

# Validação clínica de modelo de inteligência artificial para detecção de tuberculose em radiografias de tórax em casos microbiologicamente confirmados.

Maria C. B. da Silva<sup>1</sup>, Thiago F. O. de Camargo<sup>1</sup>, Joselisa P. Q. de Paiva<sup>1</sup>,  
Jaqueline D. C. Horvath<sup>1</sup>, Guilherme A. S. Ribeiro<sup>1</sup>

<sup>1</sup>Departamento de Imagem – Hospital Israelita Albert Einstein (HIAE)  
Av. Albert Einstein, 627/701, Morumbi, São Paulo-SP – Brazil

maria.cbsilva@einstein.br, thiago.ortiz@einstein.br,  
joselisa.paiva@einstein.br, jaqueline.horvath@einstein.br,  
guilherme.asr@einstein.br

**Abstract.** *This study presents an artificial intelligence (AI) algorithm based on a ResNet-50 architecture designed to identify chest X-rays with patterns suggestive of pulmonary tuberculosis (TB). The model was trained and internally validated on 13,023 images from public datasets and then clinically validated using a unique dataset of radiographs from patients with microbiologically confirmed TB, with assessments performed by a thoracic radiologist. On public data, the algorithm achieved an AUC of 0.89, sensitivity of 0.84, and specificity of 0.72, while clinical validation yielded an AUC and sensitivity of 1.00. These results underscore the critical importance of incorporating microbiological confirmation as the gold standard to ensure the accuracy and clinical reliability of AI models in TB detection.*

**Resumo.** *Este trabalho apresenta um algoritmo de inteligência artificial (IA) baseado na arquitetura ResNet-50, desenvolvido para identificar radiografias de tórax com padrões sugestivos de tuberculose (TB). O modelo foi treinado e validado internamente com 13.023 imagens de bases públicas e, posteriormente, validado clinicamente utilizando um conjunto exclusivo de radiografias de pacientes com TB confirmada microbiologicamente e revisadas por radiologista torácico. Nos dados públicos, o algoritmo obteve AUC de 0,89, sensibilidade de 0,84 e especificidade de 0,72, enquanto a validação clínica apresentou AUC e sensibilidade de 1,00. Esses resultados ressaltam a importância crucial de utilizar a confirmação microbiológica como padrão-ouro para garantir a precisão e a confiabilidade clínica dos modelos de IA na detecção da TB.*

## 1. Introdução

A tuberculose (TB) é um problema de saúde pública. A Organização Mundial de Saúde (OMS) estima que 10,6 milhões de pessoas foram diagnosticadas com TB no mundo em 2023, com cerca de 1,6 milhão de mortes atribuídas à doença [Organização Mundial da Saúde (OMS) 2024]. No Brasil, dados do Ministério da Saúde apontam cerca de 80 mil novos casos de TB. Esses números ressaltam a necessidade de diagnósticos precoces, melhoria na adesão ao tratamento e estratégias de saúde pública direcionadas para reduzir sua transmissão [Ministério da Saúde (MS) 2024]. A

OMS também tem estimulado a inovação tecnológica para otimizar o seu diagnóstico [Organização Pan-Americana da Saúde (OPAS) 2024] e diante deste cenário, ferramentas de IA aplicadas à análise de radiografias de tórax emergem como uma solução promissora.

A radiografia de tórax é um exame utilizado para a triagem e monitoramento de doenças respiratórias, como por exemplo TB. A interpretação dessas imagens, porém, pode ser desafiadora. A sobreposição de estruturas anatômicas, a inexperience do observador e as limitações técnicas podem contribuir para erros diagnósticos [Tack and Howarth 2019, Del Ciello et al. 2017].

Neste contexto, há interesse no uso de inteligência artificial (IA) em imagens médicas para a detecção automática de alterações pulmonares e novos algoritmos são desenvolvidos para suporte à decisão diagnóstica [Huy and Lin 2023, Malik et al. 2023]. Exemplos são os sistemas automatizados propostos para rastreamento de pacientes com TB pulmonar [Harris et al. 2019, Yang et al. 2023].

Entretanto é essencial validar clinicamente os algoritmos de IA antes de sua implementação, a fim de garantir segurança e eficácia dos sistemas [Park and Kressel 2018]. No contexto da TB, essas validações podem ser realizadas comparando os resultados dos algoritmos com o padrão-ouro diagnóstico, que consiste na confirmação microbiológica do *Mycobacterium tuberculosis*, agente causador da doença. Faltam estudos na literatura que desenvolvam modelos próprios e que garantam a validação em casos de TB microbiologicamente confirmados [Qin et al. 2021, Mungai et al. 2022]. A partir dessa lacuna pretende-se desenvolver e validar clinicamente um algoritmo de IA capaz de auxiliar na detecção de radiografias de tórax com padrões sugestivos de TB, utilizando casos confirmados da doença.

## 2. Materiais e Métodos

Esta seção apresenta os conceitos base, bem como o método proposto.

### 2.1. Desenvolvimento do Modelo de Inteligência Artificial

Foi desenvolvido um modelo de IA baseado em Redes Neurais Convolucionais (RNCs) para classificar radiografias de tórax em incidências frontais como “sugestivas de TB” ou “não sugestivas de TB”. O Modelo de Tuberculose Pulmonar (MTB) fornece o escore de probabilidade e o mapa de calor (ou *heatmap*) para cada uma das imagens analisadas.

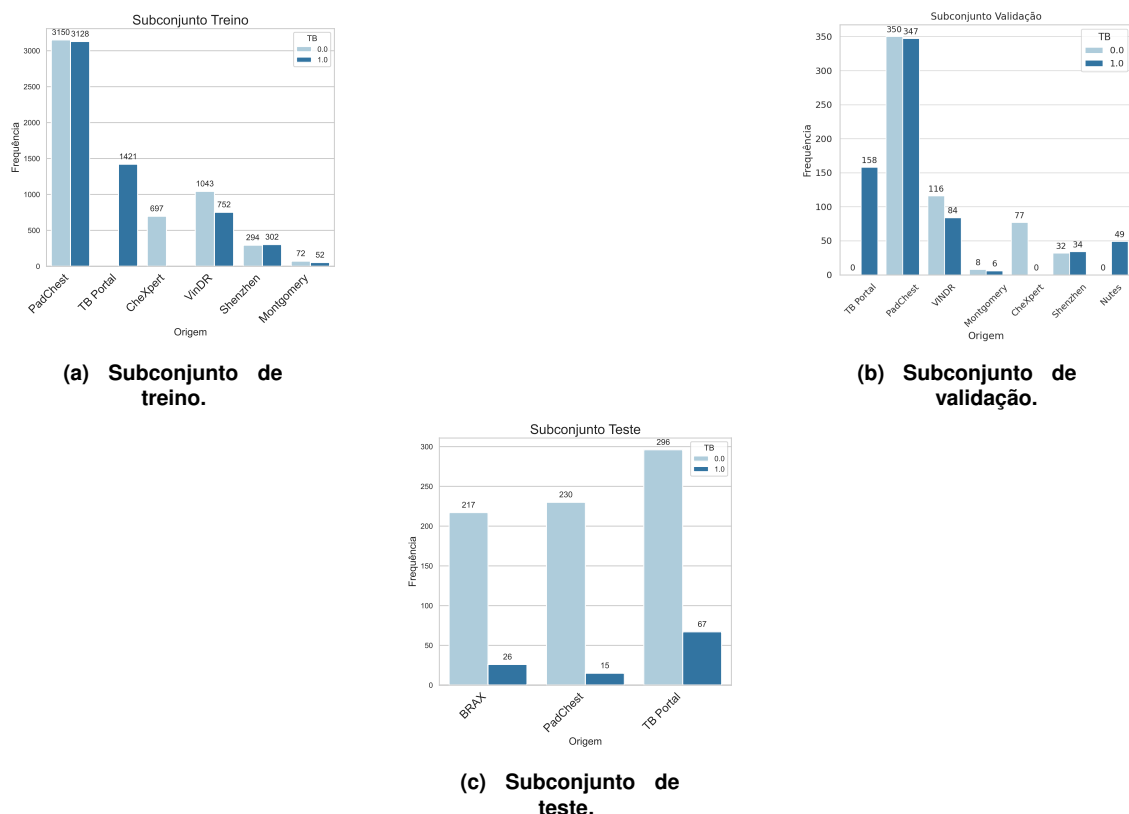
O MTB foi desenvolvido utilizando uma rede pré-treinada (ResNet-50) com pesos da biblioteca TorchXrayVision [Cohen et al. 2022], seguida de camadas densas para classificação. As imagens foram redimensionadas para 512x512 pixels, mantendo a proporção por meio de interpolação linear e preenchimento com zeros quando necessário.

Os hiperparâmetros utilizados no treinamento dos modelos foram ajustados empiricamente. Para a otimização, foi escolhido o algoritmo Adam, e a função de perda adotada foi a entropia cruzada binária. No que diz respeito ao treinamento, o MTB foi treinado durante 30 épocas, com taxa de aprendizado inicial de 0,001. O tamanho do lote (*batch size*) foi definido em 29 para o MTB. Além disso, foi aplicado um *scheduler* de uma taxa de aprendizado que reduziu a taxa de forma específica para cada modelo a cada determinado número de épocas, contribuindo para a otimização do desempenho e mitigação do *overfitting*.

Para mitigar o desbalanceamento de casos positivos e negativos de TB para treinamento do modelo, foi aplicado peso na função de perda. Os conjuntos de treino, validação interna e teste foram definidos aproximadamente em 85%, 10% e 5% dos dados, respectivamente, selecionados ao nível do paciente para evitar vazamento de informação.

## 2.2. Bases de Dados Utilizadas

Para desenvolvimento da versão inicial do modelo foram utilizados 13.023 exames provenientes das bases CheXpert [Irvin et al. 2019], TB Portals [Rosenthal et al. 2017], PadChest [Bustos et al. 2020], VinDR [Nguyen et al. 2022], Montgomery [Jaeger et al. 2014], Shenzhen [Jaeger et al. 2014], BRAX [Reis et al. 2022] e NUTES, sendo esta última uma base de dados utilizada somente na fase de validação do treinamento. Cada base foi processada para uniformizar resolução e aplicar técnicas de pré-processamento (equalização de histograma, desfoque gaussiano e normalização). Foram excluídas imagens de má qualidade, incidências não frontais e de indivíduos com idade inferior a 18 anos. Esses critérios foram aplicados de forma consistente durante as fases de treinamento (incluindo validação interna) e validação externa. A Figura 1 apresenta a divisão dos dados para o subconjunto de treino, validação e teste, separando por casos que possuem padrões sugestivos de TB (1) e casos normais para TB (0).



**Figura 1. Volumetria de imagens utilizadas na construção do MTB, divididas por rótulos positivos para compatibilidade com sugestiva de tuberculose (1) e negativos (0), bem como subconjunto e base de dados de origem.**

Para a validação externa, utilizaram-se radiografias de tórax frontais de indivíduos de ambos os sexos, acima de 18 anos, provenientes de um centro especializado no atendimento de pacientes com TB. As imagens foram compartilhadas de forma anonimizada,

juntamente com outros dados clínicos, como por exemplo co-infecção por HIV, presença de multirresistência à drogas, teste do escarro e teste molecular.

### 2.3. Estrutura do Estudo de Validação Externa

A validação externa foi conduzida de forma retrospectiva, com o objetivo de comparar o desempenho do algoritmo com o padrão-ouro, composto por radiografias de tórax de indivíduos com confirmação microbiológica de infecção pelo agente causador da TB (*Mycobacterium tuberculosis*). Um radiologista torácico, com 5 anos de experiência na área, analisou as radiografias de tórax para identificação dos padrões radiográficos sugestivos de TB pulmonar, como por exemplo cavidades em regiões apicais do pulmão.

A avaliação incluiu a análise dos índices de sensibilidade e área sob a curva ROC (AUC-ROC) do MTB, além da apresentação dos *heatmaps* gerados pelo algoritmo.

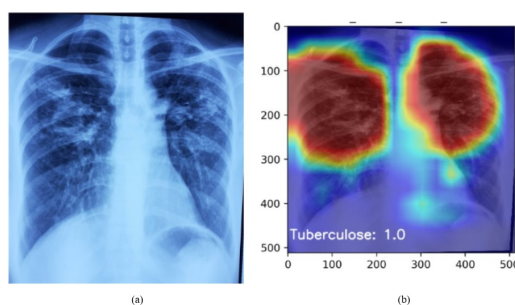
### 2.4. Aspectos Éticos

O estudo foi aprovado pelo Comitê de Ética da instituição. Para a fase retrospectiva da validação, solicitou-se isenção do Termo de Consentimento Livre e Esclarecido, uma vez que os pacientes não estavam mais em acompanhamento na instituição, não foram propostas intervenções e os dados eram totalmente anonimizados.

## 3. Resultados

Na fase de validação interna, o MTB atingiu AUC-ROC de 0,89, indicando alta capacidade de discriminar entre radiografia com e sem padrão sugestivo de TB. Os valores de sensibilidade e especificidade foram, 0,84 e 0,72, respectivamente.

Na fase de validação externa, utilizando a base com 186 radiografias com confirmação microbiológica como padrão-ouro, o modelo alcançou desempenho superior, com AUC-ROC de 1,00 e sensibilidade de 1,00 [IC 95%: 0,97–1,00], demonstrando alta precisão na detecção de padrões radiográficos sugestivos da doença. A especificidade não foi calculada pois esta base só contém valores positivos para TB. Além dos resultados em termos de métrica, os mapas de calor (*heatmaps*) são outra saída gerada pelo MTB.



**Figura 2. Representação de radiografia de tórax (a) do conjunto de validação externa com o respectivo mapa de calor (b)**

A Figura 2 apresenta o caso de uma radiografia de tórax do conjunto de validação externa de um indivíduo do sexo masculino, com 24 anos de idade, apresentando sintomas respiratórios (a). Observam-se opacidades/consolidações alveolares nos campos pulmonares superiores e provável cavidade no lobo superior direito. Paciente com cultura

de escarro positiva para *M. tuberculosis*. A radiografia foi submetida ao modelo de inteligência artificial durante a validação clínica. O modelo confirmou padrão radiográfico sugestivo de TB (escore de probabilidade elevado – neste caso igual a 1,0). O mapa de calor indicou as áreas de maior interesse na radiografia (b).

#### 4. Discussão

Os resultados demonstram que o algoritmo de IA apresentou desempenho na fase de validação externa superior ao que foi encontrado na validação interna. A maior concordância entre o modelo e os casos microbiologicamente confirmados de TB pode refletir maior abertura à integração de ferramentas digitais nas rotinas de cuidados em saúde.

A robustez da validação, realizada tanto em ambiente controlado (interna) quanto em cenários realistas (externa), reforça o potencial da ferramenta para aplicações clínicas [Kim et al. 2019, de Camargo et al. 2025], sobretudo em contextos com escassez de radiologistas e com alto índice de TB. Entretanto, desafios como a adaptação do algoritmo a variações demográficas permanecem como áreas para investigações futuras.

A validação clínica de algoritmos de IA em imagens médicas é crucial para garantir que o desempenho demonstrado em ambientes controlados se mantenha também de maneira confiável em condições reais de uso. Esse processo permite identificar possíveis vieses, limitações e variações decorrentes das diferenças entre os dados utilizados durante o treinamento e os encontrados na prática clínica, garantindo que a tecnologia auxilie o diagnóstico e evitando danos inadvertidos aos pacientes.

#### 5. Conclusão

Este estudo apresentou e validou externamente um algoritmo de IA desenvolvido para auxiliar os médicos a identificar radiografias de tórax com padrões radiográficos sugestivos de TB pulmonar. Os índices de desempenho apresentados evidenciam o potencial da ferramenta para atuar como suporte em ambientes clínicos, contribuindo para diagnósticos mais precisos em cenários com recursos limitados.

Um estudo multicêntrico em centros de todas as regiões do Brasil ampliará a volumetria para validação externa, permitindo integrar o algoritmo a fluxos clínicos, adaptá-lo a novas populações e condições radiológicas e ampliar seu impacto na prática médica.

#### Referências

- Bustos, A., Pertusa, A., Salinas, J.-M., et al. (2020). Padchest: A large chest x-ray image dataset with multi-label annotated reports. *Medical image analysis*, 66:101797.
- Cohen, J. P., Viviano, J. D., Bertin, P., et al. (2022). Torchxrayvision: A library of chest x-ray datasets and models. In *International Conference on Medical Imaging with Deep Learning*, pages 231–249. PMLR.
- de Camargo, T. F. O., Ribeiro, G. A. S., da Silva, M. C. B., et al. (2025). Clinical validation of an artificial intelligence algorithm for classifying tuberculosis and pulmonary findings in chest radiographs. *Frontiers in Artificial Intelligence*, 8:1512910.
- Del Ciello, A., Franchi, P., Contegiacomo, A., et al. (2017). Missed lung cancer: when, where, and why? *Diagnostic and interventional radiology*, 23(2):118.

- Harris, M., Qi, A., Jeagal, L., et al. (2019). A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PloS one*, 14(9):e0221339.
- Huy, V. T. Q. and Lin, C.-M. (2023). An improved densenet deep neural network model for tuberculosis detection using chest x-ray images. *IEEE Access*, 11:42839–42849.
- Irvin, J., Rajpurkar, P., Ko, M., et al. (2019). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597.
- Jaeger, S., Candemir, S., Antani, S., Wang, Y.-X. J., Lu, P.-X., and Thoma, G. (2014). Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4(6):475.
- Kim, D. W., Jang, H. Y., Kim, K. W., Shin, Y., and Park, S. H. (2019). Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean journal of radiology*, 20(3):405–410.
- Malik, H., Anees, T., Din, M., and Naeem, A. (2023). Cdc\_net: Multi-classification convolutional neural network model for detection of covid-19, pneumothorax, pneumonia, lung cancer, and tuberculosis using chest x-rays. *Multimedia Tools and Applications*, 82(9):13855–13880.
- Ministério da Saúde (MS) (2024). Brasil avança na prevenção, diagnóstico e tratamento da tuberculose. Acesso em: 24 fev. 2025.
- Mungai, B., Ong ‘angò, J., Ku, C. C., et al. (2022). Accuracy of computer-aided chest x-ray in community-based tuberculosis screening: Lessons from the 2016 kenya national tuberculosis prevalence survey. *PLOS global public health*, 2(11):e0001272.
- Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., et al. (2022). Vindr-cxr: An open dataset of chest x-rays with radiologist’s annotations. *Scientific Data*, 9(1):429.
- Organização Mundial da Saúde (OMS) (2024). *Global Tuberculosis Report 2024*. WHO, Geneva. Acesso em: 24 fev. 2025.
- Organização Pan-Americana da Saúde (OPAS) (2024). Opas pede que as américas adotem tecnologias e tratamentos inovadores para eliminar a tb. Acesso em: 24 fev. 2025.
- Park, S. H. and Kressel, H. Y. (2018). Connecting technological innovation in artificial intelligence to real-world medical practice through rigorous clinical validation: what peer-reviewed medical journals could do. *Journal of Korean Medical Science*, 33(22):e152.
- Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., et al. (2021). Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. *The Lancet Digital Health*, 3(9):e543–e554.
- Reis, E. P., de Paiva, J. P. Q., da Silva, M. C. B., Ribeiro, G. A. S., et al. (2022). Brax, brazilian labeled chest x-ray dataset. *Scientific Data*, 9(1):487.
- Rosenthal, A., Gabrielian, A., Engle, E., et al. (2017). The tb portals: an open-access, web-based platform for global drug-resistant-tuberculosis data sharing and analysis. *Am Soc Microbiol*, 55(11):3267–3282.

- Tack, D. and Howarth, N. (2019). Missed lung lesions: side-by-side comparison of chest radiography with mdct. *Diseases of the Chest, Breast, Heart and Vessels 2019-2022: Diagnostic and Interventional Imaging*, pages 17–26.
- Yang, Y., Xia, L., Liu, P., Yang, F., Wu, Y., Pan, H., Hou, D., Liu, N., and Lu, S. (2023). A prospective multicenter clinical research study validating the effectiveness and safety of a chest x-ray-based pulmonary tuberculosis screening software jf cxr-1 built on a convolutional neural network algorithm. *Frontiers in Medicine*, 10:1195451.