

PREDIÇÃO DE DIABETES ATRAVÉS DE MACHINE LEARNING: UMA REVISÃO SISTEMÁTICA

FELIPE MENDES DELPINO¹; ÂNDRIA KROLOW COSTA²; SABRINA RIBEIRO FARIAS³; ALEXANDRE DP CHIAVEGATTO FILHO⁴; RICARDO ALEXANDRE ARCÊNCIO⁵; BRUNO PEREIRA NUNES⁶

¹Universidade Federal de Pelotas – fmdsocial@outlook.com

²Universidade Federal de Pelotas – andriakc@hotmail.com

³Universidade Federal de Pelotas – sabrinarfarias@gmail.com

⁴Universidade de São Paulo – alexdiasporto@usp.br

⁵Universidade de São Paulo – ricardo@eerp.usp.br

⁶Universidade Federal de Pelotas – nunesbp@gmail.com

1. INTRODUÇÃO

A diabetes é uma doença endócrina que pode levar a problemas mais severos, como doença renal, doenças cardiovasculares e óbito (YANG *et al.*, 2019). Conviver com a diabetes pode estar associado a anos de vida perdidos, podendo chegar a cinco para homens e seis para mulheres (WRIGHT *et al.*, 2017). Os custos para tratar a diabetes são elevados. Além disso, há expectativa de aumento nos custos para tratar a diabetes pois, atualmente, as pessoas vivem mais tempo com a doença (GREGG *et al.*, 2014). Globalmente, estima-se que os custos possam passar de 1.3 trilhões em 2015 para mais de 2 trilhões de dólares até 2030 (BOMMER *et al.*, 2018).

Predizer com antecedência a diabetes pode ser útil na prevenção e manejo adequado da doença. Pensando nisso, a machine learning pode surgir com uma ideia inovadora e capaz de prever com antecedência a ocorrência de doenças crônicas, incluindo a diabetes. A machine learning é baseada na premissa de que computadores são capazes de aprender a identificar padrões a partir de dados sem programação manual (BATISTA AFM, 2019). Os métodos utilizam algoritmos computacionais aliados a elementos de um conjunto de variáveis preditoras.

Alguns estudos já vêm utilizando modelos de machine learning para ajudar no diagnóstico e tratamento de diferentes doenças, dentre as quais a diabetes se destaca por ser predita com boa acurácia (LAI *et al.*, 2019). A presente revisão teve como objetivo avaliar estudos que usaram modelos de machine learning para predição de diabetes em humanos.

2. METODOLOGIA

Trata-se de uma revisão sistemática realizada nas bases de dados do Pubmed, LILACS, Scielo, Web of Science e Scopus. A revisão seguiu as recomendações do *Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) e foi registrada no *International prospective register of systematic reviews* (PROSPERO), sob número de protocolo CRD42020215581. Foram utilizados os seguintes Mesh termos "machine learning" AND "type 2 diabetes". Foram incluídos estudos que avaliaram a diabetes usando algoritmos (modelos) de machine learning e reportaram os valores utilizando valores de área sob a curva (AUC). Os valores de AUC, variam de 0 a 1, de modo que quanto mais perto do 1 melhor é a capacidade preditiva do modelo. Estudos de revisão ou com animais foram excluídos. Dois autores realizaram o processo de seleção dos

estudos, o qual começou pela leitura dos títulos, seguido pelos resumos, até a leitura completa dos artigos. Por fim, foram revisadas as referências dos estudos incluídos para encontrar novos artigos.

Para avaliar a qualidade dos estudos, utilizou-se uma versão adaptada da escala *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis* (TRIPOD). A escala é composta por 20 itens, sendo 14 para métodos e seis para resultados. Além disso, a escala é composta por um checklist, no qual o objetivo é melhorar o relato de estudos que desenvolvem, validam ou atualizam um modelo de predição, para fins de diagnósticos ou prognósticos.

3. RESULTADOS E DISCUSSÃO

No total, 1891 títulos foram encontrados, dos quais 245 resumos foram selecionados para leitura. Destes, oito foram incluídos na revisão (BARAKAT; BRADLEY; BARAKAT, 2010; CHANGALA; RAJESWARA RAO, 2019; HAN *et al.*, 2015; HATHAWAY *et al.*, 2019; LAI *et al.*, 2019; MANIRUZZAMAN *et al.*, 2020; OLIVERA *et al.*, 2017; TIGGA; GARG, 2020; WANG *et al.*, 2021; XIE *et al.*, 2019; XIONG *et al.*, 2019; ZHANG, Lei *et al.*, 2020; ZHANG, Liying *et al.*, 2020). Entre os oito estudos, os valores de AUC variaram de 0.7524 a 1 (OLIVERA *et al.*, 2017; TIGGA; GARG, 2020).

Em um estudo, com 4682 indivíduos, foi encontrado um valor de AUC de 0.9489, através do modelo *Support vector machine* (BARAKAT; BRADLEY; BARAKAT, 2010). Han e colaboradores avaliaram mais de 79 chineses, utilizando características comportamentais e sociodemográficas, e encontraram um valor de AUC de 0.869, através do modelo *Random forest* (HAN *et al.*, 2015). Em outra publicação, com mais de 12 mil brasileiros, através de 27 fatores de risco para diabetes, os autores encontraram um valor de AUC de 0.7524, utilizando o modelo Artificial neural network (OLIVERA *et al.*, 2017). Com dados de 952 participantes Indianos, através do modelo *Random Forest*, foi encontrado um AUC de 1, demonstrando que o modelo foi capaz de predizer todos os casos de diabetes (TIGGA; GARG, 2020). Na escala TRIPOD, nenhum estudo alcançou a pontuação máxima.

Verificou-se que o modelo com predição máxima utilizou o menor tamanho de amostra. Estimativas com baixo número de indivíduos podem fornecer um viés na predição e um sobreajuste. Isso ocorre porque o número analisado de indivíduos pode ter sido muito pequeno para avaliar a real qualidade preditiva dos modelos. Por esse motivo, recomenda-se o uso de machine learning com bancos de dados grandes para resultados mais precisos e fidedignos.

4. CONCLUSÕES

Os resultados da presente revisão mostraram que a machine learning é capaz de predizer a ocorrência de diabetes com boa performance. Se bem implementados, os modelos de machine learning poderão ser úteis na tomada de decisões por profissionais de saúde, melhorando a qualidade do serviço e reduzindo custos com a saúde.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BARAKAT, N.; BRADLEY, A. P.; BARAKAT, M. N. H. Intelligible support vector machines for diagnosis of diabetes mellitus. **IEEE Transactions on Information Technology in Biomedicine**, [s. l.], v. 14, n. 4, p. 1114–1120, 2010. Available at: <https://doi.org/10.1109/TITB.2009.2039485>. Acesso em: 17 set. 2020.

BATISTA AFM, C. F. A. Machine Learning aplicado à Saúde. Workshop: Machine Learning. **19º Simpósio Brasileiro de Computação Aplicado à Saúde. Sociedade Brasileira de Computação**, [s. l.], 2019. Available at: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/29/95/245-1>. Acesso em: 6 nov. 2020.

BOMMER, C. *et al.* Global Economic Burden of Diabetes in Adults: Projections From 2015 to 2030. **Diabetes Care**, [s. l.], v. 41, n. 5, p. 963–970, 2018. Available at: <https://doi.org/10.2337/dc17-1962>. Acesso em: 22 set. 2020.

CHANGALA, R.; RAJESWARA RAO, D. Development of predictive model for medical domains to predict chronic diseases (diabetes) using machine learning algorithms and classification technique. **ARPN Journal of Engineering and Applied Sciences**, [s. l.], v. 14, n. 6, p. 1202–1212, 2019. Available at: www.arnjournals.com. Acesso em: 19 set. 2020.

GREGG, E. W. *et al.* Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985-2011: A modelling study. **The Lancet Diabetes and Endocrinology**, [s. l.], v. 2, n. 11, p. 867–874, 2014. Available at: [https://doi.org/10.1016/S2213-8587\(14\)70161-5](https://doi.org/10.1016/S2213-8587(14)70161-5). Acesso em: 22 set. 2020.

HAN, L. *et al.* Rule extraction from support vector machines using ensemble learning approach: An application for diagnosis of diabetes. **IEEE Journal of Biomedical and Health Informatics**, [s. l.], v. 19, n. 2, p. 728–734, 2015. Available at: <https://doi.org/10.1109/JBHI.2014.2325615>. Acesso em: 26 set. 2020.

HATHAWAY, Q. A. *et al.* Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. **Cardiovascular Diabetology**, [s. l.], v. 18, n. 1, p. 78, 2019. Available at: <https://doi.org/10.1186/s12933-019-0879-0>. Acesso em: 26 set. 2020.

LAI, H. *et al.* Predictive models for diabetes mellitus using machine learning techniques. **BMC Endocrine Disorders**, [s. l.], v. 19, n. 1, p. 101, 2019. Available at: <https://doi.org/10.1186/s12902-019-0436-6>. Acesso em: 23 set. 2020.

MANIRUZZAMAN, M. *et al.* Classification and prediction of diabetes disease using machine learning paradigm. **Health Information Science and Systems**, [s. l.], v. 8, n. 1, 2020. Available at: <https://doi.org/10.1007/s13755-019-0095-z>. Acesso em: 20 set. 2020.

OLIVERA, A. R. *et al.* Comparação de algoritmos de aprendizagem de máquina para construir um modelo preditivo para detecção de diabetes não diagnosticada – ELSA-Brasil: Estudo de acurácia. **Sao Paulo Medical Journal**, [s. l.], v. 135, n. 3, p. 234–246, 2017. Available at: <https://doi.org/10.1590/1516-3180.2016.0309010217>. Acesso em: 18 set. 2020.

TIGGA, N. P.; GARG, S. Prediction of Type 2 Diabetes using Machine Learning Classification Methods. **Procedia Computer Science**, [s. l.], v. 167, p. 706–716, 2020. Available at: <https://doi.org/10.1016/j.procs.2020.03.336>

WANG, Y. *et al.* Genetic Risk Score Increased Discriminant Efficiency of Predictive Models for Type 2 Diabetes Mellitus Using Machine Learning: Cohort Study. **Frontiers in Public Health**, [s. l.], v. 9, p. 606711, 2021. Available at: <https://doi.org/10.3389/fpubh.2021.606711>. Acesso em: 9 abr. 2021.

WRIGHT, A. K. *et al.* Life expectancy and cause-specific mortality in type 2 diabetes: A population-based cohort study quantifying relationships in ethnic subgroups. **Diabetes Care**, [s. l.], v. 40, n. 3, p. 338–345, 2017. Available at: <https://doi.org/10.2337/dc16-1616>. Acesso em: 22 set. 2020.

XIE, Z. *et al.* Building risk prediction models for type 2 diabetes using machine learning techniques. **Preventing Chronic Disease**, [s. l.], v. 16, n. 9, p. 190109, 2019. Available at: <https://doi.org/10.5888/pcd16.190109>. Acesso em: 26 set. 2020.

XIONG, X. lu *et al.* Machine Learning Models in Type 2 Diabetes Risk Prediction: Results from a Cross-sectional Retrospective Study in Chinese Adults. **Current Medical Science**, [s. l.], v. 39, n. 4, p. 582–588, 2019. Available at: <https://doi.org/10.1007/s11596-019-2077-4>. Acesso em: 9 abr. 2021.

YANG, J. J. *et al.* Association of Diabetes with All-Cause and Cause-Specific Mortality in Asia: A Pooled Analysis of More Than 1 Million Participants. **JAMA Network Open**, [s. l.], v. 2, n. 4, 2019. Available at: <https://doi.org/10.1001/jamanetworkopen.2019.2696>. Acesso em: 25 abr. 2021.

ZHANG, Lei *et al.* Predicting the development of type 2 diabetes in a large australian cohort using machine-learning techniques: Longitudinal survey study. **JMIR Medical Informatics**, [s. l.], v. 8, n. 7, 2020. Available at: <https://doi.org/10.2196/16850>. Acesso em: 9 abr. 2021.

ZHANG, Liying *et al.* Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan Rural Cohort Study. **Scientific Reports**, [s. l.], v. 10, n. 1, p. 1–10, 2020. Available at: <https://doi.org/10.1038/s41598-020-61123-x>. Acesso em: 9 abr. 2021.