

A Correlação de Fatores de Risco no Desenvolvimento do Ataque Cardíaco

Anderson Rodrigues dos Santos
Divinópolis - MG, Brasil
anderson.santos@aluno.cefetmg.br

Abstract—Heart attacks remain a critical public health challenge, amplified by the complex interaction of multiple risk factors. This study utilized a Pearson correlation matrix and the Modularity Class technique to identify factors associated with the occurrence of heart attacks. Patient data were statistically analyzed, revealing that while individual correlations with heart attacks were weak, subsets of factors showed relevant interpretation. These findings emphasize that nearly all risk factors have a significant impact on the development of the problem, highlighting the importance of differentiated analyses to guide more effective prevention strategies.

Index Terms—Heart attacks, Pearson correlation, Modularity Class, risk factors, prevention strategies, late-onset cardiovascular events.

I. INTRODUÇÃO

O ataque cardíaco é uma das principais causas de morte no mundo, representando uma emergência médica crítica que afeta milhões de pessoas anualmente. No Brasil, estima-se que entre 300 mil e 400 mil pessoas tenham sofrido um infarto em 2022, consolidando essa condição como um dos maiores desafios de saúde pública. Os ataques cardíacos são eventos agudos que ocorrem devido ao bloqueio das artérias coronárias por placas de gordura, impedindo o fluxo sanguíneo para o coração e resultando em danos ao músculo cardíaco. Além desse fator, diversos outros fatores de risco contribuem para a ocorrência de infartos, como níveis elevados de colesterol, diabetes, hipertensão, tabagismo, sedentarismo e histórico familiar de doenças cardíacas. A interação entre esses fatores pode amplificar significativamente o risco de desenvolvimento de doenças cardíacas, evidenciando a importância de uma análise detalhada para sua compreensão e controle. A identificação e análise dos fatores de risco são essenciais não apenas para entender o desenvolvimento de ataques cardíacos, mas também para embasar estratégias de prevenção mais eficazes. Apesar de muitos estudos já terem investigado esses fatores isoladamente, poucos exploram de forma quantitativa e sistemática as interações entre eles, utilizando abordagens estatísticas avançadas, como o coeficiente de correlação de Pearson e a técnica de Modularity Class. O coeficiente de correlação de Pearson mede a relação linear entre variáveis, enquanto a matriz de correlação, uma tabela que exibe os coeficientes de correlação entre todas as variáveis de um conjunto de dados, permite uma visualização clara e concisa dessas interações. Já a técnica de Modularity Class identifica comunidades fortemente conectadas a um nó específico, com base na correlação de Pearson, permitindo agrupar variáveis

com maior grau de associação. Este estudo busca preencher essa lacuna ao analisar os fatores de risco associados aos ataques cardíacos tardios, utilizando uma matriz de correlação e a técnica de Modularity Class para identificar os fatores que apresentam maior influência no desenvolvimento do problema. Através dessa análise, esperamos fornecer insights valiosos para embasar políticas públicas e estratégias de intervenção mais direcionadas, contribuindo para a redução da incidência de ataques cardíacos tardios e para a melhoria da saúde cardiovascular da população.

II. REFERENCIAL TEÓRICO

O ataque cardíaco, também conhecido como infarto do miocárdio, é uma das principais causas de mortalidade no mundo. De acordo com o Glossário de Saúde do Hospital Albert Einstein, essa condição ocorre devido à morte de células do músculo cardíaco em decorrência da falta de oxigenação, geralmente causada pelo bloqueio do fluxo sanguíneo nas artérias coronárias. A aterosclerose, caracterizada pelo acúmulo de placas de gordura que podem levar ao rompimento dos vasos e à formação de coágulos, é apontada como a principal causa desses bloqueios.

A Organização Mundial da Saúde (OMS) relata que as doenças cardiovasculares são responsáveis por aproximadamente 17,9 milhões de mortes por ano em todo o mundo, sendo a principal causa de óbito global. No Brasil, estima-se que entre 300 mil e 400 mil pessoas sofram infartos anualmente, evidenciando a necessidade de estratégias preventivas mais eficazes. A Organização Pan-Americana da Saúde (OPAS) destaca que fatores comportamentais, como dietas inadequadas, sedentarismo, tabagismo e consumo excessivo de álcool, estão diretamente ligados a condições como hipertensão arterial, hiperglicemia, hipercolesterolemia e obesidade. Esses fatores, segundo Fernanda Jorge Magalhães em seu estudo sobre estratégias de promoção da saúde cardiovascular, podem ser divididos em modificáveis e não modificáveis. Entre os modificáveis estão hábitos de vida, como alimentação e nível de atividade física, enquanto os não modificáveis incluem idade, sexo e histórico familiar de doenças cardiovasculares.

A interação entre múltiplos fatores de risco amplifica a probabilidade de eventos cardíacos. Estudos, como o *Individual and Joint Association between Cardiovascular Disease Risk Factors and Inadequate Lifestyle Behaviors in a Sample from*

Brazil, evidenciam que altos níveis de colesterol, frequentemente associados à obesidade, aumentam significativamente o risco de infarto. Esses achados reforçam a importância de compreender como os fatores de risco interagem e contribuem para eventos cardíacos, especialmente no contexto de ataques cardíacos tardios.

O conjunto de dados utilizado neste estudo, *Heart Attack in Youth Vs Adult in Germany*, disponibilizado pela plataforma Kaggle e pelo Massachusetts Institute of Technology (MIT), abrange informações sobre ataques cardíacos em jovens (menos de 25 anos) e adultos (25 anos ou mais) na Alemanha, entre 2015 e 2023. Os dados incluem variáveis relacionadas à saúde, estilo de vida e condições ambientais. No entanto, a distribuição desigual entre categorias, com maior predominância de registros de adultos, constitui um desafio para análises equilibradas e representativas.

Para a análise estatística, utilizou-se o coeficiente de correlação de Pearson, uma ferramenta amplamente empregada para medir a força e a direção da relação linear entre variáveis. Com valores que variam de -1 a 1, o coeficiente permite identificar correlações positivas, negativas ou ausentes. Os coeficientes foram organizados em uma matriz de correlação, permitindo uma análise integrada dos fatores de risco. Adicionalmente, a técnica de *Modularity Class* foi aplicada para identificar comunidades de variáveis fortemente conectadas. Baseada no algoritmo de Vincent Blondel *et al.*, essa técnica destaca subgrupos interligados que podem oferecer insights relevantes sobre interações entre fatores de risco.

Embora existam inúmeros estudos sobre fatores de risco cardiovasculares, poucos integram abordagens estatísticas e técnicas de análise de redes para explorar interações complexas entre variáveis. Ademais, a literatura científica apresenta uma lacuna específica no que diz respeito a ataques cardíacos tardios, resultantes de longa exposição a fatores de risco. Este estudo busca preencher essa lacuna, utilizando métodos quantitativos para identificar os fatores mais influentes e suas interações, contribuindo para estratégias de prevenção e formulação de políticas públicas mais eficazes no combate a doenças cardiovasculares.

III. METODOLOGIA

Este estudo utilizou o conjunto de dados *Heart Attack in Youth Vs Adult in Germany*, disponível na plataforma Kaggle, abrangendo informações coletadas entre 2015 e 2023. O objetivo foi investigar a relação entre fatores de risco e ataques cardíacos tardios em jovens e adultos na Alemanha, empregando métodos estatísticos e técnicas de análise de redes.

Durante o pré-processamento dos dados, registros com informações essenciais foram removidos para garantir a integridade das análises. Variáveis categóricas, como *Nível de Atividade Física* e *Qualidade da Dieta*, foram codificadas numericamente em escala binária, representando ausência ou presença de risco. As variáveis analisadas foram divididas em modificáveis, como *IMC*, *Consumo de Álcool*, *Nível de*

Atividade Física, *Qualidade da Dieta*, *Nível de Colesterol*, *Diabetes* e *Nível de Estresse*, e não modificáveis, como *Incidência de Ataque Cardíaco*, *Histórico Familiar (Genética)*, *Hipertensão*. Cada variável foi modelada de acordo com diretrizes reconhecidas, como valores de *Nível de Colesterol* iguais ou superiores a 190 mg/dl classificados como alto risco. Este critério foi uniformemente aplicado às demais variáveis.

A análise estatística foi conduzida utilizando o coeficiente de correlação de Pearson para avaliar relações lineares entre variáveis. A matriz de correlação foi calculada com a fórmula

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \times C_{jj}}} \quad (1)$$

onde R_{ij} representa o coeficiente de correlação entre as variáveis i e j , garantindo que os valores permanecessem no intervalo de -1 a 1.

Para a análise de redes, utilizou-se o software *Gephi*, aplicando a técnica de *Modularity Class*, conforme descrito no artigo *Fast Unfolding of Communities in Large Networks* de Vincent Blondel *et al.*. A modularidade foi calculada pela fórmula

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

onde Q mede a qualidade da divisão em comunidades, A_{ij} é a matriz de adjacência e m é o número total de arestas. Essa abordagem permitiu a identificação de comunidades relevantes nos grafos gerados.

Apesar de sua utilidade, a matriz de correlação apresenta limitações, incluindo sensibilidade a *outliers*, dependência de uma distribuição normal dos dados e incapacidade de detectar relações não lineares. Esses fatores podem ter influenciado os resultados, mas foram mitigados pela utilização de grafos no *Gephi*, que permitiram aplicar filtros para segmentação e análise das comunidades.

Os resultados destacaram varias comunidades de fatores de risco significativos. Intervenções voltadas para o controle dessas variáveis podem desempenhar um papel essencial na prevenção de ataques cardíacos. No entanto, a distribuição desigual dos dados, com predominância de registros de adultos, pode ter introduzido vieses na análise. Futuros estudos poderão explorar técnicas como reamostragem e regressão logística para abordar essas limitações e oferecer uma análise mais abrangente das relações entre fatores de risco e ataques cardíacos tardios.

IV. RESULTADOS

Apesar dos coeficientes de correlação de Pearson indicarem uma interação fraca ou inexistente entre os fatores de risco, a visualização em formato de grafo, gerada no *Gephi*, oferece uma perspectiva complementar sobre as possíveis conexões entre as variáveis analisadas.

Na imagem 1, a matriz de correlação evidencia que os coeficientes entre os fatores de risco e a Incidência de Ataque Cardíaco são extremamente baixos, próximos de zero. Isso

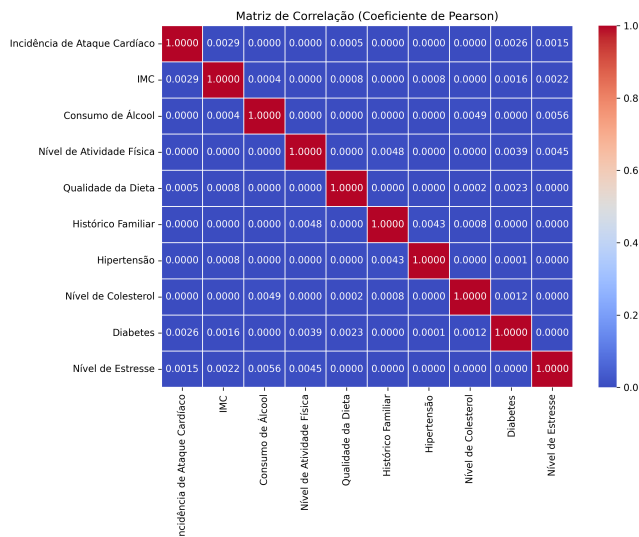


Fig. 1. A matriz de correlação com o coeficiente de Pearson.

sugere a ausência de uma relação linear significativa entre as variáveis avaliadas. Por exemplo, fatores como IMC, Consumo de Alcool, Nível de Estresse e Nível de Colesterol apresentaram coeficientes inferiores a 0,01, indicando que, isoladamente, essas variáveis não têm impacto estatisticamente significativo sobre a incidência de ataques cardíacos. Ademais, a falta de correlações fortes entre os demais fatores reforça a ideia de que essas variáveis, quando analisadas individualmente, não explicam de maneira clara o desfecho investigado. Esses resultados podem apontar para limitações na abordagem empregada, como a possível ausência de variáveis mais relevantes no conjunto de dados ou a necessidade de explorar interações não lineares ou combinatórias entre os fatores de risco. Outra hipótese é que os dados refletem características contextuais específicas, nas quais as variáveis analisadas possuem relevância reduzida para prever ataques cardíacos.

Na imagem 2, a análise em grafo revela a organização dos fatores de risco em clusters baseados na modularidade, destacando agrupamentos que podem sugerir interações indiretas ou mais complexas. Por exemplo, Nível de Atividade Física e Qualidade da Dieta formam um subgrupo coeso, enquanto Histórico Familiar e Hipertensão se agrupam em outro. Esses padrões podem indicar que, embora os coeficientes de correlação numéricos não apresentem relações evidentes, o comportamento dos dados em rede sugere conexões que merecem maior investigação com abordagens complementares.

Já na imagem 3, com ajustes na modularidade, os clusters continuam a evidenciar agrupamentos significativos, destacando interações mais específicas, como as relações entre Nível de Colesterol, Consumo de Alcool e Nível de Estresse. Essa análise reforça a utilidade da visualização em redes para complementar métricas estatísticas tradicionais, fornecendo insights qualitativos que podem passar despercebidos em abordagens convencionais. Por fim, a análise combinada da matriz

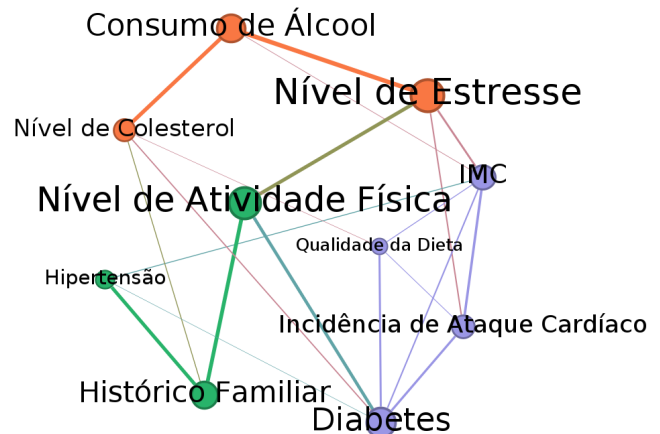


Fig. 2. Visualização da Matriz de Correlação em Grafo pelo Gephi.

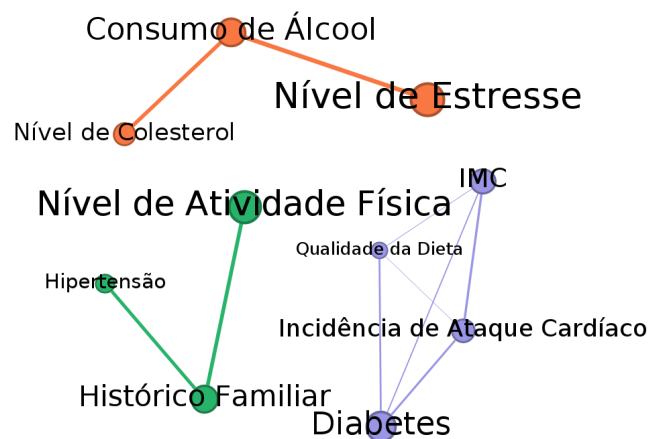


Fig. 3. Comunidade Separada pela Técnica de Modularity Class no Gephi.

de correlação e dos grafos sugere que, embora não existam evidências estatísticas diretas de associações fortes, os padrões observados na organização dos dados em redes justificam estudos mais aprofundados. Esses achados podem indicar a necessidade de reformular os modelos de coleta de dados ou incorporar variáveis adicionais que sejam capazes de captar melhor as interações entre os fatores de risco avaliados.

V. RESULTADOS E DISCUSSÃO

Os resultados obtidos a partir da análise da matriz de correlação e da visualização em forma de grafo no Gephi fornecem insights complementares sobre as interações entre os fatores de risco e a incidência de ataques cardíacos. Conforme descrito na metodologia, foram analisados os coeficientes de correlação de Pearson para identificar possíveis relações lineares, seguidos pela construção de grafos com base nos dados, utilizando a técnica de Modularity Class para identificar comunidades de fatores com maior conectividade interna.

Os coeficientes de correlação de Pearson, como previsto, indicaram uma interação linear fraca ou inexistente entre os fatores de risco. Isso reflete a limitação de análises univariadas para fenômenos complexos, onde as variáveis podem não atuar de forma isolada. No entanto, ao empregar a análise visual com grafos, foi possível identificar agrupamentos (clusters) de fatores que indicam interações não lineares e combinatórias. Por exemplo, os clusters que incluem "Nível de Atividade Física", "Histórico Familiar" e "Hipertensão", assim como "Consumo de Álcool", "Nível de Estresse" e "Nível de Colesterol", sugerem que esses fatores podem atuar em conjunto, potencializando seu impacto na incidência de ataques cardíacos.

A técnica de Modularity Class, descrita na metodologia como uma ferramenta para detectar comunidades em grafos, mostrou-se eficaz na identificação de agrupamentos estruturados. Esses agrupamentos destacam possíveis interdependências entre os fatores, mesmo na ausência de correlações lineares significativas. Por exemplo, a interligação entre "Consumo de Álcool" e "Nível de Colesterol", bem como sua influência no "Nível de Estresse", sugere a existência de relações complexas que não são capturadas por métricas tradicionais.

Comparando com a literatura existente, os resultados corroboram estudos que destacam a importância de abordagens multivariadas e visuais para a análise de fatores de risco. Enquanto métodos tradicionais, como a correlação linear, são amplamente utilizados, ferramentas de visualização em rede têm emergido como estratégias poderosas para identificar padrões ocultos e interações indiretas entre variáveis.

Por outro lado, os resultados evidenciam limitações importantes. A ausência de variáveis socioeconômicas ou comportamentais nos dados analisados pode ter limitado a abrangência das análises, subestimando possíveis interações externas. Além disso, a dependência exclusiva do coeficiente de correlação de Pearson para avaliar interações iniciais pode não capturar relações não lineares, como já discutido em abordagens recentes da literatura.

Em termos de implicações, os resultados sugerem que a combinação de técnicas estatísticas e de visualização em rede é

essencial para investigar fenômenos complexos. Esse enfoque permite identificar padrões qualitativos e interações não lineares, oferecendo uma perspectiva mais rica e abrangente sobre o comportamento dos fatores de risco. Além disso, destaca-se a importância de considerar metodologias integrativas para futuros estudos que visem explorar a dinâmica entre múltiplos fatores de risco em contextos clínicos.

VI. CONCLUSÃO

Em suma, a análise dos dados revela que a relação entre os fatores de risco e a incidência de ataque cardíaco é complexa e multifacetada. Embora os coeficientes de correlação de Pearson não evidenciem relações lineares significativas, a análise dos grafos gerados no Gephi destaca padrões de conexão que complementam a análise estatística. A organização dos fatores em clusters e as interdependências identificadas sugerem que os fatores de risco podem interagir de maneira não linear, apontando para a relevância de abordagens multivariadas na interpretação dos dados.

A análise demonstra que a visualização em grafos não substitui as análises estatísticas tradicionais, mas enriquece a interpretação ao revelar padrões qualitativos e interações complexas que podem ser negligenciadas em uma abordagem unicamente quantitativa. A combinação de técnicas analíticas, como a matriz de correlação e a análise de redes, mostra-se eficaz para explorar fenômenos complexos como a incidência de ataque cardíaco.

Para futuras pesquisas, sugere-se a inclusão de novos fatores de risco no conjunto de dados, como informações genéticas e biomarcadores, o uso de técnicas avançadas de análise de redes, como a análise de centralidade e a modelagem de redes dinâmicas, e a replicação dessa abordagem em outros contextos clínicos ou epidemiológicos. Essas estratégias podem ajudar a captar relações mais profundas e dinâmicas, contribuindo para uma melhor compreensão dos fatores associados à incidência de ataque cardíaco e, consequentemente, para o desenvolvimento de estratégias de prevenção e tratamento mais eficazes.

REFERENCES

- [1] GONÇALVES, L.; ZANLORENCI, S.; PELEGRINI, A.; LIMA, T. R. de; SILVA, D. A. S. Associação individual e simultânea entre fatores de risco para doença cardiovascular e hábitos inadequados do estilo de vida em uma amostra do Brasil. *Arquivos Brasileiros de Cardiologia*, v. 121, n. 10, p. e20240149, 2024.
- [2] NEWMAN, M. E. J. Equivalence between modularity optimization and maximum likelihood methods for community detection. *Physical Review E*, v. 94, 052315, 2016. Disponível em: <https://doi.org/10.1103/PhysRevE.94.052315>. Acesso em: 20 jan. 2025.
- [3] CLAUSET, A.; NEWMAN, M. E. J.; MOORE, C. Finding community structure in very large networks. *Physical Review E*, v. 70, n. 6, 2004. Disponível em: <https://arxiv.org/abs/cond-mat/0408187>. Acesso em: 20 jan. 2025.
- [4] HOSPITAL ISRAELITA ALBERT EINSTEIN. Glossário de saúde. Disponível em: <https://www.einstein.br/doencas-sintomas>. Acesso em: 25 jan. 2025.
- [5] ORGANIZAÇÃO MUNDIAL DA SAÚDE. Relatório mundial de saúde 2020. Disponível em: <https://www.who.int/health-topics>. Acesso em: 20 jan. 2025.

- [6] MAGALHÃES, F. J.; MENDONÇA, L. B. de A.; REBOUÇAS, C. B. de A.; LIMA, F. E. T.; CUSTÓDIO, I. L.; OLIVEIRA, S. C. de. Fatores de risco para doenças cardiovasculares em profissionais de enfermagem: estratégias de promoção da saúde. *Revista Brasileira de Enfermagem*, v. 67, n. 3, p. 394–400, 2014.
- [7] ANDERSON, R. S. A correlação de fatores de risco no desenvolvimento do ataque cardíaco: algoritmo de modelagem. Disponível em: <https://github.com/AndersonR-S/A-Correlacao-de-Fatores-de-Risco-no-Desenvolvimento-do-Ataque-Cardiaco.git>. Acesso em: 20 jan. 2025.