

## INTELIGÊNCIA ARTIFICIAL NA PREDIÇÃO DE GENGIVITE EM ADOLESCENTES: UM ESTUDO DE COORTE

**BRUNA ZEMBRUSKI GOMES<sup>1</sup>**; **LILIAN TOLEDO REYES<sup>2</sup>**; **THIAGO MACHADO ARDENGHÍ<sup>3</sup>**; **CAMILA SILVEIRA SFREDDO<sup>4</sup>**

<sup>1</sup>*Universidade Federal de Pelotas – bzg.bruna@gmail.com*

<sup>2</sup>*Universidade Federal de Santa Maria – lilitant.reyes@gmail.com*

<sup>3</sup>*Universidade Federal de Santa Maria – thiardenghi@gmail.com*

<sup>4</sup>*Universidade Federal de Pelotas – camilassfreddo@gmail.com*

### 1. INTRODUÇÃO

A gengivite é definida como uma doença inflamatória reversível, resultado do desequilíbrio entre a resposta imune-inflamatória do hospedeiro e o biofilme dental disbiótico, restrita e confinada aos tecidos gengivais marginais (MURAKAMI et al., 2018). Dados do relatório *The Global Burden of Disease* (GBD, 2017) indicaram que as doenças bucais, incluindo a gengivite, correspondem a algumas das doenças de maior prevalência na escala global. Além disso, inquéritos nacionais de saúde bucal mostraram diferenças na distribuição de gengivite, com carga de doença mais elevada em grupos populacionais menos favorecidos economicamente (BRASIL, 2011). A gengivite impacta negativamente na qualidade de vida e pode atuar como um potencial fator de risco para estágios avançados de doenças periodontais, os quais podem resultar em perda dentária (MURAKAMI et al., 2018).

Assim, uma precisa predição do risco de gengivite na adolescência, com o uso de IA, pode ser uma tecnologia útil para a área da saúde no intuito de reorientar políticas públicas, auxiliar na tomada de decisões, contribuir na prevenção da doença e, consequentemente, melhorar a qualidade de vida dos indivíduos. Além disso, a adolescência é um período crucial para a saúde uma vez que se caracteriza por importantes mudanças fisiológicas e psicossociais. Portanto, intervir em fatores de risco para doenças nesta fase em que os indivíduos estão mais vulneráveis pode permitir adoção de comportamentos e estilo de vida mais saudáveis que tendem a persistir na vida adulta. Assim, o objetivo deste estudo foi desenvolver e validar modelos de predição de gengivite em adolescentes, utilizando uma abordagem “*machine learning*”.

### 2. METODOLOGIA

Este estudo utilizou dados secundários de uma coorte prospectiva de adolescentes de Santa Maria, RS. O ponto inicial foi um levantamento epidemiológico de saúde bucal, de caráter transversal, realizado em 2012. A metodologia empregada e os resultados obtidos foram publicados em estudos prévios (TOMAZONI et al., 2014). A amostra inicial consistiu dos escolares de 12 anos cujos pais consentiram a participação na pesquisa através da assinatura do termo de consentimento livre e esclarecido. A primeira avaliação ocorreu em 2012 (T1), com 1.134 adolescentes de 12 anos de idade de escolas públicas. Posteriormente, foi realizado um acompanhamento em 2014 (T2) e 2018 (T3).

O estudo foi conduzido de acordo com a declaração Helsinki e obteve aprovação do Comitê de Ética em Pesquisa em todos os períodos de

acompanhamento (números de protocolo: 0127.0.243.000-11 em 2012; 30613714.0.0000.5421 em 2014; e 66553117.4.0000.5346 em 2018).

Para este estudo, foram utilizados os dados coletados em T2 e T3. Em T2, foram avaliados fatores sociodemográficos (renda familiar, escolaridade materna, aglomeração domiciliar, sexo e cor da pele), psicossociais (autopercepção de saúde bucal), comportamental (frequência de escovação) e fatores clínicos (biofilme dental, cálculo, histórico de gengivite, cárie dental não tratada e apinhamento dentário). O desfecho foi considerado gengivite no T3, caracterizada quando há 10% ou mais de sangramento à sondagem, de acordo com a atual Classificação da Doenças Periodontais (TROMBELL et al., 2018).

A análise da associação entre os preditores no T2 e o desfecho gengivite no T3 foi realizada considerando o modelo dos determinantes sociais de saúde (SOLAR; IRWIN, 2010). Para obter os modelos preditivos usando algoritmos de “*machine learning*”, os arquivos foram processados usando o software R Studio versão 1.4.1717. A seleção de características com o algoritmo Boruta foi adotada para determinar as variáveis a serem inseridas nos modelos “*machine learning*”. Populares algoritmos foram implementados para desenvolver modelos de “*machine learning*”: árvore de decisão, “*random forest*” e “*extreme gradiente boosting*” (XGBoost) foram utilizados, juntamente com modelos de regressão logística. Os resultados dos modelos de regressão logística foram apresentados como odds ratio (OR) e seu respectivo intervalo de confiança (IC) de 95%. A avaliação do desempenho do modelo foi baseada em medidas de discriminação e calibração. A discriminação foi avaliada usando a curva ROC com análise da área sob a curva (AUC). A precisão, sensibilidade e especificidade também foram obtidas. Diferenças entre as curvas foram analisadas pelos testes DeLong, *bootstrap* e Venkatraman. Além disso, avaliamos a sensibilidade das curvas ROC em um determinado nível de especificidade usando testes de sensibilidade e especificidade. A calibração foi avaliada usando o escore Brier (valores mais próximos de zero indicam um melhor desempenho de previsão). Além disso, calculamos os valores de Shapley usando o algoritmo “*SHapley Additive exPlanations*” (SHAP) para o modelo preditivo com melhor desempenho, a fim de interpretar a contribuição de cada variável na previsão do resultado. O nível de significância considerado foi de 5%.

### **3. RESULTADOS E DISCUSSÃO**

Ao total, 742 e 749 adolescentes foram reavaliados no T2 e T3, respectivamente (taxa de retenção de 67,8%). A média de idade foi de 14,3 anos (desvio padrão = 0,69) em 2014, e a maioria era do sexo feminino (55,9%). Além disso, 29,7% dos participantes tinham baixa renda familiar e 45,94% já haviam apresentado gengivite anteriormente. Não houve diferenças estatisticamente significativas entre os conjuntos de treinamento e teste para nenhuma das variáveis. Em nossa amostra, 440 adolescentes (58,9%) desenvolveram gengivite após um período de acompanhamento de 4 anos.

O modelo de regressão logística ajustado revelou que os adolescentes do sexo masculino tinham menor probabilidade de desenvolver gengivite (OR, 0,64; IC 95%, 0,44–0,91). No entanto, aqueles com uma quantidade considerável de biofilme dental (OR, 2,16; IC 95%, 1,36–3,46) e histórico de gengivite (OR, 1,61; IC 95%, 1,04–2,50) tinham maior probabilidade de desenvolver gengivite.

As variáveis identificadas por meio do algoritmo Boruta são mostradas na Figura 1. A Figura 2 ilustra o gráfico de resumo SHAP para o modelo XGBoost, que

forneceu informações sobre os determinantes que influenciam a probabilidade de desenvolvimento de gengivite. As classificações das variáveis preditoras foram intimamente alinhadas com o modelo de regressão logística. Valores SHAP superiores a 0 para biofilme dental, experiência prévia de gengivite, sexo masculino e presença de cárie dentária indicam que esses fatores são fatores de risco significativos para o desenvolvimento da doença.

No geral, os modelos exibiram desempenho modesto. Especificamente, durante o treinamento, apenas o modelo XGBoost alcançou um valor AUC acima de 0,70. No entanto, quando avaliado nos conjuntos de dados de teste, o desempenho deste modelo diminuiu (AUC 0,64), enquanto o modelo de regressão logística rendeu uma AUC de 0,586.

O teste de DeLong indicou diferenças estatisticamente significativas nos valores de AUC entre os modelos XGBoost e regressão logística durante o treinamento ( $P= 0,03$ ). Porém, essas diferenças não foram observadas nos conjuntos de testes ( $P= 0,11$ ). Por outro lado, o *bootstrapping* de especificidade manteve diferenças estatísticas entre esses modelos tanto para os conjuntos de treinamento quanto para os conjuntos de teste.

Diversos estudos têm buscado avaliar fatores preditivos para gengivite. Contudo, os modelos de predição geralmente investigam fatores biológicos e comportamentais individuais, difíceis de estimar risco no nível comunitário. Assim, novos modelos que também avaliem a dinâmica entre os fatores socioeconômicos e psicosociais e a saúde bucal devem ser desenvolvidos. Com os avanços da tecnologia da computação, surge a inteligência artificial, a qual pode ser definida como a habilidade não biológica de uma máquina, que tenta mimetizar o funcionamento do cérebro humano no desenvolvimento de tarefas complexas, como a resolução de problemas, reconhecimento de objetos, palavras e tomada de decisões (SHAN et al., 2021). Dentre os usos da inteligência artificial na Odontologia, está o desenvolvimento de procedimentos combinando padrões estatísticos e computacionais na análise e interpretação crítica dos conjuntos de dados disponíveis, para obter critérios válidos e fazer previsões eficazes de doenças bucais (REYES et al., 2021; SHAN et al., 2021).

Nosso estudo, pela primeira vez na literatura, desenvolveu e testou modelos de previsão de gengivite para adolescentes com base em preditores socioeconômicos, psicosociais, comportamentais e clínicos, utilizando dados coletados em um estudo de coorte de acompanhamento de 4 anos. Para isso, o método convencional de regressão logística e vários algoritmos populares de “*machine learning*” foram implementados, sendo que o algoritmo XGBoost obteve o melhor desempenho. Esses resultados promissores sugerem a viabilidade potencial da inteligência artificial para apoiar a tomada de decisões eficazes na avaliação de risco de gengivite.

#### 4. CONCLUSÕES

A presença de níveis extensos de biofilme dental, experiência prévia de gengivite, sexo masculino e presença de cárie dentária não tratada foram fatores de risco significativos para o desenvolvimento de gengivite em adolescentes. Assim, a abordagem de “*machine learning*” revelou potencial para determinar o desenvolvimento de gengivite em adolescentes, utilizando preditores fáceis de serem coletados durante a adolescência.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. Ministério da Saúde. **SB Brasil 2010**. Pesquisa Nacional de Saúde Bucal – Resultados Principais. Brasília, DF, 2011.

GBD 2017 ORAL DISORDERS COLLABORATORS *et al.* Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the global burden of disease 2017 study. **Journal of dental research**, 2020, v. 99, n. 4, p. 362-373.

MURAKAMI, S. *et al.* Dental plaque-induced gingival conditions. **Journal of Clinical Periodontology**, v. 45, suppl. 20, p. S17-S27, 2018.

REYES, L. T.; KNORST, J. K.; ORTIZ, F. R.; ARDENGHI, T. M; Scope and challenges of machine learning-based diagnosis and prognosis in clinical dentistry: A literature review. **Journal of clinical and translational research**, Singapore, v. 7, n. 4, p. 523–539, 2021.

SHAN T., TAY F. R., GU L. Application of artificial intelligence in dentistry. **Journal of dental research**, 2021 Mar, vol.100, n. 3, p. 232-44.

SOLAR, O.; IRWIN, A. **A conceptual framework for action on the social determinants of health**. WHO Document Production Services, 2010.

TOMAZONI, F. *et al.* Association of gingivitis with child oral health-related quality of life. **Journal of Periodontology**, v. 85, n. 11, p. 1557-65, Nov. 2014.

TROMBELLI, L; *et al.* Plaque-induced gingivitis: Case definition and diagnostic considerations. **Journal of clinical periodontology**, Ferrara, v. 45, n. 20, p. 44-67, 2018.