

Modelo Preditivo para Resultados Clínicos da Tuberculose com Redes Neurais

Predictive Model for Tuberculosis Clinical Outcomes with Neural Networks

Modelo Predictivo para Resultados Clínicos de Tuberculosis con Redes Neuronales

Ronilson W. S. Pereira¹, Igor W. S. Falcão², Marcos C. da R. Seruffo³, Karla Figueiredo⁴

¹Doutorando no Programa de Pós-Graduação em Ciências Computacionais – CCOMP, Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro (RJ), Brasil.

²Doutorando no Programa de Pós-Graduação em Engenharia Elétrica – PPGEE, Universidade Federal do Pará, Belém (PA), Brasil.

³Professor adjunto no Instituto de Tecnologia – ITEC, Universidade Federal do Pará – UFPA, Belém (PA), Brasil

⁴Professora adjunta no Departamento de Informática e Computação, Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro (RJ), Brasil

Autor Correspondente: **Ronilson W. S. Pereira**

E-mail: ronilsonengenharia@gmail.com

RESUMO

Objetivo: Este estudo visa desenvolver um modelo preditivo com Redes Neurais *Multilayer Perceptron* (MLP) para prever desfechos no tratamento da tuberculose, identificando abandono ou cura. **Métodos:** Foram aplicadas técnicas de imputação de dados faltantes, codificação de variáveis, normalização dos dados, balanceamento de classes usando *SMOTETomek* e *dropout* para regularização do modelo. **Resultados:** O modelo sem balanceamento obteve acurácia macro de 0.6220, precisão de 0.9105 e *recall* de 0.9779. Após balanceamento, a acurácia micro aumentou para 0.8577, mas o *recall* caiu para 0.8294. A precisão foi de 0.8749 e o F1-micro foi 0.8555. **Conclusões:** O balanceamento de classes melhorou o desempenho global, mas reduziu o *recall*. O uso de Redes Neurais *Multilayer Perceptron* e técnicas de balanceamento pode aprimorar a predição de desfechos da tuberculose, contribuindo para estratégias de saúde pública.

Descritores: Tuberculose, MLP, ML

ABSTRACT

Objective: This study aims to develop a predictive model using Multilayer Perceptron Neural Networks (MLP) to predict outcomes in tuberculosis treatment, identifying either

abandonment or cure. Methods: Techniques for imputing missing data, variable encoding, data normalization, class balancing using SMOTETomek, and dropout for model regularization were applied. Results: The model without balancing achieved a macro accuracy of 0.6220, precision of 0.9105, and recall of 0.9779. After balancing, the micro accuracy increased to 0.8577, but recall decreased to 0.8294. Precision was 0.8749, and the F1-micro score was 0.8555. Conclusions: Class balancing improved overall performance but reduced recall. The use of Multilayer Perceptron Neural Networks and balancing techniques can enhance tuberculosis outcome prediction, contributing to public health strategies.

Keywords: Tuberculosis, MLP, ML.

RESUMEN

Objetivo: Este estudio tiene como objetivo desarrollar un modelo predictivo utilizando Redes Neuronales Multilayer Perceptron (MLP) para predecir los resultados en el tratamiento de la tuberculosis, identificando abandono o cura. Métodos: Se aplicaron técnicas de imputación de datos faltantes, codificación de variables, normalización de datos, balanceo de clases utilizando SMOTETomek y dropout para la regularización del modelo. Resultados: El modelo sin balanceo obtuvo una precisión macro de 0.6220, precisión de 0.9105 y recall de 0.9779. Después del balanceo, la precisión micro aumentó a 0.8577, pero el recall disminuyó a 0.8294. La precisión fue de 0.8749 y el F1-micro fue de 0.8555. Conclusiones: El balanceo de clases mejoró el rendimiento general, pero redujo el recall. El uso de Redes Neuronales Multilayer Perceptron y técnicas de balanceo puede mejorar la predicción de los resultados de la tuberculosis, contribuyendo a las estrategias de salud pública.

Descriptores: Tuberculosis, MLP, ML

INTRODUCTION

A Tuberculose (TB) é uma doença infecciosa transmitida causada pela bactéria *Mycobacterium*, que atinge principalmente os pulmões humanos. A tuberculose continua sendo uma das enfermidades infecciosas mais letais do planeta, sendo um enfrentamento considerável para a saúde pública global ⁽¹⁾. Conforme o relatório global de tuberculose do ano de 2023, milhões de pessoas adoecem a cada ano devido à

tuberculose, e muitos sofrem complicações graves que podem resultar em morte, sobretudo em populações vulneráveis ⁽²⁾.

O tratamento da tuberculose é influenciado por diversos fatores, tornando-se um processo difícil ⁽³⁾. A determinação do sucesso do tratamento de TB é importante, dependendo de fatores como adesão à medicação, resistência aos medicamentos, comorbidades como o HIV, situação socioeconômica e outros dados clínicos ⁽⁴⁾. Prever os desfechos do tratamento de forma eficaz e eficiente é importante devido à complexidade do tratamento, que pode envolver múltiplos medicamentos e um período prolongado de administração ⁽⁵⁾.

Neste contexto, a aplicação de Inteligência Artificial (IA) e Aprendizado de Máquina (ML) tem se destacado como uma solução promissora. Redes Neurais Artificiais (RNAs), como o *Multilayer Perceptron* (MLP), são eficazes na previsão dos resultados do tratamento da tuberculose ⁽⁶⁾. Modelos de ML podem identificar pacientes em risco de falha no tratamento, permitindo intervenções rápidas e personalizadas ⁽⁷⁾.

Este estudo tem como objetivo desenvolver um modelo de aprendizado de máquina para prever os desfechos do tratamento da Tuberculose (cura ou abandono). A análise dos dados dos pacientes foi realizada com um *Multilayer Perceptron*, utilizando técnicas de balanceamento, como o *SMOTETomek*, e *dropout* para evitar *overfitting*. Os atributos de entrada foram selecionados com base na literatura, sendo classificados em categorias sociais e clínicas.

Embora as redes MLP sejam uma técnica bem estabelecida, nossa abordagem se destaca ao combinar MLP e *SMOTETomek* para o problema dos desfechos do tratamento da tuberculose, oferecendo uma solução específica para lidar com o desequilíbrio de dados. Diferente das abordagens tradicionais, nosso estudo não apenas avalia o desempenho do MLP, mas também investiga o impacto do balanceamento de dados na precisão e sensibilidade do modelo, especialmente em dados médicos, onde é crucial identificar desistências.

O restante do artigo é organizado da seguinte forma: os principais trabalhos correlatos são detalhados na Seção 2. A Seção 3 apresenta a metodologia utilizada, detalhando o modelo proposto. A seção 4 apresenta os resultados obtidos para a

previsão de desfecho de tratamento. Por fim, as principais conclusões alcançadas são resumidas e os trabalhos futuros explicados na Seção 5.

TRABALHOS CORRELATOS

Existem diversos estudos que aplicam inteligência artificial e aprendizado de máquina ao tratamento da tuberculose. Esta seção destaca alguns desses trabalhos e os relaciona com o estudo proposto neste artigo.

No estudo de Lian et al. (2023) ⁽⁸⁾, os autores propuseram modelos preditivos para hepatite aguda, insuficiência respiratória aguda e mortalidade após o tratamento da tuberculose, utilizando seis algoritmos de aprendizado de máquina (*XGBoost*, *random forest*, MLP, *light GBM*, regressão logística e SVM). Os resultados indicaram que o MLP obteve o maior valor de Área sobre a curva (AUC) (0,834) para a predição de mortalidade.

No trabalho de Janah et al. (2024) ⁽⁹⁾, os autores desenvolveram e avaliaram um MLP e *Extreme Learning Machine* (ELM) para a detecção precoce de TB utilizando dados clínicos. Foram realizados testes de balanceamento de dados com a Técnica de Sobreamostragem Minoritária (*SMOTE*). O melhor resultado obtido foi uma acurácia de 95% com o modelo MLP, sem a aplicação da técnica de *SMOTE*.

No estudo de Mohidem et al. (2021) ⁽¹⁰⁾, os autores desenvolveram um modelo para prever o número de casos de TB no distrito de Gombak, na Malásia, com base em fatores sociodemográficos e ambientais. Foram utilizadas Regressão Linear Múltipla (MLR) e Redes Neurais Artificiais (RNA) para desenvolver o modelo. A abordagem que apresentou os melhores resultados foi a RNA3, que avaliou a relação entre os aspectos sociodemográficos e os fatores ambientais, alcançando uma acurácia de 96%.

Um estudo de Orjuela et al. (2022) ⁽⁷⁾ explorou o uso de aprendizado de máquina no diagnóstico de tuberculose em cenários com recursos limitados. Foram desenvolvidos modelos com técnicas como regressão logística, árvores de classificação, *random forest*, SVM e redes neurais artificiais. O modelo de RNA obteve os melhores resultados, destacando-se em acurácia, sensibilidade e AUC.

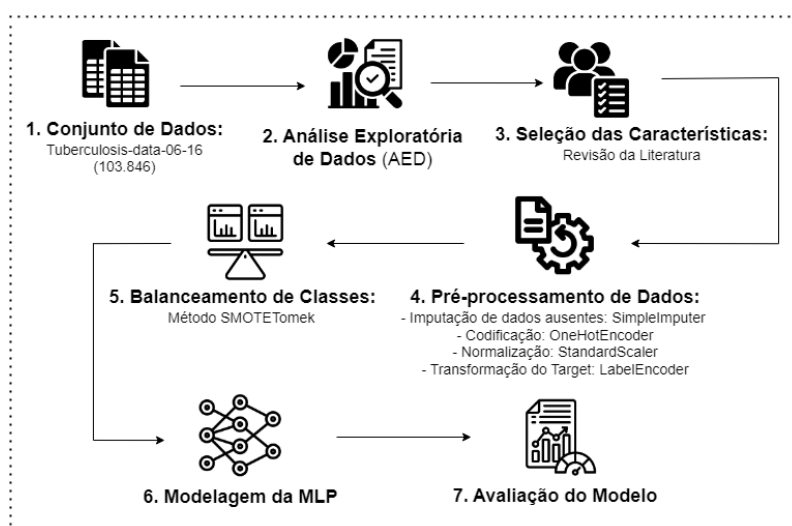
Embora existam diversos estudos sobre o uso de aprendizado de máquina para prever desfechos de doenças infecciosas, incluindo a tuberculose, poucos investigam

o impacto específico de redes MLP combinadas com técnicas de balanceamento para prever desfechos de tratamento. Este estudo foca na combinação de MLP e *SMOTETomek*, avaliando a eficácia dessas técnicas em cenários de desequilíbrio de alto nível.

MATERIAS E MÉTODOS

Nesta seção, são detalhadas as etapas e os métodos utilizados para o desenvolvimento do modelo para a predição dos desfechos de tratamento da tuberculose. Detalharam-se os passos desde o conjunto de dados, pré-processamento, modelagem e avaliação dos dados, para desenvolver um modelo robusto. O fluxo de trabalho metodológico utilizado neste estudo está resumido na Figura 1.

Figura 1. Fluxo de Trabalho da Metodologia Utilizada.



Conjunto de Dados

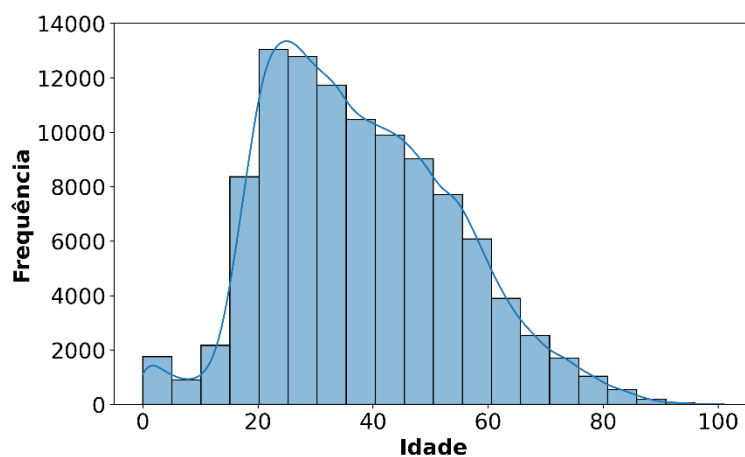
A base de dados utilizada está disponível no repositório público¹, este estudo analisou dados de 103.846 prontuários de pacientes com tuberculose no estado de São Paulo, obtidos de 2006 a 2016 por meio do TBWEB, sistema de informação utilizado para monitoramento da tuberculose. O conjunto de dados inclui variáveis demográficas, clínicas e de tratamento.

¹ https://figshare.com/articles/dataset/tuberculosis-data-06-16_csv/8066663?file=15032345

Análise Exploratória de Dados (AED)

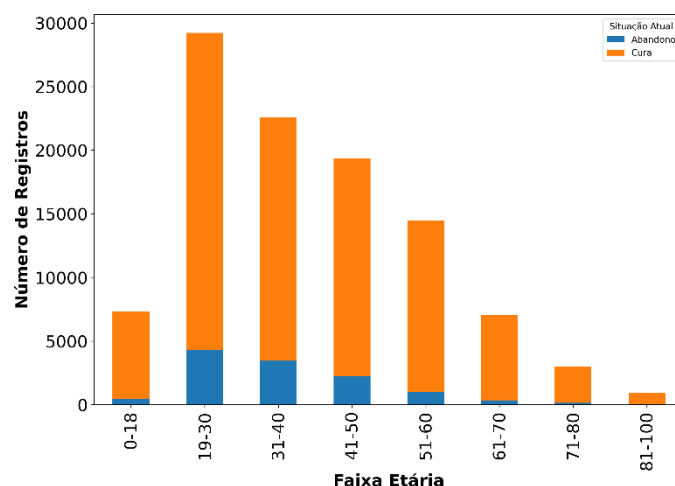
Análise Exploratória de Dados (AED) pretende examinar os dados antes da aplicação de qualquer método de aprendizado de máquina ou algum tipo de técnica estatística. Desta forma, a distribuição e características (*features*) dos atributos foram compreendidas por meio de AED. Representações gráficas como histogramas e gráficos de barras foram criadas para variáveis numéricas e categóricas, respectivamente. As análises incluíram a distribuição etária dos pacientes e o número de casos discriminados por sexo.

Figura 2. Distribuição das Idades dos Pacientes.



A Figura 2 apresenta a distribuição das idades dos pacientes no conjunto de dados. A análise mostra que a maior frequência de pacientes está na faixa etária de 20 a 25 anos. Também se observou que crianças de 5 a 10 anos têm uma menor frequência de casos de TB e, de forma similar, há uma diminuição da frequência a partir dos 60 anos. Ao verificar a contagem de casos por sexo, notou-se mais pacientes do sexo masculino, totalizando 68.755, enquanto havia 35.091 pacientes do sexo feminino.

Figura 3. Número de Registros por Faixa Etária e Situação Atual.



Por fim, os registros foram agrupados e analisados conforme a faixa etária e a situação atual do paciente (abandono ou cura). Com base nos dados da Figura 3, observou-se mais pacientes curados na faixa etária de 19 a 30 anos, totalizando 24.895 pacientes. Devido à discrepância nos dados sobre a situação atual dos pacientes, notou-se um desbalanceamento entre as classes de cura e abandono. O conjunto de dados original compreende 103.846 registros e 47 atributos.

Seleção das Características

Para facilitar a análise, as variáveis preditoras foram categorizadas com base em características (*features*) sociais e clínicas. A seleção tornou-se fundamentada na revisão da literatura, na disponibilidade de dados e na importância para a predição dos desfechos, conforme apresentado nos trabalhos de Orjuela et al. (2022) ⁽⁷⁾ e Kanesamoorthy et al. (2021) ⁽¹¹⁾.

Tabela 1. Atributos Clínicos Selecionados.

Atributos Clínicos	Descrição
tipoCaso	Tipo de caso
FORMACLIN1	Forma clínica
hiv	HIV
DIABETES	Diabetes
classif	Classificação
descoberta	Método de descoberta de casos
bac	Microscopia de esfregaço
BACOUTRO	Microscopia de esfregaço com outro material
cultEsc	Cultura de escarro
RX	Raio-x do tórax
NECROP	Necropsia
ALCOOLISMO	Alcoolismo
MENTAL	Doença mental
DROGADICAO	Dependência de drogas
instTrat	Tipo de instituição
esqIni	Esquema de tratamento inicial
HISTOPATOL	Histopatológico
faixaEtaria	Grupo de idade
TABAGISMO	Fumar

Tabela 2. Atributos Socioeconômicos Selecionados.

Atributos Sociais	Descrição
racaCor	Paciente autor-referido etnia
sexo	Gênero
ESCOLARID	Escolaridade
TIPOCUP	Tipo de ocupação

A Tabela 1 apresenta os atributos clínicos, incluindo comorbidades como diabetes, HIV e outros. O número de doses primárias e secundárias recebidas são indicadores de adesão e resposta ao tratamento, enquanto idade e tipo de caso são fatores críticos no planejamento do tratamento ⁽²⁾. O conjunto final de dados incluiu 23 atributos, divididos em duas categorias: Clínicos (Tabela 1) e Socioeconômicos (Tabela 2). Foram selecionadas 4 características (*features*) socioeconômicas: raça/cor, sexo, escolaridade e tipo de ocupação, conforme ilustrado na Tabela 2. Estudos, como o de Rodrigues et al. (2020) ⁽¹²⁾, indicam que fatores socioeconômicos e demográficos influenciam a adesão ao tratamento e os resultados clínicos.

Pré-processamento dos Dados

Para lidar com dados ausentes, empregou-se a técnica de imputação, que substitui os valores ausentes usando uma estatística descritiva (por exemplo, média, mediana ou mais frequente) ao longo de cada coluna ou usando um valor constante. Isso garante que não haja mais valores ausentes no conjunto de dados.

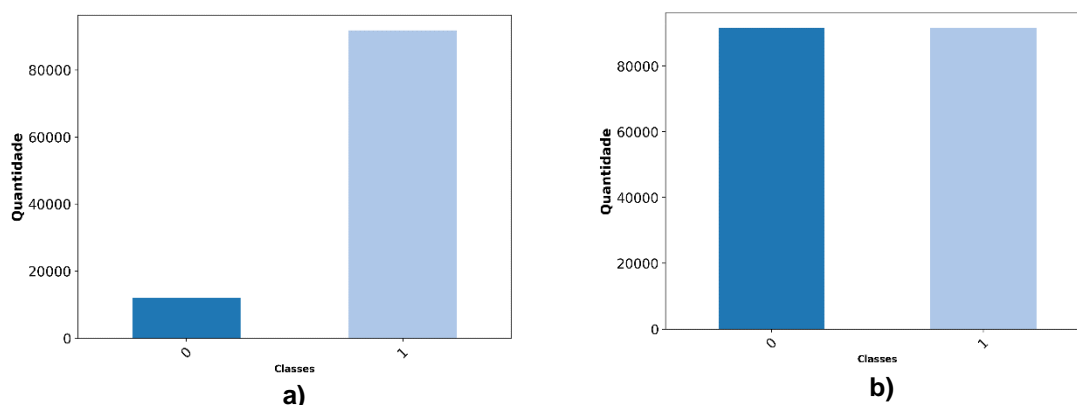
Realizou-se a codificação das características categóricas com o *OneHotEncoder*, transformando cada categoria em uma coluna binária, e a normalização das características numéricas com o *StandardScaler*, ajustando a média para zero e o desvio padrão para um ⁽¹³⁾. A variável de saída foi convertida com o *LabelEncoder* em uma binária, onde '0' representa abandono e '1' representa cura. O conjunto de dados contém 91.823 pacientes com cura ('1') e 12.023 com abandono ('0').

Balanceamento de Classes

Para abordar o desbalanceamento das classes, foi aplicado o método *SMOTETomek*, que combina o *SMOTE* (*Synthetic Minority Over-sampling Technique*) com *Tomek Links* para criar um conjunto de dados balanceado. Esse método visa

melhorar a eficiência do treinamento do modelo e permitir uma classificação mais precisa das classes.

Figura 4. (a) Distribuição da classe de saída antes do balanceamento dos dados; (b) Distribuição da classe de saída após o balanceamento dos dados com *SMOTETomek*.



Devido ao desequilíbrio no conjunto de dados de TB, que apresenta uma diferença notável na distribuição de classes no atributo de situação atual, com uma razão de classes de 7,64, conforme verificado na fase de AED e ilustrado na Figura 4 (a), utilizou-se a técnica *SMOTETomek* para resolver esse desequilíbrio. Essa técnica visa melhorar a eficiência do treinamento do modelo e permitir uma classificação mais precisa das classes.

O método combina *Tomek Links*, que sub-amostra a classe majoritária para aproximá-la da classe minoritária, com o *SMOTE*, que realiza a sobreamostragem da classe minoritária, criando pontos de dados sintéticos para se aproximar da classe majoritária ⁽¹⁴⁾. A Figura 4 (b) mostra o resultado do balanceamento, aumentando o número de registros da classe '0' e promovendo uma distribuição mais equitativa entre as classes '0' e '1'.

Modelagem e Avaliação

Um dos métodos projetado para estudar o mecanismo de doenças relacionadas a eventos como recidiva da doença em pacientes com TB é o modelo de Rede Neural Artificial (RNA), o uso de que aumentou nas últimas décadas. Atualmente, os modelos de RNA são amplamente utilizados em processos complexos e vários campos das ciências médicas, como diagnóstico e tratamento, devido ao seu alto valor de precisão preditiva e flexibilidade ⁽¹⁵⁾.

Neste estudo, utilizou-se uma MLP, uma rede neural de camadas na qual as informações fluem da camada de entrada para a camada de saída por meio de camadas ocultas. O modelo de cada neurônio da rede contém uma função de ativação não linear diferenciável. Uma rede contém uma ou mais camadas ocultas dos nós de entrada e saída. O alto grau de conectividade de uma rede é determinado pelos pesos sinápticos da rede ⁽¹⁶⁾.

Figura 5. Componentes da estrutura do MLP.

Layer (type)	Output Shape	Param #
dense_18 (Dense)	(None, 128)	14592
dropout_12 (Dropout)	(None, 128)	0
dense_19 (Dense)	(None, 64)	8256
dropout_13 (Dropout)	(None, 64)	0
dense_20 (Dense)	(None, 1)	65
Total params: 22913 (89.50 KB)		
Trainable params: 22913 (89.50 KB)		
Non-trainable params: 0 (0.00 Byte)		

A estrutura do modelo MLP neste estudo consiste em 5 camadas, conforme mostrado na Figura 5. A arquitetura inclui uma camada de entrada que recebe os recursos do conjunto de dados de treinamento, seguida por duas camadas ocultas densas com 128 e 64 neurônios, respectivamente, ambas utilizando a função de ativação ReLU. Para reduzir o risco de *overfitting*, aplicou-se a técnica de regularização de *dropout* com taxa de 50%, desligando aleatoriamente metade dos neurônios durante o treinamento.

A camada de saída do modelo consiste em um único neurônio com função de ativação Sigmoid, ideal para prever probabilidades em problemas de classificação binária. O modelo foi compilado com a função de perda binária de entropia cruzada e otimizado pelo algoritmo 'Adam' com taxa de aprendizado de 0,001. O treinamento foi realizado por 50 épocas, com tamanho de lote de 32, utilizando 80% dos dados para treinamento, 20% para teste (103.846 registros) e 20% dos dados de treinamento para validação.

As métricas de perda e acurácia foram monitoradas durante o treinamento, e o modelo foi avaliado no conjunto de teste. O impacto do balanceamento de classes foi

analisado em duas condições: não balanceado e balanceado (SMOTETomek). As métricas F1-score, acurácia e Recall foram utilizadas para comparar os desempenhos.

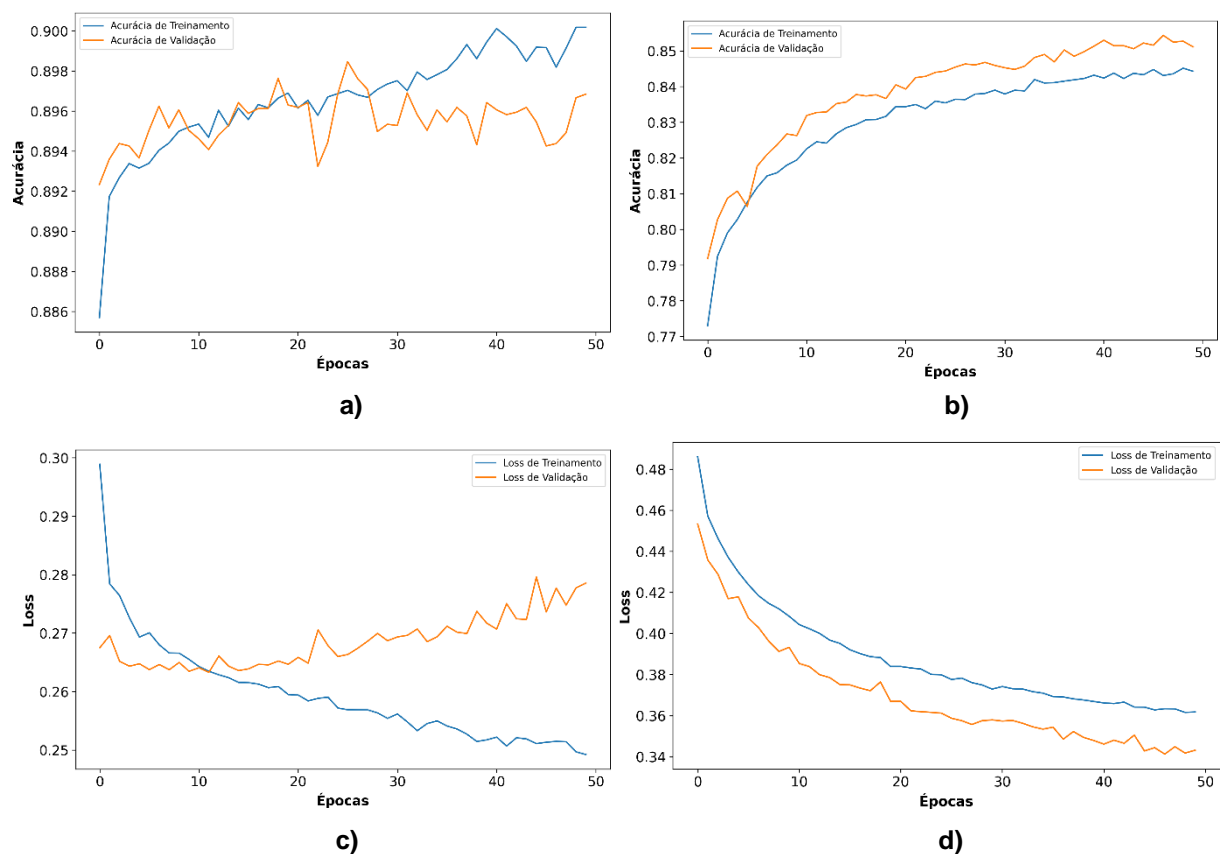
Em cenários de classificação com dados desbalanceados, a escolha das métricas F1 e acurácia (micro e macro) é crucial para uma avaliação adequada:

- Para o conjunto de dados não balanceado, utilizou-se F1-macro e acurácia macro, dando o mesmo peso a ambas as classes, já que o desbalanceamento tende a favorecer a classe majoritária.
- Para o conjunto de dados balanceado, onde as classes foram equalizadas, aplicou-se F1-micro e acurácia micro, que consideram o desempenho geral, ponderando igualmente todas as instâncias.

RESULTADOS E DISCUSSÃO

Nesta seção, são discutidos os resultados obtidos pelos modelos sem e com balanceamento de classes e discutimos o efeito do desbalanceamento de classes no desempenho e na capacidade de generalização dos modelos.

Figura 6. (a) Acurácia ao Longo das Épocas sem Balanceamento; (b) Acurácia ao Longo das Épocas com Balanceamento; (c) Curva de Aprendizado sem balanceamento; (d) Curva de Aprendizado com balanceamento.



A Figura 6 apresenta as curvas de aprendizado e de acurácia ao longo das épocas, fundamentais para analisar o comportamento do modelo durante o treinamento. Na Figura 6 (d), observa-se a curva de aprendizado com o conjunto de dados balanceado, onde as funções de perda (*loss*) para os conjuntos de treinamento e validação exibem uma tendência decrescente consistente, indicando uma melhoria contínua no ajuste do modelo. Em contrapartida, na Figura 6 (c), que mostra a curva de aprendizado sem balanceamento, embora a perda seja menor, há uma maior discrepância entre as curvas de perda do treinamento e da validação.

Tabela 3. Resultados das Métricas de Avaliação.

Métrica	Conjunto Não Balanceado	Conjunto Balanceado (<i>SMOTETomek</i>)
Acurácia Macro	0.6220	-
Precisão	0.9105	0.8749
Recall	0.9779	0.8294
F1-macro	0.6570	-
Acurácia micro	-	0.8577
F1-micro	-	0.8555

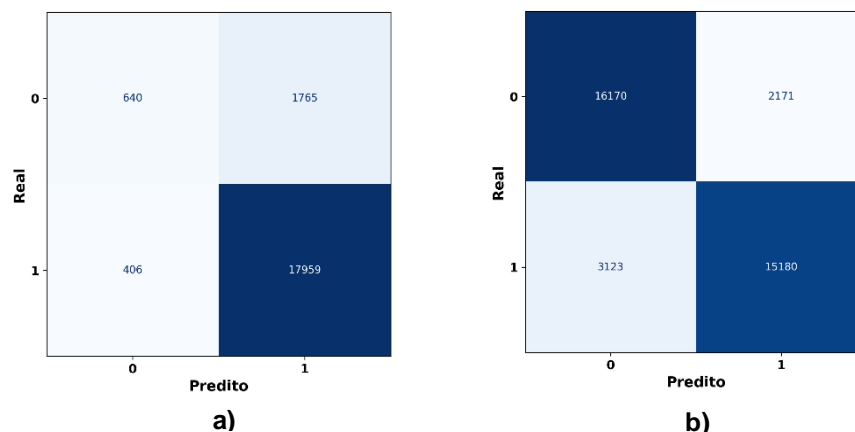
Conforme ilustrado na Tabela 3, a comparação entre os conjuntos de dados balanceado e não balanceado revela diferenças importantes no desempenho do modelo. Para o conjunto não balanceado, a acurácia macro foi de 0.6220, refletindo um desempenho moderado influenciado pelo desbalanceamento das classes. A precisão foi alta (0.9105), enquanto o recall foi ainda mais elevado (0.9779), indicando boa capacidade de identificar casos positivos. O F1-macro foi 0.6570, equilibrando precisão e *recall*.

No conjunto balanceado, utilizando a técnica *SMOTETomek*, a acurácia micro aumentou para 0.8577, demonstrando melhor desempenho global. A precisão foi de 0.8749 e o recall caiu para 0.8294, o que sugere uma leve perda na identificação de casos positivos. A F1-micro foi 0.8555, refletindo um bom equilíbrio geral. Esses resultados mostram que, enquanto o balanceamento melhorou o desempenho global, ele causou uma leve queda no *recall*, destacando o *trade-off* entre precisão e sensibilidade em contextos clínicos de pacientes com TB.

Conforme mostrado na Figura 7, as matrizes de confusão destacam o desempenho do modelo nos cenários sem balanceamento (a) e com balanceamento

(b). No conjunto não balanceado, o modelo teve alta acurácia para a classe majoritária (17.959 acertos), mas apresentou falsos negativos (406) e falsos positivos (1.765), afetando a classe minoritária. Com o balanceamento, houve uma distribuição mais equilibrada, com 16.170 acertos na classe 0 e 15.180 na classe 1, mas com aumento nos falsos positivos (2.171) e falsos negativos (3.123), refletindo o trade-off entre acurácia e sensibilidade.

Figura 7. (a) Matriz de Confusão sem Balanceamento; (b) Matriz de Confusão com Balanceamento.



As características (*features*) clínicas e socioeconômicas utilizadas no modelo foram essenciais para entender o risco de abandono ou sucesso do tratamento em pacientes com tuberculose. A inclusão de variáveis como comorbidades (ex.: HIV, diabetes) e fatores sociais (ex.: escolaridade, ocupação) proporciona uma visão abrangente dos fatores que influenciam os desfechos. O modelo, portanto, não apenas prediz, mas também auxilia profissionais da saúde na formulação de estratégias de intervenção, considerando o perfil completo dos pacientes, com o objetivo de reduzir o abandono e melhorar os resultados de pacientes com TB.

CONCLUSÃO

De modo geral, este estudo contribuiu para o desenvolvimento de um modelo preditivo para os desfechos no tratamento da tuberculose, utilizando Redes Neurais *Multilayer Perceptron* (MLP). Ao integrar dados clínicos e aplicando técnicas de balanceamento de classes, a pesquisa demonstrou ser uma abordagem eficaz na previsão de cura ou abandono do tratamento. A metodologia aplicada, que incluiu o pré-processamento dos dados, balanceamento com *SMOTETomek* e a regularização com *dropout*, possibilitou uma análise robusta dos desfechos, promovendo avanços

significativos na área da saúde pública e no gerenciamento de pacientes com tuberculose.

A análise revelou que o balanceamento de classes, apesar de melhorar a acurácia global do modelo (0.8577), levou a uma leve redução no *recall*, o que indica um *trade-off* entre precisão e sensibilidade. Os resultados mostram que o modelo sem balanceamento teve um *recall* elevado de 0.9779, porém com uma acurácia menor. Esses achados reforçam a importância de balancear o desempenho entre as métricas de avaliação para garantir uma identificação eficiente dos casos positivos. Para investigações futuras, pretende-se explorar a utilização de redes neurais mais complexas, como redes profundas ou modelos híbridos, além de aplicar técnicas de explicabilidade para uma melhor compreensão dos fatores clínicos que influenciam as previsões.

REFERÊNCIAS

1. Obeagu, E. and Obeagu, G. (2024). Understanding immune cell trafficking in tuberculosis-hiv coinfection: The role of I-selectin pathways. *Elite Journal of Immunology*, 2(2):43–59.
2. Health Organisation, W. (2023). Report 20-23, volume t/malaria/.
3. Bisht, G. S., Ganesh, D., and Kumar, S. (2024). Detection and prognosis of pulmonary tuberculosis using ml classifiers with soft computing. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, volume 5, pages 1097–1101. IEEE.
4. Motta, I., Boeree, M., Chesov, D., Dheda, K., Günther, G., Horsburgh Jr, C. R., Kherabi, Y., Lange, C., Lienhardt, C., McIlleron, H. M., et al. (2023). Recent advances in the treatment of tuberculosis. *Clinical Microbiology and Infection*.
5. Fan, S., Abulizi, A., You, Y., Huang, C., Yimit, Y., Li, Q., Zou, X., and Nijjati, M. (2024). Predicting hospitalization costs for pulmonary tuberculosis patients based on machine learning.
6. Simarmata, T. S., Isnanto, R. R., and Triwiyatno, A. (2023). Detection of pulmonary tuberculosis using neural network with feature extraction of gray level run-length matrix method on lung x-ray images. In *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 570–574. IEEE.

7. Orjuela-Canón, A. D., Jutinico, A. L., Awad, C., Vergara, E., and Palencia, A. (2022). Machine learning in the loop for tuberculosis diagnosis support. *Frontiers in Public Health*, 10:876949.
8. Liao, K.-M., Liu, C.-F., Chen, C.-J., Feng, J.-Y., Shu, C.-C., and Ma, Y.-S. (2023). Using an artificial intelligence approach to predict the adverse effects and prognosis of tuberculosis. *Diagnostics*, 13(6):1075.
9. Jannah, A. W. and Al Kindhi, B. (2024). Optimization of early detection of tuberculosis: Use of multilayer perceptron and extreme learning machine with clinical data. *Jurnal Indonesia Sosial Teknologi*, 5(5).
10. Mohidem, N. A., Osman, M., Muharam, F. M., Elias, S. M., Shaharudin, R., and Hashim, Z. (2021). Prediction of tuberculosis cases based on sociodemographic and environmental factors in gombak, selangor, malaysia: A comparative assessment of multiple linear regression and artificial neural network models. *The International Journal of Mycobacteriology*, 10(4):442–456.
11. Kanesamoorthy, K. and Dissanayake, M. B. (2021). Prediction of treatment failure of tuberculosis using support vector machine with genetic algorithm. *The International Journal of Mycobacteriology*, 10(3):279–284.
12. Rodrigues A., M. G., Sampaio, V., Lynn, T., and Endo, P. T. A brazilian classified data set for prognosis of tuberculosis, between january 2001 and april 2020.
13. Viboonsang, P. and Kosolsombat, S. (2024). Network intrusion detection system using machine learning and deep learning. In 2024 IEEE International Conference on Cybernetics and Innovations (ICCI), pages 1–6. IEEE.
14. Talukder, M. A., Sharmin, S., Uddin, M. A., Islam, M. M., and Aryal, S. (2024). Mlslwsn: machine learning-based intrusion detection using smotetomek in wsns. *International Journal of Information Security*, 23(3):2139–2158.
15. Kazempour Dizaji, M., Moniri, A., Roozbahani, R., Varahram, M., Tabarsi, P., Zare, A., Baghaei Shiva, P., Abedini, A., Marjani, M., Madani, M., et al. (2022). Application of artificial neural network model in studying the mechanism of disease relapse event in patients with tuberculosis. *Health Science Monitor*, 1(2):116–124.