Protótipo de Dashboard para mostrar explicações contrafactuais para dados de doença renal crônica

Mateus Cardoso Oliveira[1], Eduardo Coppetti Radaelli[1], Isabel Cristina Reinheimer[1], Joaquim Vinicius Carvalho Assunção[1], Carlos Eduardo Poli de Figueiredo[2], Luís Alvaro de Lima Silva[1]

[1]Centro de Tecnologia
Universidade Federal de Santa Maria (UFSM) – Santa Maria, RS – Brazil

[2]Programa de Pós-Graduação em Medicina e Ciências da Saúde
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) – Porto Alegre, RS – Brazil

oliveira.mateus@acad.ufsm.br

{ecradaelli, joaquim, luisalvaro}@inf.ufsm.br

cristinareinheimer@gmail.com, cepolif@pucrs.br

Abstract. *The use of Artificial Intelligence (AI) has been demonstrating significant potential in supporting healthcare professionals. In this study, a prototype dashboard was developed to assist physicians in clinical decision-making by incorporating counterfactual explanations based on chronic kidney disease data. The prototype provides an intuitive visualization of outcomes and enables streamlined data entry. The findings suggest that the proposed tool can improve the interpretation of results and facilitate the adoption of AI technologies within the medical domain.*

Resumo. *O uso de Inteligência Artificial (IA) tem demonstrando significativo potencial no apoio a profissionais da saúde. Neste estudo, foi desenvolvido um protótipo de dashboard para auxiliar médicos na tomada de decisão clínica, incorporando explicações contrafactuais baseadas em dados de doença renal crônica. O protótipo oferece visualização intuitiva dos resultados e possibilita a inserção simplificada de informações. Os achados sugerem que a ferramenta proposta pode aprimorar a interpretação dos resultados e facilitar a adoção de tecnologias de IA no domínio médico.*

1. Introdução

Na saúde, a Inteligência Artificial (IA) Explicável (do inglês, *Explainable Artificial Intelligence - XAI*) é impulsionada pela necessidade de transparência nos sistemas de apoio a decisão clínica, a fim de promover a confiança e aceitação pelos profissionais e pacientes. Isso garante o cumprimento de exigências legais e éticas da prática assistencial, assegurando a responsabilização profissional, a autonomia dos pacientes e a auditabilidade dos processos clínicos [Amann et al. 2020]. O estado da arte é dividido nas seguintes categorias *XAI*: (1) Explicabilidade *Post-hoc*: fornece explicações após o modelo realizar a predição, e (2) Explicabilidade *Ante-hoc*: refere-se aos modelos inerentemente interpretáveis devido seu *design* transparente. Ambas visam abordar a opacidade de modelos complexos de IA, particularmente o aprendizado profundo, que carece de transparência

em seus processos de tomada de decisão, dificultando sua ampla implantação em cenários críticos como os serviços de saúde [Loh et al. 2022].

A geração de contrafactuais é uma técnica importante dentro da explicabilidade *post-hoc*, gerando cenários alternativos que demonstram como uma pequena modificação nos recursos de entrada alteraria a predição do modelo, revelando assim a sensibilidade e os limites do raciocínio computacional. Essa abordagem é particularmente valiosa na área da saúde para entender porque o desfecho (prognóstico ou diagnóstico) foi predito e quais alterações mínimas nos dados do paciente podem impactar no modelo para alterar o resultado [Jeyaraman et al. 2023].

Utilizando um *dataset* público de doença renal crônica [Rubini and Eswaran 2015], este trabalho apresenta uma proposta de interface de saída (um *dashboard*) para contrafactuais que são geradas por modelos de IA profundos (*Deep Neural Networks - DNNs)* e seus respectivos *Nearest Unlike Neighbors (NUNs).* Ao passo que a *DNN* encontra a classificação, a *NUN* encontra o resultado contrário com o mínimo de alterações necessárias. Por fim, restrições são aplicadas para que atributos impossíveis de alterar fiquem de fora (e.g., idade), fazendo com que o contrafactual possa fornecer *insights* úteis para a prática clínica.

2. Trabalhos Relacionados

Atualmente, o impacto das pesquisas de IA têm ultrapassado a barreira entre os projetos puramente acadêmicos e aqueles com real aplicabilidade na sociedade. Assim, observamos que alguns trabalhos têm cunho bastante prático e criam ferramentas que minimizam a distância entre a academia e o usuário final [Gerlings et al. 2021, Aziz et al. 2025]

Tais ferramentas são essenciais para tornar modelos complexos de IA mais transparentes e interpretáveis para estes usuários, o que é crucial para sua aceitação e integração em fluxos de trabalho clínicos, especialmente considerando a natureza inerente de muitas técnicas avançadas de IA [Gerlings et al. 2021]. Essa explicabilidade é vital para promover a confiança dos profissionais nos resultados preditos por sistemas de IA, facilitando a implantação responsável dessas tecnologias nos ambientes clínicos de alto risco [Abgrall et al. 2024].

3. Ferramenta

A ferramenta engloba todo o processo de geração de contrafactual, mostrando de forma simples e intuitíva a saída (i.e., contrafactuais e métricas) para uma determinada entrada (i.e., conjunto de variáveis mais importantes de acordo com uma determinada métrica).

O *dashboard* desenvolvido (ver Figura 1 e 2 roda na *web* e busca oferecer uma interface interativa voltada à visualização e análise de resultados de modelos de IA. A implementação foi em Python, usando:

- Biblioteca *Streamlit* que fornece um conjunto de componentes de alto nível para a construção de interfaces responsivas, abstraindo a implementação de *front-end* tradicional (HTML, CSS e JavaScript);
- Biblioteca *Pandas* é empregada para a manipulação de dados, possibilitando a leitura de arquivos e a geração de *dataframes* otimizados para operações de filtragem, agregação e transformação. Essa integração permite que as saídas do modelo sejam apresentadas em formatos de fácil interpretação, mapeando variáveis codificadas, por exemplo, binárias (0 e 1) para “ausente” e “presente” ou categóricas numéricas (1 e 2) para “masculino” e “feminino”, em representações legíveis ao usuário final.

O *layout* da aplicação foi projetado com princípios de *design* centrado no usuário e usabilidade visual, priorizando agrupamento lógico de indicadores e utilização de elementos visuais como cores e ícones para guiar a interpretação dos dados. Essa abordagem reduz a carga cognitiva do usuário e contribui para a aceitação dos modelos de IA na prática clínica, ao transformar saídas complexas em *insights* claros e confiáveis (ver Figura 2).

Foram utilizadas as métricas *Similarity*, *Quality*, *Sparsity*, *Smoothness*, *Plausibility* e *Validation*. A *Similarity* avalia o quão próximo está o caso de consulta em relação ao contrafactual apresentado, enquanto a *Quality* mede o quanto a contrafactual gerada é coerente com os dados reais. A *Sparsity* indica a quantidade de características que diferem entre o contrafactual e o caso de consulta, e a *Smoothness* analisa o grau de variação nos dados, priorizando transições graduais. Já a *Plausibility* verifica se as alterações propostas no contrafactual são compatíveis com a realidade, evitando mudanças incoerentes para o tratamento, como a alteração da idade ou sexo. Por fim, a *Validation* identifica se houve alteração na predição do modelo DNN a partir das modificações propostas.

Figure 1. XAI Dashboard - Tela de entrada de dados com campos selecionados de acordo com o SHAP.

Figure 2. XAI Dashboard - Tela de saída de dados mostrando um exemplo de contrafactual.

4. Conclusão

Este trabalho, em andamento, pode ser considerado uma ferramenta de negação, onde o clínico pode interagir com resultados dos contrafactuais, que podem ajudar na tomada de decisões. O design focado no usuário e a usabilidade visual, com agrupamento lógico de indicadores e uso de elementos visuais, buscam reduzir a carga cognitiva e transformar saídas complexas em insights claros e confiáveis, facilitando a implantação responsável da tecnologia na prática clínica e promovendo a confiança dos profissionais nos resultados preditos. Por fim, interfaces intuitivas e explicativas podem ajudar a criar confiança em modelos profundos de IA. Consequentemente, este trabalho deve se expandir para outros datasets e realidades, unindo as constraints de cada cenário (i.e., combinações ou alterações inválidas) e desenvolvendo interfaces amigáveis e explicativas.

References

Abgrall, G., Holder, A. L., Chelly Dagdia, Z., Zeitouni, K., and Monnet, X. (2024). Should ai models be explainable to clinicians? *Critical Care*.

Amann, J., Blasimme, A., Vayena, E., Frey, D., and Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*.

Aziz, N. A., Manzoor, A., Mazhar Qureshi, M. D., Qureshi, M. A., and Rashwan, W. (2025). Unveiling explainable ai in healthcare: Current trends, challenges, and future directions. *medRxiv*.

Gerlings, J., Jensen, M. S., and Shollo, A. (2021). *Explainable AI, But Explainable to Whom? An Exploratory Case Study of xAI in Healthcare*, page 169–198. Springer International Publishing.

Jeyaraman, M., Balaji, S., Jeyaraman, N., and Yadav, S. (2023). Unraveling the ethical enigma: Artificial intelligence in healthcare. *Cureus*.

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., and Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine*, 226:107161.

Rubini, L., S. P. and Eswaran, P. (2015). Chronic Kidney Disease. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5G020.