

Desenvolvimento de Modelos Interpretáveis de Aprendizagem de Máquina Aplicáveis na Predição de Prognóstico de Pacientes com Doença de Chagas

Daniel Oliveira dos Santos Março, 2024

Resumo

A doença de Chagas, ou tripanossomíase americana, afeta entre 6 e 7 milhões de pessoas, principalmente na América Latina, incluindo Brasil e México [5, 1]. A infecção é causada pelo protozoário *Trypanosoma cruzi*, transmitido por insetos da subfamília Triatominae, mas também através da ingestão de alimentos contaminados. A doença evolui em duas fases: aguda e crônica. Cerca de 30% dos indivíduos na fase crônica desenvolvem complicações cardíacas, como arritmias e insuficiência cardíaca, que podem resultar em morte súbita cardíaca se não tratadas. A insuficiência cardíaca é uma das principais causas de morte em pacientes com doença de Chagas [4].

Medicamentos disponíveis podem melhorar a qualidade de vida e reduzir a mortalidade decorrente da insuficiência cardíaca ou de arritmias malignas associadas à doença de Chagas. Contudo, fora dos centros urbanos, a falta de serviços de saúde especializados dificulta o diagnóstico precoce. O exame eletrocardiográfico, amplamente disponível, é uma ferramenta útil para detectar sinais de insuficiência cardíaca [3].

Diante desse contexto, este projeto propõe desenvolver algoritmos de aprendizado de máquina interpretáveis para identificar automaticamente a predisposição de pacientes chagásicos cardiopatas a desfechos adversos, dentre os quais, a morte súbita cardíaca e a insuficiência cardíaca, utilizando-se de uma base de dados clínicos e de exames eletrocardiográficos de longa duração (Holter). A abordagem inclui técnicas de ciência de dados, processamento digital de sinais, extração automática de parâmetros e aprendizado de máquina, abrangendo tanto modelos clássicos quanto modernos (aprendizado profundo). Os dados para este estudo vêm do Hospital Universitário Clementino Fraga Filho (HUCFF/UFRJ) e incluem registros de eletrocardiogramas e dados clínicos de 219 pacientes com cardiomiopatia chagásica.

O objetivo central do projeto é comparar diferentes métodos de aprendizagem de máquina para predição de prognóstico nos pacientes cardiopatas chagásicos a partir de modelos interpretáveis, os quais possam indicar os parâmetros mais preponderantes para uma dada classificação. Essa análise busca aprimorar o diagnóstico e, consequentemente, melhorar a qualidade do tratamento. O projeto também pretende fornecer uma solução adequada para ambientes de saúde com recursos limitados, facilitando diagnósticos precoces e reduzindo os riscos para pacientes com doença de Chagas.

Introdução

A doença de Chagas é uma infecção parasitária que pode levar a sérias complicações cardíacas, incluindo insuficiência cardíaca e morte súbita cardíaca (MSC). O diagnóstico precoce dessas complicações em pacientes com doença de Chagas é fundamental para evitar desfechos fatais. No entanto, muitas áreas onde a doença é mais prevalente sofrem com a falta de infraestrutura médica adequada, dificultando diagnósticos precisos e oportunos [4].

Este projeto busca desenvolver e comparar modelos de aprendizado de máquina interpretáveis para analisar dados clínicos e dados de sinais eletrocardiogramas para fins de predição de prognóstico de insuficiência cardíaca, morte súbita cardíaca, bem como outras complicações associadas à cardiomiopatia chagásica. O modelo também pretende detectar desfechos específicos, como o uso de dispositivos implantáveis (como marcapassos e CDI), ablação e alta incidência de arritmias ventriculares malignas.

Os objetivos específicos do projeto incluem:

- Detecção de Insuficiência Cardíaca: Identificar indicadores de insuficiência cardíaca em pacientes com cardiopatia chagásica utilizando-se de dados de sinais eletrocardiogramas, especialmente sinais de longa duração.
- Predição de Morte Súbita Cardíaca (MSC): Descobrir padrões associados a eventos futuros de morte súbita cardíaca, permitindo respostas mais rápidas e melhor prognóstico para pacientes em risco.
- Análise de Condições Relacionadas: Explorar a relação entre outras condições médicas expressas em diversas

variáveis/parâmetros da base de dados, como AVC, hipertensão arterial, escore Rassi, dentre outros, e complicações cardíacas em pacientes com doença de Chagas.

Para alcançar esses objetivos, o modelo usará algoritmos de aprendizado de máquina com abordagens interpretáveis aplicados a um conjunto de dados clínicos e de sinais ECG de longa duração, diagnósticos médicos, procedimentos cardíacos e desfechos diversos como morte súbita cardíaca, implantações de marcapassos ou CDI, além de causas de morte diversas, como AVC e câncer.

Com essa abordagem de inteligência artificial, espera-se melhorar tanto a precisão do diagnóstico quanto a qualidade do tratamento, além de oferecer *insights* valiosos que possam contribuir para a redução de eventos fatais em pacientes com doença de Chagas e outras condições relacionadas.

Justificativa

A doença de Chagas é uma infecção parasitária que impacta milhões de pessoas na América Latina, principalmente no Brasil e no México, inclusive no interior do estado do Ceará. É uma das principais causas de cardiomiopatia e insuficiência cardíaca na região, com cerca de 30% a 40% dos pacientes com a forma crônica desenvolvendo complicações cardíacas, como arritmias, insuficiência cardíaca, aneurismas ventriculares e tromboembolismo. A insuficiência cardíaca decorrente da doença de Chagas é uma das principais causas de mortes nesses pacientes, com um alto risco de morte súbita cardíaca [1].

Apesar de medicamentos acessíveis poderem ajudar a controlar a doença, o diagnóstico precoce e o tratamento oportuno são limitados devido à escassez de serviços de saúde especializados, especialmente em regiões fora dos centros urbanos. Esta falta de infraestrutura e recursos contribui para um aumento no número de casos não diagnosticados e, conseqüentemente, para um maior risco de complicações fatais.

Dado que a detecção precoce é crítica para prevenir a progressão da doença e melhorar os resultados clínicos, há uma necessidade significativa de métodos inovadores para facilitar o diagnóstico em ambientes com recursos limitados [2]. O uso de tecnologias modernas, como o aprendizado de máquina interpretável, permite analisar grandes volumes de dados médicos, proporcionando uma oportunidade de aprimorar a identificação de padrões associados à insuficiência cardíaca e outras complicações cardíacas relacionadas à doença de Chagas.

Portanto, o desenvolvimento de um modelo de aprendizado de máquina interpretável para detectar sinais precoces de complicações cardíacas em pacientes com doença de Chagas é justificado. Esta abordagem pode oferecer uma solução eficaz para melhorar o diagnóstico precoce, facilitar o tratamento oportuno e, por fim, reduzir as taxas de mortalidade associadas à doença de Chagas e suas complicações cardíacas.

Objetivos

O objetivo geral deste projeto é desenvolver modelos de aprendizado de máquina interpretável para predição de insuficiência cardíaca e morte súbita cardíaca em pacientes com doença de Chagas, além de outros desfechos adversos, utilizando-se de uma base de dados clínicos e de registros de monitoramento cardíaco de 24 horas. Para atingir esse objetivo, os seguintes objetivos específicos foram identificados:

- Organizar um banco de dados: Reunir e estruturar dados clínicos e registros de monitoramento cardíaco de pacientes com doença de Chagas, fornecidos pelo Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro (HUCFF-UFRJ).
- Desenvolver modelos de aprendizado de máquina interpretáveis: Aplicar técnicas de aprendizado de máquina para prever insuficiência cardíaca e morte súbita cardíaca em pacientes chagásicos cardiopatas, considerando-se a camada de interpretabilidade.
- Avaliar métricas de desempenho: Analisar a precisão, sensibilidade, especificidade, entre outras métricas, para avaliar a eficácia dos modelos desenvolvidos.
- Analisar cenários de classificação: Avaliar a eficácia dos modelos em diferentes cenários, utilizando-se registros de monitoramento cardíaco de várias durações para determinar quais configurações oferecem maior precisão na predição de efeitos adversos.

Metodologia

A metodologia para atingir esses objetivos inclui os seguintes passos:

- Coleta e organização de dados: Criar um banco de dados robusto contendo informações clínicas e registros de monitoramento cardíaco de pacientes com doença de Chagas. A amostra consiste em 308 pacientes, totalizando 385 registros de monitoramento cardíaco, abrangendo um ou mais dias de exames por paciente e desfechos associados (como morte súbita, implante de marcapasso, CDI, entre outros).

- Desenvolvimento e teste de modelos de aprendizado de máquina: Testar uma variedade de algoritmos de aprendizado de máquina para criar modelos preditivos de desfechos adversos. Utilizar métodos como Redes Neurais, Florestas Aleatórias, e Máquinas de Vetores de Suporte para identificar quais são mais eficazes. Os dados de sinais serão organizados em janelas de diferentes tamanhos para fins de comparação dos resultados de predição.
- Avaliação e validação dos modelos: Aplicar métricas de desempenho para validar a precisão dos modelos e identificar quais são mais eficazes para a detecção de insuficiência cardíaca e predição de desfechos. As métricas incluem acurácia, precisão, sensibilidade e especificidade.
- Rotulação de registros de monitoramento cardíaco: Criar uma interface para a rotulação dos registros, por parte de um especialista, identificando-se características como ritmos cardíacos irregulares, bradicardia, taquicardia e outros eventos adversos. Essa rotulação será usada como suporte para treinar e validar os modelos de aprendizado de máquina.

Resultados e Impactos Esperados

O projeto tem como objetivo desenvolver um modelo de aprendizado de máquina interpretável para predição de desfechos adversos em pacientes com cardiopatia chagásica. Se bem-sucedido, este modelo pode gerar resultados significativos, entre eles:

- Melhora na detecção precoce de insuficiência cardíaca: um modelo eficaz pode facilitar a identificação precoce de insuficiência cardíaca em pacientes com doença de Chagas, permitindo uma intervenção mais rápida e eficaz.
- Ferramenta para ambientes de saúde com recursos limitados: ao criar um modelo que utiliza dados acessíveis e algoritmos eficazes, espera-se fornecer uma ferramenta que possa ser implementada em ambientes de saúde com recursos limitados, onde o acesso a diagnósticos complexos é mais difícil.
- Redução das complicações associadas à doença de Chagas: Ao identificar precocemente a insuficiência cardíaca, bem como a predisposição a efeitos cardíacos adversos, o modelo pode ajudar a reduzir o número de complicações graves, como arritmias e tromboembolismo, que frequentemente resultam em desfechos fatais.
- Diminuição da taxa de mortalidade: Ao melhorar a detecção e a predição de desfechos adversos, o projeto pode contribuir para a redução da taxa de mortalidade, resultando em melhores resultados para os pacientes.
- Avanços na pesquisa em aprendizado de máquina para a saúde: O desenvolvimento deste modelo também pode servir como base para futuros trabalhos na aplicação de aprendizado de máquina em outros contextos de saúde pública, expandindo o campo de aplicação desta tecnologia.

Os resultados e impactos esperados apontam para um potencial significativo na melhoria da qualidade de vida dos pacientes com doença de Chagas e oferecem uma oportunidade para aplicar a inovação tecnológica para enfrentar problemas de saúde críticos em regiões com infraestrutura médica limitada.

Plano de Atividades

Etapa 1: Organização do Banco de Dados (maio/2024 - agosto/2024)

Reunir e organizar os dados clínicos de pacientes com doença de Chagas fornecidos pelo Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro (HUCFF-UFRJ). Definir critérios para classificação dos pacientes e desfechos associados.

Etapa 2: Processamento de Dados (maio/2024 - dezembro/2024)

Aplicar algoritmos de pré-processamento para limpar e estruturar os dados. Extrair parâmetros essenciais para análise e caracterização dos registros de monitoramento cardíaco.

Etapa 3: Desenvolvimento de Modelos de Aprendizado de Máquina (outubro/2024 - março/2025)

Testar diferentes modelos de algoritmos de aprendizado de máquina interpretável para detecção/predição de insuficiência cardíaca e morte súbita cardíaca em pacientes com doença de Chagas. Experimentar diferentes métodos para segmentação de dados, identificando abordagens que melhorem a eficácia do modelo.

Etapa 4: Avaliação e Validação dos Modelos (janeiro/2025

- junho/2025)

Calcular métricas de desempenho, como precisão, sensibilidade e especificidade, para avaliar a eficácia dos modelos desenvolvidos. Identificar quais algoritmos ou configurações têm melhor desempenho na detecção de insuficiência cardíaca e predição de desfechos adversos.

Etapa 5: Implementação do Processo de Rotulação (outubro/2024 - março/2025)


Desenvolver uma interface para rotular manualmente os dados de sinais eletrocardiogramas, estabelecendo padrões para futuras análises. Usar algoritmos de aprendizado de máquina não supervisionados para auxiliar no processo de rotulação, identificando padrões que podem não ser imediatamente visíveis.

Etapa 6: Finalização do Projeto (julho/2025 - dezembro/2025)


Analisar os resultados obtidos durante o desenvolvimento do projeto e redigir a dissertação, bem como um artigo para submissão em congresso internacional ou periódico de impacto. Apresentar conclusões e recomendações para futuras pesquisas e possíveis aplicações práticas do modelo desenvolvido.

Referências

- [1] W. H. Organization. *Chagas Disease (also known as American Trypanoso- miasis)*. World Health Organization, [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)), 2023. Acessado em 22 de Fevereiro de 2023.
- [2] P. E. O. Primo, W. L. Caldas, J. P. V. Madeiro, D. G. Gomes, G. S. Almeida, L. P. L. Brasil, C. H. L. Cavalcante, and R. C. Pedrosa. Prediction of sudden cardiac death for chagasic patients. *Journal of Health Infomatics*, 14(1):26– 34, 2022.
- [3] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. A. Naqvi. Machine learning (ml) in medicine: Review, applications, and challenges. *Mathematics*, 9(22):2970, 2021.
- [4] A. J. Rassi, A. Rassi, and S. G. Rassi. Predictors of mortality in chronic chagas disease: a systematic review of observational studies. *Circulation*, 115(9):1101–1108, 2007.
- [5] A. Rassi Jr., A. Rassi, and J. A. Marin-Neto. Chagas disease. *Lancet*, 375(9723):1388–1402, 2010.

Documento assinado digitalmente
 DANIEL OLIVEIRA DOS SANTOS
Data: 09/05/2024 11:12:33-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do Candidato

Documento assinado digitalmente
 JOAO PAULO DO VALE MADEIRO
Data: 09/05/2024 11:06:28-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do Orientador

Assinatura do Orientador