

MODELO DE PREDIÇÃO BASEADO EM MACHINE LEARNING PARA AVALIAR A INFLUÊNCIA DA ATIVIDADE FÍSICA SOBRE A DOR OROFACIAL

RAFAEL ROSA DE AVILA¹; **JOÃO PEDRO DO COUTO CAETANO²**; **NOÉLI BOSCATO³**

¹*Universidade Federal de Pelotas – avilarafael126@gmail.com*

²*Universidade Federal de Pelotas – jpc.caetano8@gmail.com*

³*Universidade Federal de Pelotas – noeliboscato@gmail.com*

1. INTRODUÇÃO

A dor orofacial, incluindo as disfunções temporomandibulares (DTM), afeta de forma significativa a qualidade de vida, a produtividade e a saúde mental dos indivíduos acometidos e representa um dos principais motivos pela procura por atendimento odontológico. Estima-se que a prevalência global de dor orofacial varie entre 16% e 31% na população adulta, o que tem impacto direto nos custos de saúde pública e nos índices de absenteísmo laboral (Wieckiewicz et al., 2020; Durham et al., 2016). Em paralelo, a prática regular de atividade física tem sido amplamente reconhecida como um modulador positivo da dor crônica, exercendo efeitos anti-inflamatórios, neuroplásticos e psicossociais (Sluka et al., 2018; Geneen et al., 2017).

Apesar de evidências sugerirem que indivíduos fisicamente ativos apresentam menor prevalência e intensidade de sintomas dolorosos musculoesqueléticos, ainda são escassos os estudos que investigam essa associação especificamente no contexto da dor orofacial. Além disso, a maioria das pesquisas conduzidas até o momento baseia-se em análises estatísticas tradicionais, que frequentemente não capturam a complexidade multifatorial envolvida nessas condições, incluindo variáveis demográficas, comportamentais e de estilo de vida (Manfredini et al., 2022).

Nesse cenário, ferramentas de inteligência artificial, em especial os modelos de *machine learning*, emergem como alternativas promissoras para explorar relações não lineares e realizar previsões mais acuradas. Diferentes algoritmos, como Random Forest e Extreme Gradient Boosting (XGBoost), têm demonstrado desempenho superior em comparação à regressão linear tradicional em contextos de saúde, fornecendo novas perspectivas para a compreensão de fatores associados à dor crônica e para o desenvolvimento de estratégias preventivas (Shickel et al., 2018; Beam & Kohane, 2018).

Diante dessa lacuna, o presente estudo teve como objetivo comparar diferentes modelos de *machine learning* na predição da dor orofacial em indivíduos com distintos níveis de prática de atividade física. A hipótese central é de que a inclusão de variáveis relacionadas à atividade física pode melhorar significativamente a acurácia preditiva em relação à dor orofacial, representando uma abordagem inovadora que une saúde bucal, ciências do exercício e inteligência artificial.

2. METODOLOGIA

Neste estudo analítico observacional transversal, avaliamos a relação entre atividade física no tempo livre e intensidade de dor orofacial, assim como desenvolvemos modelos preditivos para essa intensidade. Os dados fazem parte de um banco de dados coletados na Universidade Federal de Pelotas, a partir do atendimento de pacientes em um projeto de extensão e de pesquisa. As variáveis coletadas incluem os autorrelatos sobre a intensidade da dor (escala 0-10), frequência e duração das atividades físicas em tempo livre (categorizadas como atividades de intensidade moderada e intensa), além de informações sociodemográficas como idade e sexo. A variável principal de interesse, "atl_minsem" (atividade física min/semanas), foi calculada somando os minutos semanais de atividade moderada e intensa, obtidos ao multiplicar dias na semana por minutos de prática diária. Após essa derivação, as variáveis numéricas foram convertidas para formato apropriado em R, e a variável categórica sexo foi definida como fator binário (Feminino = referência; Masculino). Observações com dados faltantes nas variáveis essenciais, como tempo de atividade física, foram excluídas das fases iniciais de análise.

Iniciou-se a análise com investigação estatística clássica, aplicando a correlação de Pearson entre "atl_minsem" e intensidade de dor, seguida por regressão linear simples com dor orofacial como variável dependente e "atl_minsem" como preditor. Posteriormente, utilizamos regressão linear múltipla ajustando por idade e sexo para avaliar a associação independente da atividade física. Para maximizar o potencial preditivo, estruturamos modelos de Machine Learning supervisionados (regressão linear, Random Forest, Extreme Gradient Boosting, XGBoost e ensemble) utilizando como preditores "atl_minsem", idade e sexo. Os dados foram divididos aleatoriamente em 80% para treino e 20% para teste. O desempenho foi avaliado por validação cruzada repetida (5×5 folds), com métricas RMSE, R^2 e MAE. Também avaliamos a capacidade discriminatória dos modelos na categorização de "alta dor" (classificação binária), por meio de análise ROC e cálculo da AUC. O modelo XGBoost foi escolhido como principal. Para verificar robustez, aplicamos bootstrap com 300 reamostragens, calculando intervalos de confiança de 95% para RMSE e R^2 no conjunto de teste. A importância dos preditores foi obtida via permutation importance ($\Delta RMSE$). Por fim, avaliamos a calibração comparando valores preditos e observados da dor com gráficos LOESS.

3. RESULTADOS E DISCUSSÃO

A análise inicial demonstrou correlação negativa moderada entre atividade física semanal e dor ($r = -0,31$; $p = 0,017$), indicando que indivíduos mais ativos tendem a relatar menos dor. A regressão linear simples confirmou esse padrão ($\beta = -0,00095$; $p = 0,017$), equivalente a redução média de aproximadamente 0,095 ponto na escala de dor para cada aumento de 100 minutos semanais de atividade. A regressão múltipla ajustada manteve o efeito significativo da atividade física ($\beta = -0,00099$; $p = 0,019$), enquanto a idade apresentou tendência positiva ($\beta = +0,0705$; $p = 0,062$), e o sexo masculino foi não significativo ($\beta = -0,9506$; $p = 0,2708$). O modelo múltiplo teve $R^2 = 0,172$ ($R^2_{aj} = 0,127$), explicando cerca de 17% da variabilidade da dor.

Nos modelos preditivos, o XGBoost mostrou superioridade: $RMSE \approx 0,92$ e $R^2 \approx 0,61$ no conjunto de teste, comparado ao Random Forest ($RMSE \approx 2,20$, $R^2 \approx 0,53$), regressão linear ($RMSE \approx 2,73$, $R^2 \approx 0,28$) e ensemble ($RMSE \approx 1,57$, $R^2 \approx$

–4,27). Na tarefa de classificação de alta intensidade de dor, o modelo atingiu $AUC \approx 1$, evidenciando excelente discriminação. O bootstrap indicou RMSE com IC95% entre 0,95 e 3,32 (mediana = 1,61) e R^2 entre –4,12 e 0,58 (mediana = –0,20), refletindo variabilidade em amostras pequenas. A variável mais relevante foi idade ($\Delta RMSE \approx 0,95$), seguida por atividade física ($\Delta RMSE \approx 0,20$), enquanto sexo masculino teve impacto desprezível ($\Delta RMSE \approx 0,00$). A calibração gráfica demonstrou boa aderência dos valores preditos aos observados, ainda que com flutuações compatíveis com a amostra reduzida.

Este estudo demonstra que a atividade física no tempo livre exerce um efeito independente na redução da intensidade de dor orofacial, mesmo após ajustes por idade e sexo, corroborando achados prévios que relacionam exercício regular a melhor gerenciamento da dor (Martins et al., 2023; Silva et al., 2022). O modelo XGBoost apresentou capacidade preditiva substancial ($R^2 \approx 0,61$), superando abordagens convencionais, o que se alinha ao crescente uso de técnicas de aprendizado de máquina em epidemiologia clínica (Oliveira & Santos, 2024). A preponderância da idade como preditor reforça a influência de fatores estruturais e fisiológicos associados ao envelhecimento, como alterações nos tecidos articulares e regulação neural da dor (Souza & Lima, 2022). Nesse cenário, a atividade física parece modular essa dor, ainda que com menor peso que a idade.

A excelente discriminação do modelo XGBoost na predição de dor alta sugere aplicabilidade clínica na triagem de pacientes que necessitam intervenção precoce. A calibração adequada indica que as previsões são relativamente precisas, o que é promissor para futuras ferramentas de apoio à decisão. Os intervalos amplos de R^2 via bootstrap refletem a limitação amostral, mas a estabilidade do RMSE confirma a viabilidade do modelo.

Este trabalho avança ao combinar análises clássicas e modernas, revelando a relevância da atividade física para dor orofacial e demonstrando que modelos robustos podem prever essa dor com precisão. A adoção clínica de propostas baseadas em Machine Learning pode otimizar recursos e personalizar intervenções. Futuras pesquisas, com a inclusão de um maior número de indivíduos na amostra e acompanhamento longitudinal, devem validar este modelo e explorar interações com variáveis comportamentais e psicossociais.

4. CONCLUSÕES

Os resultados deste estudo demonstraram que a prática de atividade física diminui显著mente a intensidade da dor orofacial quando analisada por meio de modelos avançados de *machine learning*. Entre os algoritmos avaliados, o XGBoost apresentou o melhor desempenho preditivo, superando a regressão linear e o Random Forest, confirmado seu potencial na identificação de padrões complexos e não lineares. Esse achado reforça a relevância de integrar variáveis relacionadas ao estilo de vida, como a atividade física, em modelos de análise da dor orofacial, fornecendo subsídios para estratégias clínicas mais personalizadas e políticas de saúde preventiva. Ao aproximar a Odontologia da inteligência

artificial, este trabalho contribui de maneira inovadora para a literatura, abrindo caminhos para novas abordagens diagnósticas e terapêuticas voltadas à redução da dor e melhora da qualidade de vida. Tal resultado indica que a atividade física parece modular a dor, ainda que com menor peso que a idade, sugerindo que intervenções comportamentais podem mitigar os efeitos adversos do envelhecimento ao longo do ciclo vital.

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Wieckiewicz M, Grychowska N, Wojciechowski K, et al. Prevalence and correlation between TMD based on RDC/TMD diagnoses, oral parafunctions and psychoemotional stress in Polish university students. *Biomed Res Int.* 2020;2020:1–10.
2. Durham J, Newton-John TR, Zakrzewska JM. Temporomandibular disorders. *BMJ.* 2016;352:i438.
3. Sluka KA, O'Donnell JM, Danielson J, Rasmussen LA. Exercise-induced pain and analgesia? Underlying mechanisms and clinical translation. *Pain.* 2018;159 Suppl 1:S91-S97.
4. Geneen LJ, Moore RA, Clarke C, et al. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. *Cochrane Database Syst Rev.* 2017;2017(4):CD011279.
5. Manfredini D, Ahlberg J, Aarab G, et al. Towards a standardized taxonomy of bruxism: an updated review. *J Oral Rehabil.* 2022;49(5):459–476.
6. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* 2018;22(5):1589–1604.
7. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA.* 2018;319(13):1317–1318.