciente como uma sequência de eventos clínicos discretos, extraídos de um vocabulário finito que inclui códigos de diagnóstico (ex: CID), procedimentos, prescrições e exames laboratoriais. As sequências são tokenizadas, incorporadas e processadas por um modelo de transformador causal, pré-treinado com o objetivo de predição do próximo token — estimando a probabilidade de um evento dado seu contexto anterior. No GPT-2, representações posicionais são somadas às dos tokens para informar a ordem na sequência, já que seu mecanismo de autoatenção processa todos os tokens em paralelo, sem noção explícita de ordem. Para lidar com isso, substituímos as codificações posicionais por deltas temporais em dias a partir do primeiro evento registrado do paciente. Essa modificação torna as pontuações de atenção sensíveis ao espaçamento temporal, permitindo capturar padrões clínicos dependentes do tempo sem arquiteturas adicionais.

Ajuste Fino: Uma vez pré-treinado, o MGP é ajustado de acordo com a tarefa final, como a predição de morte, a partir da otimização da função de perda (por exemplo, a entropia cruzada ou o erro na classificação binária). O MGP integra pré-treinamento generativo sobre as sequências de eventos clínicos com um mecanismo de codificação de dados com reconhecimento temporal. Nossa proposta aprimora a codificação contextual dos históricos de pacientes e a estrutura do espaço de incorporação. Essa inovação permite que o modelo tenha um desempenho competitivo em tarefas de previsão de desfechos, permanecendo generalizável para outras aplicações clínicas, como recomendação e recuperação.

3. METODOLOGIA E RESULTADOS

Para avaliar o desempenho e a robustez do MGP em cenários clínicos reais, conduzimos experimentos em um conjunto de dados privado e longitudinal que acompanha pacientes em tratamento de diálise ao longo de um período de 12 anos (6). A coorte incluiu milhares de pacientes com registros detalhados de eventos médicos, incluindo diagnósticos e procedimentos. Este conjunto de dados fornece uma visão abrangente da doença, caracterizada por longos períodos de acompanhamento e padrões complexos de comorbidade e hospitalizações. A coorte foi aprovada no comitê da Universidade Federal de Minas Gerais (44121315.2.0000.5149).

A Tabela 1 apresenta os valores de macro-F1 do MGP em comparação com a base de dados com dois métodos: MedBert e Pat2Vec. O MGP supera ambos os métodos por uma margem clara com um macro-F1 de 0.6107, superando o concorrente mais próximo em 6.2%. Acredita-se que o resultado é devido à capacidade do MGP em relacionar os eventos com o seu período de ocorrência, já que o conjunto de dados contém dados com uma certa esparsividade.

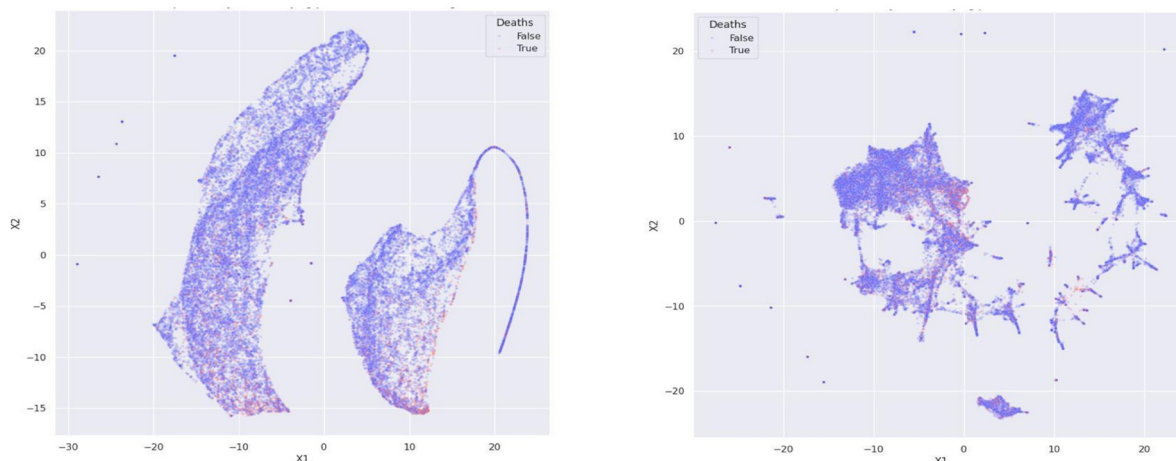
Tabela 1 - Comparação do Macro-F1 do método proposto em relação a dois métodos base.

| Método | MGP | MedBERT | Pat2Vec |
|----------|--------|---------|---------|
| F1-Macro | 0.6107 | 0.575 | 0.5154 |

O experimento demonstra de forma preliminar o potencial do pré-treinamento generativo no avanço da medicina personalizada e na otimização da alocação de recursos em saúde. Incorporar a passagem do tempo como posição aumenta a capacidade preditiva, capturando padrões temporais de forma mais eficaz. Por fim, o ajuste fino do modelo para resultados específicos refina substancialmente as incorporações aprendidas, permitindo que modelos posteriores aproveitem essas representações para uma tomada de decisão mais precisa. A Figura 1 compara a representação MGP (esquerda) com o TF-IDF (direita). Os pontos vermelhos na figura indicam pacientes que vieram a óbito. Idealmente, esses pontos deveriam estar agrupados em uma única região. A representação MGP demonstra um claro agrupamento desses pontos na parte inferior.

Em contraste, o TF-IDF forma um agrupamento central, porém com pontos dispersos em múltiplas regiões, o que pode dificultar a identificação de padrões entre os pacientes.

Figura 1 - Comparação entre a representação MGP (esquerda) e a TF-IDF (direita).



4. DESAFIOS E APRENDIZADOS

O desenvolvimento do MGP trouxe desafios referente a modelagem da abordagem, ou seja, de como acomodar as variações temporais dos eventos clínicos de forma explícita. A estratégia de delta de tempo se mostrou relevante para que o modelo gerado seja capaz de incorporar os eventos temporalmente. Entre os principais aprendizados, destaca-se a capacidade do modelo generativo em gerar um modelo mais robusto e adaptável a diferentes cenários clínicos. Os próximos passos incluem a análise em outros conjuntos de dados e desfechos, como câncer, e a análise das variações temporais para a predição e o tratamento de *drifts* (3).

Agradecimentos. Este trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unimed-BH e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).

5. REFERÊNCIAS

1. Butcher CJT, Hussain W (2022). Digital healthcare: the future. *Future Healthc J*.9(2):113-7.
2. Paiva BM, et al. (2023). Potential and limitations of machine meta-learning (ensemble) methods for predicting COVID-19 mortality in a large in-hospital Brazilian dataset. *Sci Rep*.13(1):3463.
3. Paiva BM, et al (2024). A New Natural Language Processing-Inspired Methodology (Detection, Initial Characterization, and Semantic Characterization) to Investigate Temporal Shifts (Drifts) in Health Care Data: Quantitative Study. *JMIR Med Inform*. 12:e54246.
4. Steiger, E., & Kroll, L. E. (2023). Patient Embeddings From Diagnosis Codes for Health Care Prediction Tasks: Pat2Vec Machine Learning Framework. *JMIR AI*, 2, e40755.
5. Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, 4(1), 86.
6. Guerra Jr, et al. (2018). Building the National Database of Health Centred on the Individual: Administrative and Epidemiological Record Linkage-Brazil, 2000-2015. *Int. J. Pop. D. Sci* 3(1) 446.

Uma análise exploratória da (des)informação sobre vacinação em grupos públicos no WhatsApp no Brasil

Jonatas H. dos Santos¹, João Francisco H. Olivetti², Thales H. Silva¹, Matheus Gontijo Guimaraes¹, Glaucio S. Souza^{1,3}, Philipe Melo², Julio C. S. Reis², Marco Antônio G. Rodrigues¹, Marcos André Gonçalves¹, Fabrício Benevenuto¹, Cristiano X. Lima¹

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

²Universidade Federal de Viçosa (UFV), MG, Brasil.

³Universidade Federal de Juiz de Fora (UFJF), MG, Brasil

jonatashds@ufmg.br, joao.olivetti@ufv.br, thaleshenrique@ufmg.br, guimaraesmatheus22@gmail.com, glaucio.gss@gmail.com, philipe.freitas@ufv.br, jreis@ufv.br, magro.mg@terra.com.br, mgoncalv@dcc.ufmg.br, fabricio@dcc.ufmg.br, cxlima@ufmg.br

Resumo: Este trabalho apresenta uma análise exploratória de mensagens coletadas em grupos públicos de WhatsApp com perfil político, focando nas discussões sobre vacinação entre 2020 e 2023. Utilizando técnicas de Processamento de Linguagem Natural (PLN), investigamos como conteúdos sobre vacinas, especialmente os que potencialmente contêm desinformação, circularam nesses espaços digitais durante a pandemia de COVID-19. Após uma filtragem por palavras-chave, identificamos mais de 84 mil mensagens sobre vacinas, enviadas por mais de 15 mil usuários em 941 grupos distintos. Com o modelo BERTopic, extraímos os principais tópicos debatidos, revelando forte presença de narrativas alarmistas e desinformativas, como associação entre vacinas e mortes, desconfiança na vacina chinesa e teorias sobre efeitos colaterais. Este estudo é uma etapa preliminar para o desenvolvimento de modelos de IA voltados ao combate à desinformação em saúde em ambientes digitais.

Abstract: This study presents an exploratory analysis of messages from politically-oriented public WhatsApp groups, focusing on vaccine-related discussions between 2020 and 2023. Using Natural Language Processing (NLP) techniques, we examined how vaccine content, particularly misinformation, circulated in these digital spaces during the COVID-19 pandemic. After keyword-based filtering, over 84,000 messages from more than 15,000 users across 941 groups were identified. Applying the BERTopic model, we extracted key discussion topics, which revealed the prevalence of alarmist and misleading narratives, including claims linking vaccines to deaths, distrust in Chinese vaccines, and theories about side effects. This work serves as a preliminary step toward developing AI models to combat health misinformation in digital environments.

Palavras-chave: WhatsApp; vacinação; desinformação; saúde; pandemia; COVID-19.

Nome do projeto: Contribuições da Inteligência Artificial no Estudo da Desinformação Médica em Mídias Sociais; Cristiano Xavier Lima

1. CONTEXTO E MOTIVAÇÃO

A pandemia de COVID-19 evidenciou o papel central das plataformas digitais na disseminação de (des)informação sobre saúde pública. No Brasil, o WhatsApp consolidou-se como uma das principais ferramentas de comunicação interpessoal e mobilização política, sendo amplamente utilizado para debater temas sociais e sanitários (1). Entre esses temas, a vacinação destacou-se como um dos principais focos de discussão, especialmente durante a pandemia de COVID-19. Nesse contexto, a circulação de mensagens sobre vacinas nesses grupos, muitas vezes marcada por desinformação, representa um desafio para autoridades de saúde, pesquisadores e desenvolvedores de soluções tecnológicas. Este trabalho busca

compreender como esse fenômeno se manifesta em grupos políticos públicos no WhatsApp, analisando, com técnicas de Processamento de Linguagem Natural (PLN), os principais padrões discursivos e comportamentos de disseminação de conteúdo. O objetivo é contribuir com evidências empíricas para o debate sobre saúde digital, letramento informacional e o desenvolvimento de tecnologias mais responsáveis no combate à desinformação em saúde.

2. ATIVIDADES PRINCIPAIS

As principais atividades conduzidas neste trabalho, conforme ilustrado na Figura 1, compreendem o processamento, análise e o entendimento da base de dados sobre vacinação no WhatsApp. Inicialmente, a base continha muitas mensagens fora do escopo da nossa pesquisa, como estratégia para tornar a análise mais eficiente, foi realizada filtragem com o objetivo de construir um novo conjunto de dados contendo apenas mensagens relacionadas a vacinas.

Figura 1 - Diagrama geral da metodologia de coleta e de processamento de dados do WhatsApp.

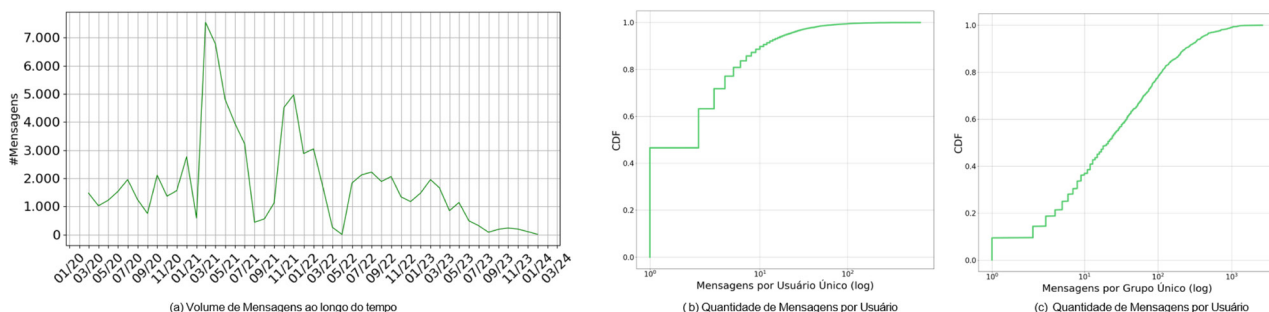


3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A base de dados utilizada contém mensagens coletadas de grupos públicos de WhatsApp entre 2020 e 2023 originalmente desenvolvidos por (2). Para filtrar conteúdos relevantes ao tema, aplicamos filtragem por palavras-chave associadas, incluindo diversas variações como "vacina", "vacinação", "vacinado" e expressões alternativas como "v4c1n4" e "vachina". Esse processo resultou em um subconjunto com 84.640 mensagens únicas, distribuídas em 941 grupos e enviadas por 15.148 usuários distintos.

Com esse recorte temático, realizamos análises, incluindo estatísticas de frequência por usuários e grupos, além da identificação de tópicos (de mensagens) com o modelo BERTopic. Também foram geradas visualizações como curvas de distribuição cumulativa (CDFs) e histogramas, com o objetivo de compreender padrões de engajamento e circulação das mensagens ao longo do tempo e entre diferentes perfis de usuários.

Figura 2 - Características da base de dados sobre vacinação no WhatsApp.



A partir deste processo, realizamos algumas análises exploratórias iniciais para entender melhor nosso conjunto de dados gerados, conforme Figura 2. Ela demonstra como o uso das palavras-chave em mensagens disseminadas no WhatsApp teve o pico ocorrendo entre 2021 e 2022, considerado "o auge da pandemia" e com a chegada da vacina. Visando investigar a maneira como tais informações eram compartilhadas, também exploramos CDF em relação às mensagens por usuário e às mensagens por grupo. Em

relação aos usuários, a análise revelou que cerca de 45% dos usuários compartilhou apenas 1 mensagem e que 90% compartilhou até 10 mensagens, mas também revelou a existência de alguns usuários que compartilham sozinhos centenas de mensagens sobre vacina. Já em relação às mensagens por grupo mostrou uma distribuição mais homogênea: cerca de 10% dos grupos continham somente uma única mensagem, enquanto 80% dos grupos considerados incluíam até 100 mensagens.

Por fim, a Figura 3 apresenta os dez tópicos mais representativos identificados por meio do modelo BERTopic, aplicado às mensagens. Essa abordagem automatiza a detecção de agrupamentos temáticos, partindo do princípio de que mensagens semanticamente próximas formam tópicos. A rotulagem dos temas foi realizada com o auxílio da IA Gemini, com base nas palavras mais frequentes de cada grupo, como morte, morreram, vacinados e vacina, no caso do tópico sobre “mortes e vacinação”. Os tópicos revelam narrativas alarmistas e potencialmente desinformativas, como a associação entre vacinas e mortes, desconfiança na vacina chinesa e supostos efeitos adversos. Também foram identificados conteúdos sobre vacinação infantil, campanhas contra gripe e número de doses aplicadas, geralmente acompanhados de dúvidas ou preocupações exageradas. Além disso, notou-se intensa circulação de vídeos e links de canais de notícias. Essa categorização evidencia o papel do WhatsApp na difusão de mensagens que podem afetar negativamente a confiança pública nas vacinas.

Figura 3 - Top-10 tópicos mais representativos identificados a partir do BERTopic.

| # | Palavras | #Msgs | Rótulo |
|----|---|-------|--------------------------------|
| 1 | mortes, morreram, morreu, morte, morrer, morrendo, vacina, vacinados, vacinas, pessoas | 1442 | Mortes e vacinação |
| 2 | chinesa, vacina chinesa, china, vacina china, chinês, vacina, chineses, vírus, chinês, dória, dória | 1098 | Vacina chinesa |
| 3 | coronavac, contra coronavirus, coronavirus, vacina coronavac, vacina contra, abaixo redes, saiba link, novo coronavirus, contra novo, sociais vacinação | 841 | Coronavac e coronavirus |
| 4 | crianças, vacinação crianças, crianças anos, vacinação infantil, vacinar crianças, infantil, filhos, vacinar, vacina crianças, covid crianças | 725 | Vacinação infantil |
| 5 | ebc, jrnews, mc, lendo jrnews, lendo, cnn, anos, im, jp, fonte | 566 | Fontes de notícias |
| 6 | gripe, contra gripe, influenza, vacinação contra, contra influenza, vacina gripe, campanha, vacinação, campanha, vacinação, contra | 540 | Vacinação contra gripe |
| 7 | dose, dose vacina, segunda dose, doses, segunda, doses vacina, duas doses, terceira dose, reforço, aplicação | 532 | Doses de reforço |
| 8 | escolas, educação, professores, aulas, trabalhadores, educação, profissionais, educação, alunos, escola, vacinação, aulas presenciais | 448 | Escolas e educação |
| 9 | presidente, vacina, comprar, presidente comprou, trump, vacinas, pra, dizer presidente, comprou, comprar vacina | 433 | Presidente e compra de vacinas |
| 10 | vídeo, vídeos, assista, vídeo, veja vídeo, assistir vídeo, assistir, veja, sobre vacinas, assista vídeo | 429 | Vídeos sobre vacinas |

4. DESAFIOS E APRENDIZADOS

Um dos principais desafios foi lidar com o grande volume e a heterogeneidade das mensagens do WhatsApp, exigindo filtragem e pré-processamento cuidadosos. A curva de aprendizado com o BERTopic também foi significativa, mas essencial para revelar padrões ocultos. O processo resultou em valiosos aprendizados sobre organização e análise de grandes bases textuais.

Agradecimentos. Esse trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unime-d-BH e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).

5. REFERÊNCIAS

1. Benevenuto F, Melo P. Misinformation Campaigns through WhatsApp and Telegram in Presidential Elections in Brazil. *Commun ACM*, 2024;67(8):72–7.
2. Resende G, Melo P, C. S. Reis J, Vasconcelos M, Almeida JM, Benevenuto F. Analyzing Textual (Mis) Information Shared in WhatsApp Groups. Em: *Proceedings of the 10th ACM Conference on Web Science*. NY, USA: ACM; 2019. p. 225–34.

Uma análise qualitativa da sumarização de diálogos de pacientes usando grandes modelos de linguagem aplicados a dados em português ruidosos e informais

Victor Macul^{5,6}, Anderson A. Ferreira^{1,3}, Leonardo Rocha², Washington Cunha¹, Ana Claudia Machado², Joao Marcos Campos¹, Gabriel Jallais¹, Adriana C. F. Viana², Elisa Tuler², Iago Araújo¹, Olivio Souza Neto⁵, Antonio Pereira⁵, Giordano de Pinho Souza^{5,7}, Joice Marques Pallone⁵, Mariana Aparecida Dumba Soares⁵, Thainá Fernandes Oliveira⁵, Welton Santos⁵, Marcos Andre Gonçalves¹

¹Universidade Federal de Minas Gerais. ²Universidade Federal de São João Del-Rei.

³Universidade Federal de Ouro Preto. ⁵Ana Health. ⁶Inspere

⁷Universidade Federal do Rio de Janeiro.

victor@anahealth.app

Resumo: Este estudo investiga o uso de Grandes Modelos de Linguagem para resumir diálogos reais no WhatsApp entre pacientes e uma equipe de saúde digital, em português. Criamos um conjunto de dados anonimizados e avaliamos os resumos gerados por LLaMA3 e Qwen2 usando prompts específicos. Avaliadores humanos analisaram os resumos quanto à cobertura, relevância, redundância e veracidade. Os resultados mostram que os LLMs conseguem produzir resumos eficazes, apesar de entradas informais, ruidosas e em uma língua de poucos recursos. O LLaMA3 teve desempenho superior ao Qwen2. A abordagem mostra potencial para melhorar a comunicação em saúde, especialmente em contextos com recursos limitados, como o Brasil.

Abstract: This study examines the application of Large Language Models (LLMs) to summarize authentic WhatsApp dialogues between patients and a digital healthcare team in Portuguese. We created an anonymized dataset and evaluated summaries generated by LLaMA3 and Qwen2 using tailored prompts. Human judges assessed them on coverage, relevance, redundancy, and veracity. Results show that LLMs can produce effective summaries despite noisy, informal, and low-resource language input. LLaMA3 slightly outperformed Qwen2. This approach shows promise for enhancing healthcare communication in resource-constrained contexts, such as Brazil.

Palavras-chave: personalização do cuidado; grandes modelos de linguagem; sumarização.

Nome do projeto: Consolidação de Estratégias de Inteligência Artificial Generativa Multimodal na Personalização de Serviços de Saúde Integral.

1. CONTEXTO E MOTIVAÇÃO

O uso crescente de prontuários eletrônicos e plataformas de mensagens instantâneas, como o WhatsApp, tem transformado a comunicação entre pacientes e equipes de saúde, especialmente em serviços digitais [1]. Esse cenário impõe um novo desafio: gerenciar de forma eficiente o alto volume de mensagens trocadas diariamente. Na atenção primária, os profissionais precisam revisar interações anteriores e outras informações relevantes para oferecer respostas precisas e personalizadas, o que consome tempo e afeta a agilidade do atendimento. No Brasil, onde grande parte da população ainda carece de acesso à atenção primária, soluções escaláveis como a Atenção Primária Digital representam uma alternativa promissora [5]. Empresas como a Ana Health têm adotado modelos de atendimento remoto via WhatsApp para ampliar o alcance dos serviços. No entanto, o volume de mensagens e a informalidade das comunicações dificultam a compreensão rápida e eficaz do histórico de cada paciente. Nesse contexto, o

uso de técnicas de sumarização automática, especialmente com Grandes Modelos de Linguagem, surge como estratégia promissora para apoiar a equipe de saúde na análise contextual das mensagens [2,3].

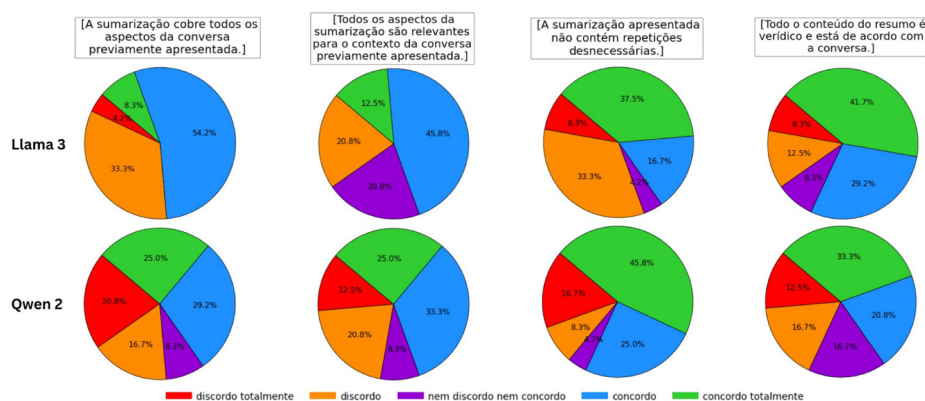
1.1 Objetivo

Avaliar o uso de LLMs de código aberto na tarefa de sumarizar mensagens de pacientes, visando apoiar profissionais de saúde com informações concisas, centradas no paciente e de fácil leitura.

2. METODOLOGIA E RESULTADOS

Este estudo foi conduzido com dados reais anonimizados de interações via WhatsApp entre pacientes e uma equipe multidisciplinar de saúde da empresa Ana Health, coletados entre outubro de 2021 e janeiro de 2024. Após aprovação do Comitê de Ética da UFMG (CAAE 80632524.4.0000.5149), foram analisadas mais de 200 mil mensagens em português. O conjunto de dados passou por um processo de pré-processamento que incluiu a remoção de caracteres especiais e mensagens padronizadas. A seguir, avaliou-se a qualidade textual das mensagens com base em três dimensões: tamanho, legibilidade (Flesch-Kincaid [4] adaptado para o português) e correção ortográfica, para compreender possíveis limitações na entrada dos LLMs. Na etapa de sumarização, agrupamos as mensagens de cada paciente em um único documento e avaliamos a capacidade de dois modelos de linguagem de código aberto na tarefa de gerar resumos informativos e precisos: LLaMA3 (8B) e Qwen2 (7B). Esses resumos foram avaliados por 24 profissionais de saúde (psicólogos e médicos) da Ana Health, por meio de um teste A/B. Cada resumo foi analisado segundo quatro critérios: cobertura, relevância, veracidade e ausência de redundância, utilizando uma escala de Likert de 5 pontos. Para complementar a análise, foram aplicadas métricas estatísticas como médias, desvios-padrão e o coeficiente Kappa-Cohen para medir o grau de concordância entre avaliadores. Todo o processamento foi realizado localmente, respeitando diretrizes de privacidade e segurança de dados.

Figura 1 - Distribuição de pontuações para cada critério, avaliando os resumos gerados pelas LLMs.



A análise da qualidade das mensagens revelou que a maioria dos textos é curta (56,12%), enquanto apenas 23,58% são longos. Textos curtos podem carecer de informações essenciais e os longos podem apresentar redundâncias que dificultam a identificação do conteúdo principal por LLMs. Os níveis médios de legibilidade indicam que as mensagens da equipe de saúde exigem maior escolaridade para compreensão em comparação às mensagens dos pacientes, que demonstram maior variabilidade. Outro fator relevante é a correção linguística: enquanto as mensagens da equipe de saúde têm vocabulário mais padronizado, os pacientes utilizam abreviações e gírias comuns ao WhatsApp, reduzindo a taxa média de palavras corretas para cerca de 44%. Essa baixa qualidade textual, especialmente nas mensagens dos pacientes, pode impactar negativamente os resultados da sumarização automática.

Apesar dessas questões, os resumos gerados por LLMs mostraram desempenho satisfatório sob

os critérios de cobertura, relevância, ausência de redundância e veracidade, conforme podemos observar nos gráficos da Figura 1. Ambos os modelos avaliados - LLaMA3 e Qwen2 - obtiveram mais de 50% de avaliações positivas ("concordo" ou "concordo totalmente") em todos os critérios. O LLaMA3 apresentou desempenho levemente superior nos aspectos de cobertura e veracidade, além de se destacar em uma análise qualitativa externa por produzir resumos objetivos, sem repetições desnecessárias ou informações incorretas. Esses resultados demonstram o potencial dos LLMs em gerar resumos úteis e contextualmente adequados, mesmo diante de mensagens de baixa qualidade, contribuindo para maior eficiência e clareza na comunicação entre pacientes e profissionais de saúde.

3. DESAFIOS E APRENDIZADOS

Apesar dos avanços dos modelos de linguagem de última geração, este estudo identificou desafios relevantes na aplicação de LLMs à sumarização de diálogos clínicos reais, especialmente quando redigidos em português informal, fragmentado e com abreviações. A baixa qualidade e estrutura dos dados exigiu avaliação criteriosa da robustez dos modelos frente a mensagens de baixa legibilidade e alta incidência de desvios linguísticos. O domínio clínico impôs restrições adicionais, dado o risco de erros comprometerem a interpretação e a tomada de decisão médica. O cenário analisado (comunicações assíncronas e temporalmente evolutivas) exigiu dos modelos manutenção de coerência contextual em interações extensas. Ainda assim, os LLMs produziram resumos clinicamente úteis, com avaliações positivas em cobertura, relevância, concisão e veracidade. Observou-se desempenho ligeiramente superior do LLaMA3, sobretudo em fidelidade e completude. O estudo ressaltou a importância de avaliações multifacetadas com especialistas para assegurar segurança e aplicabilidade clínica. Por fim, os achados reforçam o potencial dos LLMs como ferramentas estratégicas para ampliar o acesso à saúde em contextos com baixa cobertura e escassez de profissionais, promovendo maior equidade digital.

Agradecimentos. FINEP, EMBRAPPII, SEBRAE, CNPq, FAPEMIG, FAPESP, Ana Health, CIIA-Saúde e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação - INCT-TILD-IAR.

4. REFERÊNCIAS

1. Liu, S. et al. Leveraging large language models for generating responses to patient messages—a subjective analysis. *J. Am. Med. Informatics Assoc.* 31, DOI: 10.1093/jamia/ocae052 (2024)
2. El-Kassas, W. S., Salama, C. R., Rafea, A. A. & Mohamed, H. K. Automatic text summarization: A comprehensive survey. *Expert. systems with applications* 165, 113679, DOI: 10.1016/j.eswa.2020.113679 (2021).
3. Keszthelyi, D., Gaudet-Blavignac, C., Bjelogrić, M. & Lovis, C. Patient information summarization in clinical settings: Scoping review. *JMIR Med. Informatics* 11, e44639, DOI: 10.2196/44639 (2023).
4. Kincaid, J. P., Fishburne, R. P. J., Rogers, R. L. & Chissom, B. S. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Tech. Rep. 56*, Institute for Simulation and Training (1975).
5. Hone, T., Rasella, D., Barreto, M. L., Majeed, A. & Millett, C. Association between expansion of primary healthcare and racial inequalities in mortality amenable to primary care in Brazil: a national longitudinal analysis. *PLoS medicine* 14, e1002306, DOI: 10.1371/journal.pmed.1002306 (2017)

Uma revisão sistemática sobre o uso de aprendizado profundo na interpretação do exame de Papanicolaou

Franciele A. Barbosa¹, Bruna L. M. Santos², Cauan C. Marotta², Giulia Z. de Castro², Leonardo A. Ferreira², Lucas E. Alves², Lucas M. Henriques², Mateus A. da Silva², Mayara K. M. Gomides², Marcelo A. P. Xavier², Marilene V. de C. Monteiro², Frederico G. Guimarães², Walmir M. Caminhas²

¹Universidade Estadual de Montes Claros, Montes Claros, MG, Brasil.

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

{blms, cauanmarotta, giuliaz, leaufferreira, lucaseliasalves, lucasmont, mgomides, silvaamaral-mateus, mpascoal, marilenevale, fredericoguimaraes, caminhas}@ufmg.br, francielealvesb10@gmail.com.

Resumo: O câncer do colo do útero é o terceiro mais incidente e uma das principais causas de mortalidade por câncer entre mulheres no Brasil. Apesar da eficácia do exame de Papanicolaou como método de rastreamento, sua interpretação manual enfrenta limitações como análise subjetiva e elevada variabilidade inter-observador, além da escassez de citopatologistas em algumas regiões do país. Com o avanço da visão computacional, o uso de deep learning (DL), especialmente redes neurais convolucionais (CNNs), tem se mostrado promissor na detecção automática de células anormais em imagens citológicas. Este trabalho apresenta uma revisão sistemática, baseada no protocolo PRISMA, sobre o uso de DL na interpretação automatizada do Papanicolaou, considerando publicações entre 2010 e 2025. As buscas foram realizadas nas bases PubMed, IEEE Xplore, Springer e Google Scholar. Após triagens, nove artigos foram incluídos. Os resultados revelam o bom desempenho de arquiteturas como MobileNetV2, EfficientNet e DenseNet aplicadas em bases como Herlev e SIPaKMeD, indicando que as técnicas de DL oferecem avanços significativos na triagem citológica automatizada.

Palavras-chave: convolutional neural network; deep learning; Papanicolaou.

Nome do projeto: Interpretação Automatizada do Exame de Papanicolaou para Triagem do Câncer de Colo do Útero.

1. CONTEXTO E MOTIVAÇÃO

O câncer do colo uterino é o terceiro câncer mais incidente e a terceira causa de mortalidade por câncer entre as mulheres brasileiras, e está intimamente relacionado à infecção pelo vírus HPV. Para o ano de 2023, foram estimados 17.010 casos novos. O principal método de rastreamento é o exame citológico do colo uterino (Papanicolaou). Entretanto, o acesso a esse exame é desigual no Brasil, por regiões, por escolaridade e nível socioeconômico. Além disso, a análise manual do teste é sujeita à interpretação subjetiva e elevada variabilidade interobservador, o que dificulta a padronização diagnóstica, especialmente em locais com escassez de citopatologistas ^(1,2).

Nesse cenário, abordagens baseadas em visão computacional têm sido desenvolvidas para detectar e classificar automaticamente células cervicais anormais. O uso de aprendizado profundo (deep learning – DL), especialmente por meio de redes neurais convolucionais (CNNs), diferentemente dos métodos tradicionais de aprendizado de máquina, tem se mostrado eficaz por dispensar etapas manuais e complexas de pré-processamento e extração de características, permitindo o aprendizado direto das imagens e melhorando a precisão diagnóstica ⁽³⁾.

Estudos recentes demonstram que arquiteturas como ResNet, EfficientNet e outros modelos personalizados têm alcançado alta acurácia na classificação de células normais e anormais. Esses sistemas vêm sendo aplicados com sucesso em bases públicas como Herlev e SIPaKMeD, demonstrando robustez e boa generalização, além de viabilizar sistemas de diagnóstico assistido por computador (CAD) aplicá-

veis tanto a imagens isoladas quanto a lâminas inteiras, contribuindo para uma triagem mais eficiente, automatizada e acessível do câncer do colo do útero ⁽³⁾.

1.1 Objetivo

Realizar uma revisão sistemática pelo *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) ⁽⁵⁾, dos artigos encontrados na literatura que propõem técnicas de DL, para auxiliar o profissional de saúde na triagem e diagnóstico do câncer do colo do útero por meio da interpretação automatizada do exame de Papanicolaou. Portanto, o foco deste trabalho é revisar artigos que utilizam técnicas de DL para detectar/classificar células cervicais em imagens citopatológicas. Para isso, a revisão busca explorar questões relacionadas às técnicas utilizadas, formas de aplicação, disponibilidade e usabilidade de bases públicas, e os desafios para implementação clínica, visando contribuir para o aprimoramento da citopatologia cervical com métodos mais ágeis, precisos e acessíveis.

2. ATIVIDADES PRINCIPAIS

O projeto é desenvolvido por uma equipe interdisciplinar de alunos e pesquisadores das áreas de engenharia, computação e medicina, integrando conhecimentos técnicos e clínicos. As atividades são organizadas em duas reuniões semanais: uma técnica, focada no desenvolvimento dos modelos, e outra geral, para apresentação de resultados, discussão de artigos e planejamento das próximas tarefas.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A metodologia seguiu as diretrizes PRISMA, com busca nas bases *Google Scholar*, *PubMed*, *Springer* e *IEEE Xplore*, utilizando os termos “convolutional neural network”, “deep learning” e “Papanicolaou”, no período de 2010 a 2025. Foram inicialmente identificados 182 artigos, dos quais 10 duplicatas e 12 sem autoria definida foram excluídos. Após triagem por título e resumo, 135 artigos foram eliminados por não atenderem aos critérios de escopo ou disponibilidade. Na triagem de análise de métodos e conclusões, outros 18 foram excluídos. Ao final, nove artigos foram incluídos na análise, incluindo dois adicionados manualmente por relevância metodológica.

Os estudos analisados aplicaram arquiteturas como *DenseNet*, *MobileNetV2* e *EfficientNet*, utilizando principalmente as bases *Herlev* e *SIPaKMeD*, obtendo acurácias entre 83% e 99%. Modelos ensemble e técnicas de pré-processamento avançado mostraram melhor desempenho. Os resultados demonstram o potencial das redes neurais na interpretação automatizada do exame de Papanicolaou, apesar das diferenças metodológicas entre os trabalhos (5-12). Futuramente, pretende-se ampliar a revisão para outras aplicações médicas, como mamografia e citologia pulmonar, visando identificar técnicas replicáveis na triagem do câncer do colo do útero.

4. DESAFIOS E APRENDIZADOS

Durante o desenvolvimento do projeto, diversos desafios foram enfrentados, especialmente relacionados à obtenção e reprodutibilidade dos estudos. A manipulação de imagens de lâminas inteiras digitalizadas (*Whole Slide Images – WSI*) também se mostrou complexa, exigindo grande capacidade computacional e etapas de pré-processamento. Como principais aprendizados, destacam-se a familiarização com diferentes arquiteturas de DL aplicadas à citologia cervical e a valorização da padronização metodológica e do acesso aberto a dados e códigos como pilares da ciência reprodutível.

4. REFERÊNCIAS

1. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Dados e números sobre câncer do colo do útero – 2023. Rio de Janeiro: INCA; 2023. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/dados_e_numeros_col_o_22marco2023.pdf. Acesso em: 29/05/2025.

2. World Health Organization. WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention. 2nd ed. Geneva: WHO; 2021. Disponível em: <https://www.who.int/publications/i/item/9789240030824>. Acesso em: 29/05/2025.
3. Jiang P, Li X, Shen H, et al. A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis. *Artif Intell Rev*. 2023;56(Suppl 2):2687–2758. doi:10.1007/s10462-023-10588-z.
4. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269. doi:10.7326/0003-4819-151-4-200908180-00135.
5. Tan SL, Selvachandran G, Ding W, Paramesran R, Kotecha K. Cervical cancer classification from Pap smear images using deep convolutional neural network models. *Interdiscip Sci Comput Life Sci*. 2024;16:16–38. doi:10.1007/s12539-023-00589-5.
6. Prome TA, Tahiti TJ. MobileNetV2 Based Cervical Cancer Classification Using Pap Smear Images. In: *Proceedings of the 2024 IEEE International Conference on Advanced Computing and Data Science Applications (ACDSA)*. IEEE; 2024. doi:10.1109/acdsa59508.2024.10467627.
7. Archana MCP, Panicker JV. *Deep Convolutional Neural Networks for Multiclass Cervical Cell Classification*. In: *Proceedings of the 2022 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. IEEE; 2022. p. 376–80. doi:10.1109/WiSPNET54241.2022.9767129.
8. Hussain E, Mahanta LB, Das CR, Talukdar RK. A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue Cell*. 2020;65:101347. doi:10.1016/j.tice.2020.101347.
9. Manna A, Kundu R, Kaplun D, Sinitca A, Sarkar R. A fuzzy rank-based ensemble of CNN models for classification of cervical cytology. *Sci Rep*. 2021;11(1):14538. doi:10.1038/s41598-021-93783-8.
10. Gangrade J, Kuthiala R, Gangrade S, Singh YP, Manoj R, Solanki S. A deep ensemble learning approach for squamous cell classification in cervical cancer. *Sci Rep*. 2025;15(1):7266. doi:10.1038/s41598-025-91786-3.
11. Liang Y, Tang Z, Yan M, Chen J, Liu Q, Xiang Y. Comparison-Based Convolutional Neural Networks for Cervical Cell/Clumps Detection in the Limited Data Scenario. arXiv [Preprint]. 2018 Oct 14; arXiv:1810.05952. Disponível em: <https://arxiv.org/abs/1810.05952>.
12. Rahaman MM, Li C, Yao Y, Kulwa F, Wu X, Li X, Wang Q. DeepCervix: A deep learning -based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Comput Biol Med*. 2023;164:107987. doi:10.1016/j.compbimed.2023.107987.

Uso abusivo de bebidas alcoólicas: uma abordagem baseada em inquéritos e geração de dados sintéticos

Crizian S. Gomes¹, Regina T. I. Bernal¹, Larissa F. Araújo², Samuel N. Alves¹, Celso França¹, Juliana B. Souza¹, Flora V.S. O. Baldi¹, Thania M. Faria¹, Marcos A. Gonçalves¹, Jussara M. Almeida¹, Deborah C. Malta¹

¹Universidade Federal de Minas Gerais, ²Universidade Federal do Ceará

Contacting authors: criziansaar@gmail.com

Resumo: Gerar estimativas mais precisas e robustas de prevalências de consumo abusivo de bebidas alcoólicas para pequenas áreas de Belo Horizonte. Foram utilizados dados do inquérito Vigitel, sendo definido como “pequenas áreas” os nove clusters de vulnerabilidade, derivados do Índice de Vulnerabilidade da Saúde (IVS). Foram gerados dados sintéticos utilizando redes generativas adversariais condicionais (cGANs), uma técnica estado-da-arte baseada em aprendizado profundo, com o propósito de mitigar a escassez de dados, especialmente em áreas de maior vulnerabilidade, onde os dados tendem a ser mais escassos. As prevalências de consumo de bebidas alcoólicas foram calculadas para os dados reais e complementados por amostras da população sintética, nos nove clusters de vulnerabilidade. Os resultados demonstraram maior precisão nas estimativas obtidas considerando os dados sintéticos, especialmente nos grupos com maior vulnerabilidade, classificados como “elevado” e “muito elevado”.

Abstract: The goal was to generate more accurate and robust estimates of abusive alcohol consumption prevalence in small areas of Belo Horizonte. Data from the Vigitel survey were used, and “small areas” were defined as the nine vulnerability clusters derived from the Health Vulnerability Index. Synthetic data were generated using conditional generative adversarial networks (cGANs), a state-of-the-art deep learning technique, to mitigate data scarcity—especially in areas of higher vulnerability, where data tends to be sparse. Alcohol consumption prevalence was calculated using both real data and samples from the synthetic population across the nine vulnerability clusters. Results showed improved estimation accuracy with the inclusion of synthetic data, particularly in the most vulnerable groups classified as “high” and “very high”.

Palavras-chave: consumo abusivo de bebidas alcólicas; redes generativas adversariais; população sintética; inquéritos populacionais.

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

O consumo de bebidas alcoólicas constitui um importante desafio para a saúde pública, dada a magnitude de seus impactos sociais e econômicos. Esse comportamento está diretamente associado a desfechos adversos, como o aumento da violência, a ocorrência de acidentes de trânsito, a carga de doenças crônicas não transmissíveis (DCNT) e a redução da produtividade laboral¹.

O monitoramento do consumo de álcool é fundamental para orientar políticas públicas e reduzir seus danos. No Brasil, é feito por inquéritos como o Vigitel (Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico)², baseado em entrevistas por telefone fixo. Porém, essa metodologia apresenta limitações, como o viés amostral e baixa cobertura, especialmente em áreas de maior vulnerabilidade social, onde o acesso a esse tipo de serviço é reduzido, comprometendo a precisão das estimativas e dificultando a tomada de decisões para formulação de políticas públicas de saúde².

Nesse sentido, a aplicação de algoritmos de aprendizado de máquina, especialmente técnicas de aprendizado profundo, notadamente na geração de dados sintéticos, tem se consolidado como uma abordagem promissora para mitigar a escassez de dados coletados e reduzir vieses associados à distri-

buição e à representação em pequenas áreas geográficas³. Em especial, o uso de arquiteturas baseadas em Generative Adversarial Networks (GANs)⁴ permite sintetizar populações de dados com distribuições e correlações estatísticas similares às observadas nos dados reais, sem comprometer a privacidade ou a segurança da informação. Tal estratégia viabiliza avanços significativos no monitoramento das DCNT e de seus fatores de risco e proteção, possibilitando estimativas mais precisas e representativas em pequenas áreas.

1.1 Objetivo

Gerar estimativas mais precisas e robustas de prevalências de consumo abusivo de bebidas alcoólicas para pequenas áreas de Belo Horizonte.

2. ATIVIDADES PRINCIPAIS

As principais atividades foram: i) organização e tratamento do banco de dados; ii) geração de dados sintéticos para lidar com a ausência de dados; iii) atribuição das populações sintéticas aos *clusters*; iv) estimação das prevalências por *clusters*.

3. DESENVOLVIMENTO DO TRABALHO

As principais atividades foram: i) organização e tratamento do banco de dados; ii) geração de dados sintéticos para lidar com a ausência de dados; iii) atribuição das populações sintéticas aos *clusters*; iv) estimação das prevalências por *clusters*.

3.1 Metodologia

Estudo ecológico com dados do Vigitel para Belo Horizonte (MG), de 2006 a 2018. Os endereços dos entrevistados, fornecidos pelas operadoras, foram georreferenciados e vinculados ao Cadastro Nacional de Endereços para Fins Estatísticos, permitindo o linkage com os setores censitários.

Analisou-se o consumo abusivo de bebidas alcoólicas, definido como a ingestão de quatro ou mais doses padrão de álcool em uma única ocasião para mulheres, e cinco ou mais doses padrão para homens.

Utilizou-se como "pequenas áreas" nove *clusters* de vulnerabilidade, derivados do Índice de Vulnerabilidade à Saúde (IVS). Essa estratégia de reagrupamento do IVS foi proposta em estudo anterior⁵, com o objetivo de reduzir a variabilidade dos valores de IVS dentro de cada categoria.

Foram geradas populações sintéticas realistas por meio de modelos de aprendizado profundo, especificamente redes generativas adversariais condicionais (*Conditional Generative Adversarial Networks* – cGAN). O conjunto de dados utilizado inclui variáveis mistas, combinando atributos qualitativos (sexo e raça/cor) e quantitativos (idade em anos, escolaridade em anos e consumo de frutas em unidades/dia). A avaliação dos dados gerados considerou três dimensões — estrutura, qualidade e utilidade — com o objetivo de verificar sua capacidade de replicar padrões relevantes observados nos dados reais e seu potencial para subsidiar análises e políticas públicas baseadas em evidências.

A similaridade entre a distribuição marginal da variável consumo abusivo de bebidas alcoólicas nos dados reais e sintéticos foi avaliada utilizando a métrica KSComplement7 (0.968), indicando alta fidelidade da distribuição sintética em relação à original. As populações sintéticas foram atribuídas aos *clusters* por meio de modelos supervisionados, como o XGBoost⁶.

Por fim, foram estimadas as prevalências e os intervalos de confiança de 95% (IC95%) do consumo de bebidas alcoólicas, considerando tanto a população real (Vigitel) quanto a população sintética (integração entre dados reais e sintéticos), para cada um dos nove *clusters*, em três períodos distintos: T1 (2006–2009), T2 (2010–2013) e T3 (2014–2018).

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

3.2 Resultados

Ao analisar as prevalências de consumo abusivo de bebidas alcoólicas por *clusters*, nota-se uma grande variabilidade entre eles, com prevalências mais elevadas nos *clusters* classificados como de baixo risco — tanto nos dados reais (T1: 28,2%; IC95%: 25,1–31,3; T2: 26,25%; IC95%: 23,0–29,4; T3: 27,05%; IC95%: 22,9–31,0) quanto nos dados sintéticos (T1: 27,4%; IC95%: 24,7–30,0; T2: 27,7%; IC95%: 24,7–30,7; T3: 26,5%; IC95%: 22,6–30,5) (Tabela 1).

A comparação entre as estimativas baseadas exclusivamente na população real e aquelas obtidas a partir da integração com a população sintética revela maior precisão nas estimativas combinadas, sobretudo nos *clusters* com maior vulnerabilidade, como os classificados como “elevado” e “muito elevado”. No *cluster* ME-0, por exemplo, durante o período T1, o intervalo de confiança com dados reais variou de 7,3% a 24,4%, enquanto com a inclusão dos dados sintéticos essa variação foi reduzida para 12,4% a 18,2%, indicando maior precisão e confiabilidade nas estimativas geradas a partir da base integrada (Tabela 1).

Tabela 1 -Prevalência de consumo abusivo de bebidas alcoólicas nos *clusters* de vulnerabilidade para os períodos 2006 à 2009, 2010 à 2013 e 2014 à 2018.

| Clusters | População real | | | População real + sintética | | |
|----------|--------------------|--------------------|--------------------|----------------------------|---------------------|--------------------|
| | 2006 à 2009 | 2010 à 2013 | 2014 à 2018 | 2006 à 2009 | 2010 à 2013 | 2014 à 2018 |
| | % (IC 95%) | % (IC 95%) | % (IC 95%) | % (IC 95%) | % (IC 95%) | % (IC 95%) |
| BA-0 | 23,5 (21,0 - 25,9) | 23,9 (21,2 - 26,7) | 24,1 (21,2 - 26,9) | 23,3 (21,12 - 25,4) | 24,1 (21,51 - 26,6) | 24,0 (21,3 - 26,7) |
| BA-1 | 28,2 (25,1 - 31,3) | 26,2 (23,0 - 29,4) | 27,0 (22,9 - 31,0) | 27,4 (24,7 - 30,0) | 27,7 (24,7 - 30,7) | 26,5 (22,6 - 30,5) |
| MD-0 | 20,3 (17,7 - 22,9) | 20,7 (18,0 - 23,4) | 21,7 (18,9 - 24,5) | 19,5 (17,3 - 21,6) | 21,8 (19,3 - 24,3) | 21,7 (19,1 - 24,3) |
| MD-1 | 20,5 (18,0 - 22,9) | 18,9 (16,7 - 21,2) | 20,9 (18,3 - 23,4) | 20,1 (17,9 - 22,3) | 19,1 (17,0 - 21,1) | 20,6 (18,1 - 23,0) |
| EL-0 | 11,2 (7,4 - 14,9) | 18,4 (13,5 - 23,3) | 18,3 (13,1 - 23,5) | 12,0 (9,8 - 14,2) | 17,5 (14,8 - 20,2) | 18,9 (15,9 - 21,8) |
| EL-1 | 17,2 (9,3 - 25,1) | 19,6 (11,8 - 27,5) | 21,6 (14,4 - 28,9) | 17,8 (14,9 - 20,6) | 19,1 (15,9 - 22,3) | 21,3 (18,0 - 24,6) |
| EL-2 | 17,0 (13,3 - 20,7) | 15 (11,6 - 18,4) | 15,2 (11,8 - 18,5) | 17,6 (15,3 - 19,9) | 14,6 (12,4 - 16,8) | 15,2 (12,9 - 17,5) |
| ME-0 | 15,9 (7,3 - 24,4) | 20,3 (10,3 - 30,3) | 9,5 (3,2 - 15,7) | 15,3 (12,4 - 18,2) | 20,6 (17,3 - 23,9) | 9,2 (6,7 - 11,7) |
| ME-1 | 14,5 (8,8 - 20,2) | 14,0 (9,2 - 18,8) | 15,2 (9,3 - 21,1) | 14,3 (11,7 - 16,9) | 14,9 (12,1 - 17,7) | 15,2 (12,3 - 18,0) |

Nota: BA: baixo; MD: médio; EL: elevado; ME: muito elevado risco. IC95%: intervalo de confiança de 95%.

4. DESAFIOS E APRENDIZADOS

O desenvolvimento do projeto exigiu conhecimento aprofundado em aprendizado profundo, estatística e interpretação de resultados. Ao mesmo tempo, possibilitou avanços significativos na compreensão de como a integração de abordagens de inteligência artificial com sistemas tradicionais de monitoramento pode transformar a prática da saúde pública — especialmente na construção de bases de dados mais robustas e na geração de estimativas mais precisas e confiáveis para pequenas áreas.

Agradecimentos. Este trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unimed-BH e INCT- TILD-IAR

4. REFERÊNCIAS

As referências foram condensadas no link a seguir:

https://docs.google.com/document/d/1irF_1Ln5N-pBq3blbOgDvdGICMt01dDvcScKDnWKeCU/edit?tab=t.0

Uso de populações sintéticas em pequenas áreas geográficas para desfechos relacionados a Doenças Crônicas Não Transmissíveis: uma revisão de escopo

Edmar Geraldo Ribeiro, Flora Vitória Serena Oliveira Baldi, Luis Antônio Batista Tonaco, Crizian Saar Gomes, Regina Tomie Ivata Bernal, Juliana Bottoni de Souza, Leonardo Lemos Pena, Deborah Carvalho Malta

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

edmargribeiro@gmail.com

Resumo: A análise epidemiológica enfrenta desafios ao trabalhar com pequenas áreas geográficas, especialmente pelo baixo tamanho amostral, o que compromete a precisão de estimativas sobre Doenças Crônicas Não Transmissíveis (DCNTs). Neste contexto, o uso de populações sintéticas surge como alternativa para gerar dados mais robustos. Esta revisão de escopo buscou mapear o uso de populações sintéticas na modelagem de estimativas de indicadores de saúde relacionados às DCNTs em pequenas áreas, comparando com métodos tradicionais. A estratégia de busca, baseada no modelo PICO, resultou em seis estudos incluídos. Os resultados mostraram a superioridade das análises multinível em relação às estimativas de efeitos causais, usando dados sintéticos para DCNTs. As principais vantagens observadas foram a possibilidade de simular cenários hipotéticos, preencher lacunas de dados e incorporar heterogeneidade individual, reforçando o potencial dos dados sintéticos na produção de estimativas mais realistas e úteis para a tomada de decisão em saúde pública.

Palavras-chave: populações sintéticas; Doenças Crônicas Não Transmissíveis; estimativas populacionais; paciente sintético.

Abstract: Epidemiological analysis faces challenges when working with small geographic areas, particularly due to the small sample sizes, which compromise the accuracy of estimates regarding Noncommunicable Diseases (NCDs). In this context, the use of synthetic populations emerges as an alternative to generate more robust data. This scoping review aimed to map the use of synthetic populations in modeling health indicator estimates related to NCDs in small areas, comparing them to traditional methods. The search strategy, based on the PICO model, resulted in the inclusion of six studies. The findings showed the superiority of multilevel analyses in estimating causal effects using synthetic data for NCDs. The main advantages observed were the ability to simulate hypothetical scenarios, fill data gaps, and incorporate individual heterogeneity, reinforcing the potential of synthetic data in producing more realistic and useful estimates for public health decision-making.

Keywords: synthetic populations; Noncommunicable Diseases; population estimates; synthetic patient.

Nome do projeto: O uso de modelos preditivos de inteligência artificial na prevenção de doenças crônicas não transmissíveis.

1. CONTEXTO E MOTIVAÇÃO

As doenças crônicas não transmissíveis (DCNTs) representam um dos principais desafios para a saúde pública mundial. Essas condições são responsáveis por um número significativo de óbitos precoces, comprometimento do bem-estar devido a restrições funcionais importantes e impactos financeiros negativos para indivíduos, comunidades e a sociedade como um todo. Diante desse cenário, investigações epidemiológicas em saúde frequentemente recorrem a dados secundários de sistemas de informação e inquéritos populacionais. Isso ocorre porque os pesquisadores, ao analisarem áreas geográficas menores,

enfrentam limitações de tamanho amostral, o que compromete a obtenção de resultados estatisticamente robustos. Nesse contexto, o uso de populações sintéticas surge como uma alternativa promissora para ampliar o tamanho amostral e aumentar a confiabilidade das análises em pequenas áreas, atendendo à demanda por dados de qualidade na formulação de políticas públicas mais direcionadas e adequadas à realidade local.

1.1 Objetivo

Esse estudo pretende mapear e sistematizar o estado da arte sobre o uso de populações sintéticas para modelagem de estimativas de indicadores de doenças crônicas não transmissíveis em pequenas áreas geográficas.

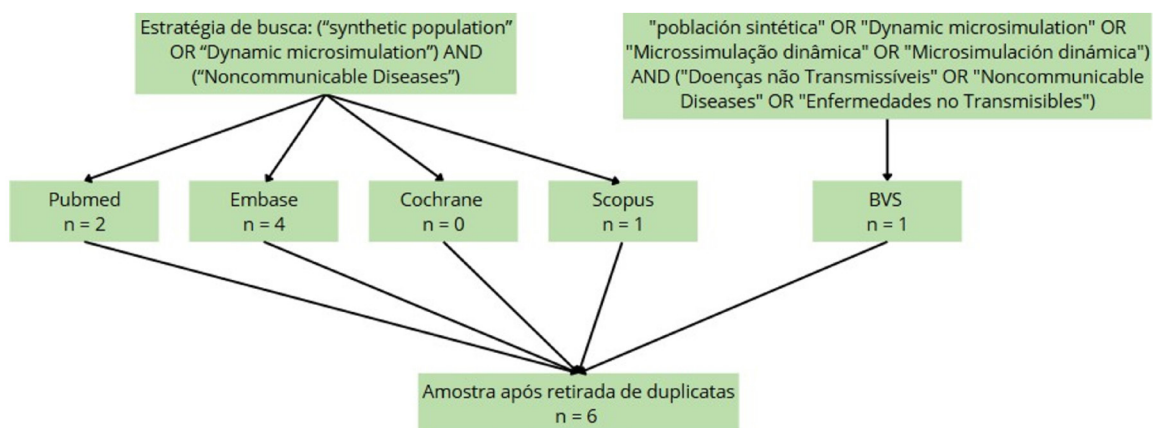
2. ATIVIDADES PRINCIPAIS

- Elaboração da estratégia PICO (Paciente/População, Intervenção, Comparação e Resultado);
- Consulta à bibliotecária para elaboração da estratégia de busca;
- Busca nas bases de dados indicadas;
- Sistematização dos estudos encontrados;
- Estabelecimento de critérios para inclusão e exclusão dos estudos encontrados;
- Síntese dos resultados dos estudos incluídos na revisão.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

A revisão de literatura é o primeiro passo para pesquisa científica e conduzida durante todo o desenvolvimento do trabalho. A escolha por uma revisão de escopo justifica-se pela ampla coleta de informações, sua característica de coleta de informações, sem distinção entre tipos de estudo e métodos, incluindo todos os trabalhos existentes, algo que contribui para uma revisão diversa ⁽²⁾. Foi elaborada a estratégia PICO para a formulação da pergunta de pesquisa: "Em pequenas áreas geográficas, como a utilização de populações sintéticas para modelagem de estimativas para indicadores de doenças crônicas não transmissíveis se compara aos métodos tradicionais de coleta de dados de saúde?". A partir dessa estratégia, com bibliotecário, foram desenvolvidas buscas e estas foram realizadas no dia 20/05/2025. O número de artigos encontrados em cada base foi sintetizado (Figura 1). Os critérios de inclusão foram: artigos publicados nos últimos 10 anos, idiomas inglês, português e espanhol, de acesso livre, que abordavam o objetivo deste trabalho. Foram excluídos os estudos duplicados. Na primeira busca foram encontrados 8 estudos; após aplicação dos critérios de inclusão e exclusão, restaram 6 estudos. Os estudos foram organizados em uma planilha no software Excel e analisadas pelo resumo, objetivo, metodologia e resultados.

Figura 1 - Estratégias de busca, bases selecionadas e tamanho da amostra.



Dos seis estudos identificados, cinco são artigos originais e um é publicação em anais de congresso. Cinco foram escritos em inglês e um em espanhol. Quatro artigos foram publicados nos últimos cinco anos e um em 2018. Três estão em revistas classificadas como A1 no Qualis-Capes (2017– 2020) e dois não foram classificados na plataforma Sucupira. Um artigo é essencialmente metodológico ⁽³⁾ e propõe uma comparação entre estudos ecológicos e análises multiníveis na estimativa de efeitos causais individuais em dados hierárquicos. Essa comparação foi feita a partir do uso de dados sintéticos para geração dos dados hierárquicos utilizados para construção de Diagramas Causais Hierárquicos (DAGs), definição de relações causais e clusterização. Esse estudo resultou na superioridade das análises multiníveis por terem menor viés residual e heterogeneidade nas estimativas.

Um artigo trata-se de uma revisão de escopo ⁽⁴⁾ que mapeia características metodológicas de Modelos de Microsimulação (MSMs) aplicados à avaliação de políticas alimentares relacionadas a doenças não transmissíveis (DCNTs). Assim, nove artigos contidos na revisão usaram dados de pesquisas e inquéritos para gerar dados sintéticos realísticos, enquanto doze artigos usaram dados sintéticos para simular um desfecho individual a partir de uma coorte guiado pela matriz de correlação de fatores de risco. Nesses estudos com uso de dados sintéticos, algumas limitações apresentadas foram o viés de recall, no qual a precisão ou integridade das lembranças dos participantes para determinar certos eventos ou fatos do ocorridos é afetada, e a dependência de estudos observacionais como ponto de partida. Sobre isso, foi apontado uma facilitação na análise de equidade considerando impacto diferenciado dos desfechos em subgrupos socioeconômicos.

Quatro estudos da estratégia de busca ⁽⁵⁻⁸⁾ realizaram microsimulações dinâmicas simularam e analisaram os resultados para seus respectivos desfechos. Um sobre o impacto da redução da obesidade na prevalência da hipertensão ⁽⁵⁾, outro sobre projeção de fatores de risco cardiometabólicos, a incidência de diabetes tipo II e de acidente vascular cerebral isquêmico ⁽⁶⁾, outro sobre saúde de idosos e expectativa de vida, considerando características demográficas, socioeconômicas, fatores de risco e transições entre estados de saúde ⁽⁷⁾, e o último avaliou impactos de intervenções simuladas para redução de peso da população do estudo ⁽⁸⁾. Em comum, os quatro estudos apontaram como vantagens do uso de dados sintéticos: I) modelagem de heterogeneidade individual (incorporação de características únicas e comportamentos dinâmicos permitem análises realistas); II) projeção de políticas públicas e seus impactos por cenários hipotéticos (não necessitando de dados longitudinais reais de longo prazo); e III) preenchimento de informações faltantes (ex.: mortalidade por status socioeconômico) via correlações estimadas e tendências externas.

4. DESAFIOS E APRENDIZADOS

Foram identificados quatro estudos experimentais, um metodológico e uma revisão de escopo. Os estudos experimentais utilizaram a técnica de microsimulação dinâmica, destacando vantagens no uso de dados sintéticos para estimar a prevalência de DCNTs em pequenas áreas, por meio da simulação de desfechos individuais. A principal limitação da revisão foi o baixo número de estudos encontrados. Como próximo passo, sugere-se testar os modelos com dados sintéticos gerados a partir de bases nacionais, comparando os resultados com dados reais e com os estudos analisados.

Agradecimentos. Este trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unimed-BH e INCT- TILD-IAR.

5. REFERÊNCIAS

As referências foram condensadas no link a seguir:

https://docs.google.com/document/d/1HI0yfDndBL_3npWlv8uBgJGFUO4foqvW6kG6PgyA-jD0/edi t?usp=sharing

Uso de Redes Generativas Adversariais (cGANs) para geração de populações sintéticas no apoio às políticas públicas em saúde

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

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

celsofranca@dcc.ufmg.br

Resumo: Este estudo propõe o uso de redes generativas adversariais condicionais (cGAN) para gerar entrevistas sintéticas do inquérito Vigitel, com base em dados de Belo Horizonte (2006-2018). O modelo foi treinado para sintetizar registros com atributos categóricos e numéricos, preservando a similaridade estatística nas distribuições marginais e correlações entre variáveis. Modelos preditivos treinados com dados sintéticos apresentaram desempenho comparável aos treinados com dados reais na classificação do status de tabagismo. A geração sintética mostrou-se uma alternativa viável para mitigar vieses e ampliar a representatividade em inquéritos de saúde, especialmente em populações vulneráveis.

Palavras-chave: população sintética; Redes Adversariais Generativas condicionais; Vigitel.

Abstract: This study proposes the use of conditional generative adversarial networks (cGANs) to generate synthetic interviews from the Vigitel survey, based on data from Belo Horizonte (2006-2018). The model was trained to synthesize records with categorical and numerical attributes, preserving statistical similarity in marginal distributions and variable correlations. Predictive models trained on synthetic data achieved performance comparable to those trained on real data in classifying smoking status. Synthetic data generation has proven to be a viable approach to mitigating bias and enhancing representativeness in health surveys, particularly for vulnerable populations.

Keywords: synthetic population; conditioned Generative Adversarial Networks; Vigitel.

Nome do projeto: Uso de Algoritmos de Machine Learning e de Inteligência Artificial para Predição em Pequenas áreas de Indicadores de Saúde Relacionados às Doenças Crônicas Não Transmissíveis (Deborah Malta, Marcos A. Gonçalves e Jussara M. Almeida).

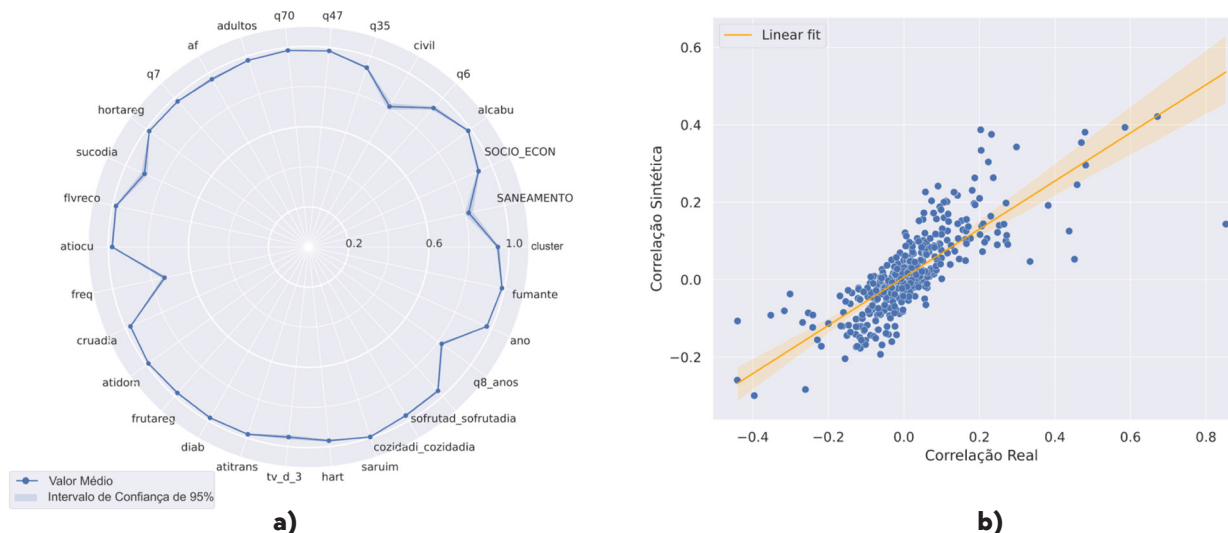
1. CONTEXTO E MOTIVAÇÃO

A geração de dados sintéticos tabulares tem ganhado destaque por sua capacidade de replicar características estatísticas e estruturais de dados reais. Essa abordagem é cada vez mais relevante em áreas como saúde, finanças, cibersegurança e sensoriamento remoto [1,3]. Dados sintéticos oferecem vantagens como: (i) treinamento e avaliação de modelos de IA sem exposição de dados sensíveis; (ii) geração de grandes volumes de dados quando os reais são escassos; (iii) balanceamento de conjuntos desbalanceados, favorecendo a representatividade e o aprendizado de casos raros; e (iv) simulação de cenários controlados para analisar limites e comportamentos específicos de modelos, muitas vezes não observáveis em dados reais [1].

No Brasil, inquéritos populacionais como o Vigitel (Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico) são realizados periodicamente para monitorar indicadores de saúde e orientar políticas públicas. No entanto, a coleta via telefone fixo introduz viés amostral e limitações de cobertura, especialmente em áreas socialmente vulneráveis, onde esse tipo de acesso é restrito. Isso leva à sub-representação de populações marginalizadas, sobretudo de baixa renda ou pertencentes a grupos racialmente discriminados, comprometendo a representatividade dos dados e a equidade das análises. Como consequência, fatores de risco tendem a ser subestimados nessas regiões, o que pode

prejudicar a formulação de políticas baseadas em evidências e dificultar a alocação eficiente de recursos em saúde [2].

Figura 1 - Similaridade estatística entre dados reais e sintéticos para diferentes atributos/eixos (a). Correlação entre pares de atributos reais e sintéticos (b).



1.1 Objetivo

Este trabalho propõe a geração de entrevistas sintéticas do sistema Vigitel por meio de redes generativas adversariais condicionais (cGANs), uma variação das GANs tradicionais que incorpora variáveis condicionais no processo de geração. Essa abordagem permite maior controle sobre as características dos dados gerados e apresenta melhor desempenho em cenários com distribuições enviesadas e variáveis discretas desbalanceadas [3]. O conjunto de dados inclui variáveis mistas, combinando atributos categóricos (como sexo e cor/raça) e discretos (como idade, escolaridade e consumo de frutas). A avaliação dos dados gerados buscou verificar sua capacidade de replicar padrões relevantes dos dados reais e seu potencial para apoiar análises e políticas públicas baseadas em evidências. Além de contribuir para a redução da variabilidade amostral - especialmente em setores com população mais vulnerável - a geração de populações sintéticas também viabiliza a criação de cenários de interesse (e.g., aumento da proporção de indivíduos em situação de vulnerabilidade), facilitando a simulação e avaliação de políticas alternativas de gestão da saúde pública em contextos fictícios, porém realistas.

2. ATIVIDADES PRINCIPAIS

As principais atividades desenvolvidas ao longo deste trabalho foram: (i) pré-processamento e engenharia de atributos; (ii) treinamento do modelo cGAN; (iii) geração de entrevistas sintéticas; e (iv) avaliação da qualidade dos dados sintéticos.

3. DESENVOLVIMENTO DO TRABALHO (METODOLOGIA E RESULTADOS)

Para os experimentos, utilizou-se uma amostra do Vigitel de Belo Horizonte (2006–2018), com 21.764 entrevistas e 30 atributos, sendo 6 numéricos e 24 categóricos. Os dados foram divididos em treinamento, validação e teste usando validação cruzada estratificada com 5 folds. Após o treinamento, a cGAN gerou entrevistas sintéticas, avaliadas em três dimensões: (i) **estrutura** —conformidade com restrições dos dados reais, como intervalos válidos, categorias corretas e nomes de colunas preservados; (ii) **qualidade** — similaridade estatística entre dados reais e sintéticos, observando distribuições marginais e correlações

entre variáveis; (iii) **utilidade** — desempenho em uma tarefa preditiva, na qual modelos de aprendizado de máquina (Regressão Logística, Random Forest e XGBoost) foram treinados com ambos os conjuntos (reais e sintéticos) para classificar o status de tabagismo. Os dados sintéticos foram considerados úteis quando os modelos apresentaram desempenho estatisticamente equivalente (Macro-F1) àquele obtido com os dados reais. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, sob o parecer nº 6.538.883.

A Figura 1(a) apresenta a similaridade estatística entre as distribuições marginais dos atributos reais e seus correspondentes sintéticos. Dos 30 atributos avaliados, 26 obtiveram escores de similaridade superiores a 0,90, 3 acima de 0.70 e 1 com score próximo a 0.67, indicando alta fidelidade na reprodução das características univariadas do conjunto de dados original. A Figura 1(b) exibe as correlações pareadas entre atributos reais (eixo x) e atributos sintéticos correspondentes (eixo y). A relação aproximadamente linear entre as correlações sugere que o modelo cGAN foi eficaz em preservar as dependências multivariadas presentes nos dados reais. Em relação à utilidade, a Tabela 1 mostra que modelos treinados com dados sintéticos obtêm desempenho estatisticamente equivalente aos treinados com dados reais. Resultados apresentados em Micro-F1 e Macro-F1 (média de 5-folds e intervalo de confiança de 95% em parênteses).

Tabela 1 - Utilidade dos dados sintéticos na tarefa de predição de tabagismo.

| Modelo | Dados Reais | | Dados Sintéticos | |
|----------------------------|-------------|-----------|------------------|-----------|
| | Micro-F1 | Macro-F1 | Micro-F1 | Macro-F1 |
| Logistic Regression | 87.7(0.0) | 47.7(1.8) | 87.7(0.0) | 46.7(0.0) |
| Random Forest | 87.5(0.2) | 50.9(1.3) | 87.6(0.1) | 46.8(0.1) |
| XGBoost | 87.7(0.1) | 47.8(1.9) | 87.7(0.0) | 46.7(0.0) |

4. DESAFIOS E APRENDIZADOS

O projeto demandou domínio em aprendizado profundo, estatística e interpretação de resultados, além de comunicação clara. A experiência consolidou habilidades técnicas e reforçou a importância da reprodutibilidade e do rigor experimental.

Agradecimentos. Este trabalho conta com o apoio do CIIA-Saúde, CNPq, FAPESP, FAPEMIG, Unime-d-BH e INCT-TILD-IAR.

5. REFERÊNCIAS

1. Lautrup AD, Hyrup T, Zimek A, Schneider-Kamp P. Systematic review of generative modelling tools and utility metrics for fully synthetic tabular data. *ACM Computing Surveys*. 2024 Dec 9;57(4):1-38.
2. Bernal RT, de Carvalho QH, Pell JP, Leyland AH, Dundas R, Barreto ML, Malta DC. A methodology for small area prevalence estimation based on survey data. *International Journal for Equity in Health*. 2020 Dec;19:1-0.
3. Majeed A, Hwang SO. Moving Conditional GAN Close to Data: Synthetic Tabular Data Generation and its Experimental Evaluation. *IEEE Transactions on Big Data*. 2024 Aug 13.

Utilização de RAG em grandes modelos de linguagem para aprimorar instruções de uso de medicamentos

Davi de Jesus¹, Antônio de Souza Júnior¹, Isaias de Oliveira², Zilma Reis²,
Leonardo Rocha¹

¹Universidade Federal de São João Del-Rei, São João del-Rei, MG. ²Universidade Federal de Minas Gerais, Belo Horizonte, MG.

davireisjesus@aluno.ufsj.edu.br, antoniopereira@aluno.ufsj.edu.br,
isaias-ramos@ufmg.br, zilma@ufmg.br, lcrocha@ufsj.edu.br

Resumo: Sistemas padronizados de prescrição de medicamentos buscam tornar as instruções mais legíveis, embora ainda enfrentem limitações quanto à personalização. Este estudo apresenta uma abordagem inovadora baseada em Grandes Modelos de Linguagem (LLMs) combinados com Geração Aumentada por Recuperação (RAG), utilizando bulas de medicamentos como fonte de informação. Avaliamos três modelos em 119 casos ambulatoriais, com as instruções geradas sendo analisadas por médicos em termos de adequação, clareza e personalização. Os resultados demonstraram melhorias na adequação (100 vs 93,0) e na clareza (95,0 vs 90,0), além da redução de erros e alucinações. A integração com fontes confiáveis aumentou a segurança das instruções, embora a validação por profissionais de saúde continue sendo indispensável.

Abstract: Standardized prescription systems aim to make instructions more readable, although they still face limitations in terms of personalization. This study presents a novel approach that combines Large Language Models (LLMs) with Retrieval Augmented Generation (RAG), utilizing drug package inserts as the primary source of information. We evaluated three models in 119 outpatient cases, with the generated instructions being analyzed by physicians regarding adequacy, clarity and personalization. The results demonstrated improvements in adequacy (100 vs. 93.0) and clarity (95.0 vs. 90.0), reducing errors and hallucinations. Integration with reliable sources increased the reliability, although validation by health professionals remains essential.

Palavras-chave: inteligência artificial, prescrição médica, Geração Aumentada de Recuperação.

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

Sistemas de prescrição eletrônica têm melhorado a legibilidade e reduzido erros em receitas manuscritas, mas ainda enfrentam limitações na comunicação clara e personalizada ao paciente. Estudos mostram que as instruções muitas vezes não explicam riscos importantes nem consideram o perfil individual, comprometendo a adesão e a segurança ⁽¹⁾. Além disso, há inconsistências entre prescritores e fragmentação nos dados sobre uso de medicamentos. Recentemente exploramos o uso de Grandes Modelos de Linguagem (LLMs) para gerar instruções personalizadas com base no padrão nacional brasileiro de prescrição eletrônica ⁽²⁾. Avaliamos diferentes estratégias de *prompting* com GPT-4 e Llama 3.1, observando melhorias na clareza e adequação, embora as limitações descritas persistam em casos clínicos complexos.

1.1 Objetivo

O objetivo deste trabalho é aprimorar a capacidade e a precisão dos modelos de LLMs na geração de prescrições de medicamentos, utilizando a técnica de Geração Aumentada por Recuperação RAG, a fim de apoiar profissionais de saúde nesse processo.

2. METODOLOGIA E RESULTADOS

Para superar as limitações identificadas nas soluções atuais (2), este trabalho propõe uma avanço introduzindo o uso de RAG (3), integrando informações externas aos LLMs para enriquecer o contexto na geração de instruções de prescrição. Especificamente, utilizamos dados das bulas da Agência Nacional de Vigilância Sanitária (Anvisa - <https://consultas.anvisa.gov.br/#/bulario/>). Considerando o alto custo computacional e o risco de ruído ao processar o conteúdo completo da bula, adotamos uma abordagem de RAG sob demanda, restringindo a recuperação apenas à seção de administração do medicamento.

Para avaliar nossa proposta, foi elaborada uma base de dados composta por 119 casos clínicos simulados, com base em 83 princípios ativos distintos, todos seguindo o padrão nacional e-SUS Atenção Primária à Saúde (APS). Esse processo envolveu cinco profissionais da saúde de diferentes especialidades, assegurando diversidade e realismo nos cenários apresentados aos modelos. Três modelos de linguagem baseados no LLaMA 3.1 8B foram utilizados para geração das 119 prescrições: **A) Modelo 1:** utilizando um prompt referência proposto em estudo anterior

(2); **B) Modelo 2:** utilizando um prompt aprimorado com uma orientação mais estruturada, desenvolvido neste estudo; e **C) Modelo 3:** proposto neste trabalho que utiliza RAG para introduzir trechos das bulas da Anvisa. A avaliação foi realizada com base nas instruções geradas para cada um dos 119 casos clínicos pelos três modelos testados. Um total de 357 respostas (119 por modelo) foi submetido à avaliação cega por cinco médicos, que consideraram quatro critérios: adequação, clareza, personalização e comparação com padrão ideal. Além disso, foi utilizada a similaridade de cosseno entre as orientações dos modelos e a referência humana, usando representações vetoriais (embeddings), que transformam os textos em vetores capazes de capturar sua semântica. A frequência de erros foi comparada entre os modelos com teste qui-quadrado de McNemar e o nível de significância utilizado foi de 5%.

Os resultados mostraram que o Modelo 3 (com RAG) obteve o melhor desempenho nos critérios de **adequação e clareza**. As medianas observadas foram de 93,0 (intervalo interquartil: 40,0) para o Modelo 1, 94,0 (53,3) para o Modelo 2 e 100,0 (15,0) para o Modelo 3, com **p = 0,001** para adequação. Para clareza, as medianas foram de 90,0 (60,0), 92,0 (55,0) e 95,0 (24,0), respectivamente, com **p = 0,012**. As análises não identificaram diferenças estatisticamente significativas em personalização ou similaridade textual. Na avaliação dos tipos de erro (4), o Modelo 3 também apresentou **reduções relevantes**, conforme observado na Tabela 1.

Tabela 1 - Erros nos resultados gerados pelos diferentes modelos.

| | Llama3.1 8b Modelo1 (n=83) [95% IC] | Llama3.1 8b Modelo2 (n=83) [95% IC] | Llama3.1 8b Modelo RAG (n=83) [95% IC] | P-valor¹ | P-valor² | P-valor³ |
|---|--|--|---|----------------------------|----------------------------|----------------------------|
| Erro tipo 1: Uso incorreto do medicamento | 31 (37.3%)[27.4 - 48.0] | 24 (28.9%)[19.9 - 39.2] | 14 (16.9%)[9.9 - 25.9] | 0.143 | <0.001 | 0.013 |
| Erro tipo 2: Instruções contraditórias ou vagas | 31 (37.3%)[27.4 - 48.0] | 18 (21.7%)[13.8 - 31.3] | 16 (19.3%)[11.8 - 28.6] | 0.019 | 0.014 | 0.832 |
| Erro tipo 3: Falta de instruções essenciais | 25 (30.1%)[21.0 - 40.5] | 28 (33.7%)[24.2 - 44.3] | 16 (19.3%)[11.8 - 28.6] | 0.678 | 0.078 | 0.017 |
| Erro tipo 4: Erros factuais (não médicos) | 17 (20.5%)[12.8 - 30.0] | 8 (9.6%)[4.5 - 17.2] | 3 (3.6%)[0.9 - 9.1] | 0.078 | 0.001 | 0.180 |
| Erro tipo 5: Instruções sem evidências científicas | 0 (0%)[0.0 - 0.23] | 0 (0%)[0.0 - 0.23] | 2 (2.4%)[0.4 - 7.3] | ... | ... | ... |
| Erro tipo 6: Instruções não solicitadas | 9 (10.8%)[5.4 - 18.7] | 1 (1.2%)[0.1 - 5.2] | 6 (7.2%)[2.9 - 14.1] | 0.021 | 0.549 | 0.125 |
| Erro tipo 7: Alucinações (informações inventadas/fabricadas) | 10 (12.0%)[6.2 - 20.2] | 6 (7.2%)[2.9 - 14.1] | 4 (4.8%)[1.5 - 10.8] | 0.388 | 0.180 | 0.687 |

Modelo 1 vs. Modelo 2. (2) Modelo 1 vs. Modelo 3. (3) Modelo 2 vs. Modelo 3.

Na avaliação dos sete tipos de erro, o Modelo 3 (Llama 3.1 com RAG), proposto neste trabalho, apresentou os melhores resultados. Houve **redução significativa nos erros tipo 1** (instruções que induzem ao uso incorreto do medicamento), considerados os mais críticos do ponto de vista clínico. Também foram observadas **melhoras relevantes nos erros tipo 2** (respostas vagas ou incorretas) e **tipo 3** (omissão de informações essenciais), em comparação aos modelos anteriores, especialmente quando comparado apenas ao prompt reestruturado (Modelo 2). Nos **erros tipo 4** (falhas pontuais), o Modelo 3 também superou estatisticamente o Modelo 1. Já nos **erros tipo 5** (falhas metodológicas específicas), a ocorrência foi baixa e sem diferenças estatísticas entre os modelos. Para os **erros tipo 6** (informações não solicitadas), o RAG não trouxe melhora significativa, possivelmente devido ao excesso de informações. Quanto aos **erros tipo 7** (alucinações), apesar de não haver diferença estatística, o Modelo 3 teve o menor número absoluto de ocorrências (4 em 83 casos), indicando maior robustez.

4. DESAFIOS E APRENDIZADOS

A aplicação de RAG em LLMs de código aberto demonstrou-se promissora, mas ainda com desafios técnicos relevantes. A extração e organização das bulas demandaram atenção especial à segmentação semântica dos conteúdos e à indexação precisa, para garantir a recuperação eficiente das informações mais relevantes durante a geração das instruções. O principal desafio foi equilibrar a riqueza informativa com a clareza textual, evitando sobrecarga de conteúdo e ambiguidades. No âmbito formativo, o aluno de Iniciação Científica envolvido no projeto desenvolveu competências em fundamentos de LLMs, técnicas de recuperação de informação e avaliação clínica, atuando em um ambiente multidisciplinar que integra inteligência artificial e saúde pública. Os aprendizados vão da curadoria de dados à avaliação crítica dos modelos, com ênfase na segurança, aplicabilidade no contexto do SUS e responsabilidade ética. Como próximos passos, busca-se avaliar a escalabilidade da abordagem em cenários reais e ampliar a base de conhecimento dos modelos, sempre com validação profissional e foco na segurança do paciente.

Agradecimentos. O estudo foi realizado com recursos do CNPq-Inova Fiocruz, Fundação Bill & Melinda Gates e FAPEMIG. A autora ZNSR é bolsista CNPq 305837/2021-4. O autor LR é bolsista do CNPq 304837/2025-3.

5. REFERÊNCIAS

1. Klepser, D., Lanham, A., and Cochran, G. (2016). Electronic prescriptions: opportunities and challenges for the patient and pharmacist. AHCT.
2. Reis, Z. S. N., Pagano, A. S., Ramos De Oliveira, I. J., Dias, C. D. S., Lage, E. M., Mineiro, E. F., Varella Pereira, G. M., De Carvalho Gomes, I., Basilio, V. A., Cruz Correia, R. J., De Jesus, D. D. R., De Souza Júnior, A. P., and Rocha, L. (2024). Evaluating Large Language Model-supported instructions for medication use: First steps toward a comprehensive model. Mayo Clinic Proceedings: Digital Health, 2.
3. Zakka, C., Shad, R., Chaurasia, A., Dalal, A. R., Kim, J. L., Moor, M., Fong, R., Phillips, C., Alexander, K., Ashley, E., Boyd, J., Boyd, K., Hirsch, S., Vogelsong, M. A., Cunningham, J. P., and Hiesinger, W. (2024). Almanac — retrieval-augmented language models for clinical medicine. NEJM AI, 1.
4. Tam, T. Y. C., Sivarajkumar, S., Kapoor, S., Stolyar, A. V., Polanska, K., McCarthy, K. R., Osterhoudt, H., Wu, X., Visweswaran, S., Fu, S., Mathur, P., Cacciamani, G. E., Sun, C., Peng, Y., and Wang, Y. (2024). A framework for human evaluation of large language models in healthcare derived from literature review. npj Digit. Med., 7.

Realização



Apoio



Parceria



Financiadores

