

TECNOLOGIAS EM SAÚDE

APLICAÇÃO DE APRENDIZADO DE MÁQUINA NO ENTENDIMENTO E DIAGNÓSTICO DA ATEROSCLEROSE: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Leonardo Gauginski

Resumo

A aterosclerose é uma doença caracterizada pelo acúmulo de placas de gordura, cálcio e outros elementos na parede das artérias, representando um desafio significativo na medicina cardiovascular devido à sua complexidade diagnóstica e prognóstica. Recentemente, o aprendizado de máquina (AM), uma vertente da inteligência artificial, tem emergido como uma ferramenta promissora para aprimorar o entendimento e o diagnóstico da aterosclerose. Este artigo explora a aplicação de técnicas de AM no diagnóstico e prognóstico da aterosclerose, comparando-as com os métodos diagnósticos tradicionais. O objetivo é investigar como o AM pode melhorar a precisão do diagnóstico, facilitar a identificação precoce de pacientes em risco e aprimorar o prognóstico da doença.

Como estratégia de busca, foi utilizado o mnemônico PICODT, que significa População, Intervenção, Comparação, Resultados, Delineamento do Estudo e Tempo. A revisão incluiu estudos que aplicaram AM e Inteligência Artificial (IA) em modelos cardiovasculares, sem restrições de desenho de estudo ou população. A pesquisa foi conduzida utilizando bases de dados médicas e científicas, com foco em publicações dos últimos cinco anos para garantir a relevância das técnicas de AM.

Os resultados indicam uma tendência crescente na aplicação de AM na medicina cardiovascular, especialmente na aterosclerose. Estudos demonstraram que o AM pode oferecer diagnósticos mais precisos e prognósticos mais confiáveis em comparação com os métodos convencionais. Além disso, o AM tem o potencial de transformar o manejo clínico da aterosclerose, proporcionando intervenções mais eficazes e personalizadas.

Este artigo contribui para a compreensão da aplicação de AM na aterosclerose, destacando seu potencial para melhorar os resultados clínicos visto que as pesquisas em AM na medicina cardiovascular estão em uma fase promissora, e espera-se que continue a evoluir, oferecendo novas perspectivas para o diagnóstico e tratamento da aterosclerose.

Palavras-chave: Aterosclerose. Aprendizado de Máquina. Diagnóstico. Prognóstico. Inteligência Artificial. Revisão Sistemática.

Abstract

Atherosclerosis is a disease characterized by the accumulation of fat plaques, calcium, and other elements in the arterial walls, representing a significant challenge in cardiovascular medicine due to its diagnostic and prognostic complexity. Recently, machine learning (ML), a branch of artificial intelligence, has emerged as a promising tool to enhance the understanding and diagnosis of atherosclerosis. This article explores the application of ML techniques in the diagnosis and prognosis of atherosclerosis, comparing them with traditional diagnostic methods. The goal is to investigate how ML can improve diagnostic accuracy, facilitate early identification of at-risk patients, and enhance disease prognosis.

The search strategy used the mnemonic PICODT, which stands for Population, Intervention, Comparison, Outcomes, Study Design, and Time.. The review included studies that applied ML and AI in cardiovascular models, without restrictions on study design or population. The research was conducted using medical and scientific databases, focusing on publications from the last five years to ensure the relevance of ML techniques.

The results indicate a growing trend in the application of ML in cardiovascular medicine, especially in atherosclerosis. Studies have shown that ML can provide more accurate diagnoses and more reliable prognoses compared to conventional methods. Additionally, ML has the potential to transform the clinical management of atherosclerosis, offering more effective and personalized interventions.

This article contributes to the understanding of ML application in atherosclerosis, highlighting its potential to improve clinical outcomes and the quality of life of patients. ML research in cardiovascular medicine is in a promising phase, and it is

expected to continue evolving, offering new perspectives for the diagnosis and treatment of atherosclerosis.

Key-words: Atherosclerosis. Machine Learning. Diagnosis. Prognosis. Artificial Intelligence. Systematic Review.

Introdução

A aterosclerose, uma das principais causas de morbidade e mortalidade cardiovascular, é caracterizada pela complexa interação de acumulação de lipídios, respostas inflamatórias e remodelação vascular nas paredes das artérias (Sánchez-Cabo et.al., 2020). A natureza assintomática da doença complica o diagnóstico precoce e o manejo eficaz, representando significativos desafios na medicina cardiovascular. Com a chegada de grandes volumes de dados e técnicas computacionais avançadas, o aprendizado de máquina (AM) emergiu como uma tecnologia fundamental, revolucionando o diagnóstico e a avaliação prognóstica das doenças ateroscleróticas.

Estudos recentes têm utilizado o AM para refinar a precisão diagnóstica e melhorar as avaliações prognósticas, superando os métodos clínicos tradicionais em precisão e poder preditivo⁴. Estudo, como o de Fan et al. demonstra a superioridade dos modelos de regressão logística sobre os algoritmos de árvore de decisão convencionais na predição de aterosclerose carotídea assintomática usando registros eletrônicos de saúde, alcançando melhorias notáveis em sensibilidade e especificidade diagnóstica⁵. De forma similar, a pesquisa de Munger et al., (2021), destacou o potencial do AM em dissecar os processos biológicos complexos subjacentes à aterosclerose, facilitando a descoberta de novos biomarcadores e estratégias de estratificação de pacientes que são essenciais para terapias personalizadas.

A aplicação de AM na medicina cardiovascular é ampla, variando desde a estratificação de risco de pacientes até análises avançadas de imagem e descoberta de biomarcadores⁶. Estudos como os conduzidos pelo grupo MESA têm usado AM para realçar a precisão preditiva de risco cardiovascular, empregando algoritmos sofisticados para analisar extensos conjuntos de dados que incluem informações clínicas, moleculares e de imagem⁷. Essas metodologias permitem uma

compreensão mais profunda dos mecanismos da doença, contribuindo para avaliações de risco mais precisas e intervenções direcionadas.

O objetivo principal deste estudo é explorar e validar a eficácia das técnicas de AM na detecção precoce e na avaliação prognóstica da aterosclerose, comparadas aos métodos diagnósticos tradicionais. Ao aproveitar uma revisão sistemática da literatura recente e aplicar análises estatísticas robustas, esta pesquisa visa elucidar o potencial dos modelos de AM em transformar a paisagem clínica dos diagnósticos cardiovasculares e do cuidado ao paciente. Por meio desta investigação, aspiramos fornecer uma avaliação abrangente do AM como uma ferramenta transformadora na identificação precoce de populações em risco e na adaptação de estratégias de tratamento personalizadas de forma eficaz.

Metodologia

Esta pesquisa caracteriza-se por ser uma revisão sistemática da literatura. Foram incluídos na revisão os estudos aqueles que avaliaram pacientes diagnosticados ou em risco de aterosclerose; pesquisas que aplicam técnicas de aprendizado de máquina (AM) para o diagnóstico ou prognóstico da aterosclerose. Além disso, foram consideradas pesquisas que aplicam técnicas de AM, comparando-as com métodos convencionais de diagnóstico e prognóstico da doença.

Em relação a seleção dos artigos, primeiramente, foram selecionados artigos que utilizam diagnósticos baseados em critérios clínicos reconhecidos e/ou exames de imagem. Em seguida, foram considerados estudos que incluem análises de risco de viés e a aplicabilidade dos estudos selecionados. Outro ponto importante foi a inclusão de artigos que discutem a aplicação de AM para melhorar os resultados diagnósticos em populações-alvo, incluindo ensaios clínicos randomizados, o que reforça a robustez dos achados.

Para garantir a atualidade e relevância das técnicas de AM discutidas, a revisão incluiu apenas estudos publicados nos últimos cinco anos. Além disso, foram considerados artigos escritos em português, inglês e espanhol, aumentando a diversidade e abrangência da pesquisa.

Os artigos foram pesquisados na base de dados científicos *PubMed*, garantindo uma ampla cobertura das publicações relevantes no campo. Dessa forma, a revisão sistemática proporciona uma visão abrangente e atualizada sobre o

uso de aprendizado de máquina no diagnóstico e prognóstico da aterosclerose, contribuindo significativamente para a área da saúde.

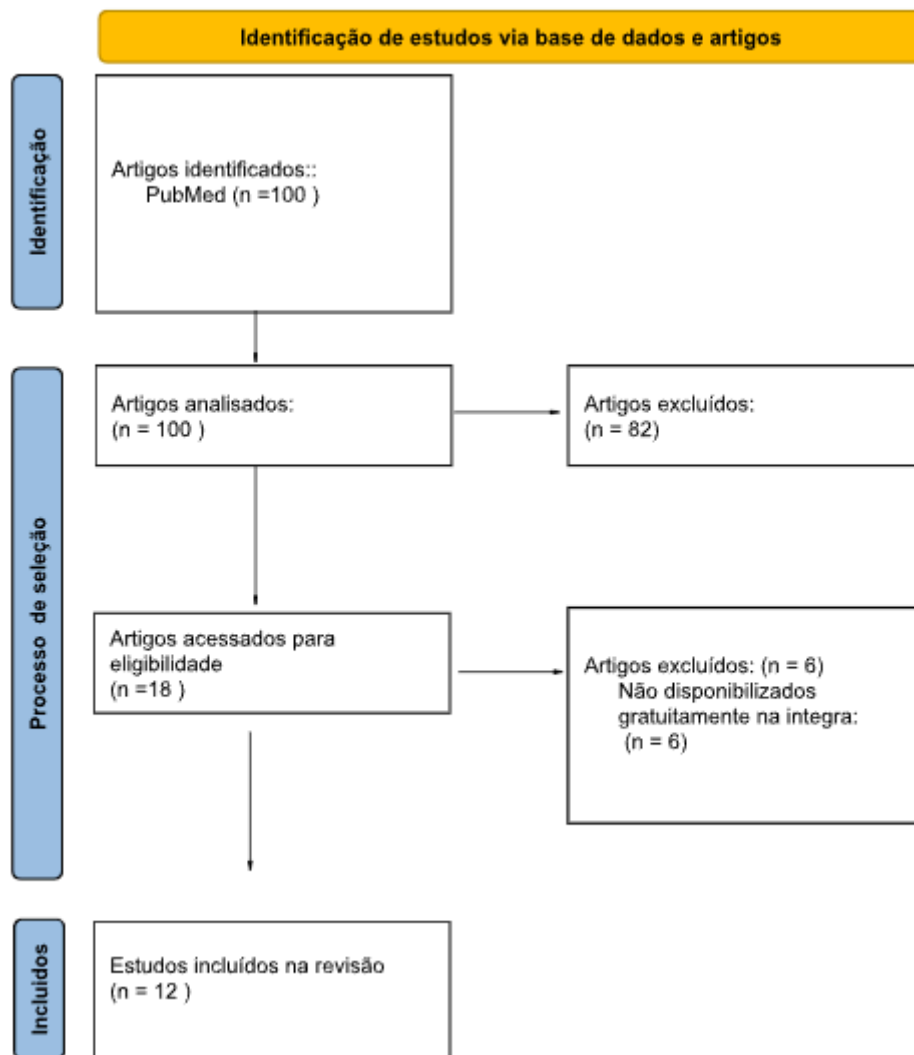
Como estratégia de busca foi utilizado o mnemônico PICODT, conforme descrito a seguir. Para P (população) foram considerados os pacientes diagnosticados ou em risco de aterosclerose; para I (intervenção) foi considerada a maneira de aplicação de técnicas de aprendizado de máquina; para C (comparação), a intervenção foi comparada com métodos tradicionais de diagnóstico e prognóstico; para O (*outcome* ou resultados) foi considerada a precisão do diagnóstico, identificação precoce de pacientes em risco e prognóstico da aterosclerose com a aplicação de técnicas de aprendizado de máquina; D (delineamento) foram incluídos estudos de diferentes delineamentos, como coortes prospectivas, coortes retrospectivas, revisões e para T (tempo) foram considerados os últimos 5 anos (2018 até 2023).

Para a pesquisa dos artigos foram empregados os seguintes termos: *Atherosclerosis OR Atheroscleroses OR Atherogenesis OR Atherogeneses; Machine Learning OR Learning, Machine OR Transfer Learning OR Learning, Transfer; Diagnosis OR Diagnoses OR Diagnose OR Diagnoses and Examinations OR Examinations and Diagnoses OR Diagnoses and Examination OR Examination and Diagnoses OR Antemortem Diagnosis OR Antemortem Diagnoses OR Diagnoses, Antemortem OR Diagnosis, Antemortem.*

Resultados

Inicialmente foram encontrados 100 artigos no *PubMed*. Em seguida, a seleção dos artigos foi realizada em *software Rayyan* e foram selecionados 18 artigos. Destes, 06 foram excluídos do estudo por não estarem disponíveis gratuitamente na íntegra (figura 1).

Figura 1. Fluxograma de seleção dos estudos.



O estudo de Kakadiaris et al. (2018), conduzido no *Multi-Ethnic Study of Atherosclerosis* (MESA) nos EUA, foi um estudo de coorte prospectiva com 5.000 participantes, cuja idade média foi de 61,0 anos. As técnicas de aprendizado de máquina utilizadas foram Rede Neural Artificial (ANN), Floresta Aleatória (RF) e Máquina de Vetores de Suporte (SVM). A RF alcançou uma AUC de 0.73, a ANN uma AUC de 0.71, e a SVM uma AUC de 0.69. O estudo comparou modelos de aprendizado de máquina com modelos de risco convencionais na previsão de doenças cardiovasculares.

No estudo de Munger et al. (2021), realizado nos Estados Unidos, foi conduzida uma revisão sistemática que analisou a aplicação de diversas técnicas de aprendizado de máquina, incluindo SVM, ANN, KNN, SVM, NB, CE, e DA. O estudo focou na análise de imagens, estratificação de risco e descoberta de biomarcadores, atingindo uma acurácia geral de até 98%.

Wang et al. (2022) realizaram um estudo de coorte nos Estados Unidos com uma população de estudo de 500 participantes, cuja idade média era de 56,4 anos. As técnicas de aprendizado de máquina utilizadas incluíram LR, RF, SVM, *XGBoost*, KNN, e CNN. O modelo de *XGBoost* alcançou um *AUCROC* de 0.87. O estudo focou na previsão de aterosclerose baseada em dados clínicos e de imagem.

Chen et al. (2022), na China, conduziram um estudo de coorte com 1.200 participantes e uma idade média de 60,2 anos. As técnicas utilizadas foram DT, RF, e SVM, com a técnica de RF alcançando um *AUCROC* de 0.81. O estudo foi centrado na previsão de desfechos cardiovasculares e análise de dados clínicos.

No estudo de Li et al. (2022), realizado nos Estados Unidos, foi conduzido um estudo de coorte com 1.300 participantes e uma idade média de 65,7 anos. As técnicas de aprendizado de máquina utilizadas incluíram RF e LR. O RF alcançou um *AUCROC* de 0.79, enquanto o LR obteve um *AUCROC* de 0.809. O estudo analisou a composição de placas ateroscleróticas usando dados de Tomografia Computadorizada com Subtração de Energia Dual (DECT).

No estudo de Liu et al. (2023), realizado na Austrália, foi conduzido um estudo de coorte com uma coleta de dados de 101.110 registros médicos e uma idade média de 62,7 anos. As técnicas de aprendizado de máquina utilizadas incluíram LR, RF, DT, XGB, e KNN. Os modelos de LR e RF apresentaram *AUCROC* de 0.92 e 0.80, respectivamente. O estudo focou na previsão de risco cardiovascular aterosclerótico utilizando abordagens de AM.

O estudo de Núñez et al. (2022), realizado nos Estados Unidos, foi um estudo de caso-controle com 1.000 participantes e uma idade média de 58,7 anos. As técnicas de aprendizado de máquina utilizadas incluíram SVM, RF, KNN, e DT, com a técnica de RF apresentando uma acurácia de 85%. O estudo focou na descoberta de biomarcadores para a detecção aprimorada de aterosclerose subclínica.

No estudo de Fan et al. (2021), realizado em Xi'an, China, foi conduzido um estudo de coorte retrospectiva com uma população de estudo composta por 18.441 participantes, dos quais 6.553 apresentavam aterosclerose carotídea assintomática (CAS). A idade média dos participantes foi de 50,86 anos. As técnicas de aprendizado de máquina utilizadas incluíram Regressão Logística (LR), Floresta Aleatória (RF), Árvore de Decisão (DT), *eXtreme Gradient Boosting* (XGB), Gaussian Naive Bayes (GNB) e K-Nearest Neighbour (KNN). A LR obteve uma

AUCROC de 0.809, uma acurácia de 74.7% e um F1 Score de 59.9%. O estudo focou na previsão de aterosclerose carotídea assintomática utilizando registros eletrônicos de saúde e técnicas de aprendizado de máquina.

Sánchez-Cabo et al. (2020) conduziram um estudo de coorte na Espanha com uma população de estudo de 1.777 participantes e uma idade média de 55,2 anos. As técnicas de aprendizado de máquina utilizadas incluíram RF, LR, e KNN. Os modelos apresentaram *AUCROC* de 0.82 (RF), 0.80 (LR), e 0.78 (KNN). O estudo focou em melhorar a definição de risco cardiovascular com AM.

No estudo de Segar et al. (2021), realizado nos Estados Unidos, foi conduzido um estudo de coorte com 12.000 participantes e uma idade média de 54,3 anos. As técnicas de aprendizado de máquina utilizadas incluíram RF, SVM, LR, ANN, e DT. Os modelos de RF e SVM apresentaram *AUCROC* de 0.83 e 0.82, respectivamente. O estudo focou na criação de modelos específicos por raça para prever risco cardiovascular de 10 anos.

Sun et al. (2023) conduziram um estudo de coorte em Taiwan com 359 participantes e uma idade média de 61,0 anos. As técnicas de aprendizado de máquina utilizadas incluíram LR, SVM, DT, RF, *XGBoost*, e *Gradient Boost*. O modelo de LR apresentou o melhor desempenho com um *AUCROC* de 0.92. O estudo focou na descoberta de biomarcadores para predizer aterosclerose de grandes artérias.

O estudo de Kampaktsis et al. (2023), realizado na Grécia, foi uma revisão sistemática que explorou várias técnicas de inteligência artificial, incluindo aprendizado de máquina. O estudo focou nas aplicações e tendências de IA na doença aterosclerótica.

Tabela 1 – Resultados dos artigos

Autor (data)	Local	Desenho epidemiológico	População de Estudo	Idade Média (anos)	Técnicas de AM utilizadas	Métricas de desempenho	Temas Abordados
Chen et al. 2022	China	Coorte	1.200	60.2	DT, RF, SVM	AUCROC 0.81 (RF)	Previsão de desfechos cardiovasculares, análise de dados clínicos
Wang et al. 2022	Estados Unidos	Coorte	500	56,4	LR, RF, SVM, XGBoost, KNN	AUCROC 0.87 (XGBoost)	Previsão de aterosclerose baseada em dados clínicos e de imagem
Munger et al. 2021	Estados Unidos	Revisão sistemática	-	-	CNN, RF, SVM, ANN, KNN, SVM, NB, CE, DA	Acurácia geral até 98%	Análise de imagens, estratificação de risco, descoberta de biomarcadores
Li et al. 2022	Estados Unidos	Coorte	1.300	65,7	LR, ANN, RF, SVM	AUCROC 0.79 (RF) (LR) AUCROC 0.809	Análise de composição de placas ateroscleróticas, uso de dados de DECT
Fan et al. 2021	China	Coorte	18.441	50.86	LR, RF, DT, XGB, GNB, KNN	LR: AUCROC 0.809, ACC 74.7%, F1 59.9%	Previsão de aterosclerose carotídea assintomática usando EHR e AM

Kakadiaris et al. 2018	Estados Unidos	Coorte	5.000	61,0	ANN, RF, SVM	RF: AUC 0.73, ANN: AUC 0.71, SVM: AUC 0.69	Comparação de modelos de AM com modelos de risco convencionais
Liu et al. 2023	Austrália	Coorte	101.110	62,7	LR, RF, DT, XGB, KNN	AUCROC 0.82 (LR), AUCROC 0.90 (RF)	Previsão de risco cardiovascular aterosclerótico usando abordagens de AM
Núñez et al. (2022)	Estados Unidos	Estudo de caso-controle	1.000	58,7	SVM, RF, KNN, DT	Acurácia 85% (RF)	Descoberta de biomarcadores para detecção aprimorada de aterosclerose subclínica
Sun et al. 2023	Taiwan	Coorte	359	61,0	LR, SVM, DT, RF, XGBoost, Gradient Boosting	AUCROC: LR 0.92, SVM 0.88, RF 0.90	Descoberta de biomarcadores para predizer aterosclerose de grandes artérias
Sánchez-Cabo et al. 2020	Espanha	Coorte	1.777	55,2	RF, LR, KNN	AUCROC 0.82 (RF), LR: AUC 0.80, KNN: AUC 0.78	Melhorar definição de risco cardiovascular com AM
Segar et al. 2021	Estados Unidos	Coorte	12.000	54,3	RF, SVM, LR, ANN, DT	AUCROC: RF 0.83, SVM 0.82	Modelos baseados em etnias para prever risco de

Kampaktsis et al. (2023)	Aleman hã	Revisão sistemática	-	-	CNN, RF, SVM, ANN, KNN, NB, DA	Variedade de métricas dependendo do estudo	10 anos de eventos cardiovasculares Aplicação de IA na imagem cardiovascular, análise de biomarcadores, estratificação de risco
--------------------------	-----------	---------------------	---	---	--------------------------------	--	--

Legenda: ANN (Artificial Neural Network); CNN (Convolutional Neural Network); DA (Discriminant Analysis); DT (Decision Tree); GNB (Gaussian Naïve Bayes); KNN (K-Nearest Neighbour); LR (Logistic Regression); RF (Random Forest); SVM (Support Vector Machine); XGB (XGBoost); SHAP (SHapley Additive exPlanations); ACC (Accuracy); AUCROC (Area Under the Receiver Operating Characteristic Curve); AM (Machine Learning); EHR (Electronic Health Record).

A análises dos vieses dos estudos observacionais incluídos nesta pesquisa estão demonstrados na figura 2.

Figura 2 - Análise dos vieses dos estudos observacionais

		Risk of bias domains							Overall
		D1	D2	D3	D4	D5	D6	D7	
Study	Chen et al. 2022	-	-	+	+	+	-	-	-
	Wang et al. 2022	-	+	+	+	+	-	+	+
	Li et al. 2022	+	-	+	+	+	-	-	-
	Fan et al. 2021	+	-	+	+	-	+	+	-
	Kakadiaris et al. 2018	-	+	+	+	+	-	+	-
	Liu et al. 2023	+	+	+	+	+	-	+	+
	Núñez et al. (2022)	-	-	+	+	-	+	+	-
	Sun et al. 2023	-	-	+	+	-	+	+	-
	Segar et al. 2021	+	+	+	+	+	-	+	+
	Sánchez-Cabo et al. 2020	+	+	+	+	+	-	+	+

Domains:
D1: Bias due to confounding.
D2: Bias due to selection of participants.
D3: Bias in classification of interventions.
D4: Bias due to deviations from intended interventions.
D5: Bias due to missing data.
D6: Bias in measurement of outcomes.
D7: Bias in selection of the reported result.

Judgement
- Moderate
+ Low

Discussão

A essência desta revisão sistemática foi explorar e descrever o uso do aprendizado de máquina na melhoria da compreensão e diagnóstico da aterosclerose em diferentes populações de pacientes e indivíduos de alto risco. Nos estudos revisados, o uso de AM está associado à aplicação geral da inteligência artificial na melhoria dos resultados. Para a maioria dos estudos, a aterosclerose é considerada uma das principais causas de mortalidade e morbidade, especialmente em países de alta renda ou desenvolvidos.

O uso de algoritmos de AM na detecção da condição e na melhoria de sua compreensão é de extrema importância para reduzir os custos gerais associados à sua prevalência. Os estudos mostram que, enquanto as abordagens tradicionais tentam analisar a associação entre uma única variável e a doença, isso leva a uma compreensão e aplicação limitadas em populações e indivíduos de alto risco. A modelagem e análise de aprendizado de máquina de tais conjuntos de dados oferecem ferramentas precisas e clinicamente práticas para diagnóstico eficaz e descoberta de biomarcadores proteômicos e metabólicos dinâmicos na medicina aplicada.

A acurácia e a AUCROC são métricas essenciais utilizadas na avaliação do desempenho dos modelos de aprendizado de máquina, especialmente no contexto da previsão e diagnóstico de condições médicas complexas como a aterosclerose. A acurácia mede a proporção de previsões corretas em relação ao total de previsões feitas, fornecendo uma visão geral da eficácia do modelo. Isso significa que se um modelo apresenta alta acurácia, ele consegue prever corretamente a maioria dos casos, sejam eles positivos ou negativos, em relação ao total de exemplos avaliados.

Entretanto, a acurácia pode ser uma métrica menos informativa em conjuntos de dados desbalanceados, onde a distribuição entre as classes não é uniforme. Por exemplo, em um cenário onde a prevalência de uma condição é muito baixa, um modelo pode atingir alta acurácia simplesmente prevendo a ausência da condição para todos os casos, sem realmente capturar a capacidade de identificar corretamente os casos positivos. Isso ocorre porque a métrica não leva em consideração a distribuição das classes, o que pode resultar em uma falsa impressão de bom desempenho do modelo.

Por outro lado, a AUCROC (Area Under the Receiver Operating Characteristic Curve) traduzido para “Área Sob a Curva Característica de Operação do Receptor” é uma métrica mais robusta que avalia a capacidade do modelo de distinguir entre classes positivas e negativas, independentemente do limiar de decisão escolhido. A curva ROC é gerada ao plotar a taxa de verdadeiros positivos (sensibilidade) contra a taxa de falsos positivos (1 - especificidade) em vários pontos de corte. A área sob essa curva (AUCROC) fornece uma medida única da capacidade discriminativa do modelo. Uma AUCROC de 1 indica perfeita

discriminação entre as classes, enquanto uma AUCROC de 0,5 indica desempenho equivalente ao acaso.

A alta AUCROC é particularmente valiosa porque reflete a capacidade do modelo em diversos cenários de limiar, tornando-a uma métrica mais informativa em situações complexas e desbalanceadas. Isso significa que um modelo com alta AUCROC é eficaz em identificar corretamente pacientes com e sem a condição de interesse, independentemente da prevalência da condição no conjunto de dados.

Em contextos médicos, como na previsão de doenças cardiovasculares, a AUCROC ajuda a determinar a eficácia de testes diagnósticos e modelos preditivos, garantindo que os profissionais de saúde possam tomar decisões informadas e precisas. Além disso, a AUCROC é especialmente útil em cenários onde há um desbalanceamento nas classes, proporcionando uma avaliação mais robusta do modelo em comparação com a acurácia simples.

Os artigos revisados destacam diversas abordagens de AM aplicadas ao diagnóstico e tratamento da aterosclerose. O estudo de Fan et al. (2021), realizado em Xi'an, China, utilizou uma coorte retrospectiva com 18.441 participantes para prever a aterosclerose carotídea assintomática usando registros eletrônicos de saúde e técnicas de aprendizado de máquina, incluindo Regressão Logística (LR), Floresta Aleatória (RF), Árvore de Decisão (DT), eXtreme Gradient Boosting (XGB), Gaussian Naïve Bayes (GNB) e K-Nearest Neighbour (KNN). A LR obteve uma AUCROC de 0.809, uma acurácia de 74.7% e um F1 Score de 59.9%.

O trabalho de Munger et al. (2021) evidencia que o aprendizado de máquina é uma ferramenta poderosa na descoberta e validação de biomarcadores para a aterosclerose. A capacidade dos algoritmos de AM de processar grandes volumes de dados e identificar padrões complexos oferece uma vantagem significativa sobre os métodos tradicionais. Isso não apenas melhora a precisão diagnóstica, mas também facilita uma melhor compreensão dos mecanismos subjacentes à aterosclerose, permitindo intervenções mais precoces e eficazes. Os biomarcadores identificados pelo estudo incluem lipoproteína de baixa densidade (LDL), apolipoproteína B (ApoB), HDL, apolipoproteína A1 (ApoA1) e proteína C-reativa (CRP). Esses biomarcadores são significativos para a inflamação vascular e são essenciais para quase todos os processos de doenças inflamatórias, incluindo a aterosclerose. Com o aumento das técnicas de biologia molecular e a

disponibilidade de grandes conjuntos de dados melhor curados, a identificação de biomarcadores pode-se tornar mais eficaz e precisa.

O artigo de Núñez et al. (2022) também foca na descoberta de biomarcadores, sendo estes proteômicos no plasma para a detecção aprimorada da aterosclerose subclínica. A pesquisa utilizou proteômica quantitativa de alta capacidade para analisar amostras de plasma de 444 indivíduos da coorte PESA, distribuídos igualmente entre aqueles com extensa aterosclerose identificada por imagem e controles correspondentes. Este estudo foi repetido após três anos para validação interna e também validado externamente em uma coorte de 350 indivíduos do estudo AWHS. Os resultados da pesquisa identificaram cinco proteínas associadas à aterosclerose subclínica: PIGR, IGHA2, APOA, HPT e HEP2. Dentre estas, três proteínas - IGHA2, APOA e HPT - formaram um painel de biomarcadores que foram posteriormente validados por imunoturbidimetria em 2999 indivíduos do estudo ILERVAS. Este painel mostrou-se eficaz na predição de aterosclerose subclínica, com uma AUCROC de 0,73 na população geral e de 0,71 na subpopulação de baixo risco cardiovascular.

A identificação e validação destes biomarcadores são de grande relevância clínica, pois podem oferecer uma alternativa acessível à imagem cardiovascular, que nem sempre está disponível universalmente. O estudo sugere que a medição dos níveis plasmáticos de IGHA2, APOA e HPT pode melhorar significativamente as estratégias de prevenção primária, especialmente em áreas onde a imagem cardiovascular não é acessível.

Os achados indicam que esses biomarcadores são capazes de prever a aterosclerose subclínica independentemente dos fatores de risco tradicionais, proporcionando uma ferramenta valiosa para a reclassificação do risco cardiovascular em indivíduos de baixo a moderado risco

O artigo de Kampaktsis et al. (2023) explora amplamente o papel da inteligência artificial (IA) e do AM na doença aterosclerótica. A pesquisa destaca a aplicação crescente de técnicas de IA para melhorar o diagnóstico, avançar na compreensão da patogênese da doença, prever desfechos, auxiliar na tomada de decisões clínicas e promover abordagens de medicina de precisão. Os algoritmos de AM, em particular, já estão sendo empregados em aplicações de imagem cardiovascular para facilitar a detecção automatizada de doenças, com a expectativa de transformar o campo nos próximos anos. Uma das aplicações

importantes discutidas é a caracterização das placas ateroscleróticas e a detecção de características de alto risco, como o sinal do anel guardanapo (Napkin-ring sign). Estudos demonstraram que algoritmos de AM podem identificar essas características com alta precisão, superando as métricas quantitativas tradicionais de TC.

Outro estudo significativo é o de Kakadiaris et al. (2018), conduzido no Multi-Ethnic Study of Atherosclerosis (MESA) nos EUA. Este estudo de coorte prospectiva com 5.000 participantes foi realizado com 6459 participantes livres de doenças cardiovasculares no início do estudo e seguidos por 13 anos e comparou modelos de AM, como Rede Neural Artificial (ANN), Floresta Aleatória (RF) e Máquina de Vetores de Suporte (SVM), com modelos de risco convencionais. Os resultados mostraram que o modelo de aprendizado de máquina superou significativamente o Calculador de Risco ACC/AHA. O modelo de AM apresentou uma sensibilidade de 0.86, especificidade de 0.95 e uma AUC de 0.92 para a previsão de "Hard CVD" (eventos cardiovasculares graves). Em comparação, o Calculador de Risco ACC/AHA teve uma sensibilidade de 0.76, especificidade de 0.56 e uma AUC de 0.71. Para a previsão de "All CVD" (todos os eventos cardiovasculares), o modelo de AM também teve um desempenho superior, com uma sensibilidade de 0.96, especificidade de 0.87 e AUC de 0.94, enquanto o Calculador de Risco ACC/AHA obteve sensibilidade de 0.75, especificidade de 0.59 e AUC de 0.72. O estudo destaca que o modelo de aprendizado de máquina não apenas recomendou menos terapia com estatinas (11.4% dos participantes) em comparação com o ACC/AHA (46.0%), mas também conseguiu identificar mais indivíduos de alto risco que o ACC/AHA teria deixado sem tratamento, reduzindo as oportunidades perdidas de tratamento.

Sánchez-Cabo et al. (2020), em Madrid, Espanha, conduziram um estudo de coorte longitudinal com 1.777 participantes, utilizando RF, LR e KNN para melhorar a definição de risco cardiovascular para jovens assintomáticos. A RF alcançou uma AUCROC de 0.82, a LR uma AUCROC de 0.80 e a KNN uma AUC de 0.78, demonstrando a capacidade do AM de fornecer previsões de risco mais robustas e precisas.

O estudo de Li et al. (2022) explorou a aplicação de técnicas de AM e dados de registros médicos eletrônicos (EMR) para melhorar a predição de doenças cardiovasculares ateroscleróticas. Est incluíam características transversais (CS) e

características longitudinais (LT) derivadas de estatísticas vitais e valores laboratoriais. A combinação dessas características permitiu uma predição mais precisa das doenças cardiovasculares ateroscleróticas.

Os resultados mostraram que todos os modelos de AM testados superaram o calculador de risco “ Pooled Cohort Equations” (PCE). A técnica de floresta aleatória (RF) aplicada à combinação de características transversais e longitudinais (RF-LTC) produziu a melhor área sob a curva (AUCROC) de 0,902 Sun et al. (2023) mostraram que o AM pode ser usado eficazmente na descoberta de biomarcadores e na predição de aterosclerose de grandes artérias (LAA). Em seu estudo retrospectivo realizado no China Medical University Hospital em Taichung, Taiwan, com 359 participantes, utilizaram LR, SVM, DT, RF, XGBoost e Gradient Boosting. A LR apresentou uma AUCROC de 0.92, uma acurácia de 0.82, recall de 0.80 e precisão de 0.81, demonstrando a eficácia do AM na identificação de biomarcadores e na predição precisa da aterosclerose. Neste estudo, os pesquisadores propuseram um novo método que integra múltiplos algoritmos de AM e métodos de seleção de características para lidar com dados multidimensionais. Entre os seis modelos de AM testados, o modelo de regressão logística (LR) exibiu o melhor desempenho preditivo, onde 62 características foram incorporadas no conjunto de validação externa. Este modelo foi capaz de prever a LAA com base em fatores de risco clínicos, como índice de massa corporal, tabagismo e medicamentos para controle de diabetes, hipertensão e hiperlipidemia, bem como metabólitos envolvidos na biossíntese de aminoacil-tRNA e metabolismo lipídico. Além disso, foram identificadas 27 características comuns entre os cinco modelos adotados (LR, SVM, RF, XGBoost e gradient boosting), que, quando usadas no modelo LR, alcançaram um valor de AUC de 0.93. Isso demonstra a eficácia da combinação de algoritmos de AM com eliminação recursiva de características e métodos de validação cruzada na identificação de biomarcadores. O estudo destacou que os métodos tradicionais de análise de biomarcadores tentam correlacionar uma única variável com a doença, resultando em uma abordagem limitada. Com o aumento na disponibilidade de conjuntos de dados biológicos maiores e melhor organizados, a modelagem e análise de aprendizado de máquina desses dados podem fornecer ferramentas práticas e precisas para diagnósticos

Segar et al. (2021), nos EUA, desenvolveram modelos baseados em AM para prever o risco de eventos cardiovasculares em 10 anos, utilizando uma

população de 13.814 participantes. Com uma AUCROC de 0.77 e uma acurácia de 0.75, o estudo sublinhou a importância de considerar variáveis raciais nos modelos preditivos, evidenciando que o AM pode oferecer previsões mais precisas e inclusivas.

Esses estudos demonstram que a integração de aprendizado de máquina na medicina cardiovascular pode transformar significativamente o diagnóstico e o tratamento da aterosclerose. Comparativamente, o calculador de risco de ACC/AHA apresenta uma AUCROC em torno de 0.70 a 0.75, dependendo da população estudada. A maior sensibilidade e especificidade dos modelos de AM, como evidenciado pelos estudos revisados, significa que menos pacientes correm o risco de receber tratamentos desnecessários ou de terem suas condições ignoradas até que se tornem críticas. Isso pode resultar em economias significativas de custos e melhorias na qualidade do atendimento ao paciente, tornando o sistema de saúde mais eficiente e responsivo. Implementar essas tecnologias em cenários clínicos, utilizando dados eletrônicos de saúde para melhorar o diagnóstico e a gestão de doenças, poderia se tornar uma ferramenta de extrema utilidade para a prática médica.

Nesta pesquisa, foram identificadas algumas limitações que devem ser consideradas ao interpretar os resultados. Em primeiro lugar, há um risco de viés inerente aos estudos incluídos. Além disso, a revisão foi conduzida por um único revisor, o que pode introduzir viés de seleção e interpretação dos dados. Outro fator limitante foi a inclusão de revisões sistemáticas, que, embora proporcionem uma visão abrangente, podem introduzir duplicação de dados e vieses próprios. Adicionalmente, alguns artigos relevantes estavam disponíveis apenas mediante pagamento, o que impossibilitou o acesso e inclusão de todos os estudos potencialmente relevantes.

Conclusão

Em resumo, a aplicação de aprendizado de máquina na aterosclerose não apenas melhora os resultados clínicos, mas também pode contribuir para um sistema de saúde mais eficiente e proativo. A incorporação dessas tecnologias avançadas pode vir a ser essencial para enfrentar os desafios complexos da medicina cardiovascular moderna.

Referências

1. CHEN, Zihan et al. Prediction of atherosclerosis using machine learning based on operations research. *Mathematical Biosciences and Engineering*, v. 19, n. 5, 2022.
2. FAN, Jiaxin et al. The prediction of asymptomatic carotid atherosclerosis with electronic health records: a comparative study of six machine learning models. *BMC Medical Informatics and Decision Making*, v. 21, 2021.
3. KAKADIARIS, Ioannis A. et al. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *Journal of the American Heart Association*, v. 7, n. 22, 2018.
4. KAMPAKTSIS, Polydoros N. et al. Artificial intelligence in atherosclerotic disease: applications and trends. *Frontiers in Cardiovascular Medicine*, v. 9, 2023.
5. LI, Qi et al. Automating and improving cardiovascular disease prediction using machine learning and EMR data features from a regional healthcare system. *International Journal of Medical Informatics*, v. 163, 2022.
6. LIU, Weber et al. Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: a systematic review and meta-analysis. *European Heart Journal-Quality of Care and Clinical Outcomes*, v. 9, n. 4, 2023.
7. MUNGER, Eric et al. Application of machine learning in understanding atherosclerosis: Emerging insights. *APL Bioengineering*, v. 5, n. 1, 2021.
8. NUNEZ, Estefania et al. Unbiased plasma proteomics discovery of biomarkers for improved detection of subclinical atherosclerosis. *EBioMedicine*, v. 76, 2022.
9. SANCHEZ-CABO, Fatima et al. Machine learning improves cardiovascular risk definition for young, asymptomatic individuals. *Journal of the American College of Cardiology*, v. 76, n. 14, 2020.
10. SEGAR, Matthew W. et al. Development and validation of machine learning–based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation*, v. 143, n. 24, 2021.

11. SUN, Ting-Hsuan et al. Machine learning approaches for biomarker discovery to predict large-artery atherosclerosis. *Scientific Reports*, v. 13, n. 1, 2023.
12. WANG, Ling-Jie et al. Machine learning-based identification of symptomatic carotid atherosclerotic plaques with dual-energy computed tomography angiography. *Journal of Stroke and Cerebrovascular Diseases*, v. 32, n. 8, 2023.