



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DA PARAÍBA - CAMPUS CAMPINA GRANDE
CURSO SUPERIOR BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

FELIPE BRAZ DA SILVA

**BENCHMARKING DE ARQUITETURAS LEVES DE DEEP
LEARNING PARA DETECÇÃO DE TUBERCULOSE EM RAIO-X:
ANÁLISE DE DESEMPENHO E CUSTO COMPUTACIONAL**

Campina Grande

Dezembro de 2025

Felipe Braz da Silva

Benchmarking de Arquiteturas Leves de Deep Learning para Detecção de Tuberculose em Raio-X: Análise de Desempenho e Custo Computacional

Artigo apresentado à Coordenação do Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Computação

Orientador: Prof.Dr.Igor Barbosa da Costa

Campina Grande
Dezembro de 2025

Catálogo na fonte:
Ficha catalográfica

S586b Silva, Felipe Braz da.

Benchmarking de arquiteturas leves de deep learning para detecção de tuberculose em raio-X: análise de desempenho e custo computacional / Felipe Braz da Silva. – Campina Grande, 2026.

18 f.: il.

Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Instituto Federal da Paraíba, 2026.

Orientador: Prof. Igor Barbosa da Costa.

1. Computação - Inteligência artificial - Deep Learning. 2. Medicina - detecção da tuberculose. 3. Redes Neurais Convolucionais. 4. Processamento de Imagem Médica. 5. Benchmarking. I. Costa, Igor Barbosa da. II. Título

CDU 004.8:004.5

Felipe Braz da Silva

Benchmarking de Arquiteturas Leves de Deep Learning para Detecção de Tuberculose em Raio-X: Análise de Desempenho e Custo Computacional

Artigo apresentado à Coordenação do Curso
de Bacharelado em Engenharia de Computa-
ção do Instituto Federal de Educação, Ciência
e Tecnologia da Paraíba, como requisito par-
cial para a obtenção do título de Bacharel em
Engenharia de Computação

Trabalho aprovado. Campina Grande, 12 de Dezembro de 2025:

Prof.Dr.Igor Barbosa da Costa
Orientador

Prof.Dr.Katyusco de Farias Santos
Membro da Banca

Prof.Me. Bruno de Brito Leite
Membro da Banca

Campina Grande
Dezembro de 2025

Dedico este trabalho à Virgem, Mãe de Deus, que sempre cuida de mim e conduz meus passos. Dedico também aos meus pais, pelo apoio incondicional, pelos ensinamentos e por acreditarem em mim em todos os momentos.

Agradecimentos

Quero expressar minha gratidão, em primeiro lugar, a Deus, a quem entreguei todos os meus planos e em quem confio a minha vida. À Nossa Senhora, Mãe de Deus, que sempre cuidou de mim e a quem recorro em todas as minhas necessidades.

Aos meus pais, Adriano e Fernanda, meu porto seguro durante toda a graduação. Muito obrigado, pelo amor e pela base sólida que vocês sempre me deram. Também agradeço a minha família, pelo apoio e por acreditarem em mim nestes cinco anos de estudos.

Aos meus amigos, especialmente Clarissa, Marcelo, Eduarda e Jeremias pelo companheirismo e por tornar o percurso mais leve e feliz. Agradeço também a Marilene, que me apoiou fortemente durante esse período.

Ao meu orientador, professor Igor, expresso minha sincera estima e gratidão. Sua orientação e conselhos foram decisivos para a realização deste trabalho, assim como também para minha formação profissional.

A todos os docentes que contribuíram para minha formação, pelos conhecimentos compartilhados e pelas discussões enriquecedoras.

Benchmarking de Arquiteturas Leves de Deep Learning para Detecção de Tuberculose em Raio-X: Análise de Desempenho e Custo Computacional

Felipe Braz da Silva¹

Instituto Federal da Paraíba (IFPB) – Campus Campina Grande
`felipe.braz@academico.ifpb.edu.br`

Igor Barbosa da Costa²

Instituto Federal da Paraíba (IFPB) – Campus Campina Grande
`igor.costa@ifpb.edu.br`

2025, v-1.0

Resumo

A tuberculose continua sendo um desafio de saúde pública global, e métodos de inteligência artificial aplicados à detecção automática em radiografias de tórax têm se mostrado promissores, possibilitando diagnósticos mais rápidos e precisos. Neste estudo, realiza-se um *benchmarking* de arquiteturas de redes neurais convolucionais para detecção de TB, com ênfase em modelos leves. Foram avaliadas as arquiteturas *MobileNet*, *NASNetMobile* e *ResNet50* sobre um conjunto público de radiografias de tórax, dividido em 70% para treino, 10% para validação e 20% para teste. Adotou-se *transfer learning* com pesos pré-treinados no *ImageNet* e adaptação da saída multiclasse para classificação binária. Dois regimes de treinamento foram considerados: (I) ajuste completo dos pesos de todas as camadas e (II) congelamento das camadas convolucionais, com treinamento apenas do topo da rede. Na abordagem (II), a *MobileNet* apresentou o melhor desempenho, com acurácia de 99,7% e área sob a curva ROC de 0,9998, mantendo baixo custo computacional. As demais arquiteturas também apresentaram resultados satisfatórios, com degradação de desempenho da *ResNet50* quando utilizada apenas como extratora de características. Os achados sugerem que modelos leves podem ser empregados com sucesso na detecção de tuberculose em cenários com recursos computacionais limitados, configurando uma alternativa viável para sistemas de apoio à decisão em contextos de baixo custo.

Palavras-chaves: Aprendizado profundo; redes convolucionais; tuberculose.

Abstract

Tuberculosis remains a major global public health challenge, and artificial intelligence methods applied to automatic detection in chest X-ray images have shown promise, enabling faster and potentially more accurate diagnoses. In this study, we conduct a benchmarking of convolutional neural network architectures for TB detection, with emphasis on lightweight models. The architectures *MobileNet*, *NASNetMobile*, and *ResNet50* were evaluated on a public chest X-ray dataset, split into 70% for training, 10% for validation, and 20% for testing. We adopted a transfer learning strategy with ImageNet-pretrained weights and adapted the original multiclass output to a binary classification setting. Two training regimes were considered: (I) full fine-tuning of all network weights and (II) freezing of convolutional layers, training only the top layers. Under regime (II), *MobileNet* achieved the best performance, with an accuracy of 99.7% and an area under the ROC curve (AUC-ROC) of 0.9998, while maintaining low computational cost. The other architectures also obtained competitive results, with a marked degradation in *ResNet50* performance when used solely as a feature extractor. Overall, the findings suggest that lightweight convolutional models can be successfully employed for tuberculosis detection in scenarios with limited computational resources, providing a viable alternative for low-cost decision support systems in healthcare.

Keywords: deep learning; convolutional neural networks; tuberculosis.

1 Introdução

A tuberculose (TB) é uma doença que afeta milhões de pessoas todos os anos. De acordo com o Relatório Global sobre TB 2024 da Organização Mundial da Saúde (OMS) ([Organização Pan-Americana da Saúde, 2025](#)), cerca de 10,8 milhões de pessoas foram acometidas pela doença em 2024, resultando em aproximadamente 1,25 milhão de mortes. O diagnóstico é feito por meio de exames laboratoriais, avaliação clínica e realização de radiografias de tórax (RX) como método complementar, por serem de baixo custo e alta disponibilidade. Entretanto, a interpretação radiológica é sujeita a variabilidade intra e interobservador, o que motiva o uso de métodos computacionais de apoio à decisão clínica ([YAYAN et al., 2024](#)).

Nos últimos anos, abordagens de aprendizado profundo, em especial as redes neurais convolucionais (CNNs), ganharam destaque na detecção automática de TB a partir de RX. O estudo de ([HANSUN et al., 2025](#)) confirma essa predominância ao mostrar que, entre 152 trabalhos analisados, 102 (67,1%) utilizaram *deep learning* para a detecção de tuberculose em radiografias de tórax.

Apesar desses avanços, alguns desafios metodológicos permanecem relevantes: (i) heterogeneidade de bases de dados e protocolos de pré-processamento, (ii) risco de *data leakage* e de sobreajuste em cenários com amostras limitadas, (iii) *domain shift* entre centros, e (iv) necessidade de soluções factíveis para ambientes com restrições de hardware, nos quais modelos leves podem ser decisivos.

Nesse contexto, considerando os estudos mais recentes, observa-se uma lacuna específica: faltam estudos de *benchmarking* que comparem, sob um *pipeline* único e reproduzível, arquiteturas leves amplamente disponíveis em *frameworks* consolidados, avaliando regimes de *transfer learning* com e sem ajuste fino das camadas convolucionais. A ausência

dessa padronização dificulta a comparação justa dos resultados reportados na literatura e a seleção informada de modelos para cenários de baixo custo.

Este trabalho tem como objetivo realizar um *benchmarking* experimental de arquiteturas leves baseadas em redes neurais convolucionais (CNNs) aplicadas à detecção de TB a partir de imagens de RX de tórax, utilizando um conjunto de dados padronizado e referenciado na literatura. Foram adotados dois regimes de treinamento: (I) ajuste completo de pesos (*fine-tuning*) e (II) congelamento das camadas convolucionais com treinamento apenas das camadas superiores (topo da rede). As questões que orientam o estudo são:

- a) Modelos leves mantêm desempenho preditivo competitivo na detecção de tuberculose em radiografias de tórax, sob o protocolo experimental adotado?
- b) O congelamento das camadas convolucionais, em uma abordagem de *transfer learning*, oferece melhor relação desempenho–custo computacional do que o ajuste completo de todos os pesos?
- c) Quais compromissos práticos emergem entre precisão preditiva e custo computacional ao comparar as diferentes arquiteturas e regimes de treinamento?

As principais contribuições desse estudo são: (i) um *benchmark* reproduzível de arquiteturas leves (*MobileNet*, *NASNetMobile* e *ResNet50*) sob dois regimes de *transfer learning*, com métricas padronizadas (acurácia, precisão, sensibilidade, *F1*-score e AUC-ROC); (ii) evidências quantitativas sobre a relação desempenho–custo computacional em cenários com recursos limitados, relevantes para aplicações de saúde pública; (iii) um protocolo claro de pré-processamento e particionamento dos dados, reduzindo vieses comuns (p. ex., *data leakage*) e facilitando a replicação e extensões futuras.

O escopo deste estudo restringe-se à análise comparativa de modelos existentes na literatura acadêmica, visando avaliar arquiteturas leves sob o protocolo experimental adotado. Assim, embora os resultados apontem potencial para apoiar processos de triagem em serviços de saúde, não se contempla, neste trabalho, o desenvolvimento de uma aplicação clínica final nem a análise de aspectos regulatórios e de responsabilização médica.

O artigo organiza-se como segue: a Seção 2 apresenta os conceitos teóricos fundamentais; a Seção 3 sintetiza pesquisas correlatas; a Seção 4 descreve o conjunto de dados, o pré-processamento, o delineamento experimental, os modelos e as métricas; a Seção 5 apresenta e discute os achados; por fim, a Seção 6 sumariza as implicações do estudo e aponta direções para trabalhos futuros.

2 Fundamentação Teórica

Esta seção apresenta os conceitos essenciais que embasam o estudo realizado, com foco em redes neurais convolucionais, arquiteturas leves, *transfer learning* e técnicas de interpretabilidade de modelos. O objetivo é fornecer o suporte teórico necessário para compreender as escolhas metodológicas adotadas nas etapas de experimentação e análise de resultados.

2.1 Redes neurais convolucionais e arquiteturas leves

Segundo (GOODFELLOW; BENGIO; COURVILLE, 2016), as redes neurais convolucionais (CNNs) podem ser definidas como um tipo especializado de rede neural projetado

para processar dados que possuem uma estrutura de grade, como imagens, nas quais os *pixels* formam uma grade bidimensional. As CNNs são particularmente eficazes em tarefas de processamento e classificação de imagens, pois são capazes de identificar padrões visuais complexos, como texturas, bordas, regiões e estruturas específicas. Por essa razão, também têm sido amplamente utilizadas na detecção de padrões em imagens médicas, apoiando o diagnóstico.

Nesse contexto, existem diversas arquiteturas de CNNs que variam quanto à profundidade da rede, ao número de neurônios e às camadas convolucionais. Entre elas, destacam-se as arquiteturas leves, com número reduzido de parâmetros e camadas, sendo consideradas neste trabalho aquelas com quantidade de parâmetros inferior a 10 milhões. Tais arquiteturas são desenvolvidas com o intuito de otimizar o custo computacional, o consumo de memória e a velocidade de inferência. Isso torna possível a execução de modelos em ambientes com recursos limitados, como dispositivos móveis e sistemas embarcados.

Diante do exposto, este trabalho adota como referência três arquiteturas: *MobileNet*, *NASNetMobile* e *ResNet50*. As duas primeiras são consideradas arquiteturas leves devido ao número reduzido de parâmetros, sendo adequadas para cenários de hardware restrito. Por outro lado, a *ResNet50* é considerada um modelo mais pesado, por possuir maior profundidade e quantidade de parâmetros, oferecendo alto desempenho em tarefas de classificação ao custo de maior demanda de processamento computacional. Os critérios adotados para essa classificação são apresentados na Seção 4.3.

2.2 *Transfer learning*, *fine-tuning* e camadas congeladas

A aprendizagem por transferência (*transfer learning*) pode ser entendida como uma técnica que consiste em transferir conhecimento de um domínio para outro domínio relacionado (WEISS; KHOSHGOFTAAAR; WANG, 2016). Em modelos de classificação de imagens, essa estratégia tornou-se amplamente empregada devido à disponibilidade de redes pré-treinadas em grandes bases de dados, permitindo que os modelos aprendam representações gerais que podem ser adaptadas a tarefas específicas.

No contexto médico, redes pré-treinadas podem ser utilizadas para tarefas de detecção mais específicas. O estudo de (MORID; BORJALI; Del Fiol, 2021) evidencia esse ponto ao demonstrar que o *transfer learning* usando o *ImageNet*, um conjunto de dados não médico, pode ser uma forma eficaz de abordar tarefas em imagens médicas. Assim, há base teórica e empírica para a adoção dessa técnica em problemas de classificação de imagens de diagnóstico.

Dado esse contexto, podem ser destacadas duas abordagens particularmente relevantes: (i) o *fine-tuning* (ajuste fino), no qual todos os pesos da rede são atualizados, permitindo adaptação mais completa ao novo domínio; e (ii) o congelamento das camadas convolucionais, em que apenas o topo da rede é treinado, utilizando as camadas iniciais como extratores fixos de características.

O *fine-tuning* oferece maior flexibilidade e capacidade de ajuste, porém está associado a maior risco de sobreajuste e a um custo computacional mais elevado. Por outro lado, o treinamento com congelamento de camadas é mais rápido e demanda menor custo computacional, sendo adequado quando se deseja transferir conhecimento previamente aprendido de um domínio para outro. Nesse caso, a parte congelada do modelo atua como extratora de características.

2.3 Modelos leves, custo computacional e aplicações médicas

As redes neurais convolucionais têm sido amplamente utilizadas na detecção de doenças a partir de exames de imagem, em especial na identificação de casos de TB em radiografias de tórax. Nesse cenário, modelos convolucionais leves ganham destaque por serem projetados para ambientes com restrições computacionais, ao mesmo tempo em que se espera que mantenham desempenho competitivo em relação a arquiteturas maiores.

No contexto de diagnóstico assistido por modelos de inteligência artificial (IA), diversos estudos têm apresentado resultados promissores. Em (NTAKOLIA et al., 2020), por exemplo, os autores propuseram um modelo de CNN leve para detectar metástases ósseas em imagens de medicina nuclear, reduzindo significativamente a quantidade de parâmetros em comparação com redes convolucionais tradicionais, sem sacrificar a eficácia. Esse tipo de abordagem é especialmente relevante em cenários de saúde com escassez de recursos computacionais.

Neste trabalho, comparam-se arquiteturas leves considerando não apenas métricas preditivas, mas também o tempo de treinamento e o número de épocas, em contraste com uma arquitetura mais profunda (*ResNet50*). Assim, justifica-se a busca por identificar a competitividade de modelos leves, como *MobileNet* e *NASNetMobile*, em contraste com modelos mais profundos, como a *ResNet50*, visando abranger cenários com restrições de poder computacional.

2.4 Interpretabilidade com Grad-CAM

A técnica Gradient-weighted Class Activation Mapping (Grad-CAM) (SELVARAJU et al., 2017) gera um mapa de ativação baseado nos gradientes das ativações da última camada convolucional, para inspecionar as regiões da imagem que mais influenciaram as decisões do modelo. Essa característica é particularmente importante em aplicações médicas, pois permite verificar se o modelo está focando em áreas anatomicamente relevantes para o diagnóstico.

No problema de detecção de tuberculose em radiografias de tórax, o Grad-CAM pode ser utilizado para avaliar se as redes convolucionais concentram sua atenção em regiões pulmonares compatíveis com alterações associadas à doença. Dessa forma, a técnica contribui para a interpretabilidade dos resultados e aumenta o potencial de aceitação clínica dos modelos.

3 Trabalhos relacionados

A seguir, apresentam-se estudos representativos que fundamentam esta pesquisa e ajudam a identificar lacunas na literatura científica recente.

Em (RAHMAN et al., 2020a), foi construído o conjunto de dados *Tuberculosis (TB) Chest X-ray* (RAHMAN; KHANDAKAR; CHOWDHURY, 2020), composto por 3.500 imagens de radiografias de tórax de pacientes infectados e 3.500 imagens de indivíduos saudáveis. As imagens foram coletadas de múltiplas fontes, como *National Library of Medicine (NLM)* dos EUA (conjuntos Montgomery e Shenzhen); *Belarus Set*, da Bielorrússia; *NIAID TB* e *RSNA CXR*. O estudo avaliou nove arquiteturas de CNNs — *ResNet18*, *ResNet50*, *ResNet101*, *CheXNet*, *InceptionV3*, *VGG19*, *DenseNet201*, *SqueezeNet* e *MobileNet* — em cenários com e sem segmentação. O trabalho se destaca pela abrangência comparativa e pelo uso de um conjunto amplamente reutilizado na literatura.

No trabalho de (VISUÑA et al., 2022), foi proposta a detecção de múltiplas doenças pulmonares utilizando um conjunto de dados composto por 119.544 imagens, das quais 3.616 são de COVID-19, 1.345 de pneumonia viral, 3.493 de tuberculose e 3.500 de indivíduos saudáveis. Foram aplicadas técnicas de normalização, redimensionamento e *data augmentation*. A pesquisa propôs uma nova arquitetura baseada em um conjunto de CNNs formado pelos modelos *VGG16*, *VGG19*, *InceptionV3*, *ResNet101V2*, *DenseNet121* e *CheXNet* pré-treinados no *ImageNet*. Foram testados os métodos de combinação: *Voting Ensemble* e *Stacking Ensemble*, usando métricas como acurácia, precisão, sensibilidade (*recall*), *F1-score*, além das curvas ROC e PR. Um diferencial do trabalho foi a aplicação da técnica de explicabilidade *Grad-CAM*, permitindo visualizar as regiões de maior atenção das redes.

A pesquisa de (NAFISAH; MUHAMMAD, 2024) utilizou as bases de Montgomery, Shenzhen e Belarus, disponibilizadas pela *National Library of Medicine for TB*, totalizando 998 imagens, sendo 692 com tuberculose e 406 sem tuberculose. As técnicas empregadas foram *data augmentation*, normalização e redução de dimensionalidade, seguidas de segmentação semântica via *U-Net*. Integraram os experimentos as arquiteturas *Xception*, *Inception-ResNet-V2*, *ResNeXt-50*, *MobileNet* e *EfficientNet* com validação empregando as métricas acurácia, sensibilidade (*recall*), precisão, *F1-score*, valor Kappa e curva ROC. O estudo também aplicou técnicas de explicabilidade (*Grad-CAM*) e visualização (*t-SNE*), ampliando a compreensão do comportamento interno dos modelos.

Em (HUY; LIN, 2023), os autores compararam o modelo proposto *CBAMWDNet* a arquiteturas clássicas como *InceptionV3*, *AlexNet*, *SqueezeNet*, *ResNet*, *DenseNet* e *WideResNet*. O conjunto de dados utilizado foi composto por 5.000 imagens (3.906 normais e 1.094 com TB), com divisão aleatória em 80% para treinamento e 20% para validação. Para avaliação incluíram as métricas acurácia, sensibilidade (*recall*), precisão, especificidade, valor preditivo negativo e *F1-score*. Esse estudo destaca-se pela análise comparativa entre modelos de diferentes portes, reforçando a relevância da discussão sobre eficiência computacional.

De forma geral, a literatura evidencia diversos avanços na aplicação de CNNs à detecção de TB, com resultados expressivos em termos de acurácia e robustez dos modelos. Contudo, observa-se que grande parte dos estudos enfatiza o desenvolvimento de novas arquiteturas ou estratégias de combinação, deixando uma lacuna na comparação padronizada de modelos leves sob condições controladas de treinamento e validação. Diante disso, este trabalho diferencia-se por concentrar-se em um *benchmarking* reproduzível e com metodologia uniforme, comparando arquiteturas leves (*MobileNet*, *NASNetMobile* e *ResNet50*) em dois regimes de *transfer learning*, o que permite avaliar a viabilidade dessas abordagens e o equilíbrio entre desempenho preditivo e custo computacional.

4 Materiais e Métodos

Esta seção descreve o delineamento metodológico adotado na pesquisa, cujo objetivo é realizar um estudo de *benchmarking* entre arquiteturas de *deep learning* aplicadas à detecção de tuberculose em radiografias de tórax. A investigação tem caráter experimental, pois envolve a execução, em ambiente controlado, de diferentes arquiteturas de CNNs em um conjunto de dados específico, e de natureza quantitativa, dado que as análises se baseiam em métricas numéricas de desempenho (Seção 4.4). O propósito é viabilizar uma comparação reproduzível e fornecer subsídios metodológicos para estudos futuros na área

de Inteligência Artificial aplicada à saúde.

Os experimentos foram conduzidos em um conjunto de radiografias de tórax amplamente utilizado na literatura, descrito na Seção 4.1. Em relação aos hiperparâmetros, adotou-se como referência o estudo de (RAHMAN et al., 2020a), uma vez que o ajuste fino (*tuning*) desses parâmetros não constitui o escopo desta pesquisa. Foram realizadas adaptações pontuais motivadas por restrições computacionais: redução do *batch size* de 32 para 16, que corresponde ao número de imagens processadas em uma única iteração durante o treinamento, bem como o uso de precisão mista (16 bits sempre que necessário). Os hiperparâmetros adotados foram: limite máximo de 50 épocas de treinamento; *batch size* igual a 16; *early stopping* com paciência de cinco épocas, monitorando a função de perda no conjunto de validação (*val_loss*); otimizador *Stochastic Gradient Descent* (SGD), com taxa de aprendizado de 0,001 e fator de *momentum* igual a 0,9.

O treinamento foi realizado em ambiente local *Jupyter Notebook*, integrado à plataforma Anaconda, com uso de uma GPU Nvidia RTX 2050 (8 GB de VRAM). Em todas as execuções, empregou-se a técnica de *transfer learning*, com modelos pré-treinados no conjunto *ImageNet* e adaptação da camada de saída para classificação binária. A avaliação do desempenho foi realizada com base nas métricas: acurácia, precisão, sensibilidade, *F1-score* e área sob a curva ROC (AUC-ROC), amplamente adotadas para avaliação de resultados na literatura de *deep learning*.

4.1 Conjunto de dados

O conjunto de dados *Tuberculosis (TB) Chest X-ray Database*, proposto por (RAHMAN et al., 2020a), foi selecionado por ser amplamente utilizado em estudos de detecção de TB, o que favorece a comparação de resultados entre pesquisas. De acordo com os autores, a versão completa contém 7.000 imagens, sendo 3.500 por classe. Neste trabalho, foi utilizada a versão pública (RAHMAN et al., 2020b) obtida via plataforma Kaggle, cujo *dataset* contém 4.200 imagens de radiografias de tórax, distribuídas em duas classes: 3.500 imagens de indivíduos saudáveis e 700 imagens de pacientes com diagnóstico confirmado.



Figura 1 – Exemplos de radiografias do conjunto de dados utilizado.

4.2 Pré-processamento

Nessa etapa, as imagens foram normalizadas pela divisão dos valores de *pixel* por 255, limitando-os ao intervalo $[0,1]$, o que facilita a convergência dos modelos durante o treinamento. Em seguida, foram redimensionadas de 512×512 para 224×224 *pixels*, tamanho padrão de entrada adotado pelas arquiteturas *ResNet50*, *MobileNet* e *NASNetMobile*. Dessa forma, esse procedimento assegura a compatibilidade com os pesos das redes pré-treinadas,

permitindo o aproveitamento da técnica *transfer learning*. O conjunto de dados foi dividido de acordo com o apresentado na Tabela 1.

Tabela 1 – Divisão do conjunto de dados utilizada nos experimentos

Partição	Percentual (%)	Quantidade de imagens
Treinamento	70%	2.940
Validação	10%	420
Teste	20%	840
Total	100%	4.200

4.3 Experimentos

A seleção das arquiteturas avaliadas — *MobileNet*, *NASNetMobile* e *ResNet50* — baseou-se em quatro critérios: (i) implementação oficial no TensorFlow/Keras; (ii) recorrência em estudos recentes da literatura; (iii) inclusão de arquiteturas leves, com número reduzido de parâmetros, adequadas para dispositivos com restrições de hardware; e (iv) presença de uma arquitetura mais profunda (*ResNet50*) para fins de comparação, todas compatíveis com entradas de dimensão (224, 224, 3). A mesma divisão em treino, validação e teste foi utilizada de forma fixa para todas as arquiteturas e abordagens, garantindo comparabilidade direta entre os modelos.

Os experimentos realizados consideraram a distribuição original das classes (3.500 normais e 700 com TB), sem técnicas adicionais de balanceamento. Por esse motivo, além da acurácia, foram utilizadas métricas como precisão, sensibilidade, *F1*-score e AUC-ROC, mais informativas em cenários de desbalanceamento.

Foram conduzidas duas abordagens experimentais baseadas em *transfer learning*. Na **Abordagem I**, todas as camadas convolucionais foram descongeladas, permitindo o ajuste completo dos pesos durante o treinamento. Na **Abordagem II**, manteve-se o congelamento das camadas convolucionais e treinou-se apenas as camadas adicionais no topo da rede.

Na Abordagem I, a camada de saída multiclasse das redes pré-treinadas foi substituída por uma camada *Dense* com um único neurônio e ativação sigmoide, adequada para classificação binária. Antes dessa camada, inseriu-se uma *Flatten* responsável por transformar o mapa de ativações em vetor unidimensional. Assim, a saída passa a representar a probabilidade de presença de TB na radiografia. A formulação matemática da função *softmax* original (Equação 1) e da função sigmoide (Equação 2) utilizada na saída binária é apresentada a seguir:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \quad i = 1, \dots, K \quad (1)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

Na Abordagem II, os modelos pré-treinados mantiveram suas camadas convolucionais congeladas, com o objetivo de atuar como extratores de características. Foram adicionadas, ao topo, uma camada *Flatten*, uma camada *Dense* com 128 neurônios e ativação *ReLU*, e uma camada final sigmoide. Dessa forma, apenas os parâmetros das

camadas superiores foram ajustados junto ao conjunto de radiografias, preservando os pesos originais do *ImageNet* e reduzindo o tempo de treinamento.

Em ambas as abordagens, foi empregada a técnica de *early stopping*, com monitoramento da perda da validação (*val_loss*). O treinamento foi interrompido nas situações em que não houvesse melhoria por 5 épocas consecutivas, restaurando automaticamente os melhores pesos obtidos. Essa estratégia permitiu otimizar o uso de recursos computacionais durante a fase de treinamento e minimizar o risco de sobreajuste (*overfitting*) nos modelos. Os resultados obtidos a partir dessas duas configurações são apresentados e discutidos na Seção 5, considerando tanto o desempenho preditivo quanto o custo computacional de cada modelo.

4.4 Métricas de avaliação

O desempenho dos modelos foi avaliado por meio das métricas acurácia, precisão, sensibilidade (*recall*), *F1*-score e área sob a curva ROC (AUC-ROC), amplamente utilizadas em tarefas de classificação binária em cenários de desbalanceamento de classes. Enquanto a acurácia resume a proporção global de acertos, precisão e sensibilidade capturam o equilíbrio entre falsos positivos e falsos negativos, e o *F1*-score combina ambas em uma métrica única. A AUC-ROC quantifica a capacidade do modelo de separar as classes em diferentes limiares de decisão.

5 Resultados e Discussão

Nesta seção, são apresentados e discutidos os resultados obtidos a partir das duas abordagens experimentais propostas: (I) ajuste completo dos pesos (*fine-tuning*) e (II) congelamento das camadas convolucionais. A análise considera, como forma de validação dos modelos, métricas preditivas amplamente utilizadas na literatura (acurácia, precisão, *recall*, *F1*-score e AUC-ROC), bem como indicadores de custo computacional (tempo de treinamento e número de épocas), permitindo uma avaliação integrada entre desempenho e eficiência.

5.1 Abordagem I: ajuste completo dos pesos

Na primeira abordagem, todas as camadas das arquiteturas foram descongeladas após o carregamento dos pesos pré-treinados no *ImageNet* e ajustadas durante o treinamento, permitindo que os modelos aprendessem representações específicas do conjunto de radiografias analisado. Os resultados encontram-se na Tabela 2.

Tabela 2 – Desempenho e custo computacional – Abordagem I (pesos descongelados)

Modelo	Acur.	Prec.	Recall	F1	AUC	Tempo	Ép.
MobileNet	0.9917	0.9650	0.9857	0.9753	0.9984	4,1 min	7
NASNetMobile	0.9845	0.9568	0.9500	0.9534	0.9943	15,6 min	14
ResNet50	0.9976	0.9929	0.9929	0.9929	0.9964	16,8 min	25

Fonte: Autoria própria.

De forma geral, observa-se que os três modelos apresentaram desempenho elevado, confirmando a capacidade das arquiteturas avaliadas na identificação de padrões associados

à tuberculose em radiografias de tórax. Além disso, verifica-se um importante *trade-off* entre desempenho e custo computacional. A *ResNet50* obteve o melhor desempenho na maioria das métricas avaliadas, porém demandou aproximadamente quatro vezes mais tempo de treinamento do que a *MobileNet*. Esta, por sua vez, alcançou excelente desempenho (AUC-ROC $\approx 0,9984$) com custo de treinamento significativamente inferior, resultado relevante para aplicações em dispositivos com limitações de processamento.

5.2 Abordagem II: camadas convolucionais congeladas

Na segunda abordagem, as camadas convolucionais dos modelos pré-treinados permaneceram congeladas, atuando como extratoras de características, enquanto apenas as camadas superiores adicionadas foram ajustadas durante o treinamento. Os resultados estão apresentados na Tabela 3.

Tabela 3 – Desempenho e custo computacional – Abordagem II (pesos congelados)

Modelo	Acur.	Prec.	Recall	F1	AUC	Tempo	Ép.
MobileNet	0.9976	0.9929	0.9929	0.9929	0.9998	4,2 min	8
NASNetMobile	0.9917	0.9716	0.9786	0.9751	0.9983	5,2 min	9
ResNet50	0.8333	0.0000	0.0000	0.0000	-	8,3 min	16

Fonte: Autoria própria.

Os resultados obtidos demonstram que os modelos *MobileNet* e *NASNetMobile* mantiveram desempenho elevado mesmo sem ajuste completo dos pesos, indicando que as representações pré-treinadas no *ImageNet* foram suficientemente expressivas para a classificação pretendida. Ademais, a *MobileNet* apresentou desempenho superior ao observado na abordagem com pesos descongelados, sugerindo que o congelamento pode funcionar como forma de regularização para modelos menores, reduzindo o risco de sobreajuste sem prejuízo de desempenho.

No entanto, observou-se que a *ResNet50* apresentou queda severa de desempenho (precisão e *recall* nulos). Embora a acurácia aparente (0,8333) seja relativamente alta, esse valor decorre do desbalanceamento entre as classes no conjunto de dados (3.500 normais e 700 com TB) e indica que o modelo passou a prever majoritariamente a classe normal. Tal condição mostra que as representações aprendidas no *ImageNet* não foram suficientes quando aplicadas ao domínio específico das radiografias sem ajuste dos pesos — fenômeno coerente com sua maior profundidade e complexidade arquitetural, que exigem adaptação mais intensa ao conjunto de dados-alvo. Esse comportamento evidencia, ainda, a importância de analisar métricas além da acurácia em cenários de desbalanceamento.

De forma comparativa entre as duas abordagens, observa-se que a *MobileNet* melhora ligeiramente seu desempenho quando as camadas convolucionais permanecem congeladas, mantendo praticamente o mesmo custo computacional. A *NASNetMobile* apresenta desempenho elevado em ambos os regimes, com leve vantagem para a abordagem com camadas congeladas. Já a *ResNet50* mostra forte dependência do ajuste completo dos pesos: quando totalmente ajustada, apresenta as melhores métricas globais, mas, ao ser utilizada apenas como extratora de características, sofre queda acentuada de desempenho. Esses resultados reforçam que o *transfer learning* com camadas congeladas é mais adequado para arquiteturas leves, enquanto modelos mais profundos demandam *fine-tuning* extensivo.

5.3 Interpretação do processo decisório via Grad-CAM

A técnica Gradient-weighted Class Activation Mapping (Grad-CAM) (SELVARAJU et al., 2017) foi utilizada para gerar um mapa de ativação, baseado nos gradientes das ativações da última camada convolucional, a fim de inspecionar as regiões das imagens que mais influenciaram as decisões dos modelos. Na Figura 2, observa-se que as áreas destacadas em cores mais quentes (como vermelho e amarelo) correspondem às regiões de maior relevância para a predição, enquanto cores mais frias (como azul) indicam menor contribuição.

As visualizações apresentadas na Figura 2 foram geradas a partir do modelo *MobileNet*, que se mostrou a configuração mais eficiente em termos de desempenho e custo computacional.

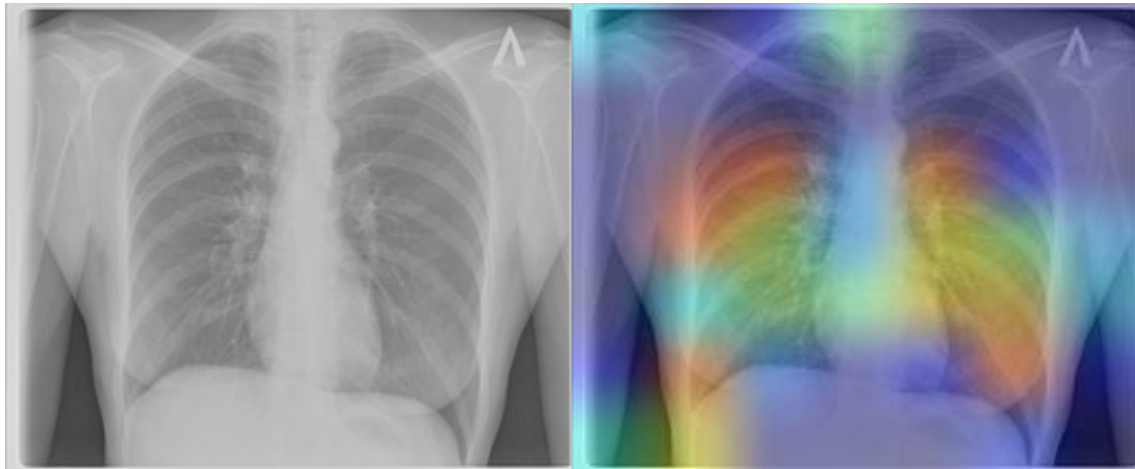


Figura 2 – Mapa de ativação gerado pelo Grad-CAM para uma radiografia positiva para TB.

Os resultados evidenciam que o modelo concentra sua atenção principalmente em áreas pulmonares estruturalmente compatíveis com alterações associadas à tuberculose, o que aumenta a confiabilidade clínica das predições e reforça o potencial uso das redes como ferramentas de apoio à decisão.

5.4 Síntese dos resultados em relação às perguntas de pesquisa

Nesta seção, apresenta-se uma síntese dos principais resultados obtidos, organizada de acordo com as perguntas de pesquisa que orientaram o desenvolvimento deste trabalho.

a) Modelos leves mantêm desempenho preditivo competitivo na detecção de tuberculose em radiografias de tórax, sob o protocolo experimental adotado?

Os resultados demonstram que sim. As arquiteturas *MobileNet* e *NASNetMobile* apresentaram métricas elevadas em ambas as abordagens, com AUC-ROC acima de 0,99. Assim, confirma-se que modelos leves são adequados para a detecção de tuberculose em cenários com recursos computacionais limitados.

b) O congelamento das camadas convolucionais, em uma abordagem de *transfer learning*, oferece melhor relação desempenho–custo computacional do

que o ajuste completo de todos os pesos?

Para modelos leves, a resposta é afirmativa. A *MobileNet* obteve seu melhor desempenho quando as camadas convolucionais permaneceram congeladas, reduzindo o tempo de treinamento e evitando sobreajuste. Entretanto, para modelos mais profundos como a *ResNet50*, o congelamento prejudicou os resultados, indicando que essas arquiteturas dependem de ajuste mais extensivo dos pesos.

c) Quais compromissos práticos emergem entre precisão preditiva e custo computacional ao comparar as diferentes arquiteturas e regimes de treinamento?

A *ResNet50* só alcança alto desempenho quando totalmente ajustada, ao custo de maior tempo de treinamento e maior número de épocas. Já a *MobileNet* oferece o melhor equilíbrio entre precisão e custo computacional, sendo a arquitetura mais indicada para aplicações em dispositivos embarcados ou sistemas com restrições de processamento.

6 Conclusão

Neste trabalho, foi conduzido um estudo de *benchmarking* de arquiteturas de aprendizado profundo aplicadas à detecção de tuberculose em radiografias de tórax. Duas abordagens de *transfer learning* foram investigadas: (I) ajuste completo dos pesos das redes pré-treinadas (*fine-tuning*) e (II) congelamento das camadas convolucionais com treinamento apenas das camadas superiores. A comparação sistemática desses regimes permitiu avaliar, de forma conjunta, desempenho preditivo e custo computacional.

Em todos os experimentos, empregou-se aprendizado por transferência a partir de modelos pré-treinados no conjunto *ImageNet*, com adaptação da saída para classificação binária. Essa estratégia viabilizou o uso de arquiteturas consolidadas em visão computacional em um domínio médico específico, reduzindo o esforço de treinamento. Adicionalmente, a técnica Grad-CAM foi aplicada para inspecionar as regiões de atenção dos modelos durante o processo de classificação, contribuindo para a interpretabilidade dos resultados e aumentando o potencial de uso como ferramenta de apoio à decisão clínica.

Os resultados obtidos demonstraram que modelos leves, em especial a *MobileNet*, apresentaram desempenho altamente competitivo em ambas as configurações, respondendo positivamente à primeira pergunta de pesquisa (Q1). Na abordagem com camadas convolucionais congeladas (Abordagem II), os modelos leves mantiveram AUC-ROC acima de 0,99 com tempo de treinamento reduzido, evidenciando uma relação favorável entre custo computacional e precisão, o que responde à segunda pergunta (Q2). Por outro lado, observou-se que modelos mais profundos, como a *ResNet50*, dependem de ajuste completo dos pesos para alcançar alto desempenho, mostrando forte sensibilidade ao regime de treinamento. Esse comportamento é particularmente relevante em cenários com restrições de hardware, contribuindo para a resposta à terceira pergunta (Q3) sobre os compromissos práticos entre desempenho e custo computacional.

Um dos diferenciais deste estudo foi a ênfase em arquiteturas leves, que apresentaram menor custo computacional e tempo de treinamento significativamente reduzido, sem perda substancial de desempenho. Essas características tornam tais modelos particularmente adequados para sistemas de apoio ao diagnóstico em locais com infraestrutura computacional limitada, alinhando-se às demandas de saúde pública em contextos de baixo custo e sugerindo a viabilidade de soluções embarcadas para triagem automatizada.

Como limitação, destaca-se que os experimentos foram conduzidos com uma única divisão treino–validação–teste e uma execução por configuração de modelo, sem cálculo de média e desvio padrão entre múltiplas inicializações aleatórias. Essa decisão decorreu de restrições computacionais e do foco em uma comparação controlada entre arquiteturas sob um protocolo fixo de *transfer learning*. Ainda assim, o uso da mesma partição de dados e dos mesmos hiperparâmetros de base para todos os modelos garante uma comparação justa, e as diferenças mais acentuadas observadas — como o colapso da *ResNet50* com camadas congeladas — dificilmente se explicam apenas por variação aleatória.

Como trabalhos futuros, pretende-se ampliar o conjunto de dados de treinamento, incorporando novas bases e técnicas de *data augmentation* para aumentar a representatividade dos exemplos e a robustez dos modelos. Além disso, considera-se a investigação detalhada das arquiteturas leves avaliadas, a fim de identificar quais componentes contribuem de forma mais decisiva para a redução do tempo de treinamento e a manutenção do desempenho. Também se propõe explorar métodos adicionais de explicabilidade e realizar estudos de implementação e validação em dispositivos embarcados ou plataformas de baixa potência, aproximando os resultados aqui obtidos de cenários reais de aplicação clínica.

Referências

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Acesso em: 1 out. 2025. Disponível em: <<http://www.deeplearningbook.org>>. Citado na página 3.

HANSUN, S. et al. Diagnostic performance of artificial intelligence–based methods for tuberculosis detection: Systematic review. *J Med Internet Res*, v. 27, p. e69068, mar. 2025. ISSN 1438-8871. Acesso em: 12 abr. 2025. Disponível em: <<https://www.jmir.org/2025/1/e69068>>. Citado na página 2.

HUY, V. T. Q.; LIN, C.-M. An improved densenet deep neural network model for tuberculosis detection using chest x-ray images. *IEEE Access*, v. 11, p. 42839–42849, 2023. Acesso em: 12 abr. 2025. Disponível em: <<https://ieeexplore.ieee.org/>>. Citado na página 6.

MORID, M. A.; BORJALI, A.; Del Fiol, G. A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in Biology and Medicine*, v. 128, p. 104115, 2021. ISSN 0010-4825. Acesso em: 1 out. 2025. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0010482520304467>>. Citado na página 4.

NAFISAH, S. I.; MUHAMMAD, G. Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. *Neural Computing and Applications*, v. 36, n. 1, p. 111–131, 2024. Acesso em: 12 abr. 2025. Disponível em: <<https://doi.org/10.1007/s00521-022-07258-6>>. Citado na página 6.

NTAKOLIA, C. et al. A lightweight convolutional neural network architecture applied for bone metastasis classification in nuclear medicine: A case study on prostate cancer patients. *Healthcare*, v. 8, n. 4, 2020. ISSN 2227-9032. Acesso em: 1 out. 2025. Disponível em: <<https://www.mdpi.com/2227-9032/8/4/493>>. Citado na página 5.

Organização Pan-Americana da Saúde. *Dia Mundial da Tuberculose 2025: Sim! Podemos acabar com a TB! Compromisso, Investimento, Resultados*. 2025. Acesso em: 19 mai. 2025. Disponível em: <<https://www.paho.org/pt/eventos>>. Citado na página 2.

RAHMAN, T.; KHANDAKAR, A.; CHOWDHURY, M. E. H. *Tuberculosis (TB) Chest X-ray Database*. IEEE Dataport, 2020. Acesso em: 12 abr. 2025. Disponível em: <<https://dx.doi.org/10.21227/mps8-kb56>>. Citado na página 5.

RAHMAN, T. et al. Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization. *IEEE Access*, v. 8, p. 191586–191601, 2020. Acesso em: 12 abr. 2025. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9224622>>. Citado 2 vezes nas páginas 5 e 7.

RAHMAN, T. et al. *Tuberculosis (TB) Chest X-ray Dataset*. 2020. <<https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset>>. Disponível em: Kaggle. Acesso em: 12 abr. 2025. Citado na página 7.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [s.n.], 2017. p. 618–626. Acesso em: 1 out. 2025. Disponível em: <https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html>. Citado 2 vezes nas páginas 5 e 11.

VISUÑA, L. et al. Computer-aided diagnostic for classifying chest x-ray images using deep ensemble learning. *BMC Medical Imaging*, v. 22, n. 1, p. 178, 2022. Acesso em: 12 abr. 2025. Disponível em: <<https://doi.org/10.1186/s12880-022-00904-4>>. Citado na página 6.

WEISS, K.; KHOSHGOFTAAR, T. M.; WANG, D. A survey of transfer learning. *Journal of Big data*, Springer, v. 3, n. 1, p. 9, 2016. Acesso em: 1 out. 2025. Disponível em: <<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6>>. Citado na página 4.

YAYAN, J. et al. Early detection of tuberculosis: a systematic review. *Pneumonia*, v. 16, n. 1, p. 11, 2024. ISSN 2200-6133. Acesso em: 19 mai. 2025. Disponível em: <<https://doi.org/10.1186/s41479-024-00133-z>>. Citado na página 2.

	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
	Campus Campina Grande - Código INEP: 25137409
	R. Tranquilino Coelho Lemos, 671, Dinamérica, CEP 58432-300, Campina Grande (PB)
	CNPJ: 10.783.898/0003-37 - Telefone: (83) 2102.6200

Documento Digitalizado Ostensivo (Público)

TCC corrigido com a data correta

Assunto:	TCC corrigido com a data correta
Assinado por:	Henrique Nascimento
Tipo do Documento:	Tese
Situação:	Finalizado
Nível de Acesso:	Ostensivo (Público)
Tipo do Conferência:	Documento Original

Documento assinado eletronicamente por:

- Henrique do Nascimento Cunha, COORDENADOR(A) DE CURSO - FUC1 - CCEC-CG, em 25/02/2026 17:02:22.

Este documento foi armazenado no SUAP em 25/02/2026. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

Código Verificador: 1779267
Código de Autenticação: 5e603cfef1

