

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: Computação Inteligente ou Modelagem Computacional

Título: Modelagem Preditiva Estratificada por Determinantes Sociais da Saúde para Análise de Desfechos da Tuberculose no Brasil com Aplicação de Inteligência Artificial

Orientador: Maicon Herverton Lino Ferreira da Silva Barros
(maicon.herverton@upe.br)

A tuberculose permanece como uma das doenças infecciosas de maior impacto no mundo e sua dinâmica no Brasil caracteriza-se por forte heterogeneidade regional e profunda influência de determinantes sociais da saúde, como escolaridade, renda, densidade domiciliar, condições de moradia e desigualdades de acesso aos serviços (CORTEZ *et al.*, 2021; WHO, 2023; WHO, 2024). Estudos recentes demonstram que além de fatores clínicos e laboratoriais, aspectos sociais e contextuais estão diretamente associados a desfechos desfavoráveis como abandono, falha terapêutica, morte e perda de seguimento (CHAVES TORRES *et al.*, 2019; PENG *et al.*, 2024; CHEN *et al.*, 2024). Entretanto, a maior parte dos modelos de aprendizado de máquina voltados para tuberculose ainda opera de maneira global, sem considerar estratificações populacionais específicas ou cenários diferenciados por perfis de vulnerabilidade.

A presente proposta visa desenvolver modelos preditivos estratificados por determinantes sociais da saúde, capazes de identificar padrões de risco segundo perfis populacionais distintos, como população privada de liberdade, população em situação de rua, adolescentes, idosos, indivíduos com comorbidades e grupos expostos a maior vulnerabilidade estrutural. O trabalho utilizará dados clínicos e sociodemográficos provenientes do SINAN-TB, integrados a variáveis derivadas de indicadores sociais e territoriais. Serão formados diferentes cenários experimentais, comparando modelos treinados globalmente com modelos específicos para cada grupo populacional, examinando sua capacidade de generalização, explicabilidade e utilidade prática para vigilância epidemiológica.

A literatura mostra que métodos de aprendizado de máquina apresentam desempenho robusto na predição de desfechos da tuberculose (HUSSAIN e JUNEJO, 2019; SAUER *et al.*, 2018; AHAMED FAYAZ *et al.*, 2024; RODRIGUES *et al.*, 2024), mas ainda há escassez de estudos focados em estratificação e análise orientada a determinantes sociais. As contribuições brasileiras mais relevantes, incluindo Lino Ferreira da Silva Barros *et al.* (2021; 2022; 2023; 2024a; 2024b; 2025), estabeleceram avanços metodológicos importantes ao consolidar bases padronizadas e propor benchmarks rigorosos. A presente pesquisa avança além desses trabalhos ao incorporar camadas contextuais, estratificação populacional, comparação entre macrocenários e microcenários e interpretação de impacto social dos atributos.

Do ponto de vista metodológico, o projeto utilizará técnicas de pré-processamento, imputação, engenharia de atributos derivados de indicadores sociais, algoritmos de aprendizado de máquina de diferentes famílias (árvores de decisão, *boosting*, SVM, regressão logística, TabPFN e modelos interpretáveis) e métodos explicáveis como SHAP e PDP. A avaliação seguirá métricas adequadas a bases desbalanceadas (F1-macro, MCC, AUC) em combinação com validação cruzada e validação temporal. Além disso, serão comparados modelos globais e modelos específicos para cada subgrupo, permitindo avaliar se há ganhos reais de desempenho e interpretabilidade quando a modelagem é contextualizada.

O resultado esperado é a criação de um conjunto de modelos contextualizados que ofereçam suporte mais preciso aos serviços de vigilância epidemiológica, permitindo identificar vulnerabilidades específicas em grupos populacionais distintos. A pesquisa também contribuirá cientificamente ao estabelecer metodologia sistemática de modelagem preditiva estratificada por determinantes sociais, alinhada às necessidades do Sistema Único de Saúde (SUS) e do cenário epidemiológico brasileiro.

Referências Bibliográficas:

- AHAMED FAYAZ, Shaik *et al.* **Machine learning algorithms to predict treatment success for patients with pulmonary tuberculosis.** PLoS One, v. 19, n. 10, p. e0309151, 2024.
- CHAVES TORRES, Ninfa Marlen *et al.* **Factors predictive of the success of tuberculosis treatment: A systematic review with meta-analysis.** PloS one, v. 14, n. 12, p. e0226507, 2019.
- CHEN, Jingfang *et al.* **Predictive machine learning models for anticipating loss to follow-up in tuberculosis patients throughout anti-TB treatment journey.** Scientific Reports, v. 14, n. 1, p. 24685, 2024.
- CHINAGUDABA, SessaSai Nath *et al.* **Predictive Analysis of Tuberculosis Treatment Outcomes Using Machine Learning: A Karnataka TB Data Study at a Scale.** arXiv preprint arXiv:2403.08834, 2024.
- CORTEZ, Andreza Oliveira *et al.* **Tuberculosis in Brazil: one country, multiple realities.** Jornal Brasileiro de Pneumologia, v. 47, p. e20200119, 2021.
- HOLLMANN, Noah *et al.* **Tabpfn: A transformer that solves small tabular classification problems in a second.** arXiv preprint arXiv:2207.01848, 2022.
- HUSSAIN, Owais A.; JUNEJO, Khurum N. **Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models.** Informatics for Health and Social Care, v. 44, n. 2, p. 135-151, 2019.
- KILLIAN, Jackson A. *et al.* **Learning to prescribe interventions for tuberculosis patients using digital adherence data.** In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 2019. p. 2430-2438.
- LINO FERREIRA DA SILVA BARROS, Maicon Herverton *et al.* **Benchmarking machine learning models to assist in the prognosis of tuberculosis.** In: *Informatics.* MDPI, 2021. p. 27. <https://doi.org/10.3390/informatics8020027>.
- LINO FERREIRA DA SILVA BARROS, Maicon Herverton *et al.* **A Brazilian classified data set for prognosis of tuberculosis, between January 2001 and April 2020.** Sci Data 9, 771 (2022). <https://doi.org/10.1038/s41597-022-01892-4>.
- LINO FERREIRA DA SILVA BARROS, Maicon Herverton *et al.* **Machine-learning model for classification of the prognosis of tuberculosis using real data from Brazil.** In: *2023 18th Iberian Conference on Information Systems and Technologies (CISTI).* IEEE, 2023. p. 1-6. <https://doi.org/10.23919/CISTI58278.2023.10211510>
- LINO FERREIRA DA SILVA BARROS, Maicon Herverton *et al.* **Evaluating the Usability of an AI-Powered Application for Tuberculosis Mortality Prediction.** In: *Proceedings of the 19th Iberian Conference on Information Systems and Technologies – CISTI 2024a, Volume 1.* Cham: Springer, 2026. p. 1-6. (Lecture Notes in Networks and Systems). <https://link.springer.com/book/9783032128751>.
- LINO FERREIRA DA SILVA BARROS, Maicon Herverton *et al.* **Machine-learning for classification of the prognosis of tuberculosis using real data from Brazil.** Tese (doutorado em engenharia de computação), Universidade de Pernambuco, Recife (2024b).
- LINO FERREIRA DA SILVA BARROS, M. H. *et al.* **Machine Learning Classification of Favorable vs Unfavorable Tuberculosis Treatment Outcomes Using Clinical and Sociodemographic Data from Brazil's SINAN-TB (2001–2023),** 03 September 2025, PREPRINT (Version 1) available at Research Square <https://doi.org/10.21203/rs.3.rs-7502054/v1>.
- PENG, An-zhou *et al.* **Explainable machine learning for early predicting treatment failure risk among patients with TB-diabetes comorbidity.** Scientific Reports, v. 14, n. 1, p. 6814, 2024.
- RODRIGUES, Moreno MS *et al.* **Machine learning algorithms using national registry data to predict loss to follow-up during tuberculosis treatment.** BMC Public Health, v. 24, n. 1, p. 1385, 2024.
- RUSDAH, Rusdah; PAINEM, Painem; KUSUMANINGSIH, Dewi. **Enhancing Prediction of Treatment Duration in New Tuberculosis Cases: A Comprehensive Approach with Ensemble Methods and Medication Adherence.** Jurnal Teknik Informatika (Jutif), v. 6, n. 2, p. 891-904, 2025.
- SAUER, Christopher Martin *et al.* **Feature selection and prediction of treatment failure in tuberculosis.** PloS one, v. 13, n. 11, p. e0207491, 2018.
- WORLD HEALTH ORGANIZATION (WHO). **Global Tuberculosis Report 2023.** Geneva: World Health Organization, 2023. Disponível em: <https://www.who.int/publications/i/item/9789240083851>. Acesso em: 22 de nov. de 2025.

WORLD HEALTH ORGANIZATION (WHO). *Global Tuberculosis Report 2024*. Geneva: WHO, 2024.
Disponível em: <https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024>. Acesso em: 22 de nov. de 2025.