UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
CIÊNCIA DA COMPUTAÇÃO

Leonardo Luiz Gambalonga Alves de Oliveira

# Desenvolvimento de um modelo híbrido de Deep Learning para a detecção de doenças pulmonares em imagens de raio-X

Florianópolis
12 de julho de 2026

Leonardo Luiz Gambalonga Alves de Oliveira

# Desenvolvimento de um modelo híbrido de Deep Learning para a detecção de doenças pulmonares em imagens de raio-X

Trabalho de Conclusão de Curso submetido ao Curso de Graduação em Ciência da Computação do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito para obtenção do título de Bacharel em Ciência da Computação.
Orientador: Prof. Alexandre Gonçalves Silva, Dr.

Florianópolis
12 de julho de 2026

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Oliveira, Leonardo Luiz Gambalonga Alves de
Desenvolvimento de um modelo híbrido de Deep Learning para a detecção de doenças pulmonares em imagens de raio-X. / Leonardo Luiz Gambalonga Alves de Oliveira ; orientador, Alexandre Gonçalves Silva, 2026.
154 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Ciências da Computação, Florianópolis, 2026.

Inclui referências.

1. Ciências da Computação. 2. Aprendizado de Máquina. 3. Redes Neurais. 4. Radiografia do Tórax. 5. Doenças Pulmonares. I. Silva, Alexandre Gonçalves. II. Universidade Federal de Santa Catarina. Graduação em Ciências da Computação. III. Título.

Dedico este trabalho à minha avó, Salete, mulher de muita força e perseverança, símbolo de bondade e sabedoria, e que será sempre a minha inspiração. *In memoriam*.

## AGRADECIMENTOS

Agradeço a Deus, pelas bençãos, pela vida e por me permitir fazer tudo isso possível.

Agradeço à minha mãe, por todo o apoio, incentivo, compreensão e paciência. Foram muitos dias fora de casa e, mesmo quando eu estava por perto, eu tive que dedicar meu tempo para elaboração deste e de muitos outros trabalhos ao longo da graduação. Sem ela, nada disso seria possível.

À minha avó, que durante toda a minha vida me apoiou, me incentivou a seguir sempre nos estudos e me ensinou a ser uma pessoa melhor, e que, infelizmente, não está mais conosco.

Ao meu pai, pelo suporte necessário para que eu pudesse concluir esta etapa com tranquilidade. E, claro, aos meus tios, tias e primos, que sempre estiveram presentes e me apoiaram em todos os momentos.

Ao meu orientador, Prof. Alexandre Gonçalves Silva, pela confiança, pela disponibilidade e pelos ensinamentos que tornaram este trabalho possível.

Aos meus amigos do curso, pela companhia, pelas trocas e por tornarem esta caminhada mais leve. Com certeza eu não teria chegado tão longe sem cada dica, esclarecimento e recomendação de vocês. Parafraseando um grande amigo, “ninguém se forma sozinho em Computação”, e é verdade: cada um de vocês contribuiu para que eu chegasse até aqui.

Aos meus amigos que não são do curso, que, mesmo por vezes sem saber muito sobre o que eu estava fazendo, colaboraram de alguma forma, seja com algum detalhe técnico de suas respectivas áreas, seja com a companhia. Agradeço muito por tirarem um pouco do seu tempo para me dar uma mão. Seja em Criciúma ou em Lyon, sempre estive rodeado de pessoas dispostas a me ajudar e que, de fato, fizeram a diferença. Toda a jornada foi incrível e, sem cada um de vocês, eu não teria vivido essa experiência que, por vezes, foi difícil e até dolorosa, mas gratificante e inesquecível no melhor sentido.

À Universidade Federal de Santa Catarina, principalmente pelas oportunidades, mas também pela formação e pela estrutura oferecidas.

A todos que, de alguma forma, contribuíram para a realização deste trabalho, o meu sincero agradecimento. Sou muito grato a todos.

*“Why not risk life when it’s gon’ keep goin’?”*
(Young Thug)

## RESUMO

As doenças pulmonares representam um desafio crescente para os sistemas de saúde, cenário agravado pela pandemia de COVID-19, que sobrecarregou os recursos hospitalares e comprometeu o diagnóstico de outras patologias, como a tuberculose, cujos óbitos no Brasil cresceram 85,71% entre 2018 e 2023. Embora a radiografia de tórax seja o exame de imagem mais acessível para a triagem dessas doenças, sua interpretação depende da disponibilidade de profissionais especializados e está sujeita a erros. Nesse contexto, este trabalho teve como objetivo desenvolver um sistema de apoio ao diagnóstico de patologias pulmonares em radiografias torácicas, baseado em um *pipeline* de aprendizado profundo composto por três módulos: segmentação pulmonar com a arquitetura CMUNeXt, classificação binária (normal ou patológico) e classificação multiclasse (normal, pneumonia, COVID-19 e tuberculose), ambas por meio de *ensembles* de cinco redes neurais convolucionais (modelo híbrido ResNet152 + EfficientNet-B5, ResNet50, DenseNet121, VGG16 e EfficientNet-B0), avaliados em duas arquiteturas de integração: paralela e em cascata. Para o treinamento, foi composta uma base de dados a partir de conjuntos públicos, disponibilizada em repositório aberto, e foram aplicadas técnicas de *data augmentation*, *fine-tuning* em duas fases e *Test Time Augmentation*, cujas contribuições foram quantificadas por análise ablativa. O módulo de segmentação atingiu IoU de 95,16% e coeficiente Dice de 97,52%. Na classificação, o *ensemble* multiclasse sem segmentação alcançou acurácia de 99,43% e *F1-score* de 99,40% na base interna, e o *pipeline* em cascata obteve *F1-score* de 94,20% em uma base externa não vista durante o treinamento, evidenciando boa capacidade de generalização. Observou-se, ainda, que as configurações sem segmentação superaram consistentemente aquelas com segmentação. Assim, concluiu-se que a combinação de arquiteturas heterogêneas por *ensemble learning* mitigou os erros individuais dos modelos e produziu predições mais robustas, demonstrando o potencial do sistema proposto como ferramenta de apoio ao diagnóstico de doenças pulmonares.



**Palavras-chave:** Aprendizado de Máquina. Inteligência Artificial. Redes Neurais. Doenças Pulmonares. Radiografia do Tórax.

## ABSTRACT

Lung diseases represent a growing challenge for health systems, a scenario aggravated by the COVID-19 pandemic, which overloaded hospital resources and compromised the diagnosis of other pathologies, such as tuberculosis, whose deaths in Brazil grew 85.71% between 2018 and 2023. Although chest radiography is the most accessible imaging exam for screening these diseases, its interpretation depends on the availability of specialized professionals and is prone to errors. In this context, this work aimed to develop a decision-support system for the diagnosis of lung pathologies in chest radiographs, based on a deep-learning *pipeline* composed of three modules: lung segmentation with the CMUNeXt architecture, binary classification (normal or pathological) and multiclass classification (normal, pneumonia, COVID-19 and tuberculosis), the latter two by means of *ensembles* of five convolutional neural networks (a hybrid ResNet152 + EfficientNet-B5 model, ResNet50, DenseNet121, VGG16 and EfficientNet-B0), evaluated in two integration architectures: parallel and cascade. For training, a dataset was composed from public collections and made available in an open repository, and *data augmentation*, two-phase *fine-tuning* and *Test Time Augmentation* techniques were applied, whose contributions were quantified through an ablation analysis. The segmentation module reached an IoU of 95.16% and a Dice coefficient of 97.52%. In classification, the multiclass *ensemble* without segmentation achieved 99.43% accuracy and 99.40% *F1-score* on the internal set, and the cascade *pipeline* obtained a 94.20% *F1-score* on an external set unseen during training, evidencing good generalization capability. It was also observed that configurations without segmentation consistently outperformed those with segmentation. Thus, it was concluded that combining heterogeneous architectures through *ensemble learning* mitigated the individual errors of the models and produced more robust predictions, demonstrating the potential of the proposed system as a decision-support tool for the diagnosis of lung diseases.



**Keywords:** Machine Learning. Artificial Intelligence. Neural Networks. Lung Diseases. Chest Radiography.

## LISTA DE FIGURAS


Figura 1 – Diagrama das modalidades de imagiologia médica. . . . . 29
Figura 2 – Exemplos de técnicas de imagiologia médica: (a) radiografia do tórax, (b) tomografia computadorizada do cólon, (c) ressonância magnética do cérebro, (d) imagem de uma retinografia, (e) imagem de medicina nuclear do cérebro e (f) imagem de ultrassonografia de um feto. . . . 30
Figura 3 – Estruturas ósseas presentes na radiografia do tórax. . . . . 32
Figura 4 – Estrutura do mediastino na radiografia do tórax. . . . . 33
Figura 5 – Estruturas pulmonares na radiografia do tórax. . . . . 34
Figura 6 – Exemplos de radiografias de pneumonia, Covid-19 e tuberculose. . . 37
Figura 7 – Diagrama Inteligência Artificial . . . . . 40
Figura 8 – Ilustração das diferentes estratégias de aprendizado - Os círculos vermelho e azul representam dados rotulados de diferentes classes. O círculo cinza representa dados não rotulados. As linhas pretas definem o limite de decisão subjacente entre classes. Os círculos riscados representam dados que ignoram e usam as informações do rótulo em diferentes estágios. No diagrama mais a direita, o Agente (triângulo) interage com o ambiente (círculo), explorando (linha tracejada) e otimizando ações para maximizar recompensas (estrelas). . . . . 42
Figura 9 – Arquitetura de (a) rede rasa e (b) rede neural profunda. . . . . 44
Figura 10 – As diferenças entre etapas de ML e DL. . . . . 45
Figura 11 – Ilustração dos diferentes tipos de redes - Azul e vermelho representam classes, e amostras preenchidas com cada cor indicam dados rotulados, e amostras brancas indicam dados não rotulados. A cor de fundo representa o valor de probabilidade posterior ou a distribuição condicional de classe obtida por cada modelo. O modelo híbrido integra o poder discriminativo do modelo discriminativo e a disponibilidade de dados não supervisionados do modelo generativo. 47
Figura 12 – Rede Recorrente. . . . . 48
Figura 13 – Arquitetura CNN . . . . . 50
Figura 14 – Diagrama de convolução . . . . . 51

Figura 15 – Diagrama de exemplo de Pooling . . . . . 53
Figura 16 – Tipos de funções de ativação: (a) Tangente Hiperbólica, (b) Sigmoide (Logística), (c) ReLU e (d) PreLU. . . . . 56
Figura 17 – Matriz de confusão . . . . . 58
Figura 18 – Exemplo de Curva ROC com diferentes modelos . . . . . 61
Figura 19 – Exemplos de segmentação: (a) Imagem do pulmão em raio-X, (b) Máscara e (c) Máscara aplicada. . . . . 63
Figura 20 – Arquitetura U-Net (exemplo para $32 \times 32$ pixels na resolução mais baixa) . . . . . 63
Figura 21 – Arquitetura CMUNeXt . . . . . 67
Figura 22 – Exemplos de radiografias para as classes: Normal, Tuberculose, COVID-19 e Pneumonia. . . . . 75
Figura 23 – Diagrama do exemplo da segmentação do pulmão . . . . . 79
Figura 24 – Diagrama do modelo de arquitetura em cascata proposta . . . . . 90
Figura 25 – Diagrama do modelo de arquitetura paralela proposta . . . . . 91
Figura 26 – Gráfico de Curvas de perda e *IoU* de validação do *CMUNeXt*. . . . . 94
Figura 27 – Matriz de confusão do CMUNeXt. . . . . 95
Figura 28 – Gráfico de perda e acurácia. . . . . 97
Figura 29 – Matriz de confusão do modelo multiclasse no conjunto de teste. . . . 97
Figura 30 – Matriz de confusão do modelo multiclasse no conjunto do *dataset* externo. . . . . 98
Figura 31 – Matriz de confusão do modelo multiclasse no conjunto de teste com *data augmentation*. . . . . 100
Figura 32 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com *data augmentation*. . . . . 101
Figura 33 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com TTA. . . . . 102
Figura 34 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com TTA e *Harmonization*. . . . . 103
Figura 35 – Gráfico de perda e acurácia do treinamento do *fine-tuning*. . . . . 104
Figura 36 – Matriz de confusão do modelo multiclasse no conjunto de teste com *fine-tuning*. . . . . 105

Figura 37 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com *fine-tuning*. . . . . 106
Figura 38 – Matriz de confusão do modelo multiclasse no conjunto de teste com segmentação. . . . . 108
Figura 39 – Gráfico de perda e acurácia do treinamento do *fine-tuning* com segmentação. . . . . 108
Figura 40 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com segmentação. . . . . 109
Figura 41 – Matriz de confusão do modelo binário. . . . . 111
Figura 42 – Matriz de confusão do modelo binário com segmentação. . . . . 112
Figura 43 – Matriz de confusão do modelo multiclasse. . . . . 114
Figura 44 – Matriz de confusão do modelo multiclasse com segmentação. . . . . 115
Figura 45 – Matriz de confusão do modelo multiclasse em cascata. . . . . 117
Figura 46 – Matriz de confusão do modelo multiclasse em cascata. . . . . 119
Figura 47 – Matriz de confusão no *dataset* externo. . . . . 120
Figura 48 – Matriz de confusão no *dataset* externo com segmentação. . . . . 122
Figura 49 – Comparação de desempenho entre as arquiteturas propostas sem (*raw*) e com segmentação (*masked*). . . . . 123
Figura 50 – Estrutura de diretórios do *dataset*. . . . . 133

## LISTA DE TABELAS


Tabela 1 – Configurações do *hardware* utilizado para o desenvolvimento do trabalho. . . . . 73
Tabela 2 – Distribuição dos casos nos *datasets*. . . . . 76
Tabela 3 – Transformações de *data augmentation* aplicadas durante o treinamento. 80
Tabela 4 – Métricas de treinamento do *CMUNeXt* por época. (✓= melhor modelo selecionado, *threshold* = 0,50) . . . . . 94
Tabela 5 – Métricas de treinamento do *CMUNeXt* por época (* Estimativa com proporção de foreground de 35% dos pixels). . . . . 95
Tabela 6 – Resultados obtidos no conjunto de teste . . . . . 96
Tabela 7 – Resultados obtidos no conjunto do *dataset* externo . . . . . 98
Tabela 8 – Resultados obtidos com e sem *data augmentation* . . . . . 99
Tabela 9 – Resultados obtidos com e sem *data augmentation* . . . . . 100
Tabela 10 – Resultados obtidos com e sem TTA e *Harmonization* . . . . . 102
Tabela 11 – Resultados obtidos com e sem *fine-tuning* . . . . . 104
Tabela 12 – Resultados obtidos com e sem *Fine-Tuning* no conjunto de teste externo 105
Tabela 13 – Resultados obtidos com e sem segmentação . . . . . 107
Tabela 14 – Resultados obtidos com e sem segmentação . . . . . 109
Tabela 15 – Métricas de desempenho por estratégia no modelo binário. (✓= melhor modelo) . . . . . 112
Tabela 16 – Métricas de desempenho por estratégia no modelo binário com segmentação. (✓= melhor modelo) . . . . . 112
Tabela 17 – Métricas de desempenho por estratégia no modelo multiclasse. (✓= melhor modelo) . . . . . 115
Tabela 18 – Métricas de desempenho por estratégia no modelo multiclasse com segmentação. (✓= melhor modelo) . . . . . 116
Tabela 19 – Métricas de desempenho do Resultado em Cascata. . . . . 117
Tabela 20 – Métricas de desempenho do Resultado em Cascata. . . . . 119
Tabela 21 – Métricas de desempenho no dataset externo. . . . . 121
Tabela 22 – Métricas de desempenho no *dataset* externo com segmentação. . . . 122
Tabela 23 – Comparação com trabalhos relacionados da literatura. Seg. = Segmentação; Ens. = *Ensemble*; FT = *Fine-Tuning*; — = não reportado. 125

**SUMÁRIO**


**Lista de Figuras** . . . . . **13**

**Lista de tabelas** . . . . . **17**

**1 INTRODUÇÃO** . . . . . **23**
1.1 OBJETIVOS . . . . . 24
**1.1.1 Objetivo Geral** . . . . . **24**
**1.1.2 Objetivos Específicos** . . . . . **24**
1.2 MÉTODO DE PESQUISA . . . . . 25
1.3 ESTRUTURAÇÃO DO TRABALHO . . . . . 25

**2 FUNDAMENTAÇÃO TEÓRICA** . . . . . **27**
2.1 DIAGNÓSTICO POR IMAGEM . . . . . 27
**2.1.1 Modalidades de Imagem** . . . . . **27**
**2.1.2 Radiologia** . . . . . **30**
**2.1.3 Radiografia** . . . . . **30**
**2.1.4 Radiografia do tórax** . . . . . **31**
2.2 DOENÇAS PULMONARES . . . . . 35
2.3 INTELIGÊNCIA ARTIFICIAL . . . . . 38
2.4 APRENDIZAGEM DE MÁQUINA . . . . . 39
**2.4.1 Tipos de aprendizado** . . . . . **40**
**2.4.2 Redes Neurais Artificiais** . . . . . **42**
2.5 APRENDIZAGEM PROFUNDA . . . . . 43
**2.5.1 Rede Neural Convolucional** . . . . . **49**
*A Modelo Geral* . . . . . *49*
*B Camada de Convolução* . . . . . *51*
*C Pooling* . . . . . *52*
*D Camadas fully connected (FC)* . . . . . *54*
*E Função de Ativação* . . . . . *54*
2.6 MEDIDAS DE DESEMPENHO . . . . . 56
2.7 SEGMENTAÇÃO DE IMAGENS . . . . . 62

2.8 ARQUITETURA *U-NET* . . . . . 62
2.9 ARQUITETURA *CMUNEXT* . . . . . 65

**3 ESTADO DA ARTE . . . . . 69**
3.1 CLASSIFICAÇÃO DAS IMAGENS DE RAIO-X . . . . . 70
3.2 POSICIONAMENTO DA PESQUISA PROPOSTA . . . . . 71

**4 METODOLOGIA . . . . . 73**
4.1 MATERIAIS E DADOS . . . . . 73
**4.1.1 *Hardware* e ferramentas de suporte . . . . . 73**
**4.1.2 Bases de dados . . . . . 74**
4.2 PRÉ-PROCESSAMENTO . . . . . 77
**4.2.1 Segmentação dos pulmões . . . . . 78**
**4.2.2 *Data Augmentation* . . . . . 79**
4.3 ARQUITETURAS DE CNN . . . . . 80
**4.3.1 Híbrido (ResNet152 + EfficientNet-B5) . . . . . 80**
**4.3.2 ResNet50 . . . . . 82**
**4.3.3 DenseNet121 . . . . . 82**
**4.3.4 Vgg16 . . . . . 82**
**4.3.5 EfficientNet_B0 . . . . . 83**
4.4 MÉTODOS DE AGREGAÇÃO . . . . . 83
**4.4.1 Votação por maioria (*Hard Voting*) . . . . . 84**
**4.4.2 Média das previsões (*Soft Voting*) . . . . . 84**
**4.4.3 Média ponderada pela AUC (*Weighted Average by AUC*) . . . . . 84**
4.5 ESTRATÉGIAS DE REFINAMENTO . . . . . 85
**4.5.1 Fine-Tuning . . . . . 85**
*A Primeira fase: ajuste das camadas finais . . . . . 86*
*B Segunda fase: ajuste de camadas convolucionais . . . . . 86*
**4.5.2 Harmonização . . . . . 87**
**4.5.3 *Test Time Augmentation* . . . . . 88**
4.6 MODELO DE ARQUITETURA PROPOSTO . . . . . 89

**5 RESULTADOS E DISCUSSÃO . . . . . 93**
5.1 SEGMENTAÇÃO DOS PULMÕES . . . . . 93

5.2 ANÁLISE ABLATIVA . . . . . 96
**5.2.1 Impacto do *Data Augmentation* . . . . . 99**
**5.2.2 Impacto do TTA e *Harmonization* . . . . . 102**
**5.2.3 Impacto do *Fine-Tuning* . . . . . 104**
**5.2.4 Impacto da Segmentação . . . . . 107**
5.3 ESTÁGIO BINÁRIO . . . . . 110
**5.3.1 Sem segmentação . . . . . 110**
**5.3.2 Com segmentação . . . . . 112**
5.4 ARQUITETURA PARALELA . . . . . 113
**5.4.1 Estágio Multiclasse - Sem Segmentação . . . . . 113**
**5.4.2 Estágio Multiclasse - Com Segmentação . . . . . 115**
5.5 ARQUITETURA CASCATA . . . . . 116
**5.5.1 *Pipeline* - Sem Segmentação . . . . . 117**
**5.5.2 *Pipeline* - Com Segmentação . . . . . 118**
5.6 AVALIAÇÃO COM DATASET EXTERNO . . . . . 120
**5.6.1 Sem segmentação . . . . . 120**
**5.6.2 Com segmentação . . . . . 121**
5.7 COMPARAÇÃO ENTRE ARQUITETURAS . . . . . 123
5.8 COMPARAÇÃO COM A LITERATURA . . . . . 124

**6 CONCLUSÃO . . . . . 127**
6.1 PRINCIPAIS CONCLUSÕES . . . . . 127
6.2 TRABALHOS FUTUROS . . . . . 129
**6.2.1 Ampliação e Diversificação do *Dataset* . . . . . 129**
**6.2.2 Aprimoramento do Módulo de Segmentação . . . . . 130**
**6.2.3 Representação por *Patches* Internos ao Pulmão . . . . . 130**
**6.2.4 Otimização de Hiperparâmetros e Arquiteturas . . . . . 131**
**6.2.5 Validação Clínica . . . . . 131**

**APÊNDICE A – ESTRUTURA DE DIRETÓRIOS DO *DATASET* 133**

**REFERÊNCIAS . . . . . 135**

# 1 INTRODUÇÃO

No Brasil, as doenças pulmonares são uma das maiores causas de morbilidade, afligindo cerca de 27,3% da população com algum tipo de enfermidade (Telemedicina Morsch, 2024). Atualmente, com a pandemia iniciada em dezembro de 2019 de rápida propagação espalhando o vírus responsável por casos de pneumonia, SARS-CoV-2 ou coronavírus, muitas doenças se tornaram mais mortais pelo impacto prejudicial da covid-19, já que muitos recursos de saúde foram direcionados para o combate ao novo vírus, o que reduziu os tratamentos e diagnósticos destes (BBC News Brasil, 2020). Dessa forma, enquanto em 2018 a tuberculose, por exemplo, foi responsável por aproximadamente 7 mil mortes, em 2023 ela causou em torno de 13 mil óbitos, assim atingindo um aumento de 85,71% (World Health Organization, 2023).

Com isso, apesar do esforço do governo brasileiro em reduzir o tempo de tratamento, no caso dessa enfermidade pulmonar houve uma diminuição de 18 para 6 meses (Ministério da Saúde do Brasil, 2023), ainda se faz necessário a eficácia, a velocidade e a precisão no diagnóstico dessas patologias, para que os pacientes possam iniciar o tratamento o mais rápido possível.

Dito isso, dentre as diversas formas de diagnósticos, a radiografia do tórax, pelo seu custo baixo e natureza não invasiva, é a mais requisitada (ROCHA; MENDONÇA; CAMPILHO, 2021). Entretanto, o tempo não trivial de espera de análise e a suscetibilidade a interpretações dúbias, mesmo contando com a atuação de profissionais experientes, evidenciam os problemas persistentes na aplicação dessa técnica e destacam a urgência de soluções complementares que agilizem e aprimorem a precisão diagnóstica. (ROCHA; MENDONÇA; CAMPILHO, 2021). Por essa razão, o apoio e a aplicação de sistemas computacionais auxiliares para o diagnóstico em âmbito clínico e hospitalar tornam-se uma alternativa de grande interesse para ajudar na decisão e diagnóstico.

Nesse caminho, a inteligência artificial (IA) vem numa crescente de notoriedade, podendo o seu uso ser visto além da comunidade científica, mas, inclusive, no dia a dia das pessoas em âmbitos diversos, como jogos, finanças e varejo. Além disso, também vem sendo uma ferramenta de desenvolvimento de softwares capazes de aplicar a segmentação de imagens de tomografias computadorizadas e obter bons desempenhos referentes às tarefas de interpretação e decisão, conseguindo identificar e classificar

doenças (CHICHANOSKI, 2022).

Assim, a utilização de ferramentas de processamento de imagens constitui uma solução e simplificação tecnológica com grandes benefícios (COLAÇO, 2022), tanto para diminuir o tempo de diagnóstico, o que poderia iniciar os tratamentos mais cedo, como servir de triagem, detectando casos que precisem ser analisados com urgência. Além disso, o desenvolvimento da área de IA, juntamente e particularmente das subáreas Machine Learning (ML) e Deep Learning (DL), provou um aumento de estudos neste tópico de análise médica. Por conseguinte, ampliaram-se as possibilidades de aplicações e de caminhos a se seguir por esses estudos, como, por exemplo, segmentação, localização, classificação e detecção (COLAÇO, 2022).

Então, este trabalho tem o propósito de desenvolver um modelo de detecção de patologias em imagens radiográficas do tórax, que combine diferentes redes CNN, a fim de criar um modelo mais preciso, diminuindo equívocos na matriz de confusão, ferramenta capaz de medir o sucesso da classificação de modelos (KAMILA et al., 2024).

## 1.1 OBJETIVOS

### 1.1.1 Objetivo Geral

O objetivo geral deste trabalho é desenvolver um modelo híbrido de Deep Learning para a detecção de doenças pulmonares em imagens radiográficas do tórax, integrando diferentes abordagens para melhorar a precisão e reduzir erros na análise.

### 1.1.2 Objetivos Específicos

- Buscar e organizar um conjunto de dados público abrangente de imagens radiográficas de tórax, garantindo qualidade e diversidade para o treinamento do modelo.
- Implementar e treinar modelos individuais de Deep Learning, como CNNs (Redes Neurais Convolucionais), para segmentação e classificação de doenças pulmonares entre 4 classes: não patológico, pneumonia, covid-19 e tuberculose.

- Desenvolver um modelo híbrido, combinando as saídas de modelos individuais para aprimorar a detecção e classificação.
- Avaliar o desempenho do modelo híbrido por meio de métricas como precisão, sensibilidade e especificidade, comparando-o a abordagens tradicionais.
- Validar o modelo em um conjunto de dados realista para analisar sua aplicabilidade clínica e hospitalar.
- Explorar técnicas de otimização, como regularização e ajuste de hiperparâmetros, para melhorar a robustez do modelo.

## 1.2 MÉTODO DE PESQUISA

O método de pesquisa para o desenvolvimento de um modelo híbrido de Deep Learning para detecção de doenças pulmonares foi dividido em etapas. Primeiramente, foi realizada uma revisão bibliográfica aprofundada para identificar lacunas, metodologias existentes e potencialidades dos modelos híbridos no diagnóstico radiológico.

Na sequência, foram feitos a coleta e o pré-processamento de dados utilizando datasets públicos, como o disponibilizado pelo *Guangzhou Women and Children's Medical Center* e o *COVID-19 Radiography Database*, aplicando técnicas como normalização e aumento de dados para preparar o conjunto para o treinamento.

A implementação incluiu a construção de modelos individuais de Deep Learning (ex.: CNNs) e sua combinação em um modelo híbrido, explorando estratégias como ensemble learning. O modelo foi avaliado com métricas como acurácia e F1-Score, validando sua eficácia e generalização com dados externos.

Por fim, o desempenho do modelo foi comparado a métodos existentes, e sua aplicabilidade prática foi explorada, considerando desafios técnicos, éticos e legais em ambientes clínicos.

## 1.3 ESTRUTURAÇÃO DO TRABALHO

Este documento de monografia foi dividido em 6 capítulos: Introdução, Fundamentação Teórica, Estado da Arte, Metodologia, Resultados e Discussão, e

Conclusão. O presente capítulo, Introdução, tem o propósito de apresentar e contextualizar a problemática e os objetivos. O Capítulo 2 trata dos principais conceitos teóricos e fundamentais da literatura para a compreensão deste trabalho. No Capítulo 3 é apresentada a revisão bibliográfica e em que se encontra o conhecimento sistematicamente sobre o assunto. O Capítulo 4 abrange os métodos utilizados e de que forma os dados foram analisados. Já o Capítulo 5 exibe os resultados obtidos no desenvolvimento em conjunto de discussões a partir desses. Finalmente, o Capítulo 6 conclui o trabalho, apresentando as limitações encontradas no sistema e caminhos futuros.

# 2 FUNDAMENTAÇÃO TEÓRICA

## 2.1 DIAGNÓSTICO POR IMAGEM

A imagiologia, ou diagnóstico por imagem, é uma especialidade da medicina em que consiste na obtenção de imagens que representam visualmente órgãos, tecidos e estruturas fisiológicas tanto sob a pele quanto sob os ossos em diferentes partes do corpo humano ou animal (HUSSAIN et al., 2022). Essas representações podem ser usadas para inúmeros fins como clínicos, monitoramento, tratamento e científicos (ISLAM et al., 2023). Essa área ao longo dos anos se manteve em constante evolução, com o desenvolvimento de novas técnicas e tecnologias que permitem a obtenção de imagens cada vez mais detalhadas e precisas. Em adição à crescente da IA, também a possibilitou avanços significativos na análise e interpretação dessas imagens.

### 2.1.1 Modalidades de Imagem

Além disso, esse campo contém diversas modalidades de imagem. Como mostrado na Figura 1, o diagnóstico por imagem é composto por cinco categorias, cada uma usando diferentes tecnologias e princípios físicos. Assim, por consequência, obtendo diferentes resultados e que, por possuírem características próprias, ajudam a produzir informações distintas. Essas subclasses são:

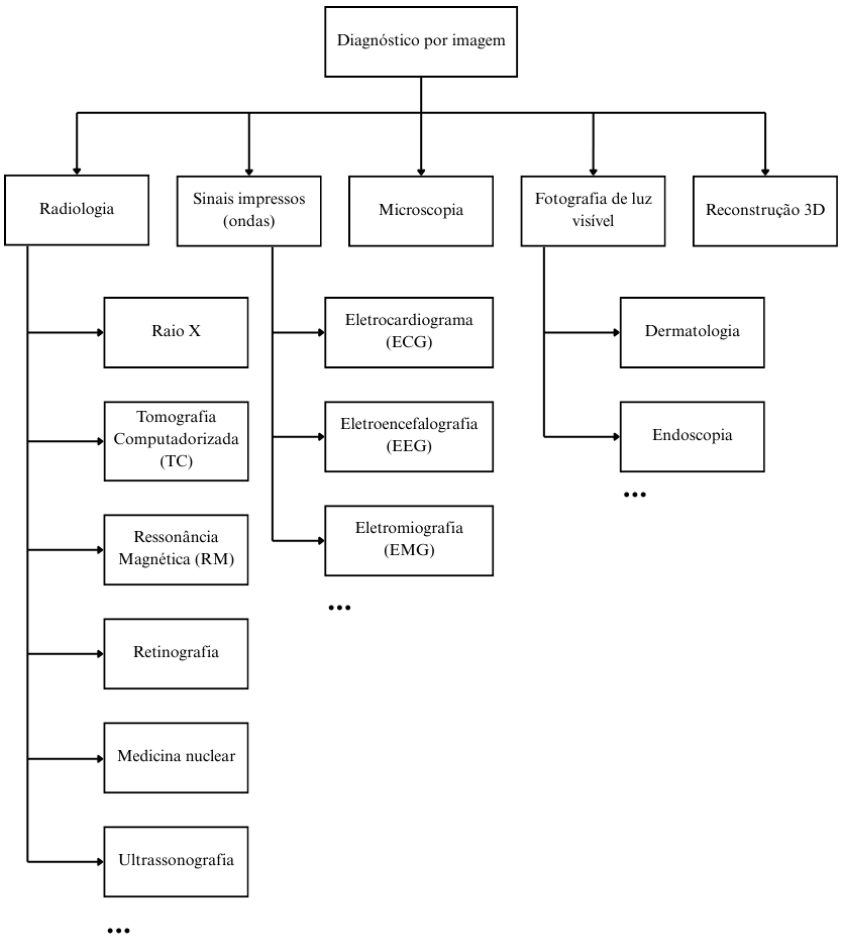
- Radiologia: A radiologia é a especialidade médica que utiliza técnicas de imagem para diagnosticar e tratar doenças.
- Sinais impressos (ondas): Também conhecido por eletrogramas, os sinais impressos, aplicam técnicas que utilizam ondas eletrônicas com o objetivo de capturar funções fisiológicas e cognitivas do corpo, como a atividade elétrica do coração (eletrocardiograma), do cérebro (eletroencefalograma) ou dos músculos esqueléticos e nervos periféricos (eletromiograma). Dessa forma, registrando a atividade elétrica do corpo, o que fornece informações sobre a saúde e o funcionamento dos órgãos e sistemas do corpo (EL-GAMAL; ELMOGY; ATWAN, 2016).

- Microscopia: A microscopia é uma técnica de imagem que utiliza um microscópio para ampliar objetos pequenos, como células e tecidos, permitindo a visualização de detalhes que não são visíveis a olho nu (University of Edinburgh – Edinburgh Imaging, 2025). Existem diferentes tipos de imagens que podem ser capturadas dependendo das ondas do microscópio que são focalizadas na amostra (EL-GAMAL; ELMOGY; ATWAN, 2016).

- Fotografia: A fotografia é uma técnica de imagem que utiliza uma câmera para capturar imagens visuais de objetos ou cenas, podendo ser colorido ou em preto e branco (EL-GAMAL; ELMOGY; ATWAN, 2016). Na medicina, a fotografia é frequentemente usada para documentar lesões, feridas, condições dermatológicas (dermatologia) ou ainda estruturas internas (endoscopia), além de outros aspectos visuais relacionados à saúde (EL-GAMAL; ELMOGY; ATWAN, 2016; PETERSILGE et al., 2022).

- Reconstrução 3D: Essa categoria de imagem é uma técnica que utiliza algoritmos, modelos em *deep learning* e software para criar modelos tridimensionais a partir de dados de imagem bidimensionais, como tomografias computadorizadas (TC) ou ressonâncias magnéticas (RM) (MCMILLIAN; BANERJEE; BUENO-OROVIO, 2025). Esses modelos 3D permitem uma visualização mais detalhada e realista de estruturas anatômicas, auxiliando no planejamento cirúrgico com medições precisas, na avaliação de lesões e na compreensão da anatomia complexa (ZUKIć et al., 2025).

Figura 1 – Diagrama das modalidades de imagiologia médica.




Fonte: Adaptado de (ISLAM et al., 2023).

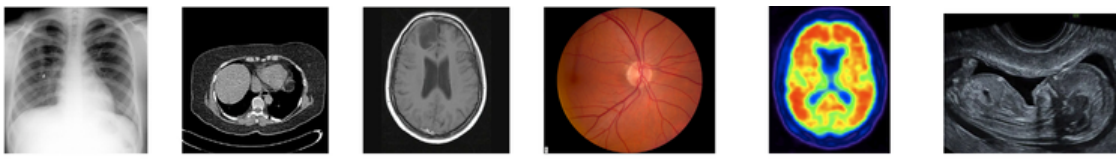
Dessa forma, as imagens médicas obtidas desempenham um papel essencial para a avaliação de um paciente, já que contribuem para o diagnóstico precoce de doenças, auxiliam no planejamento de tratamentos e monitoramento da evolução clínica, minimizando a necessidade de outros procedimentos que possam oferecer algum risco (HUSSAIN et al., 2022). Além disso, a imagiologia médica também contribui crucialmente na pesquisa científica, permitindo o estudo detalhado da anatomia e fisiologia humana, bem como o desenvolvimento de novas técnicas e tecnologias para melhorar a prática médica.

### 2.1.2 Radiologia

A origem da radiologia se dá ao descobrimento dos raios X, radiação de forma eletromagnética ionizante, por Wilhelm Röntgen em 1895, sendo a primeira técnica de imagiologia médica a ser desenvolvida (THOMAS; BANERJEE, 2013). Contudo atualmente, as técnicas pertencentes a radiologia não necessariamente precisam utilizar radiação ionizante, e sim visualizar o corpo por algum meio indireto (National Cancer Institute, 2025). Dessa forma, se tornando um campo vasto e diversificado que abrange uma variedade de técnicas e modalidades de imagem, cada uma com suas próprias aplicações clínicas e características específicas.

Dentre as diversas técnicas pertencentes a radiologia, destacam-se a radiografia, a tomografia computadorizada (TC), a ressonância magnética (RM), a retinografia, a medicina nuclear e a ultrassonografia. Cada uma dessas técnicas possui suas próprias propriedades, vantagens e limitações, sendo escolhidas de acordo com a região do corpo e são adaptadas conforme o objetivo clínico e as necessidades e características do paciente (ISLAM et al., 2023). Na Figura 2 são apresentados os exemplos das técnicas citadas anteriormente.

Figura 2 – Exemplos de técnicas de imagiologia médica: (a) radiografia do tórax, (b) tomografia computadorizada do cólon, (c) ressonância magnética do cérebro, (d) imagem de uma retinografia, (e) imagem de medicina nuclear do cérebro e (f) imagem de ultrassonografia de um feto.




Fonte: Adaptado de (REMESEIRO; BOLóN-CANEDO, 2019; Tesla Imagem, 2025; National Institute of Biomedical Imaging and Bioengineering (NIBIB), 2025).

### 2.1.3 Radiografia

As imagens de raio-X ou radiografias são obtidas por meio pela exposição de uma pequena dose de radiação ionizante capaz de atravessar objetos sólidos. No corpo humano, o feixe de radiação é absorvido de maneira diferente pelos tecidos, conforme a densidade, resultando em uma imagem bidimensional. Os ossos, por exemplo, absorvem

mais radiação e aparecem como áreas brancas na radiografia, enquanto os tecidos moles e menos densos absorvem menos radiação e aparecem como áreas mais escuras (HUSSAIN et al., 2022). Para um uso correto, seguro e em busca de uma imagem clara existe uma quantidade certa de radiação e um posicionamento adequado em relação a área a ser examinada. Assim, as radiografias são amplamente utilizadas para diagnosticar fraturas ósseas, doenças pulmonares, problemas dentários e outras condições médicas (HUSSAIN et al., 2022).

Essa é uma das técnicas mais amplamente utilizadas na prática médica sua popularidade se deve a sua rapidez, baixo custo e ampla disponibilidade. Outra vantagem comparado aos outros métodos é a qualidade da imagem em alta resolução capaz de mostrar detalhes anatômicos em estruturas internas como pulmões. Além disso, ela não é invasiva, o que significa que não requer a inserção de quaisquer instrumento no corpo, tornando-a uma opção segura para muitos pacientes (ISLAM et al., 2023).

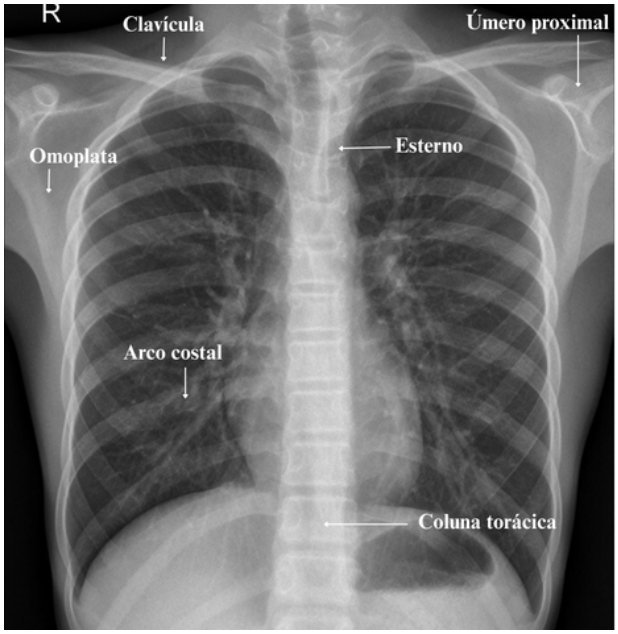
### 2.1.4 Radiografia do tórax

O exame de radiografia na avaliação do tórax deve ser realizado em pelos ao menos duas incidências, as mais comuns são a posteroanterior (PA) e a lateral (perfil). A incidência PA consiste no posicionamento do paciente de pé com o peito encostado na placa de imagem, enquanto a incidência lateral é realizada com o paciente de lado, permitindo uma visão diferente do tórax (WADA; RODRIGUES; SANTOS, 2019). Como o resultado da radiografia é uma imagem bidimensional, a realização de múltiplas incidências é fundamental para obter uma visão mais completa e precisa do tórax, permitindo a identificação de anormalidades que podem não ser visíveis em apenas uma incidência (WADA; RODRIGUES; SANTOS, 2019). É preferível que a incidência seja em PA do que em anteroposterior (AP) para evitar a distorção da imagem, já que a incidência AP é realizada com o paciente deitado, o que pode resultar em uma imagem menos nítida e mais difícil de interpretar por conta da divergência dos raios X. (WADA; RODRIGUES; SANTOS, 2019).

Com uma qualidade boa de imagem, a radiografia do tórax considerada normal é capaz de mostrar detalhes anatômicos em estruturas internas e órgãos, mesmo que sobrepostos, elementos moles e ossos, como os pulmões, coração, vasos sanguíneos, diafragma, parte da caixa torácica e parte da coluna vertebral.

O tórax ósseo do peito é composto por 12 pares de costelas, o esterno e a coluna vertebral. O esterno é um osso plano localizado na parte frontal do tórax, que se conecta às costelas por meio de cartilagens costais. As costelas são ossos curvos que se estendem a partir da coluna vertebral e se conectam ao esterno na parte frontal do tórax. A coluna vertebral é composta por vértebras que se estendem ao longo da parte posterior, fornecendo suporte estrutural e proteção para a medula espinhal (HOBBS, 2007). Essa estrutura óssea é fundamental para a proteção dos órgãos internos, como os pulmões e o coração, além de desempenhar um papel importante na respiração, permitindo a expansão pulmonar, e no movimento, dando suporte para ombros e braços (). Na Figura 3 é possível observar as estruturas ósseas citadas anteriormente, além de outros ossos como a clavícula, omoplata e úmero proximal, componentes do complexo do ombro.

Figura 3 – Estruturas ósseas presentes na radiografia do tórax.

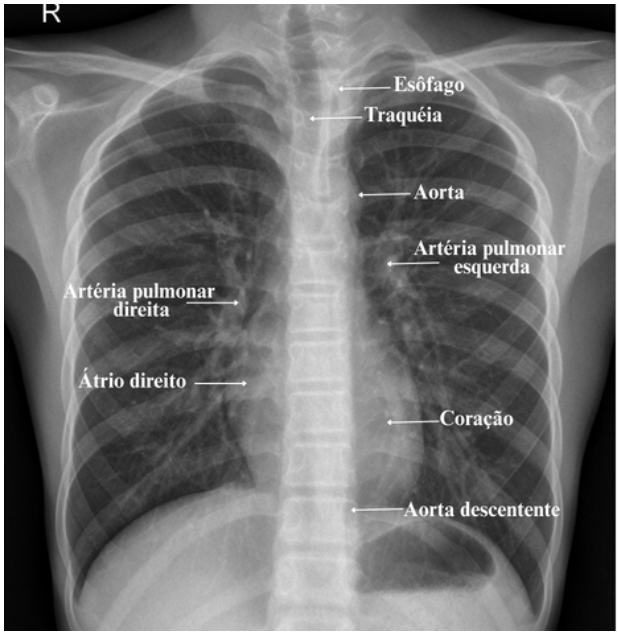



Fonte: Adaptado de (COLAÇO, 2022).

Além disso, é presente também o mediastino, espaço localizado entre os pulmões, que contém o coração, grandes vasos sanguíneos, incluindo a raiz da aorta e artérias pulmonares proximais, traqueia e brônquios principais, esôfago e linfonodos (HOBBS, 2007). Há também a glândula timo localizada acima do coração, que é mais proeminente em crianças e adolescentes (WADA; RODRIGUES; SANTOS, 2019). Na

Figura 4 é possível observar a silhueta do mediastino, com os órgãos e estruturas citados anteriormente, na radiografia. Entretanto para melhor visualização e identificação das 4 seções do mediastino (superior, anterior, médio e posterior), é necessário na grafia em perfil ou em exames seccionais (TC e RM) (WADA; RODRIGUES; SANTOS, 2019)

Figura 4 – Estrutura do mediastino na radiografia do tórax.

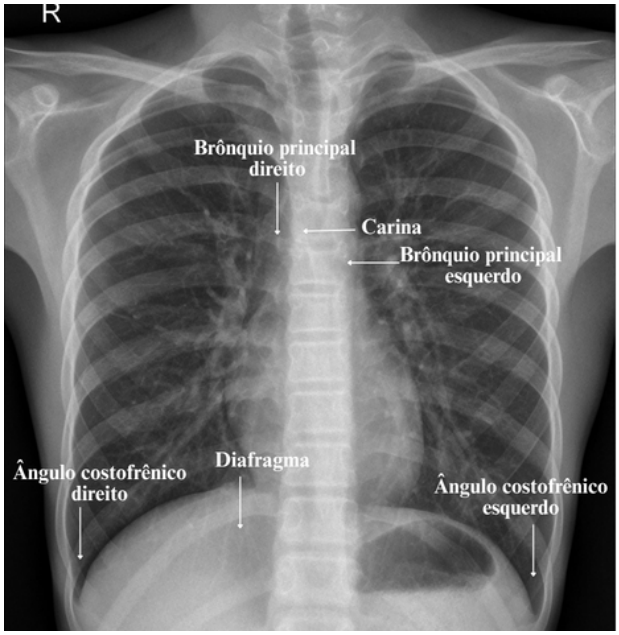



Fonte: Adaptado de (COLAÇO, 2022).

Os pulmões são órgãos compostos por tecido esponjoso e elástico, chamado parênquima pulmonar, que é responsável pela troca gasosa. Esse tecido contém as ramificações terminais da via aérea, como os bronquíolos e os milhões de sacos alveolares. Por conta dessa composição aerada, os pulmões aparecem como áreas escuras na radiografia.

Ainda assim, é possível observar, conforme na Figura 5, estruturas centrais e periféricas que compõem o sistema respiratório, como a traqueia que divide-se na carina para dar origem aos brônquios principais. À medida que penetram nos pulmões, essas vias aéreas e os vasos pulmonares ramificam-se progressivamente, criando um padrão de linhas finas que contrastam com o fundo radiotransparente (escuro) do órgão.

Figura 5 – Estruturas pulmonares na radiografia do tórax.




Fonte: Adaptado de (COLAÇO, 2022).

Além disso, os pulmões são divididos em lobos, nomeados conforme a localização, sendo o pulmão direito composto por três lobos (superior, médio e inferior) e o pulmão esquerdo por dois lobos (superior e inferior), devido à presença do coração no lado esquerdo do tórax (HOBBS, 2007). A fissura horizontal separa o lobo superior do lobo médio no pulmão direito, enquanto a fissura oblíqua separa o lobo inferior dos outros lobos em ambos os pulmões (HOBBS, 2007). Essas fissuras podem ser visualizadas na radiografia como linhas finas que se estendem entre os lobos, embora sua visibilidade possa variar dependendo da qualidade da imagem e da posição do paciente (WADA; RODRIGUES; SANTOS, 2019).

Mais abaixo na radiografia, é possível observar o diafragma, um músculo em forma de cúpula que separa a cavidade torácica da cavidade abdominal. O diafragma é responsável pela respiração, contraindo-se para permitir a entrada de ar nos pulmões e relaxando-se para expelir o ar durante a expiração. Na radiografia, como apontado na Figura 5, o diafragma aparece como uma linha curva que se estende de um lado do tórax ao outro, podendo ser dividido em duas seções chamadas de hemidiáfragmas direito e esquerdo, sendo o lado direito, pela presença do fígado abaixo, geralmente mais alto do

que o hemidiáfragma esquerdo (HOBBS, 2007).

Outra estrutura importante é a cavidade pleural, um espaço virtual entre as camadas da pleura que revestem os pulmões e a parede torácica. Normalmente, esse espaço contém uma pequena quantidade de líquido pleural que lubrifica as superfícies e permite o movimento suave dos pulmões durante a respiração (WADA; RODRIGUES; SANTOS, 2019). Na radiografia, a cavidade pleural é representada por uma linha fina que segue a borda dos pulmões, e sua presença ou ausência pode indicar condições como derrame pleural (acúmulo de líquido), pneumotórax (presença de ar) ou hemotórax (acúmulo de sangue) (HOBBS, 2007).

Um dos locais em que são importantes para avaliar algum desses tipos de derrame é na porção inferior dos pulmões, onde o diafragma encontra as costelas, formando os ângulos costofrênicos, indicado na Figura 5. No momento que isso acontece, esses ângulos aparecem achatados e atenuados como resultado do acúmulo de líquido ou ar (HOBBS, 2007).

## 2.2 DOENÇAS PULMONARES

Na análise das radiografias do tórax, é possível identificar uma variedade de alterações na aparência normal das estruturas pulmonares e adjacentes que podem indicar a presença de doenças ou condições patológicas. Essas alterações podem se manifestar de diversas formas, como:

- Opacidades: Áreas de aumento da densidade que aparecem mais esbranquiçadas na radiografia, indicando a substituição do ar por líquidos, células inflamatórias ou tecido anormal (ROUND, 2007).

- Nódulos ou massas: Pequenas massas ou lesões que podem ser benignas ou malignas, aparecendo como áreas arredondadas de aumento de densidade, podendo ser relacionadas a processos inflamatórios, infecciosos ou neoplásicos (ROUND, 2007).

- Consolidações: Tipo de opacidade, caracterizado pro áreas de aumento difuso da densidade pulmonar, onde os alvéolos estão preenchidos com líquido ou células inflamatórias (ROUND, 2007).

- Cavitações: Áreas radiotransparentes, por conta da destruição do tecido pulmonar que, por consequência, formam cavidades, frequentemente associadas a infecções crônicas, ou a processos neoplásicos (BEHZADI-KHORMOUJI et al., 2020; ROUND, 2007).

- Padrão intersticial: Caracterizado por um aumento da densidade do interstício pulmonar, que pode se manifestar como linhas finas, reticuladas ou em forma de rede, indicando a presença de doenças intersticiais (ROUND, 2007).

A seguir estão algumas das principais doenças pulmonares que mais tiveram aumento de casos pós-pandemia e que podem ser identificadas por meio da radiografia do tórax:

- Tuberculose: Doença infecciosa causada pela bactéria Mycobacterium tuberculosis, caracterizada principalmente pelo acometimento do sistema respiratório, embora possa atingir outros órgãos. Inicia-se quando o bacilo é inalado e alcança os alvéolos pulmonares, onde é fagocitado por macrófagos, mas consegue sobreviver e se multiplicar no interior dessas células. Assim, esse mecanismo permite a disseminação da infecção e o estabelecimento de uma resposta inflamatória local, com formação de granulomas, estruturas típicas da doença que tentam conter a proliferação bacteriana. Após a infecção inicial, o organismo desencadeia uma resposta imunológica que pode controlar a doença ou permitir sua progressão. Nos pulmões, essa resposta leva à formação de áreas de inflamação, necrose caseosa e destruição do parênquima pulmonar.

  Essas alterações podem ser observadas em exames de imagem na forma de opacidades, não tão frequentemente com padrão nodular ou reticulonodular (intersticial), além de consolidações e, de maneira bastante característica, cavitações, resultantes da necrose do tecido pulmonar. Essas lesões predominam, em geral, nos lobos superiores e refletem a evolução do processo infeccioso.

- Pneumonia: Caracterizada como uma infecção que atinge o parênquima pulmonar, sendo compreendida não apenas como uma doença isolada, mas como um conjunto de síndromes que variam conforme o agente causador e o estado de saúde do paciente. Elas podem ser causadas por diversos patógenos, incluindo causas virais, bacterianas ou até fúngicas. O processo se inicia quando ocorre um desequilíbrio entre os microrganismos que tentam colonizar o trato respiratório e as defesas

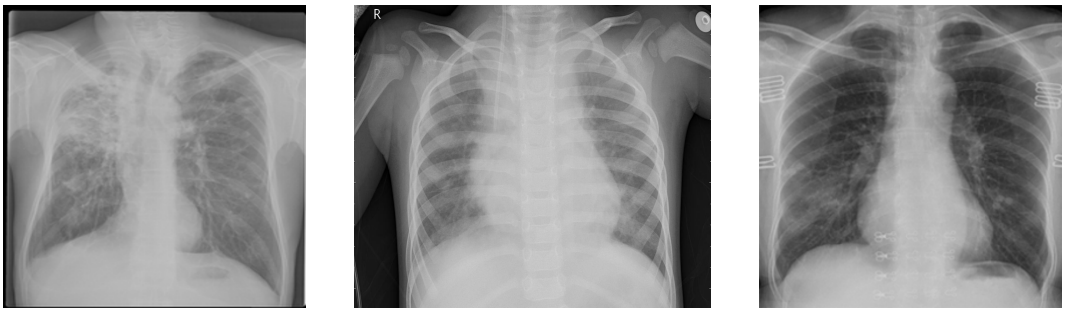
naturais do corpo. Quando essas barreiras falham, os agentes infecciosos atingem os alvéolos e desencadeiam uma resposta inflamatória intensa.

É nessa reação do sistema imunológico que torna os capilares sanguíneos mais permeáveis, permitindo que fluidos, proteínas e células de defesa preencham os espaços aéreos, o que gera as opacidades e consolidações vistas em exames de imagem e dificulta a troca de oxigênio.

- Covid-19: Doença infecciosa causada pelo vírus SARS-CoV-2, caracterizada principalmente pelo acometimento do sistema respiratório, mas com potencial de afetar múltiplos órgãos. O processo é iniciado quando o vírus entra no organismo e se liga aos receptores presentes nas células, especialmente no epitélio pulmonar, permitindo sua entrada e replicação. Após a infecção celular, o organismo desencadeia uma resposta imunológica que pode variar de leve a intensa. Em muitos casos, essa resposta é responsável pelos principais danos teciduais, especialmente nos pulmões, onde ocorre destruição das células alveolares, inflamação difusa e prejuízo na troca gasosa.

  Esse processo leva ao acúmulo de líquido e células inflamatórias nos alvéolos e no interstício pulmonar, o que pode ser observado em exames de imagem, como radiografias, na forma de opacidades, consolidações bilaterais e padrão intersticial.

Figura 6 – Exemplos de radiografias de pneumonia, Covid-19 e tuberculose.



(a) Tuberculose (b) Pneumonia (c) Covid-19

Fonte: Elaborado pelo autor.

## 2.3 INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial (IA) é uma área da ciência, que por muito tempo já foi estudada principalmente por pesquisadores de ciências humanas e naturais, mas que atualmente se localiza mais em computação. O seu foco é a criação de sistemas capazes de realizar tarefas que normalmente utilizam inteligência humana.

Nesse sentido, a IA, embora seja uma área recente — com o termo cunhado em 1956 por John McCarthy e outros pesquisadores no Dartmouth College —, herda questões muito mais antigas. Há mais de 2000 anos, filósofos e cientistas já investigavam processos como o ver, aprender, recordar e raciocinar (RUSSELL; NORVIG, 2020). Aristóteles, por exemplo, explorou a lógica como base do pensamento no século IV a.C., e no século XIX, pioneiros da psicologia experimental, como Wilhelm Wundt, estudaram a estrutura da mente. Até mesmo descobertas neurofisiológicas, como as de Camillo Golgi sobre os neurônios em 1875, ajudaram a entender a base biológica da cognição — tema que hoje dialoga diretamente com a IA. Seu marco moderno, afinal, consolida-se com trabalhos como os de McCulloch e Pitts, que nos anos 1940 propuseram modelos de neurônios artificiais, e com Alan Turing, que em 1950 formulou seu célebre teste para avaliar a inteligência das máquinas (RUSSELL; NORVIG, 2020).

Além disso, é necessário compreender, também, a complexidade da definição de IA, já que existem diversas respostas para a pergunta "O que é Inteligência Artificial?". Contudo, pode-se dividi-las em quatro grandes linhas de pensamento ao longo do tempo, que diferenciam o foco no processo mental ou no comportamental, e na imitação humana ou na racionalidade ideal:

I. Sistemas que pensam como seres humanos:

"A automatização das atividades que associamos com o pensamento humano, atividades tais como tomada de decisões, resolução de problemas, aprendizado,..." (BELLMAN, 1978).

II. Sistemas que atuam como seres humanos:

“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (KURZWEIL, 1990).

III. Sistemas que pensam racionalmente:

“O estudo das faculdades mentais pelo seu uso de modelos computacionais.” (CHARNIAK; MCDERMOTT, 1985).

IV. Sistemas que atuam racionalmente:

"A Inteligência Computacional é o estudo do projeto de agentes inteligentes." (POOLE; MACKWORTH; GOEBEL, 1998).

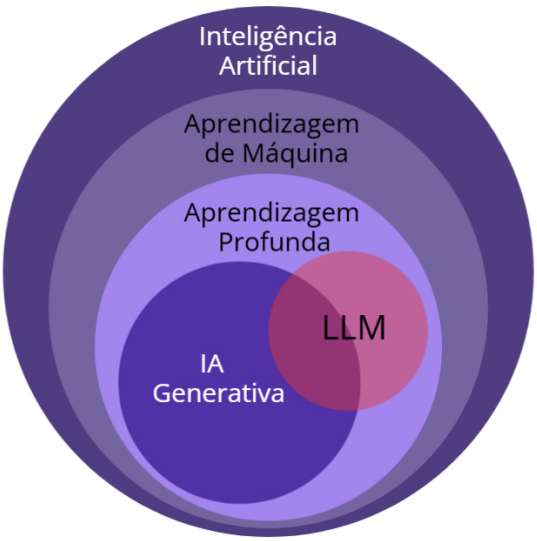
As duas primeiras linhas medem o sucesso em termos de fidelidade ao desempenho humano, enquanto as duas últimas medem a racionalidade e se baseiam na otimização das decisões. Outrossim, as linhas de pensamento I e III referem-se ao processo de pensamento e raciocínio, já as II e IV ao comportamento. Enquanto os processos centrados no homem são baseados em ciência empírica, envolvendo confirmações experimentais e hipóteses, a abordagem racional está fundamentada em uma combinação de matemática e engenharia (RUSSELL; NORVIG, 2020).

## 2.4 APRENDIZAGEM DE MÁQUINA

A aprendizagem de máquina ou *Machine Learning* (ML) é uma área da inteligência artificial que permite que computadores aprendam a partir de dados, sem serem explicitamente programados para realizar tarefas específicas. Dessa forma, a aprendizagem se dá pelo ajuste iterativo dos modelos em busca de alcançar um nível aceitável de desempenho. Com o aumento da complexidade dos problemas computacionais e do volume de dados gerados, tornou-se essencial desenvolver métodos que criem hipóteses ou modelos capazes de resolver problemas com mínima intervenção humana (FACELI et al., 2011).

A inteligência artificial, pela crescente evolução acentuada nos últimos anos, tornou-se um nome muito amplo. Dessa forma, dividiu-se em diversos subcampos, tendo como principal a aprendizagem de máquina. Como ilustrado na figura 7, nele está inserida a aprendizagem profunda ou *Deep Learning* (DL), que, por sua vez, inclui outras duas subáreas que intersectam entre si: a *Large Language Model* (LLM) ou Modelo de Linguagem Grande, que tem o intuito de referir-se ao tipo de modelo de inteligência artificial que utiliza aprendizado de máquina para entender e gerar linguagem natural, e a inteligência artificial generativa, cujo foco é na criação de um novo conteúdo, como texto, imagem, áudio e vídeo, com base nos dados fornecidos.

Figura 7 – Diagrama Inteligência Artificial




Fonte: Adaptado de (MEHTA et al., 2024).

### 2.4.1 Tipos de aprendizado

Existem diferentes tipos de algoritmos de ML, podendo ser divididos em 5 grandes grupos:

- Aprendizado supervisionado: No aprendizado supervisionado, o modelo é treinado para aprender a função que mapeia uma entrada para uma saída, baseada em amostras de outros pares de entradas e saídas de *datasets* rotulados (TAMAAZOUSTI; BORGNE; HUDELOT, 2021). Dessa forma, o processo de aprendizagem desse paradigma orientado a tarefas é com base em comparar a saída calculada e a saída projetada; assim, referindo-se a computar o erro e ajustá-lo até um valor aceitável (PANESAR et al., 2021). Portanto, o uso mais comum são tarefas de classificação e regressão com o objetivo de separar e ajustar os dados, respectivamente.

- Aprendizado não supervisionado: De forma oposta ao aprendizado anterior, o não supervisionado não usa *datasets* rotulados, ou seja, sem nenhuma interferência humana nos dados. Assim, desconsidera padrões pré-determinados, o que, para

alguns problemas, é uma vantagem. De maneira geral, é usado amplamente na análise exploratória de dados, reconhecimento de padrões e modelagem preditiva (IBM, 2021). Dessa forma, o exemplo mais comum é a identificação automática de amigos para um usuário em uma rede social (PANESAR et al., 2021).

- Aprendizado semi-supervisionado: É uma técnica que junta ambos os aprendizados supervisionado e não supervisionado (IBM, 2021). Já que emprega dados rotulados e não rotulados, isto é, são parcialmente classificados. O objetivo final é conseguir resultados melhores do que apenas usar o paradigma supervisionado com os mesmos dados rotulados, ou seja, quando os dados não têm informação suficiente para criar uma norma geral (HAAS, 2020). Como exemplo de aplicação, tem-se a verificação de fraude, rotulagem de *clusters*, tradução automática e classificação de textos (PANESAR et al., 2021).

- Aprendizado por reforço: Também chamado de Aprendizado por reforço com feedback humano ou *Reinforcement learning from human feedback* (RLHF) é um tipo de programação dinâmica que treina os algoritmos usando um sistema que avalia o comportamento, gerando recompensa ou penalidade. A cada etapa, um algoritmo de reforço, também conhecido como "agente", atua e prevê as características em uma etapa futura com base em características passadas e presentes, e uma recompensa ou penalidade é atribuída com base na previsão (PANESAR et al., 2021).

  Desse jeito, com uma série de repetições, o agente aprende as melhores estratégias. Portanto, é uma ferramenta poderosa para treinar modelos de IA que ajudam a aumentar a autonomia ou otimizar a eficiência operacional de sistemas sofisticados, assim sendo recomendada para carros autônomos, robótica e desenvolvimento de jogos; portanto, não sendo interessante aplicar para a resolução de problemas básicos ou diretos (TAMAAZOUSTI; BORGNE; HUDELOT, 2021).

- Aprendizado autossupervisionado: Em inglês, *Self-supervised learning* (SSL) é uma aprendizagem que permite que os modelos se treinem sozinhos em dados não rotulados. Ao aproveitar a estrutura inerente dos próprios dados de entrada, os algoritmos SSL, também conhecidos como métodos de aprendizagem preditiva ou de pretexto, geram rótulos de forma autônoma, aprendendo a prever uma parte da

entrada a partir de outra. Assim, converte de forma eficiente problemas de aprendizagem não supervisionada em aprendizagem supervisionada, criando estruturas de treinamento escaláveis e eficientes. O SSL tem se mostrado particularmente útil em domínios com uso intensivo de dados, como visão computacional e processamento de linguagem natural (PLN) (IBM, 2021).

Figura 8 – Ilustração das diferentes estratégias de aprendizado - Os círculos vermelho e azul representam dados rotulados de diferentes classes. O círculo cinza representa dados não rotulados. As linhas pretas definem o limite de decisão subjacente entre classes. Os círculos riscados representam dados que ignoram e usam as informações do rótulo em diferentes estágios. No diagrama mais a direita, o Agente (triângulo) interage com o ambiente (círculo), explorando (linha tracejada) e otimizando ações para maximizar recompensas (estrelas).




Fonte: Adaptado de (DASHA, 2022).

### 2.4.2 Redes Neurais Artificiais

As Redes Neurais Artificiais, conhecidas como *Artificial Neural Networks* (ANNs), foram inspiradas no funcionamento do cérebro humano. O primeiro modelo criado foi o *Perceptron*, desenvolvido com o objetivo de simular o processo da visão humana e a forma como o cérebro reconhece padrões (CHICHANOSKI, 2022).

Uma rede neural artificial é composta por neurônios artificiais, que são organizados em camadas: uma camada de entrada, uma ou mais camadas ocultas (ou intermediárias) e uma camada de saída.

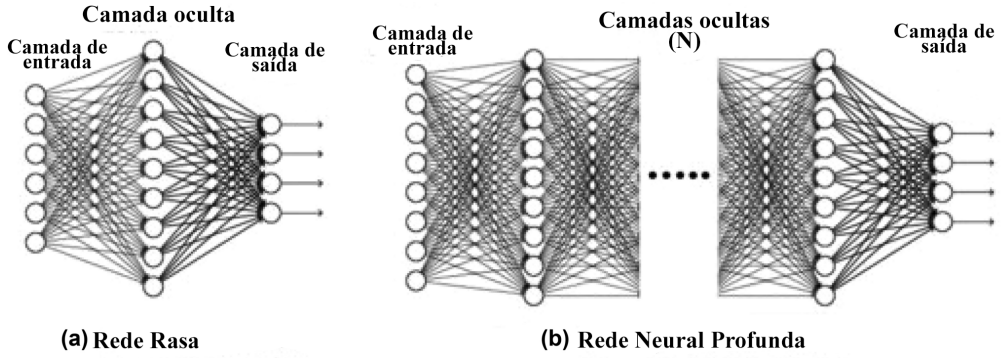
Cada neurônio recebe valores de entrada, realiza operações matemáticas (como soma ponderada), e gera um valor de saída. Quando todos os neurônios de uma camada estão conectados a todos os neurônios da próxima camada, essa estrutura é considerada uma camada densa (ou totalmente conectada).

Para melhorar a flexibilidade do modelo, adiciona-se a cada neurônio um parâmetro extra chamado *bias* (ou viés), que é somado à combinação linear das entradas ponderadas pelos pesos. Diferentemente dos pesos, o *bias* não está associado a nenhuma entrada específica, funcionando como um termo constante que permite ao modelo ajustar o resultado mesmo quando todas as entradas forem zero. Dessa forma, o *bias* desloca a função de ativação e amplia a capacidade da rede de aprender representações mais complexas.

## 2.5 APRENDIZAGEM PROFUNDA

Como mostrado anteriormente, a aprendizagem profunda pertence como subárea à aprendizagem de máquina. Ela utiliza redes neurais profundas, *Deep Neural Networks* (DNN) para a resolução de problemas complexos, aprendendo automaticamente padrões a partir de grandes volumes de dados (GOODFELLOW; BENGIO; COURVILLE, 2016). Com isso, comumente, a principal diferença entre ML e DL é a estrutura da arquitetura da ANN, relacionada à profundidade. Não existe uma definição exata que diga quantas camadas ocultas são necessárias para que o modelo possa ser classificado como DL (CHICHANOSKI, 2022). Contudo, de maneira geral, considera-se no mínimo duas dessas camadas e sem máximo, sendo comuns modelos com mais de 100 camadas. A figura 9 exemplifica como ambas as arquiteturas são. Além disso, frequentemente utiliza-se da especialização das técnicas da rede, em conjunto com a quantidade de camadas, como forma de diferenciar ambos os campos de aprendizado. Assim, o uso de blocos especializados para extrair padrões hierárquicos automaticamente, como o uso de convolução e recorrência, configura um modelo como DL (GOODFELLOW; BENGIO; COURVILLE, 2016).

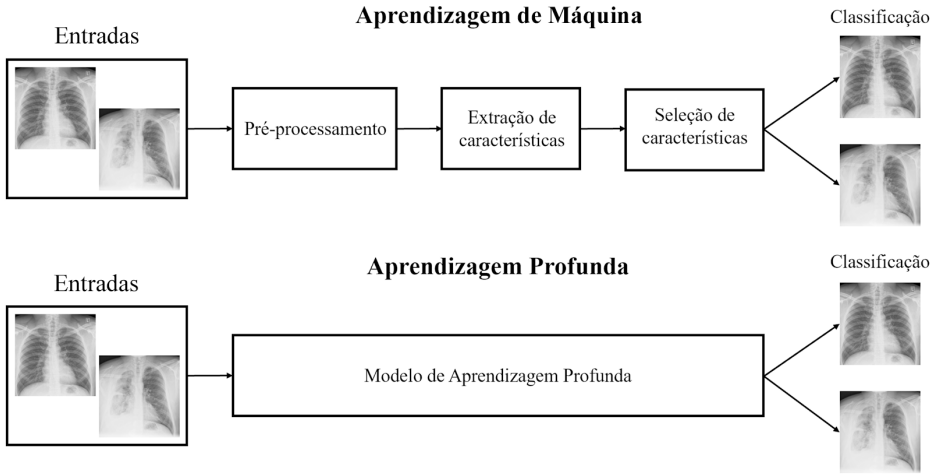
Figura 9 – Arquitetura de (a) rede rasa e (b) rede neural profunda.




Fonte: Adaptado de (SARKER, 2021).

Nesse sentido, diferentemente das abordagens convencionais de aprendizado de máquina, que requerem etapas sequenciais como pré-processamento, extração e seleção criteriosa de características, os algoritmos de DL automatizam o aprendizado de representações diretamente a partir dos dados. Isso permite que o processo de aprendizado e classificação seja realizado de forma integrada, sem necessidade de intervenção manual nas fases intermediárias (Figura 10) . No entanto, essa automatização vem acompanhada de um desafio: os modelos de DL são altamente complexos e muitas vezes funcionam como "caixas-pretas", dificultando a interpretação de seus mecanismos internos. Apesar disso, os benefícios em termos de precisão e escalabilidade geralmente superam essa limitação, especialmente em tarefas que envolvem grandes volumes de dados e alta variabilidade. Consequentemente, outro diferencial é a dependência de extração manual de características (como cor ou posição de objetos em imagens) e algoritmos como árvores de decisão; o *Deep Learning* elimina essa etapa, aprendendo diretamente de dados brutos (imagens, textos, áudios), trabalhando com nível de abstração maior (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 10 – As diferenças entre etapas de ML e DL.




Fonte: Adaptado de (ALZUBAIDI et al., 2021), conforme adaptado por (COLAÇO, 2022).

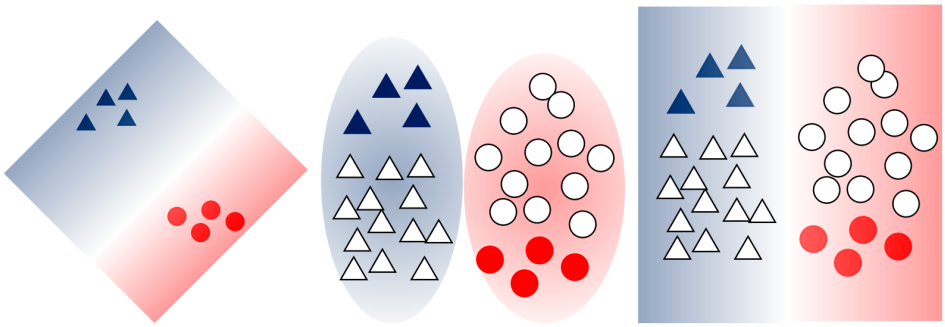
As camadas que constituem as camadas ocultas na Figura 10 podem ser CNNs, RNNs, GANs, entre outras, que podem também seguir diferentes tipos de arquitetura, como generativa, discriminativa ou híbrida (combinações de diferentes tipos de redes).

- Generativa: Essa técnica aprende a distribuição conjunta das variáveis observadas e da variável-alvo, com o objetivo de gerar novos dados que imitam os padrões presentes nos dados de treinamento, sendo uma das abordagens mais promissoras do aprendizado de máquina (BISHOP, 2006). Essa abordagem utiliza redes adversárias generativas, ou *Generative Adversarial Networks* (GANs), *autoencoders* variacionais, ou *Variational Autoencoders* (VAEs) e redes invertíveis. Os modelos generativos podem ser aplicados a diversas tarefas relacionadas à classificação, como aprendizado semi-supervisionado, detecção de *outliers* e calibração de confiança, uma vez que permitem estimar a distribuição dos dados de entrada (HAYASHI, 2023). Em adição, novos modelos obtém mais sucesso em tarefas abertas que requerem mais criatividade e versatilidade (NGUYEN; TULABANDHULA; WATSON-MANHEIM, 2023). Dessa forma, entre os principais casos de uso estão a geração de arte, sintetização de música e aumento de dados (*data augmentation*).
- Discriminativa: Modelos discriminativos são baseados em redes neurais

supervisionadas e focam na modelagem da probabilidade condicional de um resultado dado um determinado contexto (como um vetor de características) (NGUYEN; TULABANDHULA; WATSON-MANHEIM, 2023). O objetivo é minimizar os erros de classificação ou maximizar a probabilidade de previsões de classe precisas. Essa abordagem emprega em redes como CNNs tradicionais e *Transformers* técnicas como retro propagação (*backpropagation*), otimização por gradiente descendente e regularização para ajustar os limites de decisão e alcançar uma precisão de classificação ideal (BISHOP, 2006). Historicamente, modelos discriminativos têm sido amplamente utilizados em tarefas de classificação e regressão em diversas áreas, como finanças, saúde, setor automotivo, entre outras.

- Híbrida: Em suma, essa abordagem combina as características dos modelos generativos e discriminativos. Ou seja, ela permite estimar a distribuição dos dados de entrada (como nos modelos generativos), ao mesmo tempo em que mantém uma alta capacidade de classificação (característica dos modelos discriminativos). Essa configuração é especialmente vantajosa em cenários de aprendizado semi-supervisionado (HAYASHI, 2023). Um exemplo de rede híbrida é a *Deep Convolutional Generative Adversarial Network* (DCGAN), que junta CNN e GAN. Ela aprende e gera estruturas e padrões espaciais em imagens, resultando em imagens de maior qualidade.

Figura 11 – Ilustração dos diferentes tipos de redes - Azul e vermelho representam classes, e amostras preenchidas com cada cor indicam dados rotulados, e amostras brancas indicam dados não rotulados. A cor de fundo representa o valor de probabilidade posterior ou a distribuição condicional de classe obtida por cada modelo. O modelo híbrido integra o poder discriminativo do modelo discriminativo e a disponibilidade de dados não supervisionados do modelo generativo.



(a) Discriminativa (b) Generativa (c) Híbrida

Fonte: Adaptado de (HAYASHI, 2023)

.

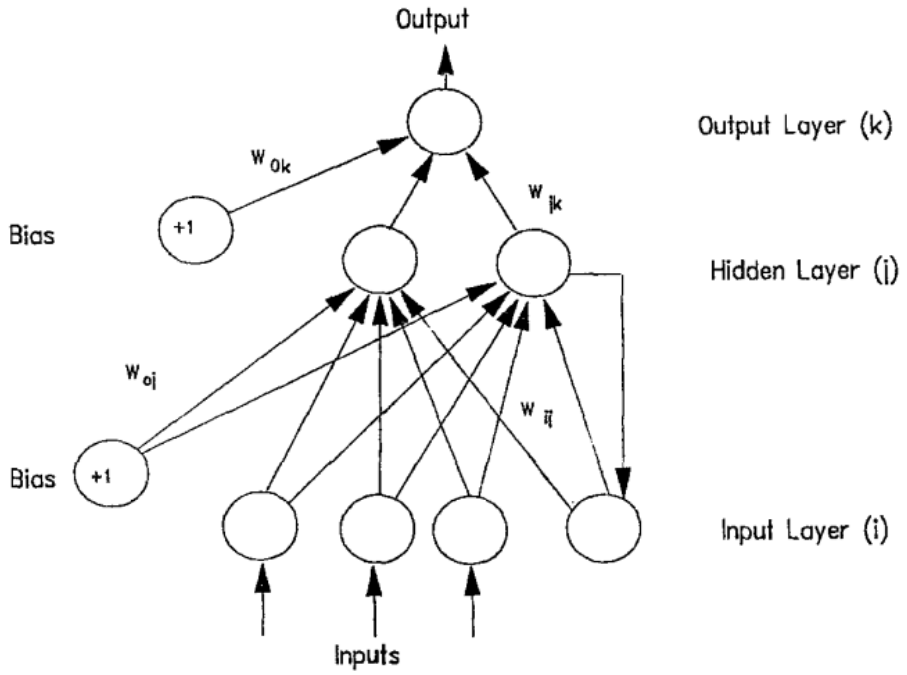
Em questão de fluxo de dados ainda há três principais classificações em redes neurais; são elas:

- Rede Neural *Feed-Forward*: Como ilustra na Figura 9, essa rede consiste numa camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada é composta por um número diferente de neurônios. Dessa forma, os que compõe a primeira camada são variáveis independentes no modelo. Por outro lado os que estão na última camada são variáveis dependentes. Já a camada do meio representa a interação entre as variáveis de entrada. Assim, atividades dos neurônios em cada camada são uma função não-linear das atividades na camada anterior.

- Rede Recorrente: Essa arquitetura é bem similar à rede *Feed-Forward* com a principal diferença sendo que ela tem a camada de saída, oculta ou ambas retroalimentando a camada de entrada. Ou seja, inclui qualquer rede que o design em que a atividade deve passar pela rede mais de uma vez antes que os pesos (coeficientes de saída) sejam ajustados e a saída seja produzida (DEMATOS et al., 1996).

Dessa forma, forma-se ciclos por meio de conexões recorrentes em que as informações podem voltar a etapas anteriores, como ilustra a Figura 12. Essa característica torna possível modelar sequências e memórias temporais, mas também introduz complexidade à dinâmica do treinamento. Como resultado, essas redes podem ser mais difíceis de otimizar (GOODFELLOW; BENGIO; COURVILLE, 2016). No entanto, sua estrutura é considerada mais próxima do funcionamento do cérebro humano, tornando-as mais biologicamente realistas. Um exemplo clássico dessa arquitetura é a Rede Neural Recorrente (RNN), amplamente utilizada em tarefas como processamento de linguagem natural e reconhecimento de fala, onde a ordem dos dados importa.

Figura 12 – Rede Recorrente.




Fonte: (DEMATOS et al., 1996).

- Rede Conectada Simetricamente: Semelhante a rede recorrente, com a única diferença sendo a simetria funcional presente nesta rede, ou seja, os pesos das conexões são iguais nas duas direções. Com isso, se tornam mais fáceis de analisar e mais restritas. Existem ainda dois tipos de classificação: as redes simétricas sem

camada oculta que são as "Redes de Hopfield"e com camada oculta que são as "Máquinas de Boltzmann".

### 2.5.1 Rede Neural Convolucional

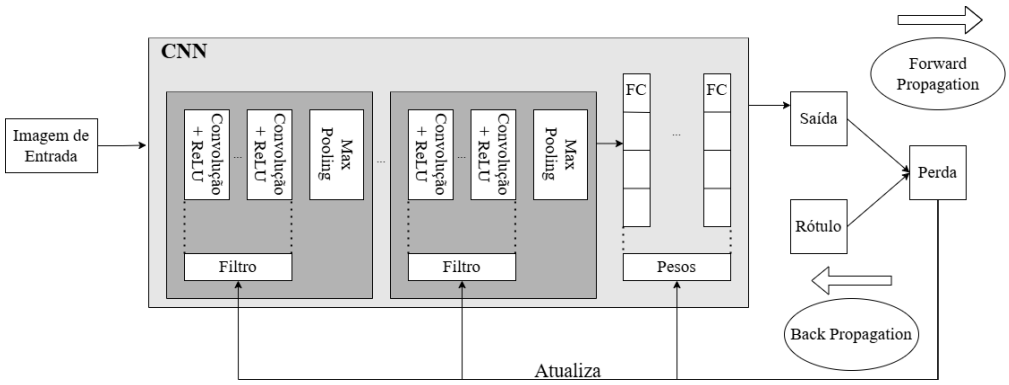
As redes neurais convolucionais, *Convolutional Neural Networks* (CNNs), ou ainda ConvNet, são uma classe de ANNs do tipo *feed-forward*. Inspiradas na organização do córtex visual dos animais, as CNNs replicam o modo como neurônios corticais respondem a estímulos em regiões específicas do campo visual, denominadas campos receptivos, permitindo reconhecimento de padrões locais (como bordas, texturas ou formas) de forma hierárquica, que, vindos de diferentes neurônios, cobrem o campo de visão. Dessa forma, em conjunto com seu ótimo desempenho, são mais utilizadas em aplicações que pertencem ao domínio de visão computacional, como classificação de imagens e detecção de objetos.

A maior vantagem da CNN, que faz com que esse modelo seja considerado superior a outros clássicos pertencentes à arquitetura *feed-forward* é o compartilhamento de pesos. Isso leva a uma redução de parâmetros que precisam ser treinados. Dessa forma, ajuda a desenvolver modelos mais complexos, aprimorar a generalização e sem acontecer *overfitting* - fenômeno em que o modelo se adapta excessivamente aos dados de treinamento, capturando até ruídos e variações irrelevantes, o que prejudica sua performance em dados não vistos. Outros benefícios são: a etapa de classificação ser incorporada à etapa de extração de características, ambas por meio do processo de aprendizado, deixando a saída mais organizada e dependente dos recursos extraídos; e ser mais fácil de implementar grandes redes usando modelos de CNN do que outros modelos gerais de ANN.

#### *A Modelo Geral*

O modelo de CNN, de maneira geral, possui 4 componentes: camada de convolução, camada de *pooling* (na Figura 13 representada pelo *Max Pooling*), função de ativação, camada totalmente conectada, ou *fully connected* (FC) *layers*. A arquitetura típica é composta por uma pilha de camadas de convolução e *pooling* seguida por uma ou mais camadas totalmente conectadas.

Figura 13 – Arquitetura CNN




Fonte: Adaptado de (YAMASHITA et al., 2018).

A imagem de entrada em redes neurais tradicionais tem o formato de um vetor unidimensional. Contudo, nas CNNs, a entrada do modelo é geralmente uma imagem tridimensional, representada por altura, largura e número de canais. Por exemplo, imagens em escala de cinza possuem apenas um canal, enquanto imagens RGB possuem três. Assim, a entrada de cada camada é um tensor tridimensional com dimensão $(m \times m \times r)$, onde $m$ representa a altura e largura (normalmente iguais) e $r$ representa a profundidade (número de canais). As camadas convolucionais aplicam filtros (ou *kernels*) denotados por $k$ e que também são tridimensionais $(n \times n \times q)$, onde $n < m$ e $q <= r$. Cada filtro percorre a imagem computando produtos escalares locais (convoluções) que geram um mapa de características (*feature map*). Em seguida, uma função de ativação não linear (como ReLU) é aplicada, introduzindo não linearidades no modelo, transformando todos os valores negativos em zero e mantendo os valores positivos inalterados (GOODFELLOW; BENGIO; COURVILLE, 2016). Após a extração de características, utiliza-se camadas de *pooling* (subamostragem) comumente usando filtros de dimensão $(p \times p)$. Essas camadas reduzem a dimensionalidade espacial dos mapas de características, diminuindo o número de parâmetros da rede e, consequentemente, o custo computacional.

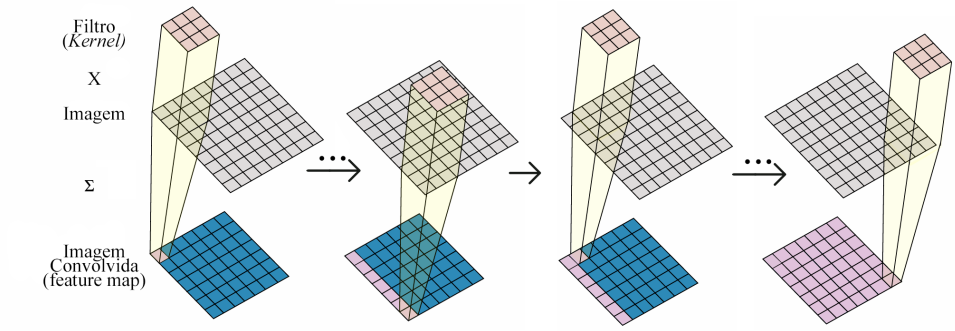
Finalmente, as FCs interpretam as características extraídas e realizam a classificação final com base nas ativações das camadas anteriores .

## *B Camada de Convolução*

A camada de convolução é o primeiro bloco de construção de uma CNN. Como o nome sugere, a principal tarefa matemática realizada é chamada de convolução, que consiste na representação matemática de como um sistema linear opera sobre um sinal (GOODFELLOW; BENGIO; COURVILLE, 2016). Historicamente, a convolução já era muito aplicada em processamento de sinais e imagens, desde filtros passa-baixo, circuito eletrônico que permite a passagem de baixas frequências, ao filtro gaussiano, sendo usada para remoção ou aplicação de bordas, redução de brilho e outras tarefas que auxiliavam no tratamento de imagens (CHICHANOSKI, 2022).

Dessa forma, nessa camada, através dessas operações lineares, ou convoluções, são extraídas características dos dados. Cada camada é constituída por vários filtros de tamanho igual, em que cada um dos seus valores representa um peso do *kernel*. Assim, cada filtro é responsável por reconhecer um padrão específico na imagem, como cor, bordas, formas dos objetos e texturas. No início do processo, os pesos desses filtros são inicializados aleatoriamente e, a cada época de treinamento, são ajustados para aprimorar a extração das características mais relevantes dos dados de entrada. No primeiro passo da operação, o *kernel* passa sobre toda a imagem horizontal e verticalmente, obtendo o mapa de características. Além disso, o produto escalar entre a imagem de entrada e o *kernel* é determinado, onde seus valores correspondentes são multiplicados e somados para criar um único valor escalar, calculado simultaneamente. Então, todo esse processo é repetido até que não seja possível fazer mais nenhuma varredura na matriz (Figura 14).

Figura 14 – Diagrama de convolução




Fonte: Adaptado de (CARIN et al., 2019).

A parte processada pelo filtro da entrada é denominada de campo receptivo local (*local receptive field*). Assim, a combinação dos pixels desse campo resulta em um pixel como valor de saída (YAMASHITA et al., 2018). Contudo, todos os campos são filtrados com os mesmos pesos locais para todo o pixel, no caso do exemplo da Figura 14, definindo os filtros de tamanho $k \times k \times d$, em que $k$ é a dimensão espacial do filtro, sendo igual a 3; e $d$ a dimensão de profundidade, que depende da entrada da camada, que também é igual a 3, já que possui 3 canais (RGB), logo os pesos resultam em 27. Dessa forma, obtém-se uma combinação linear de 9 pixels da vizinhança para gerar um único pixel de saída.

Além disso, é adotado um passo, ou *stride*, de um. O *stride* é a distância em pixels que o filtro se desloca a cada movimento, tanto na horizontal quanto na vertical. Um *stride* de 1 (S=1) significa que o filtro desliza um pixel de cada vez. Isso resulta em uma grande sobreposição entre os campos receptivos vizinhos. Já um *stride* de 2 (S=2) significa que o filtro pula dois pixels a cada movimento, o que resulta em uma menor sobreposição e, assim, um mapa de características menor.

Isso, portanto, torna a camada convolucional diferente da camada FC, em que cada neurônio se liga a todos os neurônios da camada seguinte. Já que apenas alguns pesos estão disponíveis entre duas camadas adjacentes. Dessa forma, o uso de filtros nas convolucionais significa que cada ponto no mapa de características resultante é calculado usando apenas uma pequena região local da imagem de entrada (o campo receptivo), e não a imagem inteira.

Além disso, o princípio do compartilhamento de pesos garante que o mesmo filtro, com os mesmos valores (pesos), seja aplicado repetidamente em todas as regiões da imagem. Com essas características juntas (localidade e compartilhamento de pesos), há uma grande redução no número total de parâmetros que a rede precisa aprender durante o treinamento, tornando-a mais leve e rápida de treinar, ou seja, eficiente (COLAÇO, 2022).

### *C Pooling*

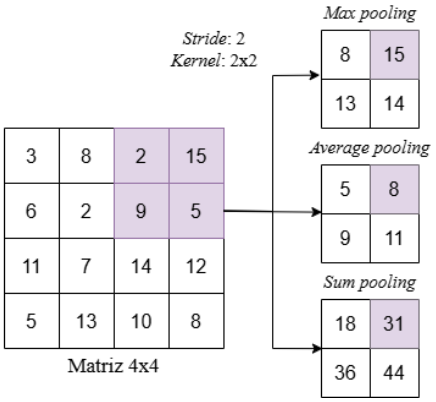
A operação de *pooling* tem o objetivo de reduzir as dimensões espaciais (altura e largura) dos mapas ao longo das camadas da rede. Dessa forma, há uma necessária redução do custo computacional, já que compensa o aumento comum da profundidade

dos tensores, *d*, ao longo das camadas (PONTI; COSTA, 2017). Além do mais, também permite um tipo de composição de banco de filtros (conjunto de detectores de características) multirresolução que processa imagens em diferentes espaços-escala e ajuda na mitigação de *overfitting* (GOODFELLOW; BENGIO; COURVILLE, 2016).

Existem diferentes operações de *pooling*, os mais comuns são o *max pooling*, utilizado no exemplo da Figura 13, o *average pooling* e o *sum pooling*. A operação de *max pooling* tem como saída o maior valor do conjunto, útil para casos em que se necessita extrair características como bordas ou pontos, por exemplo. O *average pooling* refere-se ao que a saída é a média dos valores do conjunto, melhor para manter o contexto espacial geral (GOODFELLOW; BENGIO; COURVILLE, 2016). Já o *sum pooling* apresenta, como saída, a soma dos valores do *feature map* (PONTI; COSTA, 2017), o que ajuda em casos de quantificação total de ativação de características em determinada área. A Figura 15 demonstra esses tipos. Além desses, há outros métodos como: *min pooling*, *global average pooling* e *global max pooling* (COLAÇO, 2022).

Essas camadas, de forma semelhante às camadas convolucionais, empregam um mecanismo de varredura que utiliza um *kernel* e um *stride* previamente definidos. Portanto, pode-se dizer que elas 'convoluem' apenas no sentido de como percorrem a entrada, mas a operação matemática interna é distinta. Ademais, é possível ao invés de utilizar o *pooling*, aumentar o *stride* nas convoluções, o que substitui como redutor (PONTI; COSTA, 2017).

Figura 15 – Diagrama de exemplo de Pooling




Fonte: Adaptado de (COLAÇO, 2022).

### *D Camadas fully connected (FC)*

As camadas totalmente conectadas, ou camadas FC, em CNNs estão localizadas após numerosas camadas convolucionais, no final do modelo. A saída do mapa de características, tipicamente proveniente da última camada de *pooling*, é reduzida, transformada em um vetor unidimensional e conectada a uma ou mais camadas FC, em que toda entrada é conectada a toda saída com um peso associado a cada elemento, ou neurônio, do vetor de entrada (YAMASHITA et al., 2018) (PONTI; COSTA, 2017).

Dessa forma, se, por exemplo, a camada convolucional anterior a uma camada FC gera um tensor $5 \times 5 \times 3$, ele é redimensionado para que possua tamanho $1 \times (5 \times 5 \times 3) = 1 \times 75$. Portanto, cada elemento terá 75 pesos usados para calcular a combinação linear do vetor. A última dessas camadas determina a saída final da rede, com a probabilidade para cada classe nas tarefas de classificação.

### *E Função de Ativação*

Após cada operação linear, como convolução, é aplicada uma função de ativação. Essa função ajuda a rede a aprender relações não lineares entre as características e as imagens, o que faz com que a rede seja mais robusta para a identificação de diferentes padrões.

A entrada de uma função de ativação é o resultado da soma ponderada entre as entradas do próprio neurônio e os *bias*. Após receber esse valor, a função determina se ativa ou não esse neurônio. Além disso, há a necessidade dessas funções serem diferenciáveis, pois, caso não tenha essa propriedade, o treinamento por *backpropagation* não seria possível.

Dessa forma, para cada demanda, uma função de ativação apropriada precisa ser escolhida. Por exemplo, para a tarefa de multiclassificação, é utilizada a função *softmax* na camada final. Essa função normaliza os valores reais da saída da camada FC até as probabilidades da classe do objetivo, em que cada valor varia entre 0 e 1, e todos os valores somados são iguais a 1. Entretanto, para as camadas ocultas, utiliza-se a função retificadora linear, ou *rectified linear function* (ReLU), como denotado na Figura 13.

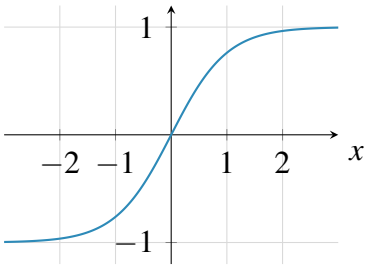
A ReLU (Equação 2.1) é útil para introduzir não linearidade, acelerar o

treinamento e evitar o problema do gradiente desaparecendo, ou *vanishing gradients*. Ela ativa apenas neurônios com entradas positivas, não precisando de operações mais complexas como multiplicação, divisão e potência, o que torna-a uma rede esparsa e eficiente. Contudo, por retificar os valores inferiores a zero, há a possibilidade de acontecer o inverso do *vanishing gradients*, o *exploding gradients*, caso os valores dos pesos sejam suficientemente grandes (COLAÇO, 2022).

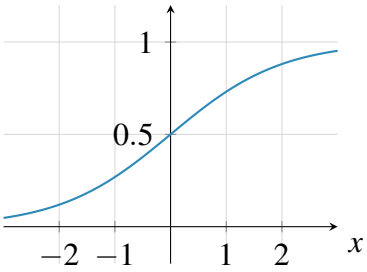
$$f(x) = \max(0,x) = \begin{cases} x_i, & x_i \geq 0 \\ 0, & x_i < 0 \end{cases} \tag{2.1}$$

Além da função ReLU, há ainda outras frequentemente utilizadas em modelos de redes neurais, como mostra a Figura 16. A (a) é a tangente hiperbólica que mapeia qualquer valor de entrada em um intervalo entre [-1, 1], o que ajuda a centralizar os dados; são utilizadas mais em redes recorrentes. Já a (b) é a sigmoide, ou também conhecida como logística, que admite valores que estão dentro do intervalo [0, 1], funciona de forma semelhante à função anterior, porém não admite valores negativos. Ambas as funções sofrem com o problema do sumiço de gradiente. Por fim, tem-se a (d) que é PreLU, a ReLU paramétrica, uma variação que permite valores negativos com menor ponderação, parametrizada com uma variável $0 \leq a \leq 1$. Uma vantagem particular dessa versão é a possibilidade de aprender $a$ durante o treinamento (PONTI; COSTA, 2017).

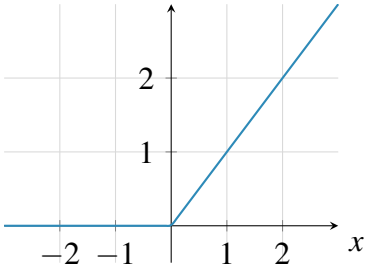
Figura 16 – Tipos de funções de ativação: (a) Tangente Hiperbólica, (b) Sigmoide (Logística), (c) ReLU e (d) PreLU.



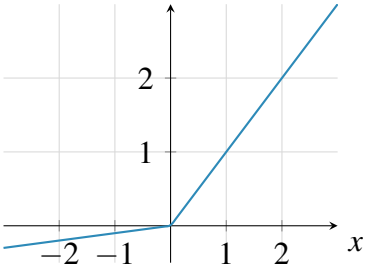

(a) Tangente Hiperbólica: $tanh(x)$




(b) Sigmoide (Logística): $logistic(x)$




(c) ReLU: $\max[0,x]$




(d) PreLU: $\max[ax,x], a = 0.1$

Fonte: Adaptado de (PONTI; COSTA, 2017).

## 2.6 MEDIDAS DE DESEMPENHO

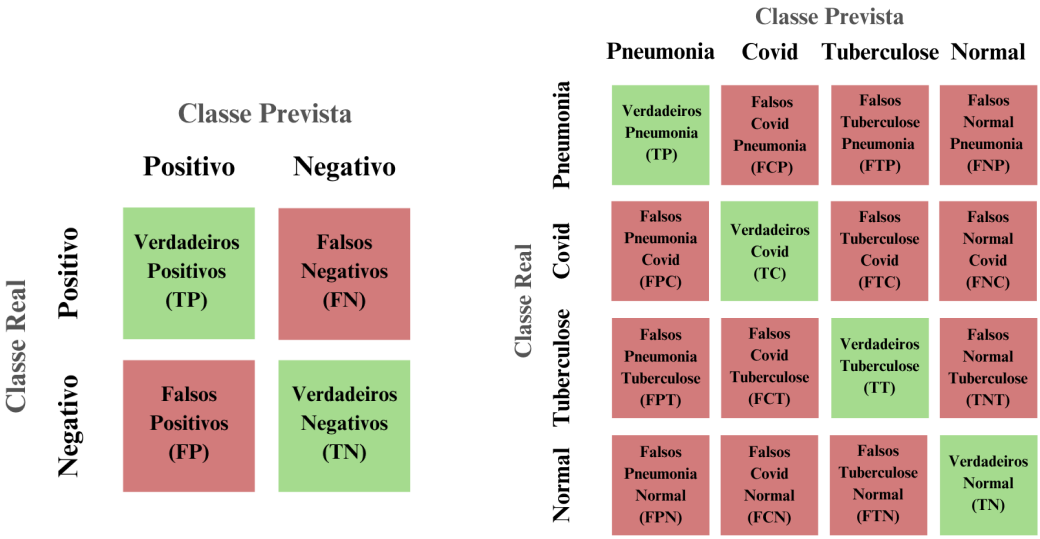
Após a execução de um modelo de *Machine Learning* é necessário aplicar métricas que avaliem seu desempenho. Essas métricas são essenciais para entender a eficácia do modelo em realizar a tarefa para a qual foi treinado. A escolha das métricas apropriadas depende do tipo de problema (classificação, regressão, etc.) e dos objetivos específicos do projeto, nenhuma métrica é universalmente aplicável (STEHMAN, 1997). Antes de entrar em detalhes sobre as métricas, é importante entender a matriz de confusão, ou tabela de confusão, ferramenta fundamental de avaliação. Ela é uma tabela que descreve o desempenho de um modelo de classificação, mostrando as previsões feitas pelo modelo em comparação com os valores reais. Dessa forma, ela é composta por quatro elementos principais:

- **Verdadeiros Positivos (TP)**: Casos em que o modelo previu corretamente a classe positiva.
- **Falsos Positivos (FP)**: Casos em que o modelo previu a classe positiva, mas o valor real é negativo.
- **Verdadeiros Negativos (TN)**: Casos em que o modelo previu corretamente a classe negativa.
- **Falsos Negativos (FN)**: Casos em que o modelo previu a classe negativa, mas o valor real é positivo.

Na Figura 17 é possível observar a estrutura da matriz de confusão binária, onde as linhas representam as classes reais e as colunas representam as classes previstas pelo modelo, com cada quadrante representando cada combinação descrita anteriormente.

Além disso, há ainda os casos de matrizes de confusão para problemas de classificação multiclasse, onde as classes reais e previstas podem ser mais de duas, resultando em uma matriz maior, mas o conceito fundamental permanece o mesmo. Numa matriz $n \times n$, nesse caso $n = 4$, para qualquer classe $i$, os elementos que estão na diagonal $TP_i$ representam as previsões corretas, enquanto os elementos fora da diagonal representam os erros de classificação, incluindo falsos positivos $FP_i$ que estão na coluna fora da diagonal e falsos negativos $FN_i$ que se encontram na linha fora da diagonal.

Figura 17 – Matriz de confusão




(a) Classificação binária (b) Classificação multiclasse

Fonte: Adaptado de (COLAÇO, 2022).

Com as medidas definidas, outras métricas gerais podem ser calculadas a partir desses valores, como:

- **Acurácia (ACC)**: Mede a proporção de previsões corretas em relação ao total de previsões feitas. É calculada como:

$$\text{Acurácia} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Acurácia Multiclasse} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FN_i)}$$

  A acurácia é uma métrica útil quando as classes estão balanceadas, mas pode ser enganosa em casos de classes desbalanceadas. Dessa forma, ela sozinha não pode fornecer uma visão holística do desempenho do modelo.

- **Precisão (*Precision*)**: Mede a proporção de verdadeiros positivos em relação ao total de positivos previstos pelo modelo. É calculada como:

$$\text{Precisão} = \frac{TP}{TP+FP}$$

$$\text{Precisão Multiclasse} = \frac{TP_i}{TP_i + FP_i}$$

A precisão é importante quando o custo de falsos positivos é alto.

- **Revocação (*Recall*)**: Conhecido também como taxa de sensibilidade ou taxa de verdadeiros positivos (TPR). Mede a proporção de verdadeiros positivos em relação ao total de positivos reais. É calculada como:

$$Recall = \frac{TP}{TP + FN}$$

$$Recall\text{Multiclasse} = \frac{TP_i}{TP_i + FN_i}$$

A revocação é crucial quando o custo de falsos negativos é alto, essencial para aplicações em que a detecção de positivos é crítica, como em diagnósticos médicos.

- ***F1-Score***: É a média harmônica entre precisão e revocação, fornecendo a junção entre as duas métricas com o mesmo peso cada. É calculada como:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i}$$

O F1-Score é útil quando se deseja um equilíbrio entre precisão e revocação, especialmente em casos de classes desbalanceadas.

- **Especificidade (*Specificity*)**: Também conhecido como taxa de verdadeiros negativos (TNV). Mede a proporção de verdadeiros negativos em relação ao total de negativos reais. É calculada como:

$$\text{Especificidade} = \frac{TN}{TN + FP}$$

$$\text{Especificidade Multiclasse} = \frac{TN_i}{TN_i + FP_i}$$

A especificidade é importante quando o custo de falsos positivos é alto, ajudando a entender a capacidade do modelo em identificar corretamente os negativos.

- **Taxa de Falsos Positivos (FPR)**: Mede a proporção de falsos positivos em relação ao total de negativos reais. É calculada como:

$$\text{FPR} = \frac{FP}{FP + TN}$$

$$\text{FPR Multiclasse} = \frac{FP_i}{FP_i + TN_i}$$

A FPR é crucial em situações onde o custo de falsos positivos é elevado, como em testes de triagem médica.

- **Área sob a Curva ROC (AUC-ROC)**: Mede a capacidade do modelo de distinguir entre classes positivas e negativas. A curva ROC (*Receiver Operating Characteristic*) é um gráfico que mostra a taxa de verdadeiros positivos contra a taxa de falsos positivos em diferentes limiares de decisão. A AUC (*Area Under the Curve*) varia de 0 a 1, onde 1 indica um modelo perfeito e 0,5 indica um modelo que não tem capacidade discriminativa. Caso, ainda, ocorra uma curva negativa, significa que a performance do classificador não é aceitável. A AUC-ROC é calculada como a integral:

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(t)\, d\text{FPR}(t)$$

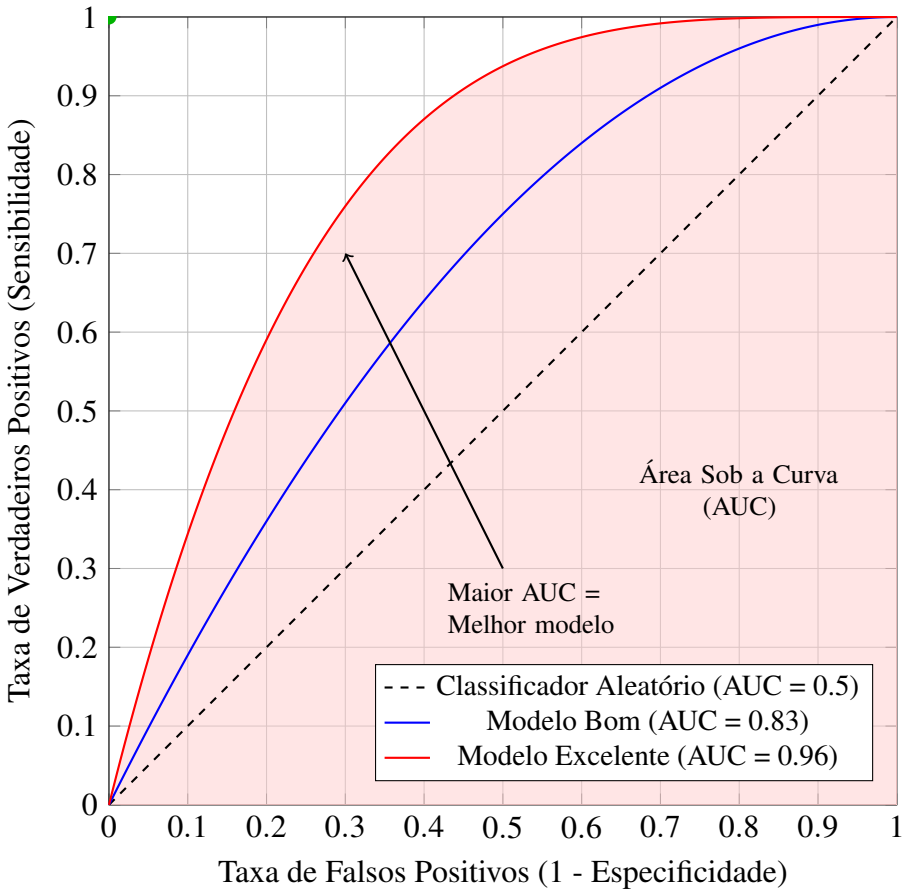
A AUC-ROC é especialmente útil para avaliar modelos em problemas de classificação binária. Ele fornece uma visão geral do desempenho do modelo em todos os limiares possíveis, sendo bastante apropriado para *datasets* tanto balanceados quanto desbalanceados (CATAL, 2012).

- ***Log Loss* (*Logarithmic Loss*)**: Conhecido também como perda logarítmica, mede a incerteza das previsões feitas pelo modelo, penalizando previsões incorretas com maior severidade. É especialmente útil para problemas de classificação probabilística, onde o modelo retorna uma probabilidade para cada classe. O Log Loss é calculado como:

$$\textit{Log Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

onde $N$ é o número total de amostras, $y_i$ é o valor real (0 ou 1) e $p_i$ é a probabilidade prevista para a classe positiva. A saída é um valor de erro entre 0 e 1, onde

Figura 18 – Exemplo de Curva ROC com diferentes modelos




Fonte: Autor.

valores mais próximos de 0 indicam menor erro e assim melhores previsões. Contudo, como os valores da verdade absoluta para as previsões corretas são exatamente “0” ou “1”, o *Log Loss* se torna particularmente útil para avaliar a qualidade das previsões probabilísticas, sendo sensível a erros de classificação, especialmente quando o modelo está confiante em suas previsões incorretas, já que cresce exponencialmente. Dessa forma, reforçando previsões que não são apenas corretas, mas também confiáveis (BERGMANN; STRYKER, 2025).

## 2.7 SEGMENTAÇÃO DE IMAGENS

A segmentação de imagens é uma técnica de processamento muito útil para aplicações de imagens médicas, automatizando, facilitando o delineamento de estruturas anatômicas e as regiões de maior interesse, ou *Regions of Interest* (ROI). O ROI pode ser os pulmões em sua totalidade ou alguma parte deles, como lóbulos, brônquios pulmonares ou ainda regiões infectadas ou lesionadas. Essa delimitação aumenta a eficiência e reduz a complexidade do modelo de IA, já que extrai de maneira rápida a informação útil para a tarefa-alvo e consegue uma representação simples e suficiente dos dados, tornando-se uma etapa essencial de processamento.

Dessa forma, a segmentação de imagem tem como definição o processo de partição de uma imagem digital em várias partes, ou seja, em conjuntos de píxeis. Cada uma dessas regiões é não sobreposta e homogênea referente a uma certa característica, como, por exemplo, intensidade, textura ou cor . Como resultado, obtém-se uma imagem em que o objeto é diferenciado do fundo ou restante da imagem, com píxeis da mesma região (objeto) apresentando propriedades semelhantes entre si, ao mesmo tempo que diferem dos píxeis de outras regiões (fundo), atingindo o objetivo principal desse processo.

Além disso, são diversos os algoritmos para delimitação do pulmão em imagens de raio-x. Dentre os métodos tradicionais apresentados na literatura inclui-se: *thresholding*, *region growing* e contornos ativos (COLAÇO, 2022). Contudo, pelo avanço de DL recentemente, os mais utilizados são os que se baseiam em redes neurais para segmentar uma imagem, o que causou uma mudança significativa no paradigma dessa área.

Dessa forma, de maneira mais frequente, o resultado gerado são máscaras de valor de saída binária da imagem de entrada da rede neural (Figura 19). Assim, pode ser possível a multiplicação das imagens de entrada e saída, isolando a área de interesse do restante.

## 2.8 ARQUITETURA *U-NET*

A arquitetura *U-Net* representa um marco significativo no desenvolvimento de redes neurais convolucionais voltadas para tarefas de segmentação semântica, tendo

Figura 19 – Exemplos de segmentação: (a) Imagem do pulmão em raio-X, (b) Máscara e (c) Máscara aplicada.




(a) Imagem do pulmão em raio-X

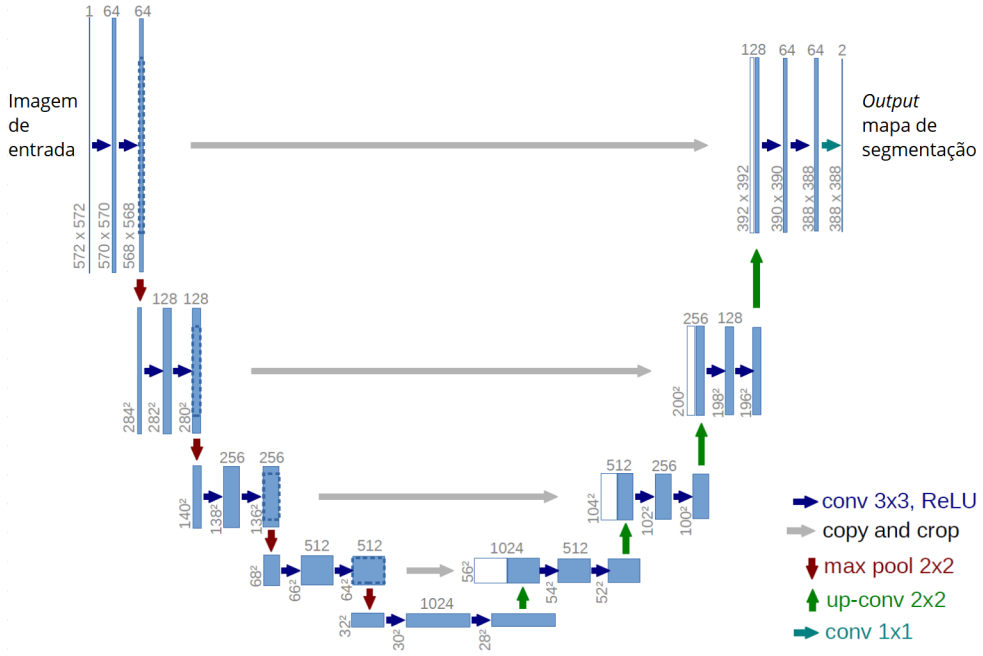
(b) Máscara

(c) Máscara aplicada

Fonte: Autor.

sido originalmente proposta por (RONNEBERGER; FISCHER; BROX, 2015) para segmentação de imagens biomédicas. Esta arquitetura recebeu sua denominação devido à sua estrutura característica em forma de "U"como mostra a Figura 20, que combina um caminho de contração (*encoder*) com um caminho de expansão (*decoder*), permitindo a captura de informações contextuais e a localização precisa de características relevantes.

Figura 20 – Arquitetura U-Net (exemplo para 32 × 32 pixels na resolução mais baixa)




Fonte: Adaptado de (RONNEBERGER; FISCHER; BROX, 2015).

A concepção da *U-Net* surgiu da necessidade de superar as limitações das arquiteturas tradicionais de segmentação, particularmente em aplicações médicas onde a precisão na delimitação de estruturas anatômicas é crucial. Diferentemente das redes neurais convolucionais convencionais focadas em classificação, a *U-Net* foi especificamente projetada para produzir segmentações densas, onde cada pixel da imagem de entrada recebe uma classificação específica. Esta capacidade de segmentação *pixel-wise* tornou a arquitetura especialmente adequada para aplicações em processamento de imagens médicas, onde a identificação precisa de regiões de interesse é fundamental para o diagnóstico.

A estrutura da *U-Net* é composta por duas trajetórias principais: o caminho de contração (*contracting path*) no lado direito e o caminho de expansão (*expansive path*) no lado esquerdo. O caminho de contração segue a arquitetura típica de uma rede neural convolucional, consistindo em aplicações repetidas de duas convoluções $3 \times 3$ (demonstrado pelas setas azuis na Figura 20), seguidas por funções de ativação ReLU em cada uma e operações de *max pooling* $2 \times 2$ com *stride* de 2 para *downsampling* (demonstrado pelas setas vermelhas na Figura 20). Contudo, sem as camadas FC, o que a torna uma rede totalmente convolucional (FCN). Esta sequência permite a captura de características em múltiplas escalas e a redução progressiva da resolução espacial, concentrando-se na extração de informações contextuais relevantes (RONNEBERGER; FISCHER; BROX, 2015).

O caminho de expansão, por sua vez, utiliza convoluções transpostas (*transpose convolutions*) para restaurar progressivamente a resolução espacial da representação, permitindo a localização precisa das características detectadas através da camada de *upsampling* (demonstrado pelas setas verdes na Figura 20), com um passo de $2 \times 2$, dobrando o número de neurônios da camada anterior. Uma característica distintiva da *U-Net* é a presença de conexões de salto (*skip connections*) que conectam diretamente camadas correspondentes entre os caminhos de contração e expansão (demonstrado pelas setas cinzas na Figura 20). Estas conexões permitem a transferência de informações de alta resolução do *encoder* para o *decoder*, preservando detalhes espaciais que poderiam ser perdidos durante o processo de redução dimensional.

As conexões *skip* combinam mapas de características de diferentes níveis de resolução através de operações de concatenação. Esta estratégia permite que a rede mantenha tanto informações contextuais globais quanto detalhes locais precisos,

resultando em segmentações mais acuradas e na preservação de bordas finas. A combinação dessas informações em múltiplas escalas torna a *U-Net* particularmente eficaz na identificação de estruturas complexas com variações significativas em tamanho e forma.

A versatilidade da arquitetura *U-Net* tem levado ao desenvolvimento de numerosas variações e melhorias, incluindo *U-Net++*, *Attention U-Net*, *CMUNeXt* e *Dense U-Net*, cada uma introduzindo refinamentos específicos para diferentes aplicações. A *U-Net++*, por exemplo, incorpora conexões *skip* mais densas e caminhos de decodificação aninhados, enquanto a *CMUNeXt* integra o bloco *CMUNeXt* com uma estrutura de convoluções diferentes.

## 2.9 ARQUITETURA *CMUNEXT*

A arquitetura CMUNeXt representa uma evolução recente das redes neurais convolucionais aplicadas à segmentação de imagens, proposta com o objetivo de melhorar a eficiência computacional e a capacidade de extração de características em múltiplas escalas. Inspirada na estrutura da U-Net, a CMUNeXt mantém o paradigma encoder–decoder, mas introduz modificações significativas nos blocos convolucionais e nos mecanismos de processamento de características.

A principal motivação para o desenvolvimento da *CMUNeXt* está na limitação das arquiteturas tradicionais, que frequentemente apresentam alto custo computacional e redundância na extração de características. Nesse contexto, a *CMUNeXt* propõe uma abordagem mais leve e eficiente, mantendo ou até superando o desempenho em tarefas de segmentação, especialmente em cenários com restrições de recursos computacionais.

A arquitetura é composta por um *encoder* responsável pela extração hierárquica de características e um *decoder* que reconstrói a resolução espacial da imagem. Assim como na *U-Net*, conexões de salto (*skip connections*) são utilizadas para preservar informações espaciais detalhadas. No entanto, a principal inovação dessa arquitetura é implementação de blocos *CMUNeXt*, que substituem as convoluções tradicionais. Esse tipo de convolução possui um campo receptivo local, característica que dificulta a captura de informações globais da imagem, essenciais para localizar corretamente regiões de interesse.

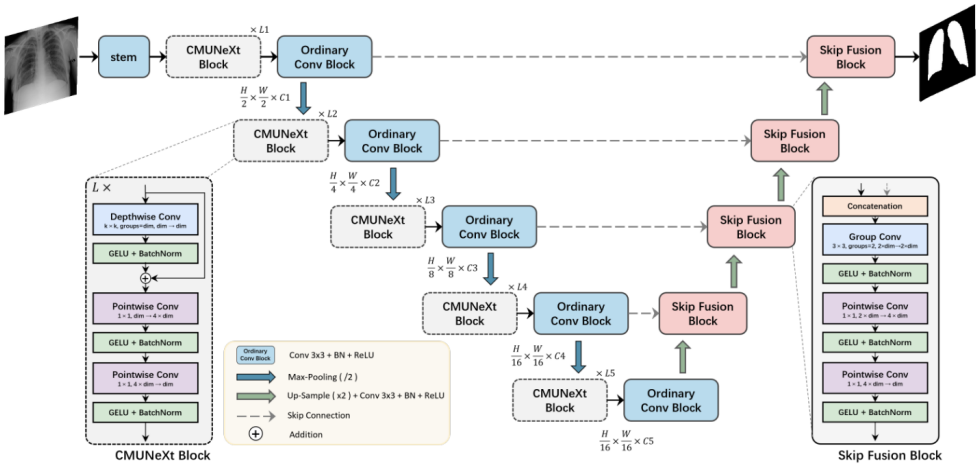
Os blocos *CMUNeXt* são projetados para capturar informações em múltiplas

escalas de forma eficiente, com foco na captura do contexto global. Diferentemente de abordagens que utilizam múltiplos tamanhos de kernel em paralelo, esses blocos combinam convoluções com diferentes tamanhos de kernel para ampliar o campo receptivo. Essa estrutura é baseada em convolução separável em profundidade (*depthwise separable convolution*) que decompõe a convolução tradicional em duas etapas, a primeira é a convolução em profundidade (*depthwise*), responsáveis pela extração de características espaciais em cada canal (número de canais iguais aos grupos) e a segunda é a convolução pontual (*pointwise*, $1 \times 1$), que combina as informações entre canais. Dessa forma, esses blocos são otimizados para reduzir a redundância e o número de parâmetros.

Além disso, nesse bloco é implementado o *inverted bottleneck*, no qual a dimensão interna das características é expandida antes de ser reduzida novamente, permitindo uma melhor captura de informações contextuais, o que melhora a mistura de informações espaciais e de canal. Essa combinação de técnicas torna a *CMUNeXt* mais eficiente na extração de características, capturando padrões locais e relações globais, melhorando a eficiência computacional.

Na figura 21 é possível observar a estrutura geral da *CMUNeXt*, destacando os blocos CMUNeXt e as conexões de salto. A arquitetura é projetada para ser flexível, permitindo ajustes no número de blocos e na profundidade da rede, o que a torna adequada para uma variedade de tarefas de segmentação em diferentes domínios.

Figura 21 – Arquitetura CMUNeXt




Fonte: Adaptado de (TANG et al., 2023).

# 3 ESTADO DA ARTE

A aplicação de técnicas de Inteligência Artificial, e mais especificamente de DL, tem redefinido os paradigmas no campo da imagiologia médica. A análise de radiografias de tórax (CXR), um dos exames de imagem mais comuns e acessíveis globalmente, representa uma área de imenso potencial para o desenvolvimento de Sistemas de Apoio ao Diagnóstico, ou *Computer-Aided Diagnosis* (CAD). Já que, apesar de sua centralidade, a interpretação de radiografias torácicas é uma tarefa inerentemente complexa e suscetível a erros, mesmo para radiologistas experientes (NASSER; AKHLOUFI, 2023). A natureza bidimensional da imagem resulta na sobreposição de múltiplas estruturas anatômicas, como o coração, os pulmões, a vasculatura e o esqueleto torácico, o que pode obscurecer ou mimetizar achados patológicos. Lesões sutis, como nódulos incipientes ou infiltrações discretas, podem ser facilmente negligenciadas. Adicionalmente, fatores humanos como a fadiga do especialista, a variabilidade interobservador e o enorme volume de exames a serem analisados diariamente em ambientes clínicos sobrecarregados aumentam significativamente o risco de diagnósticos incorretos ou tardios (COLAÇO, 2022).

A capacidade dos algoritmos de *Deep Learning* de aprender automaticamente representações hierárquicas e características complexas a partir de dados brutos, como pixels de uma imagem, os distingue fundamentalmente das abordagens de ML tradicionais. Enquanto os métodos clássicos demandam uma etapa manual e intensiva de extração de características (*handcrafted features*), que depende do conhecimento de especialistas para definir descritores de textura, forma ou intensidade, os modelos de DL automatizam esse processo, permitindo a descoberta de padrões sutis que podem ser imperceptíveis ao olho humano. Esta automação não apenas otimiza o fluxo de trabalho, mas também abre a porta para um novo patamar de precisão e eficiência diagnóstica, como destacado por diversas pesquisas na área (ALINSAIF, 2024).

O desenvolvimento de um sistema CAD robusto para análise de radiografias de tórax geralmente se desdobra em etapas sequenciais ou tarefas interdependentes. Uma das etapas de pré-processamento mais cruciais e frequentemente citadas na literatura é a segmentação pulmonar. O objetivo desta tarefa é isolar os campos pulmonares do restante da imagem, que inclui estruturas como o coração, a caixa torácica e o diafragma. Arquiteturas CNNs do tipo codificador-decodificador, com a *U-Net* sendo o exemplo mais

proeminente, tornaram-se o padrão-ouro para esta tarefa. A estrutura da U-Net, permite reconstruir uma máscara de segmentação precisa, preservando os detalhes espaciais das bordas pulmonares, conforme demonstrado nos trabalhos de (COLAÇO, 2022) e (CHICHANOSKI, 2022).

## 3.1 CLASSIFICAÇÃO DAS IMAGENS DE RAIO-X

Após o pré-processamento e a segmentação, a tarefa central é a classificação das imagens, que pode ser abordada de diferentes maneiras. A literatura descreve desde a classificação binária, como a diferenciação entre radiografias normais e patológicas, até a classificação multiclasse, onde o objetivo é identificar patologias específicas como pneumonia, câncer pulmonar, tuberculose e COVID-19. A complexidade do problema de classificação direciona a escolha das arquiteturas e das estratégias de modelagem.

A evolução das metodologias para esta tarefa de classificação é notável. As primeiras abordagens, fundamentadas no ML tradicional, utilizavam extratores de características como *Local Binary Patterns* (LBP) ou *Histogram of Oriented Gradients* (HOG) para quantificar os padrões da imagem, que eram então fornecidos a classificadores como *Support Vector Machines* (SVM) ou *Random Forests*. Embora eficazes até certo ponto, a performance desses sistemas era intrinsecamente limitada pela qualidade e relevância das características projetadas manualmente.

O advento do DL com as CNNs marcou um ponto de inflexão. Modelos seminais como a *AlexNet* e a *VGG* foram os primeiros a demonstrar a eficácia avassaladora de empilhar camadas convolucionais para tarefas de reconhecimento visual em larga escala, estabelecendo as bases para a subsequente explosão de aplicações de DL em visão computacional. Além das arquiteturais já citadas, outras fundamentais como a *ResNet* e *DenseNet*, pré-treinadas em vastos conjuntos de dados de imagens naturais como o *ImageNet* e depois ajustadas (*fine-tuned*) para a tarefa médica, rapidamente superaram os métodos tradicionais. Esses modelos demonstraram uma capacidade superior de aprender representações de características relevantes diretamente dos dados radiográficos. Pesquisas mais recentes, como a de (RAVI; ACHARYA; ALAZAB, 2023) buscaram otimizar ainda mais esse desempenho, explorando arquiteturas mais eficientes, como a família *EfficientNet*. Esses modelos utilizam um escalonamento composto para balancear a profundidade, a largura e a resolução da rede de forma ótima, alcançando alta acurácia

com menor custo computacional, uma consideração importante para a aplicação clínica.

Para lidar com a sutileza e a sobreposição de características entre diferentes patologias pulmonares, estratégias mais avançadas foram propostas, movendo-se além do uso de um único modelo. As técnicas de Aprendizagem por Comitês (*Ensemble Learning*) surgiram como uma abordagem poderosa para aumentar a robustez e a precisão. Métodos como *voting*, onde a decisão final é tomada pela maioria das previsões de múltiplos modelos, ou *stacking*, onde um meta-modelo aprende a combinar de forma inteligente as previsões de vários modelos base, têm mostrado resultados promissores. Por exemplo, a combinação de diferentes modelos da família *EfficientNet*, cada um treinado para focar em aspectos distintos da imagem, pode criar um sistema classificador mais resiliente e preciso do que qualquer modelo individualmente, uma direção explorada por trabalhos como o de (SHARMA; MITTAL; GOYAL, 2022). Esta abordagem, que utiliza múltiplos especialistas em paralelo, alinha-se diretamente com a proposta de um fluxo de trabalho diagnóstico hierárquico.

## 3.2 POSICIONAMENTO DA PESQUISA PROPOSTA

Em síntese, o estado da arte na análise de radiografias de tórax evoluiu de métodos dependentes de extração manual de características para sofisticadas arquiteturas de DL. A literatura demonstra uma clara trajetória de progresso: da segmentação como um passo de otimização com modelos como a *U-Net*, passando pelo uso de CNNs cada vez mais eficientes, até a exploração de abordagens híbridas e de comitês (*ensembles*) para maximizar a acurácia.

No entanto, uma lacuna identificada na literatura é a exploração de um modelo híbrido e hierárquico que combine essas tarefas de forma explícita e sequencial, simulando um fluxo de raciocínio diagnóstico. Muitas pesquisas focam ou na classificação binária (normal vs. patológico) ou na classificação multiclasse direta, mas poucos trabalhos propõem uma arquitetura em cascata: um primeiro estágio que atua como um triador robusto para detectar a presença de qualquer anomalia e, somente para os casos positivos, um segundo estágio composto por um comitê de especialistas para determinar a patologia específica entre um conjunto de possibilidades (como tuberculose, COVID-19 e pneumonia). É precisamente nesta lacuna que o presente trabalho se insere, propondo o desenvolvimento e a avaliação de um modelo com essa estrutura hierárquica,

que integra segmentação, uma classificação binária inicial e uma classificação multiclasse subsequente por meio de um comitê de CNNs, buscando um sistema de diagnóstico mais alinhado com a prática clínica e potencialmente mais preciso.

# 4 METODOLOGIA

Neste capítulo, é apresentada a metodologia adotada para a execução desse projeto. Portanto, são apresentados os materiais, bases de dados, modelos e métodos que foram aplicados para as tarefas que integram o sistema de análise de raios-X do tórax.

## 4.1 MATERIAIS E DADOS

### 4.1.1 *Hardware* e ferramentas de suporte

Para o desenvolvimento deste trabalho, foram utilizadas diversas ferramentas que se pudessem concluir os objetivos propostos. Em questão de linguagem, foi selecionada *Python* para o desenvolvimento de algoritmos, scripts, preparação e tratamento de dados. Essa linguagem, além de ser apropriada à manipulação de dados, possui um grande leque de bibliotecas para ML e DL. Portanto, como biblioteca, recorreu-se ao *PyTorch* para o desenvolvimento dos algoritmos de treinamento e construção dos modelos propostos. Além disso, foi utilizada a *scikit-learn*, biblioteca para a implementação de métricas e avaliação dos modelos. Já no quesito de *hardware*, as configurações da máquina usada para este projeto estão na Tabela 1. Contudo, como o treinamento dos modelos propostos demanda mais recursos de placa de vídeo, recorreu-se a ambientes virtuais como o notebook do *Google Colab* e *Vlab* disponibilizado pela UFSC, ambos em *Jupyter*, interface de renderização.

Tabela 1 – Configurações do *hardware* utilizado para o desenvolvimento do trabalho.

| **Detalhes** | **Máquina** |
|---|---|
| Sistema Operativo | Windows 11 |
| CPU | Intel Core i5-11400H @ 2.70GHz |
| GPU | NVIDIA GeForce RTX 3050 4GB |
| RAM | 24GB |

Fonte: Autor.

### 4.1.2 Bases de dados

Neste trabalho, a base de dados utilizada para o treinamento, teste e validação da rede neural é resultado da composição de outras bases de dados. Os critérios para a escolha foram a grande frequência na utilização na literatura, bom volume de imagens, rotulação precisa das mesmas e, principalmente, a disponibilidade pública.

Dessa forma, a composição se trata de três bases diferentes referentes a classes diferentes, ou seja, uma base para pneumonia, outra para covid-19 e mais uma para tuberculose, em que de cada uma foram retiradas imagens para compor a classe de normal.

As imagens de raio-X relativas à doença de tuberculose foram disponibilizadas em (KIRAN; JABEEN, 2024) e contém 2494 imagens de pacientes com essa doença e 514 na classe normal, coletadas num hospital local do Paquistão. A resolução delas é de $256 \times 256$ pixels.

No caso da classe de COVID-19, a base de dados foi fornecida pelos mesmos que a base anterior (RAHMAN et al., 2020)(RAHMAN et al., 2021), porém com fonte em diversas publicações e associações, incluindo a *Italian Society of Medical and Interventional Radiology* (SIRM), e disponibilizadas em (CHOWDHURY et al., 2020). Contém 3.616 raios-X classificados como COVID-19 e 10.192 como sem patologia e resolução de $256 \times 256$.

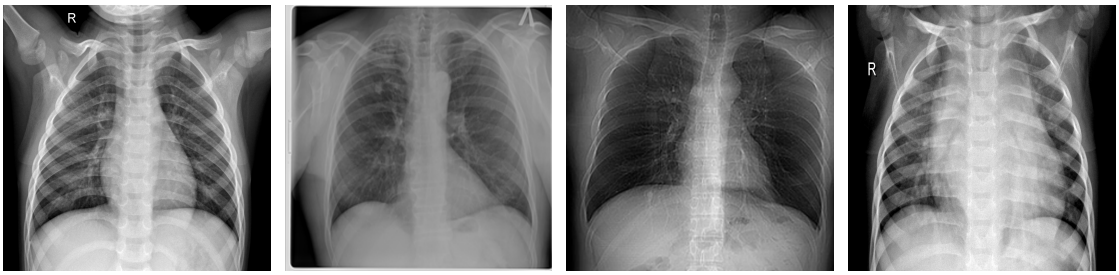
Além disso, para a pneumonia, foi encontrado um *dataset* com fonte em *Guangzhou Women and Children's Medical Center*, China, fornecido pela Universidade de San Diego, Califórnia, e disponibilizado em (KERMANY; ZHANG; GOLDBAUM, 2018). Contudo, os dados foram acessados através da plataforma Kaggle [1], que hospeda uma cópia do *dataset* original. São 5.863 radiografias, nas quais estão separadas em 3 diferentes pastas: val, train, test; ainda, em cada pasta há outra separação: Normal, Pneumonia. As dimensões variam, sendo mais comum $1048 \times 604$.

Por fim, foi utilizado um dataset de tuberculose fornecidas publicamente pela Universidade do Qatar em conjunto com a Universidade de Dhaka, Bangladesh, foram disponibilizadas em (RAHMAN et al., 2020) e provenientes de publicações no *National Library of Medicine* (NIH), Estados Unidos. É composta por 4.200 radiografias torácicas, nas quais 700 delas são classificadas como diagnosticadas com tuberculose, e o restante,

[1] https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data

3.500, sem patologia pulmonar. A resolução delas é de $512 \times 512$ pixels. Essas imagens não foram utilizadas para compor o dataset usado neste trabalho, mas sim exclusivo para a etapa de teste, a fim de avaliar a capacidade de generalização do modelo proposto, uma vez que elas não foram vistas durante o treinamento e validação, além das imagens serem visivelmente dissemelhantes de quaisquer outras utilizadas.

Figura 22 – Exemplos de radiografias para as classes: Normal, Tuberculose, COVID-19 e Pneumonia.



Normal Tuberculose COVID-19 Pneumonia

Fonte: Autor.

Para a seleção das imagens, utilizou-se um script (descrito no algoritmo 1) em que se tinha como parâmetros a pasta de origem do *dataset*, a pasta de destino e a quantidade desejada. Assim, esse código, de forma aleatória, seleciona o número antes pedido de imagens, faz uma cópia delas, separa em treino, validação e teste na proporção de 70%, 15% e 15%, respectivamente, e encaminha-as à pasta de destino. Como resultado, obteve-se a Tabela 2. O resultado final da composição foi disponibilizado num repositório público do Zenodo em (OLIVEIRA; SILVA, 2026), para que outros pesquisadores possam utilizá-lo e, principalmente, para que seja possível a reprodução dos resultados apresentados neste trabalho. A estrutura de diretórios do *dataset* resultante é apresentada no Apêndice A.

Tabela 2 – Distribuição dos casos nos *datasets*.

| **Classe** | **Total** | **Treino (70%)** | **Validação (15%)** | **Teste (15%)** |
|---|---|---|---|---|
| COVID-19 | 2.300 | 1.610 | 345 | 345 |
| Tuberculose | 2.300 | 1.610 | 345 | 345 |
| Pneumonia | 2.300 | 1.610 | 345 | 345 |
| Normal | 7.100 | 4.970 | 1.065 | 1.065 |
| **Total** | **13.900** | **9.800** | **2.100** | **2.100** |

Fonte: Autor.

O processo de divisão do conjunto de dados segue um algoritmo determinístico que garante reprodutibilidade através de uma semente aleatória. A estratégia consiste em: (1) consolidar imagens selecionadas das múltiplas fontes em um dataset intermediário, organizando-as por classe; (2) embaralhar o conjunto consolidado de forma determinística; (3) dividir cada classe em proporções definidas mantendo a distribuição de classes balanceada em cada conjunto. O algoritmo a seguir formaliza esse processo.

**Algorithm 1** Algoritmo para divisão estratificada do dataset em conjuntos de treinamento, validação e teste.

**Require:** $M$: mapa que associa cada classe $c$ a uma lista de pares $(p,q)$, em que $p$ é a pasta de origem e $q$ a quantidade de imagens a selecionar de $p$
$D$: pasta de destino
$s$: semente aleatória para reprodutibilidade
$r_{train} = 0{,}70$: proporção de imagens para treinamento
$r_{val} = 0{,}15$: proporção de imagens para validação

**Ensure:** Diretórios $D$/`train/`, $D$/`val/` e $D$/`test/`, cada um contendo subpastas por classe $c$ (*covid*, *tuberculose*, etc.)

Inicializar gerador aleatório com semente $s$
**for** cada classe $c$ em $M$ **do**
    **for** cada par $(p,q)$ associado a $c$ em $M$ **do**
        $imgs \leftarrow$ listar arquivos de imagem em $p$
        $selecionadas \leftarrow$ amostra aleatória de $q$ imagens de $imgs$
        Copiar $selecionadas$ para $D$/`dataset/`$c$/ ▷ consolidação intermediária
    **end for**
**end for**
**for** cada classe $c$ em $D$/`dataset/` **do**
    $imgs \leftarrow$ listar imagens em $D$/`dataset/`$c$/
    Embaralhar $imgs$ com semente $s$ ▷ garante reprodutibilidade
    ▷ Calcular número de amostras por conjunto (piso da proporção)
    $n_{train} \leftarrow \lfloor |imgs| \times r_{train} \rfloor$ ▷ $|imgs|$ = total de imagens de $c$
    $n_{val} \leftarrow \lfloor |imgs| \times r_{val} \rfloor$
    $n_{test} \leftarrow |imgs| - n_{train} - n_{val}$ ▷ restante
    Copiar primeiras $n_{train}$ imagens para $D$/`train/`$c$/
    Copiar próximas $n_{val}$ imagens para $D$/`val/`$c$/
    Copiar últimas $n_{test}$ imagens para $D$/`test/`$c$/
**end for**

Fonte: Autor.

## 4.2 PRÉ-PROCESSAMENTO

A etapa de *Pre-processing* é crucial em qualquer projeto de aprendizado de máquina, especialmente quando se trata de análise de imagens médicas, onde a qualidade e a relevância dos dados podem impactar significativamente o desempenho do modelo. Nesta seção, são descritos os procedimentos de pré-processamento aplicados às imagens

de raio-X do tórax para preparar os dados para as etapas subsequentes de segmentação e classificação.

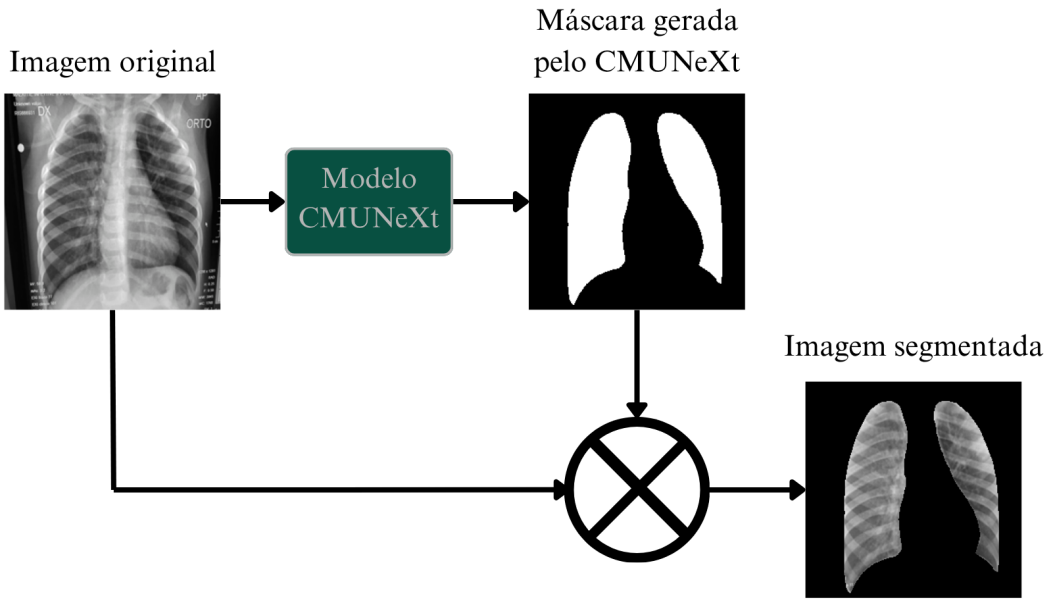
### 4.2.1 Segmentação dos pulmões

Na etapa de segmentação de imagens, a base de dados utilizada é a mesma que para a classe de COVID-19 na Seção 4.1.2, disponibilizada em (CHOWDHURY et al., 2020). Foram consideradas todas as radiografias desse *dataset*, já que para todas elas há sua respectiva máscara, além de que a classe para segmentação não importa. Dessa forma, totalizando 21.165 raios-X com máscara.

A arquitetura a ser implementada é a *CMUNeXt*, baseada em *U-Net*, já que demonstra enorme sucesso na literatura na resolução de problemas de segmentação. Desse jeito, a imagem original de raio-X é submetida ao modelo de segmentação, que processa os dados de entrada e gera uma máscara binária correspondente às regiões de interesse, conforme a Seção 2.7 explica.

Assim, para obter a imagem segmentada final, é realizada uma operação de multiplicação elemento a elemento entre a imagem original e a máscara binária gerada. Esta operação matemática resulta na preservação apenas das regiões correspondentes aos pulmões na imagem original, enquanto as demais áreas são suprimidas. O resultado é uma imagem onde apenas as estruturas pulmonares são mantidas visíveis, facilitando análises posteriores focadas especificamente nessas regiões anatômicas. Essa arquitetura está ilustrada na Figura 23.

Figura 23 – Diagrama do exemplo da segmentação do pulmão




Fonte: Adaptado de (CHICHANOSKI, 2022).

### 4.2.2 *Data Augmentation*

O *data augmentation* é uma técnica de regularização amplamente empregada no treinamento de redes neurais convolucionais que consiste na geração artificial de novas amostras de treinamento por meio da aplicação de transformações sobre as imagens originais (SHORTEN; KHOSHGOFTAAR, 2019). O objetivo é aumentar a diversidade do conjunto de treinamento sem a necessidade de coletar novas amostras, reduzindo o risco de *overfitting* e melhorando a capacidade de generalização do modelo para imagens não vistas durante o treinamento.

Em imagens médicas, a aplicação de *data augmentation* deve ser cuidadosa, pois transformações excessivamente agressivas podem introduzir artefatos que não ocorrem na prática clínica ou distorcer estruturas anatomicamente relevantes para o diagnóstico (MUMUNI; MUMUNI, 2022). Neste trabalho, as transformações foram selecionadas de forma a simular variações realistas que podem ocorrer na aquisição de radiografias torácicas, como diferenças de posicionamento do paciente, variações de exposição e ruídos do sensor.

A implementação foi realizada com a biblioteca Albumentations (BUSLAEV et al., 2020), aplicando um conjunto de transformações exclusivamente durante o treinamento, de forma aleatória a cada época. A Tabela 3 descreve cada transformação utilizada com seus respectivos parâmetros e probabilidade de aplicação.

Tabela 3 – Transformações de *data augmentation* aplicadas durante o treinamento.

| **Transformação** | **Parâmetros** | $p$ | **Descrição** |
|---|---|---|---|
| *Resize* | $224 \times 224$ px | 1,0 | Redimensiona a imagem para o tamanho de entrada das CNNs. |
| *HorizontalFlip* | — | 0,5 | Espelhamento horizontal para simular variação de posicionamento. |
| *RandomBrightnessContrast* | brilho $\pm 0{,}2$; contraste $\pm 0{,}2$ | 0,5 | Variação de brilho e contraste para simular diferenças de exposição. |
| *Rotate* | limite $= 15$ | 0,5 | Rotação aleatória dentro de $[-15, +15]$ para simular inclinação do paciente. |
| *GaussNoise* | — | 0,3 | Adição de ruído gaussiano para simular artefatos do sensor. |
| *CLAHE* | — | 0,3 | Equalização adaptativa de histograma com limite de contraste para realce de estruturas pulmonares. |
| *Normalize* | $\mu$=[0,485; 0,456; 0,406]; $\sigma$=[0,229; 0,224; 0,225] | 1,0 | Normalização com média e desvio padrão do ImageNet. |

Fonte: Autor.

## 4.3 ARQUITETURAS DE CNN

### 4.3.1 Híbrido (ResNet152 + EfficientNet-B5)

A arquitetura híbrida proposta combina duas redes neurais convolucionais profundas que são a ResNet152 e a EfficientNet-B5, por meio da concatenação de suas representações latentes, aproveitando as propriedades complementares de cada uma delas. A implementação foi adaptada a partir do trabalho de (HASSAN, 2026), e modificada para atender às particularidades do presente estudo.

A ResNet152 é uma variante profunda da família ResNet (*Residual Network*), proposta por (HE et al., 2016), composta por 152 camadas e fundamentada no conceito de conexões residuais (*skip connections*). Essas conexões permitem que o gradiente flua diretamente entre camadas não adjacentes durante o retropropagação, mitigando o problema do desvanecimento do gradiente em redes muito profundas. No modelo implementado, o bloco convolucional da ResNet152 é utilizado como extrator de características, com a camada de classificação original removida, produzindo um vetor de características de dimensão 2048.

A EfficientNet-B5 pertence à família EfficientNet, introduzida por (TAN; LE, 2019), cujo diferencial é o escalonamento composto (*compound scaling*) das dimensões de largura, profundidade e resolução da rede de forma proporcional e simultânea, resultando em modelos mais eficientes em termos computacionais sem perda de desempenho. A variante B5 apresenta maior capacidade representativa em relação aos modelos menores da família, sendo adequada para tarefas que exigem extração de características detalhadas. Analogamente à ResNet152, o bloco de características da EfficientNet-B5 é utilizado sem a camada de classificação original, gerando também um vetor de dimensão 2048.

Com isso, os vetores extraídos pelas duas redes são concatenados, resultando em uma representação combinada de dimensão 4096, que é então processada por um módulo classificador composto por três camadas totalmente conectadas (FC), com normalização de lotes (*Batch Normalization*), ativação ReLU e camadas de Dropout para regularização. A saída final corresponde ao número de classes da tarefa, no caso, 2 para o estágio de classificação binária e 4 para o estágio multiclasse. A combinação das duas arquiteturas permite que o modelo capture simultaneamente padrões de alto nível aprendidos pela profundidade da ResNet152 e representações eficientes aprendidas pela EfficientNet-B5, conferindo maior capacidade discriminatória ao conjunto.

No presente trabalho, os seguintes modelos foram carregados com pesos pré-treinados no ImageNet (IMAGENET1K_V1), e suas camadas classificadoras originais foram substituídas por *Dropout* (taxa 0,4) seguido de camada linear com o número de classes da tarefa.

### 4.3.2 ResNet50

A ResNet50 é uma arquitetura de rede neural residual de 50 camadas, também proposta por (HE et al., 2016) como uma versão mais compacta em relação à ResNet152. Sua estrutura é baseada em blocos residuais do tipo *Bottleneck*, cada um composto por três camadas convolucionais, com filtros de dimensões $1 \times 1$, $3 \times 3$ e $1 \times 1$, respectivamente, seguidas de normalização de lotes e ativação ReLU. As conexões de atalho (*skip connections*) somam a entrada do bloco à sua saída, permitindo o treinamento eficiente de redes profundas ao preservar o fluxo do gradiente.

Os pesos pré-treinados permitem ao modelo transferir conhecimento adquirido em larga escala para o domínio de imagens médicas, acelerando a convergência e melhorando a generalização com conjuntos de dados menores.

### 4.3.3 DenseNet121

A DenseNet121 (*Densely Connected Convolutional Network*) foi proposta por (HUANG et al., 2017) e caracteriza-se pela conectividade densa entre camadas: cada camada recebe como entrada a concatenação dos mapas de características de todas as camadas anteriores do mesmo bloco, em vez de apenas a camada imediatamente anterior. Essa estratégia favorece o reaproveitamento de características, reduz o número de parâmetros e promove a regularização implícita da rede.

A arquitetura é organizada em quatro blocos densos (*dense blocks*), intercalados por blocos de transição responsáveis pela redução espacial dos mapas de características por meio de convoluções $1 \times 1$ e *pooling* médio. O primeiro bloco denso contém 6 camadas densas, o segundo 12, o terceiro 24 e o quarto 16. Cada camada densa é composta por normalização de lotes, ativação ReLU, convolução $1 \times 1$, nova normalização de lotes, nova ativação ReLU e convolução $3 \times 3$, seguida de *Dropout* para regularização.

### 4.3.4 Vgg16

A VGG16 foi proposta por (SIMONYAN; ZISSERMAN, 2015) e é caracterizada por sua arquitetura uniforme e relativamente simples, composta por 13

camadas convolucionais com filtros de tamanho $3 \times 3$ e 3 camadas totalmente conectadas, totalizando 16 camadas com parâmetros treináveis, assim vindo a denominação VGG16. As camadas convolucionais são agrupadas em blocos separados por camadas de *max pooling* para redução espacial progressiva. A simplicidade e uniformidade da arquitetura facilitam a interpretação e o ajuste fino do modelo.

Além disso, neste trabalho, foi utilizada a variante VGG16 com normalização de lotes (VGG16-BN) que tem a vantagem de convergir mais rápido e atingir métricas de acurácia melhores. Contudo, devido ao elevado número de parâmetros da arquitetura (aproximadamente 138 milhões), foi necessário reduzir o tamanho de lote (*batch size*) durante o treinamento para evitar erros de acesso à memória da GPU, utilizando-se lotes de 8 amostras para este modelo.

### 4.3.5 EfficientNet_B0

A EfficientNet-B0 é a variante base da família EfficientNet, proposta por (TAN; LE, 2019). Diferentemente das arquiteturas tradicionais que escalonam apenas uma das dimensões da rede (profundidade, largura ou resolução), a EfficientNet adota o escalonamento composto, que aumenta as três dimensões de forma balanceada e proporcional por meio de coeficientes derivados de uma busca de arquitetura neural, ou *Neural Architecture Search* (NAS). A unidade fundamental da EfficientNet é o bloco MBConv (*Mobile Inverted Bottleneck Convolution*), herdado da MobileNetV2, que combina convoluções *depthwise* separáveis com conexões residuais e um mecanismo de atenção de canal (*Squeeze-and-Excitation*).

Assim, esse modelo representa o ponto de partida do escalonamento composto, sendo o modelo mais leve da família, com aproximadamente 5,3 milhões de parâmetros. Sua inclusão no *ensemble* tem como objetivo introduzir um modelo de menor complexidade computacional e diferente viés indutivo em relação às arquiteturas mais profundas, contribuindo para a diversidade do *ensemble*.

## 4.4 MÉTODOS DE AGREGAÇÃO

Para combinar as predições dos modelos CNN utilizados na arquitetura proposta, serão implementados três métodos distintos de agregação, cada um com características

específicas para otimizar o desempenho do *ensemble*. Já que a utilização de apenas um modelo resulta em um maior risco de *overfitting*. Dessa forma, estes métodos representam diferentes abordagens para consolidar as decisões individuais dos modelos em uma predição final mais robusta e confiável.

### 4.4.1 Votação por maioria (*Hard Voting*)

O primeiro método consiste na previsão mais votada pelos modelos, também conhecida como *hard voting* ou votação por maioria. Este método é particularmente eficaz quando os modelos apresentam alta concordância nas predições corretas, pois permite que a decisão coletiva supere eventuais erros individuais de modelos isolados. A simplicidade desta abordagem facilita a interpretabilidade dos resultados, uma vez que a decisão final é baseada puramente no consenso majoritário.

### 4.4.2 Média das previsões (*Soft Voting*)

O segundo método implementado utiliza a média das previsões, caracterizando uma abordagem de *soft voting*, considerando as probabilidades de cada classe. As probabilidades de cada classe são calculadas como a média aritmética das probabilidades correspondentes de todos os modelos. Esta estratégia permite incorporar o grau de confiança de cada modelo individual na decisão final, resultando em predições mais suaves e potencialmente mais precisas, especialmente em casos em que as decisões individuais são menos categóricas.

### 4.4.3 Média ponderada pela AUC (*Weighted Average by AUC*)

O terceiro método representa uma evolução da abordagem anterior, implementando a média ponderada das previsões dos modelos, onde o peso de cada modelo corresponde à sua respectiva AUC (*Area Under the Curve*) obtida durante a fase de validação. Esta estratégia reconhece que nem todos os modelos possuem a mesma capacidade discriminativa, atribuindo maior influência aos modelos que demonstraram melhor desempenho individual. O cálculo da predição final é realizado através da soma ponderada das probabilidades de cada modelo, onde os pesos são normalizados para

garantir que sua soma seja igual a um. Esta abordagem é teoricamente mais robusta, pois permite que modelos com melhor desempenho individual tenham maior influência na decisão final, enquanto modelos com desempenho inferior contribuem proporcionalmente menos para o resultado agregado.

A implementação destes métodos de agregação permite uma análise comparativa do desempenho de diferentes estratégias de *ensemble*, possibilitando a seleção da abordagem mais adequada para o problema específico de classificação de patologias pulmonares. A escolha do método de agregação pode ser influenciada por fatores como a diversidade dos modelos individuais, a distribuição das classes no *dataset* e os requisitos específicos de sensibilidade e especificidade do sistema diagnóstico proposto.

## 4.5 ESTRATÉGIAS DE REFINAMENTO

### 4.5.1 Fine-Tuning

O *fine-tuning* é uma técnica de aprendizado por transferência (*transfer learning*) que consiste em ajustar os pesos de um modelo pré-treinado para uma nova tarefa, aproveitando o conhecimento representacional adquirido durante o treinamento original em grandes conjuntos de dados, como o ImageNet. Em vez de treinar uma rede do zero, o *fine-tuning* parte de uma inicialização favorável e adapta gradualmente as representações internas ao domínio específico da nova tarefa. Neste caso, a classificação de patologias pulmonares em imagens de raio-X.

A justificativa para o uso do *fine-tuning* em imagens médicas é bem estabelecida na literatura: modelos pré-treinados em conjuntos de dados de imagens naturais desenvolvem detectores de bordas, texturas e padrões hierárquicos nas camadas iniciais que são amplamente transferíveis para domínios distintos (KIM et al., 2022). As camadas mais profundas, por outro lado, capturam representações mais específicas ao domínio de origem e, portanto, beneficiam-se de ajuste direcionado ao novo conjunto de dados.

No presente trabalho, o *fine-tuning* foi aplicado a todos os cinco modelos do *ensemble*, em ambos os estágios de classificação (binário e multiclasse) e em ambas as baterias de dados (com e sem segmentação). O processo foi estruturado em duas fases

progressivas, descritas a seguir.

*A Primeira fase: ajuste das camadas finais*

Na primeira fase, todos os parâmetros dos *backbones* (blocos convolucionais pré-treinados) são congelados, ou seja, seus gradientes são desativados (`requires_grad = False`), impedindo que seus pesos sejam atualizados durante a retropropagação. Apenas os parâmetros do módulo classificador, camadas totalmente conectadas inseridas na etapa de adaptação, são mantidos treináveis.

Essa estratégia tem como objetivo estabilizar o classificador antes de qualquer ajuste nas camadas convolucionais. Como os pesos do classificador são inicializados aleatoriamente, permitir a atualização simultânea do *backbone* nessa fase poderia propagar gradientes de grande magnitude e degradar as representações pré-treinadas, fenômeno conhecido como *catastrophic forgetting*. A Fase 1 foi executada por 15 épocas com taxa de aprendizado de $5 \times 10^{-4}$ e *early stopping* com paciência de 5 épocas.

*B Segunda fase: ajuste de camadas convolucionais*

Na segunda fase, as camadas finais dos *backbones* são descongeladas seletivamente, permitindo que as representações de alto nível sejam ajustadas ao domínio médico. O descongelamento é realizado de forma progressiva e apenas os blocos mais profundos de cada arquitetura são liberados, enquanto as camadas iniciais, responsáveis por detectores de baixo nível (bordas, texturas), permanecem congeladas, pois essas características são amplamente transferíveis independentemente do domínio.

Especificamente, para a ResNet152 e ResNet50, são descongelados os blocos residuais finais (*layer4* e o último bloco do *layer3*); para a EfficientNet-B5 e EfficientNet-B0, os blocos MBConv correspondentes às últimas etapas da rede; para a DenseNet121, o quarto bloco denso (*denseblock4*) e a camada de normalização final (*norm5*); e para a VGG16, as últimas camadas convolucionais do bloco de características. Para evitar a degradação das representações pré-treinadas durante o descongelamento, são utilizadas taxas de aprendizado diferenciadas por grupo de parâmetros: os parâmetros do *backbone* descongelados recebem uma taxa de aprendizado de $1 \times 10^{-5}$, significativamente menor do que a aplicada ao classificador, que recebe $5 \times 10^{-5}$. Essa assimetria garante que os

ajustes nos pesos convolucionais sejam sutis e direcionados, preservando o conhecimento pré-treinado enquanto adapta gradualmente as representações ao domínio-alvo. A Fase 2 foi executada por 20 épocas com *early stopping* de paciência 7 épocas.

A combinação das duas fases de *fine-tuning*, que significa a estabilização do classificador seguida de ajuste progressivo do *backbone* com taxas de aprendizado diferenciadas, representa uma abordagem consolidada na literatura para maximizar a transferência de conhecimento em conjuntos de dados de tamanho moderado (ELIWA, 2025), como é o caso das imagens de raio-X utilizadas neste trabalho.

### 4.5.2 Harmonização

A harmonização de imagens é uma técnica de pré-processamento aplicada durante a inferência com o objetivo de reduzir a discrepância visual entre imagens provenientes de fontes ou equipamentos distintos dos utilizados no treinamento. Em contextos de avaliação com dados externos, diferenças de brilho, contraste e distribuição de intensidade entre os domínios de treinamento e teste podem degradar o desempenho do modelo, mesmo que as estruturas anatômicas de interesse sejam visualmente semelhantes (KHAN et al., 2026).

Com isso, neste trabalho, a harmonização foi implementada por meio da equalização adaptativa de histograma com limite de contraste (CLAHE) aplicada ao canal de luminância no espaço de cor LAB. O espaço LAB separa a informação de luminosidade (canal L) das informações de cromaticidade (canais A e B), permitindo que o ajuste de contraste seja realizado exclusivamente sobre a componente de brilho da imagem, sem alterar sua tonalidade. A aplicação do CLAHE sobre o canal L equaliza localmente o histograma de luminância, aproximando a distribuição de intensidades das imagens externas ao padrão predominante no conjunto de treinamento e reduzindo variações de exposição decorrentes de diferenças entre equipamentos radiológicos.

Além disso, essa técnica foi aplicada em conjunto com o TTA durante a avaliação no *dataset* externo, sendo processada previamente às transformações de aumento de dados de cada passagem. Assim como o TTA, a técnica atua exclusivamente na fase de inferência, sem qualquer modificação no processo de treinamento dos modelos.

### 4.5.3 *Test Time Augmentation*

O *Test Time Augmentation (TTA)* é uma técnica de inferência que consiste em aplicar múltiplas transformações a uma mesma imagem durante a fase de teste e combinar as predições resultantes por meio de agregação, tipicamente pela média das probabilidades (MOLCHANOV et al., 2020). Ao contrário do data augmentation, que é aplicado durante o treinamento para aumentar a diversidade dos dados, o TTA atua na inferência com o objetivo de reduzir a variância das predições individuais e produzir estimativas mais robustas e estáveis.

Dessa forma, a intuição por trás do TTA é que diferentes versões transformadas de uma mesma imagem podem enfatizar aspectos distintos das características visuais relevantes para a classificação. A média das probabilidades sobre essas versões funciona como uma forma de ensemble no espaço de augmentações, aproveitando a invariância do modelo a pequenas variações para obter uma estimativa mais confiável da classe verdadeira (SHERKATGHANAD et al., 2024).

No presente trabalho, o TTA foi aplicado durante a avaliação no dataset externo. Para cada imagem, foram geradas 5 versões augmentadas utilizando o mesmo conjunto de transformações empregado no treinamento que são o espelhamento horizontal, a variação de brilho e contraste, a rotação de até 15 graus e o ruído gaussiano e que em seguida finalizada com a normalização. Desse jeito, as 5 versões foram processadas simultaneamente em um único *forward pass* em lote (*batch*), aproveitando o paralelismo da GPU para minimizar o custo computacional adicional. As probabilidades resultantes das 5 versões foram então agregadas pela média aritmética, produzindo a predição final para cada imagem.

A implementação em *batch* é uma distinção importante em relação à abordagem ingênua de processar cada versão augmentada individualmente: ao empilhar as $N$ versões de cada imagem em um único tensor de dimensão $(B \times N, C, H, W)$ e processar em um único *forward pass*, o custo computacional total é significativamente reduzido em comparação ao processamento sequencial imagem a imagem, tornando o TTA viável mesmo para conjuntos de dados de tamanho moderado.

## 4.6 MODELO DE ARQUITETURA PROPOSTO

O modelo híbrido de *deep learning* proposto neste trabalho integra diferentes paradigmas de aprendizado profundo para maximizar as vantagens específicas de cada abordagem, resultando em uma arquitetura robusta e eficiente para a detecção de patologias pulmonares em imagens de raio-X. A arquitetura proposta, ilustrada na Figura 24, é composta por três etapas principais interconectadas: o módulo de segmentação, o módulo de classificação binária e o módulo de classificação multiclasse.

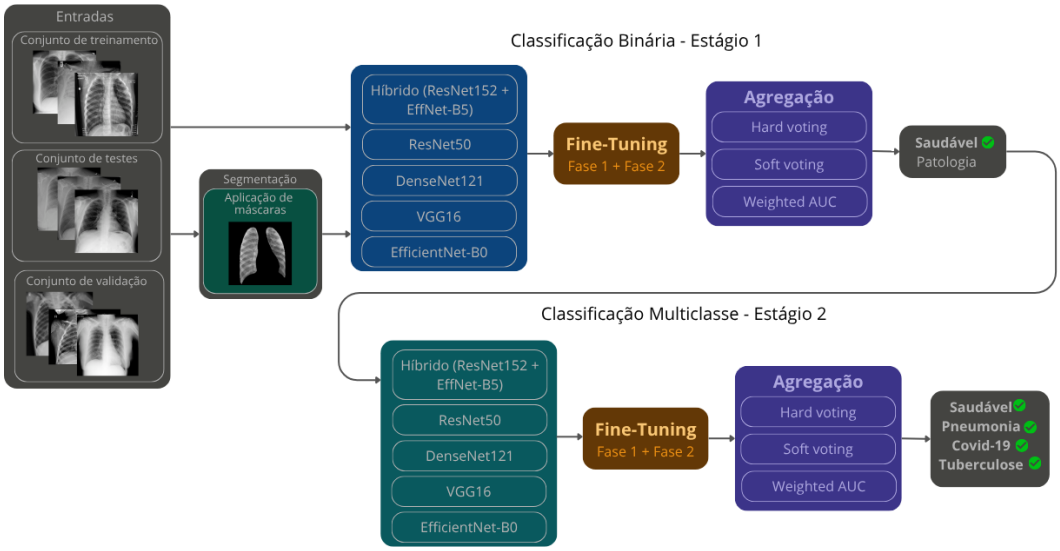
O primeiro módulo de segmentação inclui a confecção de um *dataset* com imagens de raio-X de pulmões de classes diferentes, conforme critérios selecionados (Seção 4.1.2), seguido pela segmentação dessas radiografias para realçar apenas as áreas de interesse pulmonar. Esta etapa de pré-processamento é fundamental para reduzir ruídos e focar a análise nas regiões anatomicamente relevantes para o diagnóstico.

O segundo módulo consiste na aplicação de diferentes arquiteturas de CNN, que são Híbrido (ResNet152 em conjunto com EfficientNet-B5), ResNet50, DenseNet121, Vgg16 e EfficientNet_B0, para classificação binária, determinando se a imagem é “Saudável” ou “Patológica”. O conjunto de imagens que foram classificadas como "Patológica"passa então para o terceiro módulo.

O terceiro módulo, por sua vez, realiza a classificação multiclasse, utilizando o mesmo conjunto de modelos CNN para categorizar especificamente as patologias detectadas, o que inclui pneumonia, COVID-19 e tuberculose, ou também o estado saudável, caso não tenha sido devidamente detectado no estágio anterior.

Nesses últimos dois módulos, o *fine-tuning* em duas fases é aplicado individualmente a cada modelo, e as predições são combinadas pelos três métodos de agregação por *ensemble*: *hard voting*, *soft voting* e média ponderada pela AUC.

Figura 24 – Diagrama do modelo de arquitetura em cascata proposta




Fonte: Autor.

Além disso, foi proposta e avaliada uma segunda arquitetura denominada paralela, ilustrada na Figura 25. Nesta abordagem, os módulos de classificação binária e multiclasse são treinados e executados de forma independente, sem que o resultado de um estágio influencie a entrada do outro. Ambos os módulos recebem diretamente as imagens provenientes do módulo de segmentação e produzem suas predições de forma autônoma.

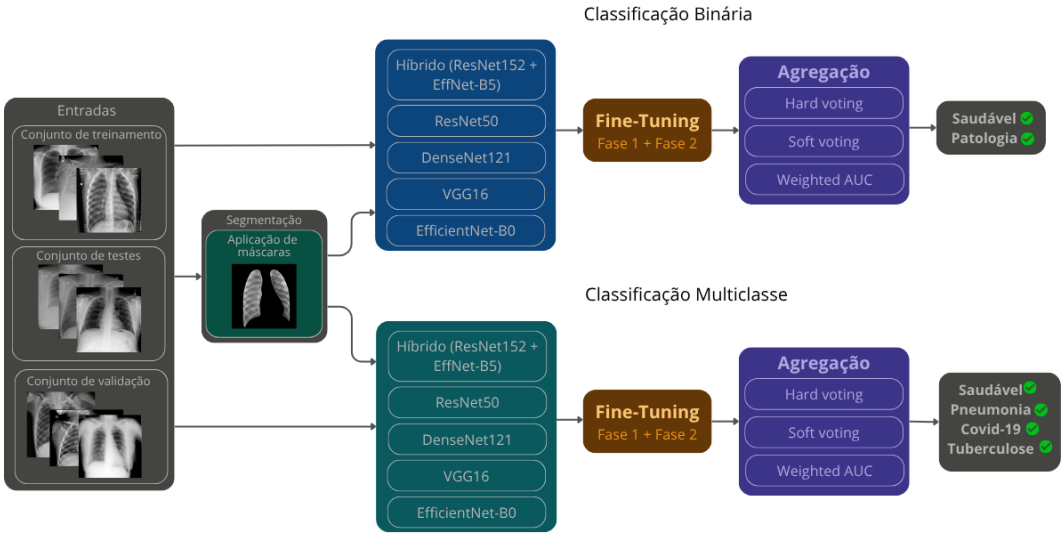
Assim como na arquitetura em cascata, o módulo de classificação binária é composto pelo mesmo conjunto de cinco modelos CNN, bem como a aplicação de *fine-tuning* e a combinação feita pelos métodos de agregação.

Dito isso, essa abordagem apresenta como vantagem a eliminação do risco de propagação de erros entre estágios, fenômeno característico das arquiteturas em cascata no qual um erro de classificação no módulo binário compromete necessariamente o resultado do módulo multiclasse. Por outro lado, a ausência de triagem binária implica uma tarefa de maior complexidade para o módulo multiclasse, que precisa lidar simultaneamente com a distinção entre casos normais e patológicos e com a diferenciação entre as patologias específicas.

A comparação entre as arquiteturas em cascata e em paralela permite avaliar o impacto da estratégia de decomposição da tarefa de classificação no desempenho geral do

sistema, bem como identificar qual abordagem apresenta melhor equilíbrio entre precisão e robustez para o diagnóstico de patologias pulmonares em imagens de raio-X.

Figura 25 – Diagrama do modelo de arquitetura paralela proposta




Fonte: Autor.

# 5 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados obtidos a partir da aplicação dos métodos descritos na seção de metodologia, bem como uma discussão crítica sobre o desempenho dos modelos e as implicações dos resultados para a análise de raios-X do tórax. A avaliação foi realizada utilizando métricas de desempenho discutidos anteriormente, além de análises qualitativas das predições em casos específicos.

## 5.1 SEGMENTAÇÃO DOS PULMÕES

O modelo *CMUNeXt* apresentado na Seção 2.9, foi utilizado para a tarefa de segmentação pulmonar das imagens do dataset do projeto, já que por ter a arquitetura leve e eficiente torna-o adequado para aplicações com conjuntos de dados de tamanho moderado, como o presente case. Assim, ele foi treinado com validação cruzada e avaliado com base nas métricas de *IoU* (*Intersection over Union*), *F1-Score*, *recall*, especificidade, precisão e acurácia. As máscaras geradas pelo *CMUNeXt* foram utilizadas como entrada para o módulo de pré-processamento do *pipeline* de classificação descrito nas seções anteriores.

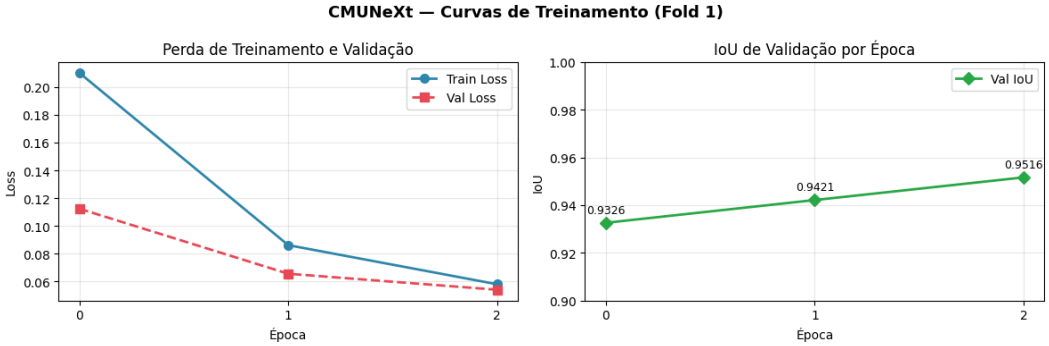
O treinamento foi realizado por 3 épocas, com monitoramento do *IoU* de validação para seleção do melhor modelo. Esse número reduzido de épocas foi suficiente para o objetivo pretendido: já na primeira época o modelo alcançou *IoU* de validação superior a 0,93, e ao final da terceira época atingiu 0,9516, qualidade adequada para as máscaras que seriam utilizadas como entrada do *pipeline* de classificação. Como a tarefa de segmentação pulmonar é anatomicamente bem definida e a arquitetura *CMUNeXt* é leve, a rede convergiu rapidamente para um patamar elevado de desempenho, o que motivou a interrupção do treinamento nesse ponto. A Tabela 4 apresenta a evolução das métricas de treinamento ao longo das épocas, e a Figura 26 ilustra as curvas de perda e *IoU* de validação.

Tabela 4 – Métricas de treinamento do *CMUNeXt* por época. (✓= melhor modelo selecionado, *threshold* = 0,50)

| **Época** | ***Train Loss*** | ***Val Loss*** | ***Val IoU*** |
|---|---|---|---|
| 0 | 0,2102 | 0,1124 | 0,9326 |
| 1 | 0,0862 | 0,0656 | 0,9421 |
| 2✓ | 0,0580 | 0,0541 | 0,9516 |

Fonte: Autor.

Figura 26 – Gráfico de Curvas de perda e *IoU* de validação do *CMUNeXt*.




Fonte: Autor.

O modelo atingiu o melhor *IoU* de validação de 0,9516 na terceira época (época 2), com *threshold* de decisão 0,50, sendo esse o modelo selecionado para geração das máscaras do dataset final. A redução consistente da perda de treinamento e validação ao longo das épocas, sem divergência entre as duas curvas, indica ausência de *overfitting* no processo de treinamento. Cabe ressaltar, no entanto, que o *IoU* de validação ainda apresentava tendência de crescimento ao final da terceira época, o que sugere que um treinamento mais prolongado poderia elevar ainda mais o desempenho da segmentação.

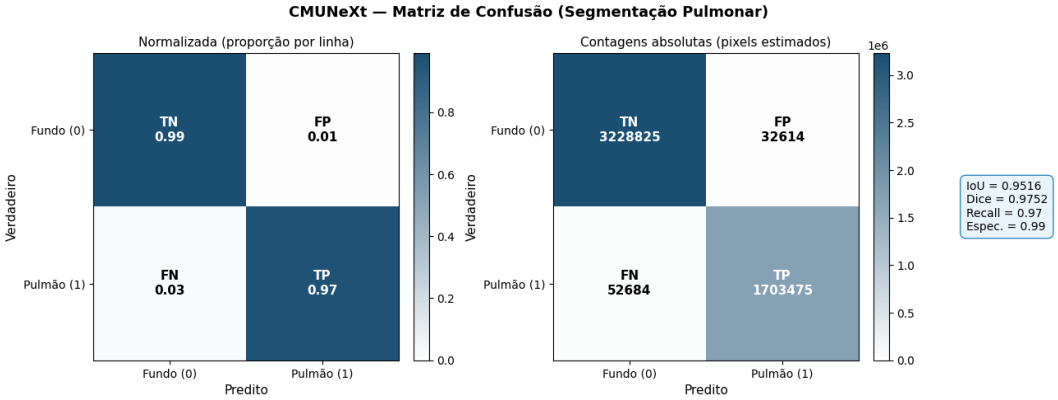
A Tabela 5 apresenta as métricas de desempenho do modelo *CMUNeXt* na segmentação pulmonar, derivadas a partir do *IoU* de validação e da matriz de confusão normalizada obtida sobre o conjunto de teste.

Tabela 5 – Métricas de treinamento do *CMUNeXt* por época (* Estimativa com proporção de foreground de 35% dos pixels).

| **Métrica** | ***IoU*** | **Acurácia*** | **Precisão*** | ***Recall*** | ***F1-Score*** | **Especificidade** |
|---|---|---|---|---|---|---|
| Valor | 95,16% | 98,60% | 98,49% | 97,00% | 97,52% | 99,00% |

Fonte: Autor.

Figura 27 – Matriz de confusão do CMUNeXt.




Fonte: Autor.

Os resultados demonstram que o *CMUNeXt* apresenta desempenho expressivo na delimitação da região pulmonar: o *F1-Score* de 97,52% e o *IoU* de 95,16% indicam elevada sobreposição entre as máscaras preditas e as máscaras de referência (*ground truth*). O *recall* de 97,00% revela que o modelo identifica corretamente 97 de cada 100 pixels pertencentes ao pulmão, enquanto a especificidade de 99,00% indica que apenas 1% dos pixels de fundo é erroneamente classificado como região pulmonar, resultado relevante para evitar a inclusão de estruturas não pulmonares na etapa de classificação subsequente.

A matriz de confusão (Figura 27) confirma esse comportamento: os erros residuais concentram-se na fronteira entre a região pulmonar e o mediastino, onde a transição entre as classes é anatomicamente gradual, representando o principal desafio para modelos de segmentação em raio-X torácico. A ausência de erros grosseiros, como, por exemplo, a classificação de grandes regiões de fundo como pulmão ou vice-versa, valida a qualidade das máscaras geradas para uso no *pipeline* de classificação.

## 5.2 ANÁLISE ABLATIVA

A análise ablativa tem como objetivo quantificar a contribuição individual de cada componente do *pipeline* proposto. Para cada técnica avaliada, são apresentados os resultados no *dataset* interno (conjunto de teste) e no *dataset* externo, permitindo verificar se os ganhos de desempenho se sustentam em dados fora da distribuição de treinamento.

Assim, esses primeiros testes foram feitos apenas com o modelo multiclasse, a princípio apenas com arquitetura híbrida (ResNet152 + EfficientNet-B5) e, por consequência, sem a aplicação de métodos de agregação. O pré-processamento aplicado foi apenas de normalização e redimensionamento, mantendo todas as imagens em $224 \times 224$ pixels.

Além disso, manteve-se para todos os testes a mesma configuração de treinamento, utilizando o 60 épocas e *early stopping* com paciência de 10 épocas, e o otimizador AdamW com taxa de aprendizado inicial de $3 \times 10^{-4}$ e decaimento de peso de $1 \times 10^{-2}$. O modelo foi selecionado com base na melhor AUC obtida no conjunto de validação durante o treinamento.

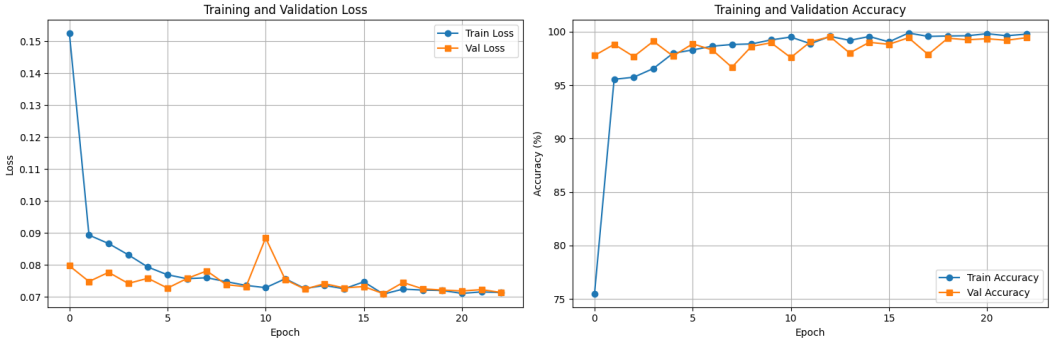
Considerando essas primeiras configurações apresentadas, os resultados obtidos foram as seguintes:

Tabela 6 – Resultados obtidos no conjunto de teste

| **Métricas** | **Acurácia** | **Precisão** | ***Recall*** | ***F1-Score*** |
|---|---|---|---|---|
| Valor | 99,52% | 98,50% | 99,25% | 98,75% |

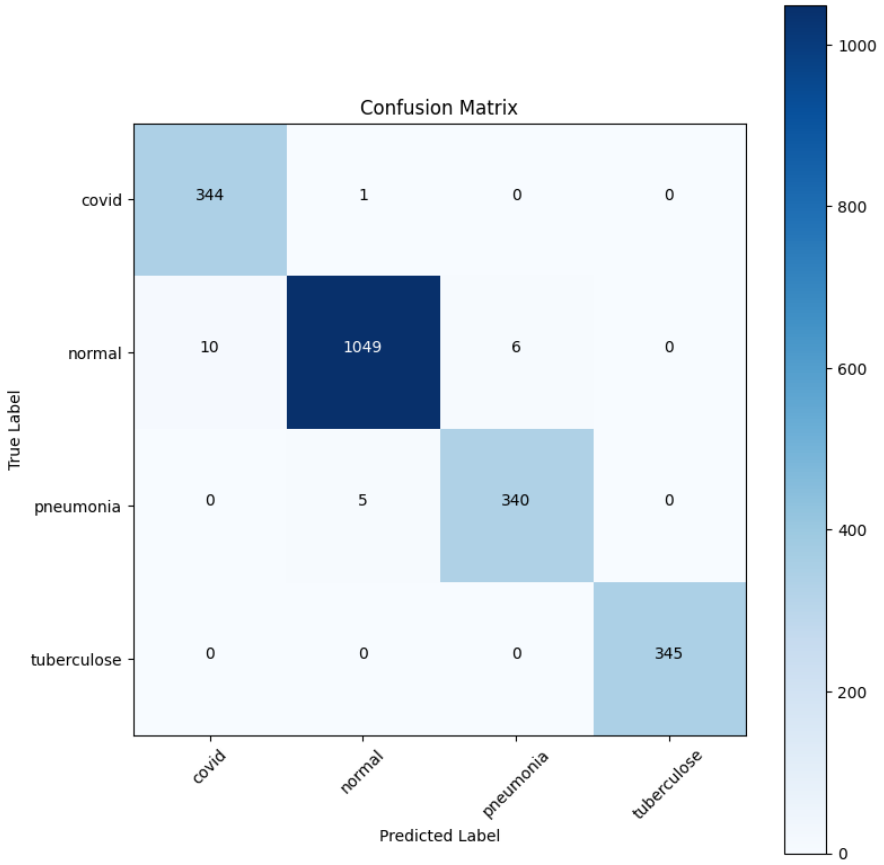
Fonte: Autor.

Figura 28 – Gráfico de perda e acurácia.




Fonte: Autor.

Figura 29 – Matriz de confusão do modelo multiclasse no conjunto de teste.




Fonte: Autor.

Os resultados no conjunto de teste interno revelam um desempenho expressivo do modelo base, com acurácia de 99,52% e *F1-score* de 98,75%, demonstrando que a arquitetura híbrida é capaz de aprender representações altamente discriminativas a partir do *dataset* de treinamento. A matriz de confusão, como mostrado na Figura 29, confirma que a maior parte dos erros se concentra na confusão entre as classes covid e normal, com 10 casos de normal classificados incorretamente como covid e 6 casos de normal classificados como pneumonia, enquanto as classes tuberculose e pneumonia apresentaram classificação praticamente perfeita.

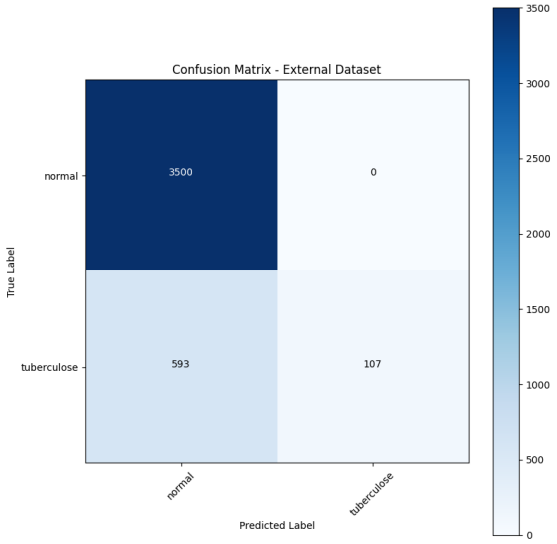
Para o *dataset* externo, os resultados foram os seguintes:

Tabela 7 – Resultados obtidos no conjunto do *dataset* externo

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Valor | 85,88% | 92,76% | 57,64% | 59,35% |

Fonte: Autor.

Figura 30 – Matriz de confusão do modelo multiclasse no conjunto do *dataset* externo.




Fonte: Autor.

Entretanto, ao avaliar o modelo no *dataset* externo, observa-se uma degradação significativa de desempenho, com F1-score de apenas 59,35% e recall de 57,64%, conforme apresentado na Tabela 7. A matriz de confusão do *dataset* externo (Figura

30) revela que o modelo classificou corretamente todos os 3.500 casos normais, porém apresentou dificuldade acentuada na detecção de tuberculose, com apenas 107 casos corretamente identificados em um total de 700, ou seja, 593 casos de tuberculose foram classificados erroneamente como normal.

Dessa forma, essa discrepância entre o desempenho no *dataset* interno e externo indica que o modelo, sem técnicas de regularização e aumento de dados, apresenta limitada capacidade de generalização para imagens provenientes de fontes distintas da distribuição de treinamento, motivando a aplicação das técnicas descritas nas seções seguintes.

### 5.2.1 Impacto do *Data Augmentation*

A segunda etapa da análise ablativa consistiu na aplicação das técnicas de *data augmentation*, descritas na Seção 4.2.2, para verificar seu impacto no desempenho do modelo e, especialmente, em sua capacidade de generalização para dados externos. A Tabela 8 apresenta a comparação de desempenho dos modelos treinados sem *data augmentation* e com a aplicação do aumento de dados, utilizando as mesmas configurações de treinamento e arquitetura.

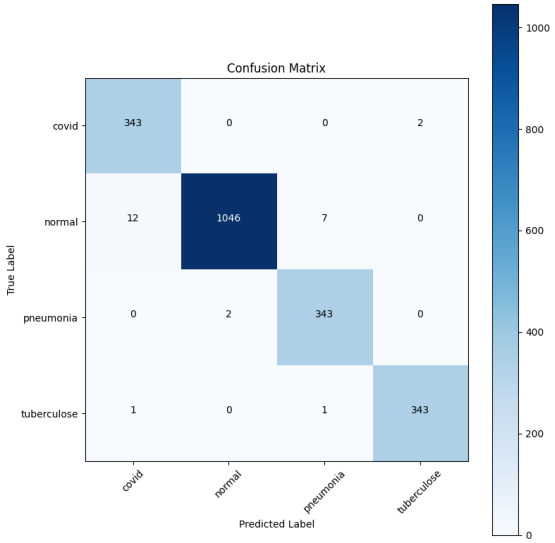
Tabela 8 – Resultados obtidos com e sem *data augmentation*

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Sem *Data Augmentation* | 99,52% | 98,50% | 99,25% | 98,75% |
| Com *Data Augmentation* | 99,57% | 98,85% | 98,49% | 98,85% |

Fonte: Autor.

A aplicação de *data augmentation* resultou em um leve aumento na acurácia e precisão, embora tenha havido uma pequena queda no recall. Esses resultados sugerem que o aumento de dados pode ajudar a melhorar a capacidade do modelo de generalizar para novos exemplos, mas também pode introduzir mais falsos negativos, o que é refletido na redução do recall.

Figura 31 – Matriz de confusão do modelo multiclasse no conjunto de teste com *data augmentation*.



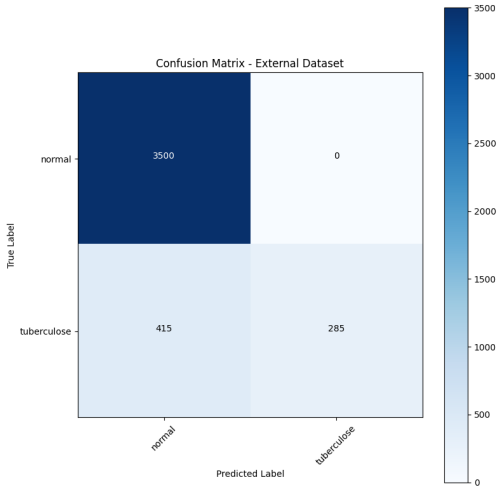

Fonte: Autor.

Tabela 9 – Resultados obtidos com e sem *data augmentation*

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Sem *Data Augmentation* | 85,88% | 92,76% | 57,64% | 59,35% |
| Com *Data Augmentation* | 90,12% | 94,70% | 70,36% | 76,14% |

Fonte: Autor.

Figura 32 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com *data augmentation*.




Fonte: Autor.

No conjunto de teste interno (Tabela 8), a aplicação de *data augmentation* resultou em um leve incremento de acurácia (de 99,52% para 99,57%) e de precisão (de 98,50% para 98,85%), com uma pequena queda no recall (de 99,25% para 98,49%). Esse comportamento é esperado: ao expor o modelo a versões transformadas das imagens durante o treinamento, introduz-se uma forma de regularização que reduz o *overfitting* mas pode aumentar marginalmente os falsos negativos em amostras mais simples do conjunto de teste. O F1-score praticamente se manteve estável (98,75% $\rightarrow$ 98,85%), indicando que o equilíbrio geral entre precisão e recall foi preservado.

Entretanto, o impacto mais relevante do *data augmentation*, contudo, foi observado no *dataset* externo (Tabela 9), onde o F1-score saltou de 59,35% para 76,14%, o que significa um ganho de 16,79 pontos percentuais. A acurácia passou de 85,88% para 90,12% e o recall de 57,64% para 70,36%, refletindo uma melhora substancial na detecção de tuberculose em imagens de fonte externa. A matriz de confusão, mostrada na Figura 32, confirma que o número de casos de tuberculose corretamente identificados aumentou de 107 para 285, em 700 amostras, ainda que uma parcela significativa (415 casos) tenha sido erroneamente classificada como normal.

### 5.2.2 Impacto do TTA e *Harmonization*

Já na terceira etapa dessa análise, foram implementados os métodos que visam transformar as imagens de teste para alcançar uma melhor correspondência com o que foi treinado, ou seja, buscar uma maior generalização do modelo para dados fora da distribuição de treinamento, como é esse o caso. Para isso, foram aplicados o TTA, técnica descrita na Seção 4.5.3 e o *Harmonization*, detalhada em Seção 4.5.2, utilizando as mesmas configurações de treinamento e arquitetura.

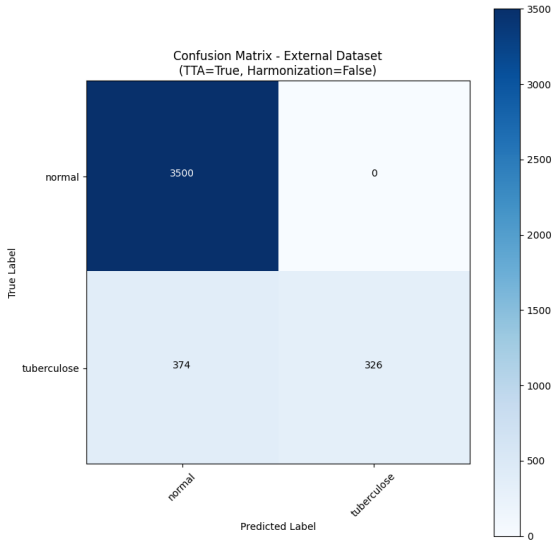
A Tabela 10 apresenta a comparação de desempenho dos modelos treinados sem e com a aplicação do TTA.

Tabela 10 – Resultados obtidos com e sem TTA e *Harmonization*

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Sem TTA e *Harmonization* | 90,12% | 94,70% | 70,36% | 76,14% |
| Com TTA | 91,10% | 95,17% | 73,29% | 79,24% |
| Com TTA e *Harmonization* | 91,24% | 95,24% | 73,71% | 79,67% |

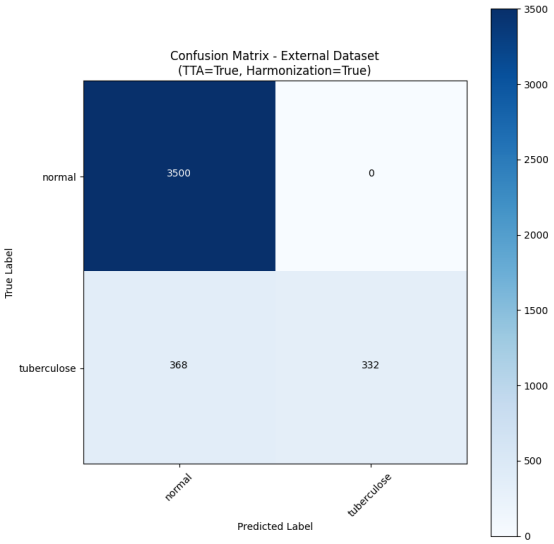
Fonte: Autor.

Figura 33 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com TTA.




Fonte: Autor.

Figura 34 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com TTA e *Harmonization*.




Fonte: Autor.

A aplicação do TTA isoladamente resultou em um incremento de F1-score de 76,14% para 79,24%, com ganhos consistentes em acurácia (90,12% $\rightarrow$ 91,10%), precisão (94,70% $\rightarrow$ 95,17%) e recall (70,36% $\rightarrow$ 73,29%). Esses resultados confirmam que a média de múltiplas predições sobre versões aumentadas da mesma imagem produz estimativas mais robustas e estáveis, reduzindo a variância associada a fatores como posicionamento e condições de iluminação presentes no *dataset* externo. A matriz de confusão correspondente (Figura 33) mostra que o número de casos de tuberculose corretamente identificados aumentou de 285 para 326, em 700 amostras.

Além do mais, a combinação de TTA com *Harmonization* resultou em uma melhora adicional, ainda que discreta, atingindo F1-score de 79,67%, acurácia de 91,24% e recall de 73,71%. A técnica de harmonização, que equaliza o histograma do canal de luminância no espaço LAB para aproximar o brilho e contraste das imagens externas ao padrão do *dataset* de treinamento, contribuiu com um ganho incremental de 0,43 pontos percentuais no F1-score em relação ao TTA isolado. A Figura 34 ilustra a matriz de confusão com TTA e harmonização, onde 332 casos de tuberculose foram corretamente identificados.

### 5.2.3 Impacto do *Fine-Tuning*

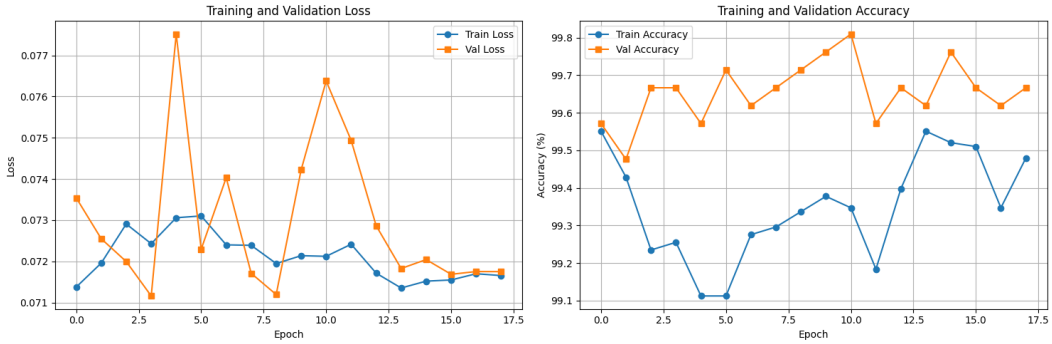
Em seguida foram realizados testes com a aplicação do método de *fine-tuning* em duas fases, conforme descrito na Seção 4.5.1. A Tabela 11 apresenta a comparação de desempenho no conjunto de teste interno antes e após o *fine-tuning*, enquanto a Figura 35 ilustra as curvas de perda e acurácia durante o processo de ajuste fino. A Tabela 12 e a Figura 37 apresentam os resultados correspondentes no *dataset* externo.

Tabela 11 – Resultados obtidos com e sem *fine-tuning*

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Sem *Fine-Tuning* | 99,57% | 98,85% | 98,49% | 98,85% |
| Com *Fine-Tuning* | 99,81% | 99,67% | 99,33% | 99,16% |

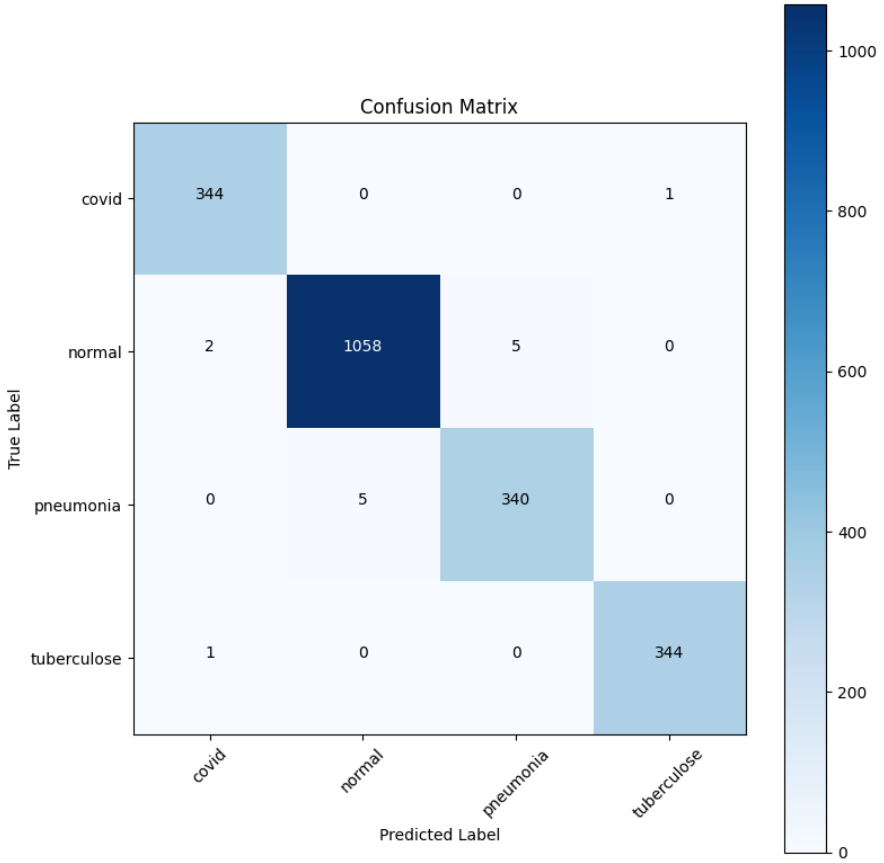
Fonte: Autor.

Figura 35 – Gráfico de perda e acurácia do treinamento do *fine-tuning*.




Fonte: Autor.

Figura 36 – Matriz de confusão do modelo multiclasse no conjunto de teste com *fine-tuning*.



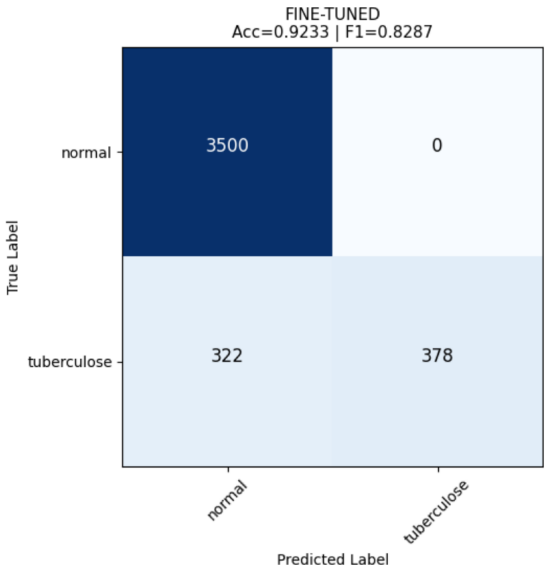

Fonte: Autor.

Tabela 12 – Resultados obtidos com e sem *Fine-Tuning* no conjunto de teste externo

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Sem *Fine-Tuning* | 91,24% | 95,24% | 73,71% | 79,67% |
| Com *Fine-Tuning* | 92,33% | 95,79% | 77,00% | 82,87% |

Fonte: Autor.

Figura 37 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com *fine-tuning*.




Fonte: Autor.

No conjunto de teste interno (Figura 36), o *fine-tuning* promoveu melhorias consistentes em todas as métricas: a acurácia aumentou de 99,57% para 99,81%, a precisão de 98,85% para 99,67%, o recall de 98,49% para 99,33% e o F1-score de 98,85% para 99,16%. A matriz de confusão, como ilustra a Figura 36, evidencia que os erros residuais após o *fine-tuning* são pontuais e distribuídos entre as classes, sem concentração em nenhum par específico. Isso, portanto, indica que o ajuste fino das camadas superiores do *backbone* contribuiu para uma melhor discriminação entre as classes mais próximas. O gráfico de perda e acurácia (Figura 35) mostra que o *fine-tuning* convergiu em aproximadamente 17 épocas, com a perda de validação mantendo-se estável e próxima à perda de treinamento, sugerindo boa capacidade de generalização sem sinais de *overfitting*.

No *dataset* externo (Figura 37), o impacto do *fine-tuning* foi igualmente positivo: o F1-score aumentou de 79,67% para 82,87%, a acurácia de 91,24% para 92,33% e o recall de 73,71% para 77,00%. A Figura 37 mostra que 378 dos 700 casos de tuberculose foram corretamente identificados após o *fine-tuning*, contra 332 na configuração anterior.

### 5.2.4 Impacto da Segmentação

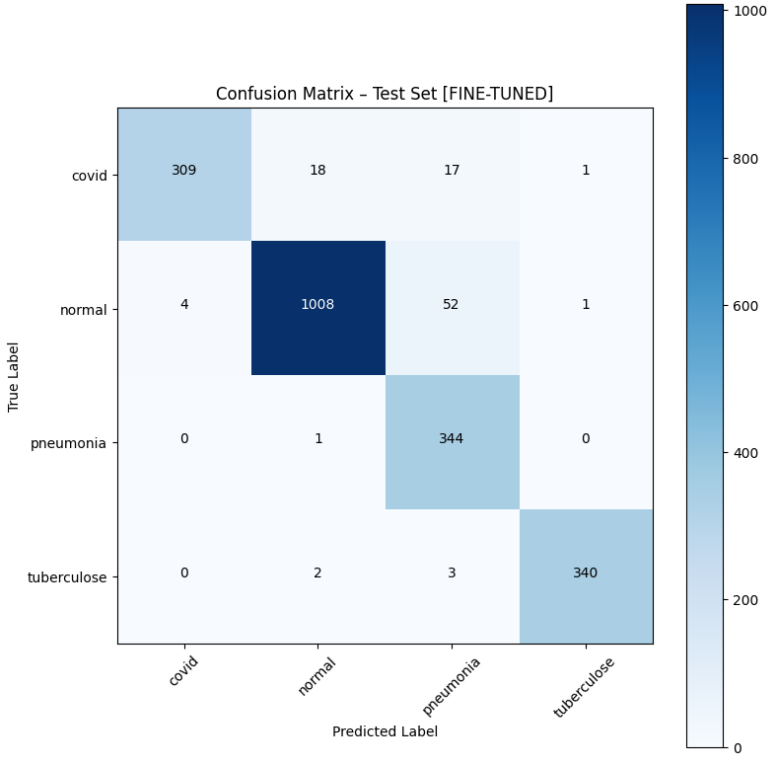
A última etapa da análise ablativa consistiu na avaliação do impacto da aplicação das máscaras de segmentação pulmonar sobre as imagens do *dataset*. O objetivo é verificar se a restrição da análise à região de interesse anatômica, ou seja, se eliminar o fundo e estruturas não pulmonares contribui para o desempenho de classificação. A Tabela 13 apresenta a comparação entre o modelo treinado sem segmentação (utilizando as imagens originais) e o modelo treinado com segmentação (*dataset_masked*), ambos com *fine-tuning* aplicado. As Figuras 38 e 39 apresentam a matriz de confusão e as curvas de treinamento do modelo com segmentação, respectivamente.

Tabela 13 – Resultados obtidos com e sem segmentação

| **Métricas** | **Acurácia** | **Precisão** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Sem Segmentação | 99,81% | 99,67% | 99,33% | 99,16% |
| Com Segmentação | 95,29% | 94,70% | 95,62% | 94,90% |

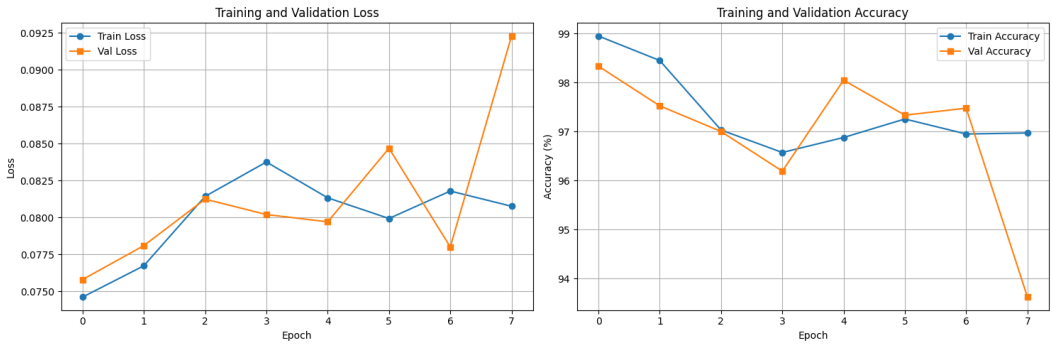
Fonte: Autor.

Figura 38 – Matriz de confusão do modelo multiclasse no conjunto de teste com segmentação.




Fonte: Autor.

Figura 39 – Gráfico de perda e acurácia do treinamento do *fine-tuning* com segmentação.



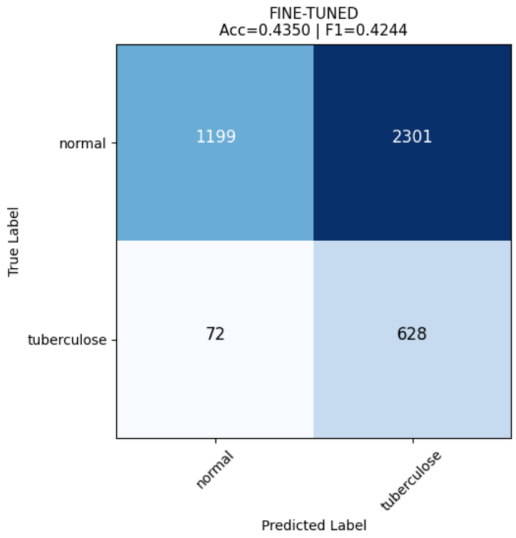

Fonte: Autor.

Tabela 14 – Resultados obtidos com e sem segmentação

| Métricas | Acurácia | Precisão | Recall | F1-Score |
|---|---|---|---|---|
| Sem segmentação | 92,33% | 95,79% | 77,00% | 82,87% |
| Com segmentação | 43,50% | 57,89% | 61,99% | 42,43% |

Fonte: Autor.

Figura 40 – Matriz de confusão do modelo multiclasse no conjunto de teste externo com segmentação.




Fonte: Autor.

Os resultados no conjunto de teste interno (Tabela 13) revelam que a aplicação da segmentação resultou em uma queda de desempenho em relação ao modelo sem segmentação: a acurácia reduziu de 99,81% para 95,29% e o F1-score de 99,16% para 94,90%. A matriz de confusão (Figura 38) indica que, com a segmentação, os principais erros se concentram na confusão entre as classes covid e normal (18 casos) e entre covid e pneumonia (17 casos), além de um número maior de erros na classificação da classe normal (52 casos classificados como pneumonia). O gráfico de treinamento (Figura 39) mostra que o modelo com segmentação convergiu em apenas 8 épocas, com acurácia de validação atingindo aproximadamente 97%, inferior à obtida sem segmentação. Além da maior variância nas curvas, sugerindo que a tarefa de classificação sobre imagens segmentadas apresenta maior dificuldade de otimização.

Uma hipótese para esse comportamento é que a aplicação das máscaras, ao zerar o fundo das imagens, elimina informações contextuais que o modelo aprendeu a explorar durante o treinamento sem segmentação, como padrões de textura nas regiões periféricas do pulmão e estruturas do mediastino, que podem ser relevantes para a distinção entre covid e pneumonia. Além disso, possíveis imprecisões nas máscaras de segmentação podem ter suprimido regiões pulmonares relevantes em algumas amostras, introduzindo ruído no processo de treinamento. Esses resultados sugerem que, embora a segmentação seja uma etapa de pré-processamento amplamente adotada na literatura para imagens médicas, seu benefício não é universal e depende da qualidade das máscaras e da arquitetura de classificação utilizada. No contexto do presente trabalho, o modelo sem segmentação demonstrou desempenho superior no *dataset* interno, enquanto a análise do impacto da segmentação no *dataset* externo será apresentada nas seções seguintes, no contexto das arquiteturas de *ensemble*.

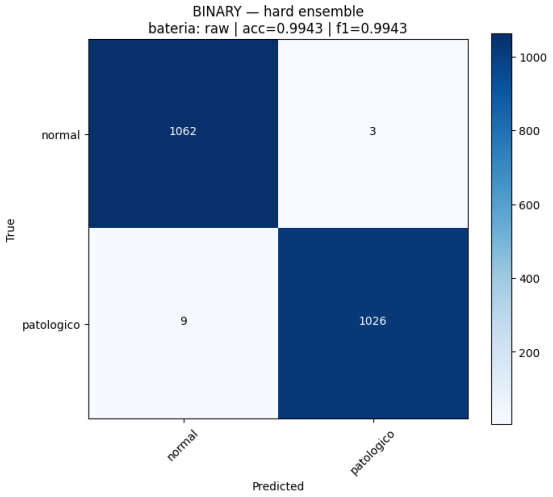
## 5.3 ESTÁGIO BINÁRIO

O estágio binário é responsável pela triagem inicial das imagens, classificando cada radiografia torácica como pertencente à classe normal ou patológico. Por ser compartilhado pelas duas arquiteturas propostas, paralela e em cascata, seus resultados são apresentados nesta seção de forma independente. O *ensemble* foi avaliado com os métodos descritos na Seção 4.4 para as duas baterias de dados.

### 5.3.1 Sem segmentação

A Figura 41 apresenta a matriz de confusão do modelo binário sem segmentação, utilizando o método de agregação *hard voting*, que obteve o melhor desempenho nessa configuração. A Tabela 15 apresenta as métricas dos três métodos de *ensemble*.

Figura 41 – Matriz de confusão do modelo binário.




Fonte: Autor.

O *ensemble* binário sem segmentação demonstrou desempenho expressivo, com acurácia e F1-score de 0,9943 para o método *hard voting*. A matriz de confusão (Figura 41) evidencia que, em um total de 2.100 amostras de teste, apenas 12 foram classificadas incorretamente: 3 casos normais foram identificados como patológicos e 9 casos patológicos foram classificados como normais. Esses 9 falsos negativos representam o erro mais crítico do sistema, pois correspondem a imagens patológicas que seriam encaminhadas como normais ao usuário final, sem prosseguir para a etapa de identificação da patologia específica.

A Tabela 15 mostra que os três métodos de *ensemble* apresentaram desempenho muito próximo, com a AUC igual a 0,9998 para todos eles, indicando elevada capacidade discriminativa do conjunto de modelos independentemente do método de agregação. O *hard voting* obteve leve vantagem em acurácia e F1-score (0,9943) em relação ao *soft* e *weighted voting* (0,9938), sendo selecionado como o método de referência para esta configuração.

Tabela 15 – Métricas de desempenho por estratégia no modelo binário. (✓= melhor modelo)

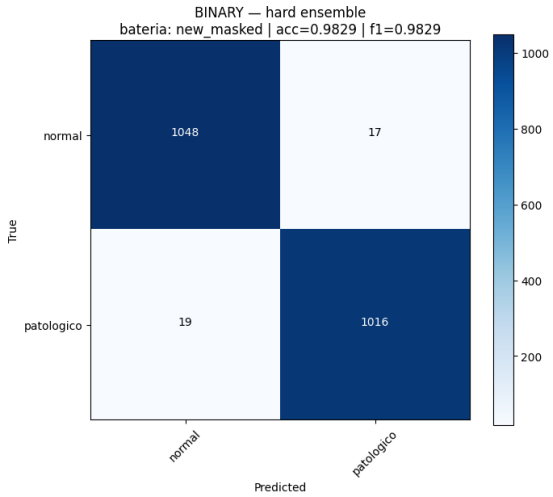
| **Estratégia** | ***Accuracy*** | ***F1-Score*** | ***AUC*** |
|---|---|---|---|
| *Hard*✓ | 0,9943 | 0,9943 | 0,9998 |
| *Soft* | 0,9938 | 0,9938 | 0,9998 |
| *Weighted* | 0,9938 | 0,9938 | 0,9998 |

Fonte: Autor.

### 5.3.2 Com segmentação

A Figura 42 apresenta a matriz de confusão do modelo binário com segmentação (dataset masked), e a Tabela 16 compara os três métodos de ensemble nessa configuração.

Figura 42 – Matriz de confusão do modelo binário com segmentação.




Fonte: Autor.

Tabela 16 – Métricas de desempenho por estratégia no modelo binário com segmentação. (✓= melhor modelo)

| **Estratégia** | **Acurácia** | ***F1-Score*** | ***AUC*** |
|---|---|---|---|
| Hard✓ | 0,9829 | 0,9829 | 0,9983 |
| Soft | 0,9819 | 0,9819 | 0,9983 |
| Weighted | 0,9819 | 0,9819 | 0,9983 |

Fonte: Autor.

Com a aplicação das máscaras de segmentação pulmonar, o ensemble binário obteve acurácia e F1-score de 0,9829 com *hard voting*, demonstrando um desempenho inferior ao obtido sem segmentação (0,9943). A matriz de confusão (Figura 42) mostra que o número de erros aumentou: 17 casos normais foram classificados como patológicos e 19 casos patológicos foram classificados como normais, totalizando 36 erros em 2.100 amostras. Esse resultado é consistente com o observado na análise ablativa (Seção 5.2), onde a segmentação também resultou em queda de desempenho no conjunto de teste interno do modelo individual.

Uma hipótese para esse comportamento é que a segmentação, ao zerar as regiões externas ao pulmão, elimina informações de contexto anatômico, como o contorno cardíaco e a forma do diafragma, que podem ser relevantes para a distinção binária entre imagens normais e patológicas em alguns casos. Além disso, possíveis imprecisões nas máscaras geradas pelo *CMUNeXt* podem ter introduzido variações artificiais nas regiões de fronteira pulmonar. A AUC de 0,9983 para todos os métodos, no entanto, indica que o modelo mantém elevada capacidade discriminativa, e a queda de desempenho se manifesta principalmente nos casos fronteiriços próximos ao limiar de decisão.

Em relação aos métodos de ensemble, o *hard voting* novamente obteve o melhor desempenho (acurácia e F1-score = 0,9829), enquanto *soft* e *weighted* voting resultaram em métricas idênticas (0,9819). Isso sugere que, para a tarefa binária com imagens segmentadas, a decisão majoritária discreta entre os modelos é mais robusta do que a média das probabilidades.

## 5.4 ARQUITETURA PARALELA

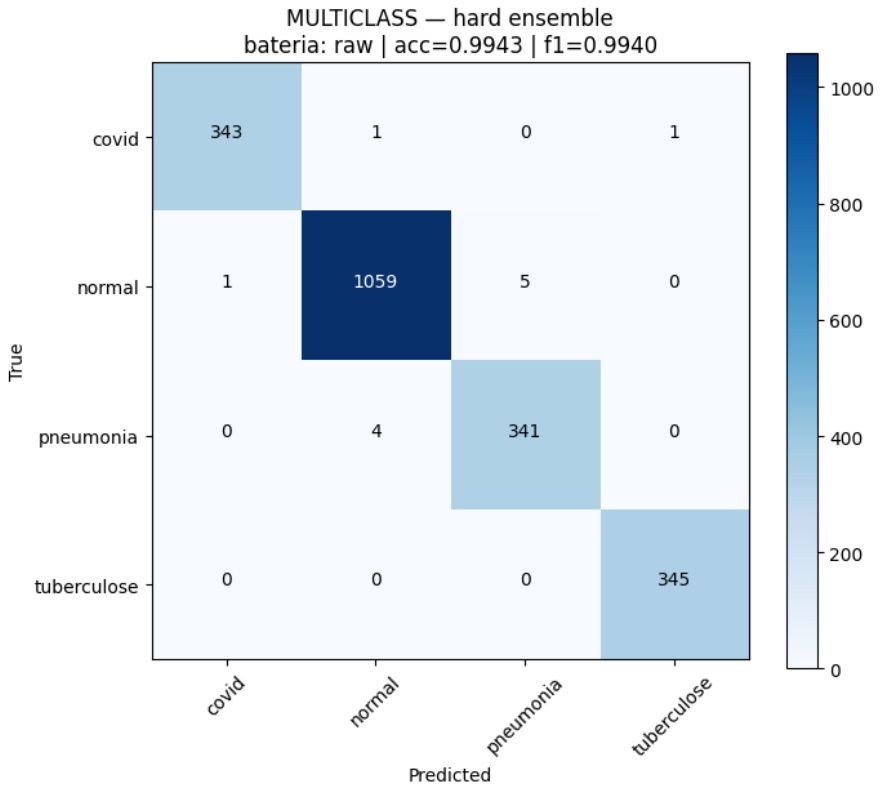
Nesta seção são apresentados os resultados do estágio multiclasse da arquitetura paralela, que opera de forma independente sobre as imagens sem condicionamento ao resultado do estágio binário. O módulo multiclasse classifica diretamente cada imagem em uma das quatro classes: covid, normal, pneumonia e tuberculose.

### 5.4.1 Estágio Multiclasse - Sem Segmentação

A Figura 43 apresenta a matriz de confusão do modelo multiclasse sem segmentação, com o método *hard voting*, e a Tabela 17 apresenta as métricas dos três

métodos de ensemble.

Figura 43 – Matriz de confusão do modelo multiclasse.




Fonte: Autor.

O estágio multiclasse sem segmentação apresentou desempenho elevado, com acurácia de 0,9943 e *F1-score* de 0,9940 para o método *hard voting*. A matriz de confusão (Figura 43) mostra que os erros são pontuais e distribuídos entre as classes: 1 caso de covid classificado como normal, 1 como tuberculose, 1 caso de normal classificado como covid e 5 como pneumonia, e 4 casos de pneumonia classificados como normal. A classe tuberculose foi classificada corretamente em todos os 345 casos, demonstrando que os padrões radiológicos da tuberculose são altamente discriminativos para o ensemble proposto.

A covid apresentou o maior número de confusões relativas, com 2 erros em 345 amostras (taxa de acerto de 99,42%). A confusão entre covid e normal é esperada na literatura, uma vez que as manifestações radiológicas da COVID-19 em estágios iniciais

podem ser sutis e assemelhar-se a imagens normais. Os três métodos de agregação apresentaram desempenho equivalente em acurácia (0,9943), com a votação por maioria obtendo leve vantagem no *F1-score* (0,9940 vs 0,9938).

Tabela 17 – Métricas de desempenho por estratégia no modelo multiclasse. (✓= melhor modelo)

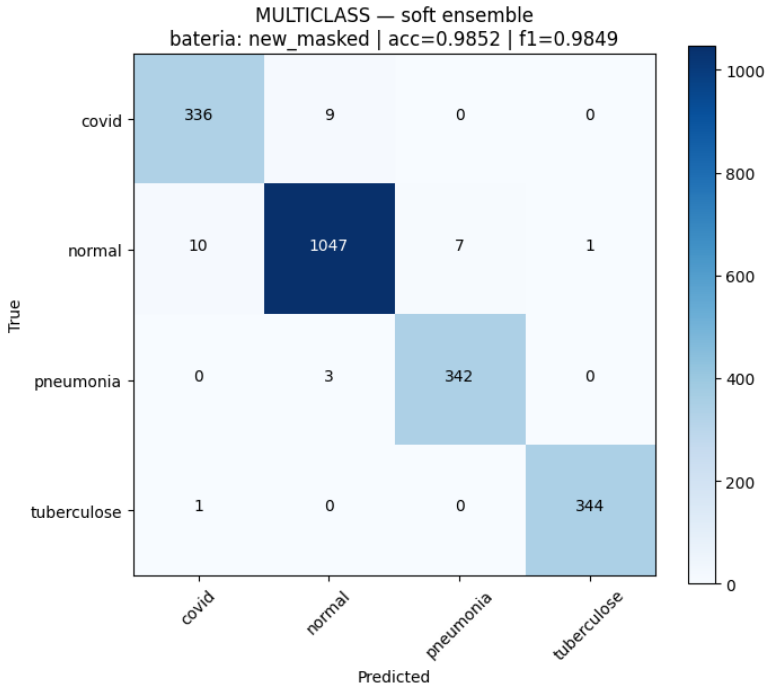
| **Estratégia** | **Acurácia** | ***F1-Score*** | ***AUC*** |
|---|---|---|---|
| Hard✓ | 0,9943 | 0,9940 | 0,9996 |
| Soft | 0,9943 | 0,9938 | 0,9996 |
| Weighted | 0,9943 | 0,9938 | 0,9996 |

Fonte: Autor.

### 5.4.2 Estágio Multiclasse - Com Segmentação

A Figura 44 apresenta a matriz de confusão do modelo multiclasse com segmentação e a Tabela 18 compara os três métodos de ensemble nessa configuração.

Figura 44 – Matriz de confusão do modelo multiclasse com segmentação.




Fonte: Autor.

Tabela 18 – Métricas de desempenho por estratégia no modelo multiclasse com segmentação. (✓= melhor modelo)

| **Estratégia** | **Acurácia** | ***F1-Score*** | ***AUC*** |
|---|---|---|---|
| Hard | 0,9833 | 0,9830 | 0,9993 |
| Soft✓ | 0,9852 | 0,9849 | 0,9993 |
| Weighted | 0,9852 | 0,9849 | 0,9993 |

Fonte: Autor.

Com a aplicação das máscaras de segmentação, o ensemble multiclasse obteve acurácia de 0,9852 e *F1-score* de 0,9849 com *soft voting*, resultado inferior ao obtido sem segmentação (0,9943), mas ainda altamente competitivo. A matriz de confusão (Figura 44) revela que os principais erros se concentram na classe covid (9 casos classificados como normal) e na classe normal (10 casos classificados como covid e 7 como pneumonia), padrão de confusão semelhante ao observado no modelo individual com segmentação (Seção 2.7).

Em contraste com a configuração sem segmentação, onde o *hard voting* foi o melhor método, com segmentação o *soft voting* e o *weighted* voting obtiveram melhor desempenho (acurácia e *F1-score* = 0,9852) em comparação ao *hard voting* (0,9833). Isso sugere que, com imagens segmentadas, a incorporação do grau de confiança de cada modelo na decisão final, característica do *soft* e *weighted voting*, é mais benéfica do que a votação discreta, possivelmente porque as máscaras introduzem maior variabilidade nas predições individuais, tornando mais informativa a distribuição de probabilidades de cada modelo. A classe tuberculose novamente apresentou desempenho próximo ao perfeito (344/345 corretos), e a classe pneumonia manteve alta taxa de acerto (342/345).

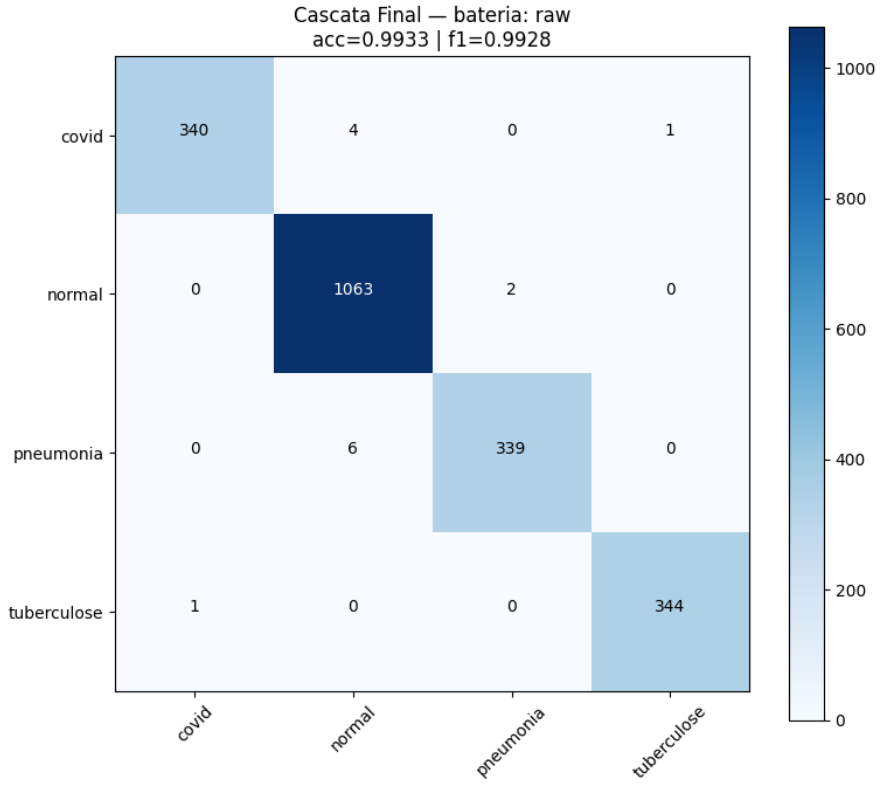
## 5.5 ARQUITETURA CASCATA

A arquitetura em cascata conecta os dois estágios sequencialmente: as imagens identificadas como normais pelo estágio binário são diretamente classificadas como normais no resultado final, enquanto apenas as imagens identificadas como patológicas são encaminhadas ao estágio multiclasse para identificação da patologia específica. Esse mecanismo de dupla verificação tem como objetivo reduzir falsos positivos no resultado final, ao condicionar a classificação multiclasse à confirmação prévia de anormalidade.

### 5.5.1 *Pipeline* - Sem Segmentação

A Figura 45 apresenta a matriz de confusão do pipeline em cascata sem segmentação, e a Tabela 19 apresenta as métricas do resultado final (método *soft voting*).

Figura 45 – Matriz de confusão do modelo multiclasse em cascata.




Fonte: Autor.

Tabela 19 – Métricas de desempenho do Resultado em Cascata.

| **Métrica** | **Acurácia** | **Precisão** | ***Recall*** | ***F1-Score*** |
|---|---|---|---|---|
| Valor | 0,9933 | 0,9947 | 0,9908 | 0,9928 |

Fonte: Autor.

O *pipeline* em cascata sem segmentação alcançou acurácia de 0,9933, precisão de 0,9947, *recall* de 0,9908 e *F1-score* de 0,9928, utilizando *soft voting* como método de agregação. A matriz de confusão (Figura 45) mostra que os principais erros residuais

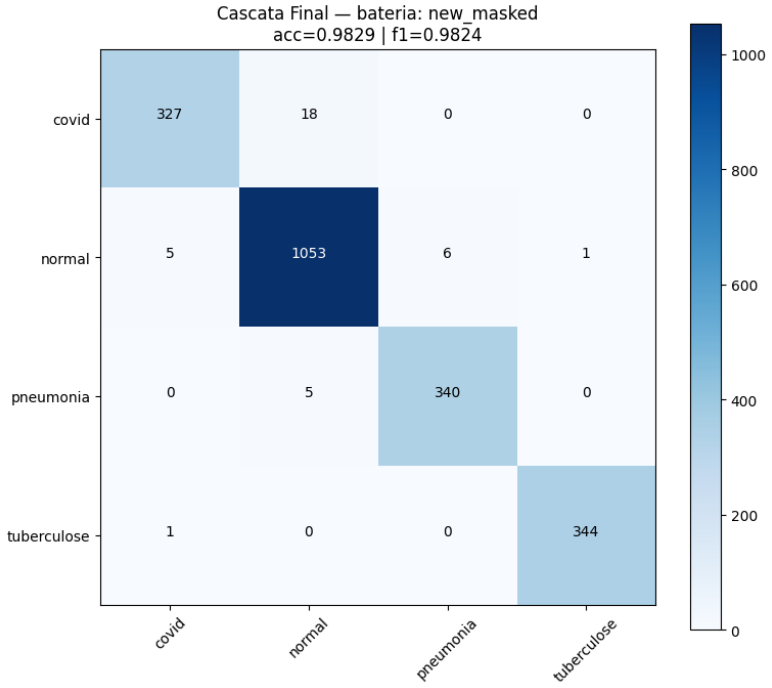
da cascata se concentram na classe covid: 18 casos de covid foram classificados como normal e 5 casos de normal foram classificados como covid. Esse padrão indica que o estágio binário, ao encaminhar corretamente a maioria das imagens patológicas ao estágio 2, ainda deixa escapar alguns casos de covid que apresentam manifestações radiológicas sutis e são identificados como normais já na triagem inicial.

Comparando com a arquitetura paralela sem segmentação (*F1-score* = 0,9940), a cascata apresentou leve queda no *F1-score* (0,9928), mas com *recall* mais equilibrado entre as classes. Esse resultado é esperado, pois na cascata os erros do estágio binário são irreversíveis, uma imagem patológica classificada como normal pelo E1 nunca chega ao E2, enquanto na arquitetura paralela o modelo multiclasse tem a oportunidade de classificar diretamente todos os casos. A vantagem da cascata, contudo, está na maior precisão para os casos em que o E1 detecta patologia, reduzindo o risco de classificações multiclasse errôneas em imagens essencialmente normais.

### 5.5.2 *Pipeline* - Com Segmentação

A Figura 46 apresenta a matriz de confusão do *pipeline* em cascata com segmentação, e a Tabela 20 apresenta as métricas do resultado final (método *hard voting*).

Figura 46 – Matriz de confusão do modelo multiclasse em cascata.




Fonte: Autor.

Tabela 20 – Métricas de desempenho do Resultado em Cascata.

| **Métrica** | **Acurácia** | **Precisão** | ***Recall*** | ***F1-Score*** |
|---|---|---|---|---|
| Valor | 0,9829 | 0,9851 | 0,9798 | 0,9824 |

Fonte: Autor.

O *workflow* em cascata com segmentação alcançou acurácia de 0,9829, precisão de 0,9851, *recall* de 0,9798 e *F1-score* de 0,9824, com *hard voting*. A matriz de confusão (Figura 46) mostra uma distribuição de erros mais concentrada nas classes covid e pneumonia: 4 casos de covid foram classificados como normal e 1 como tuberculose; 6 casos de pneumonia foram classificados como normal; e 2 casos de normal foram classificados como pneumonia.

Em relação à cascata sem segmentação, o *pipeline* com segmentação apresentou queda em todas as métricas, com redução de 0,0104 pontos no *F1-score* (0,9928 $\rightarrow$ 0,9824). Esse resultado reforça o padrão observado ao longo de toda a análise: a segmentação, embora conceitualmente adequada para restringir a análise à

região pulmonar, reduz o desempenho no dataset interno em todas as configurações avaliadas. Uma possível explicação é que o dataset de treinamento, composto majoritariamente por imagens de fontes públicas e com variação de qualidade nas máscaras, pode ter introduzido inconsistências que prejudicam a generalização do modelo treinado com imagens segmentadas.

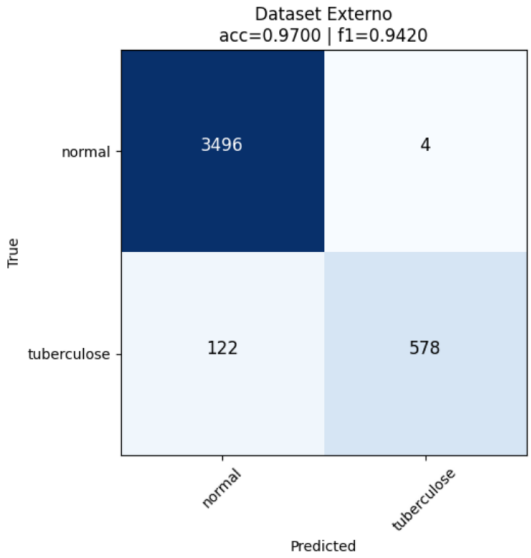
## 5.6 AVALIAÇÃO COM DATASET EXTERNO

A avaliação com o dataset externo tem como objetivo verificar a capacidade de generalização do sistema proposto para imagens provenientes de uma fonte de dados completamente distinta da utilizada no treinamento. A avaliação foi realizada utilizando o pipeline em cascata com soft voting, com e sem a aplicação das máscaras de segmentação.

### 5.6.1 Sem segmentação

A Figura 47 apresenta a matriz de confusão do pipeline em cascata no dataset externo sem segmentação, e a Tabela 21 apresenta as métricas correspondentes.

Figura 47 – Matriz de confusão no *dataset* externo.




Fonte: Autor.

Tabela 21 – Métricas de desempenho no dataset externo.

| **Métrica** | **Acurácia** | **Precisão** | ***Recall*** | ***F1-Score*** |
|---|---|---|---|---|
| Modelo | 0,9700 | 0,9797 | 0,9123 | 0,9420 |

Fonte: Autor.

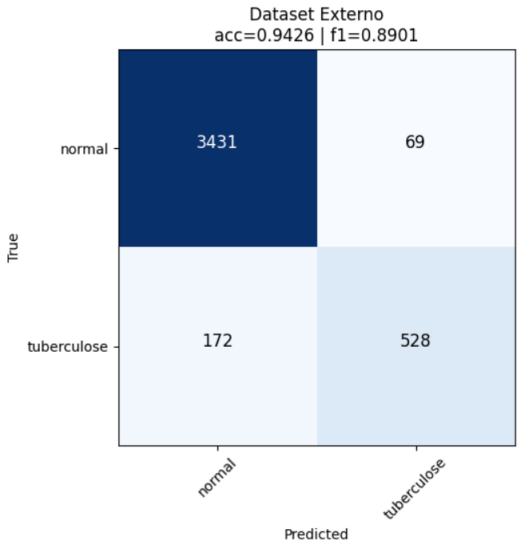
No dataset externo sem segmentação, o modelo alcançou acurácia de 0,9700, precisão de 0,9797, recall de 0,9123 e F1-score de 0,9420. A matriz de confusão (Figura 47) mostra que, dos 3.500 casos normais, 3.496 foram corretamente identificados e apenas 4 foram classificados incorretamente como tuberculose. Para a classe tuberculose, 578 dos 700 casos foram corretamente identificados, enquanto 122 foram classificados como normal, o que representa a principal fonte de erro do sistema nessa base de dados.

O desempenho no dataset externo, embora inferior ao obtido no dataset interno (F1-score = 0,9928), demonstra boa capacidade de generalização do sistema para dados fora da distribuição de treinamento. A queda de recall para a classe tuberculose (578/700 = 82,57%) é esperada, dado que as imagens externas foram adquiridas em condições distintas das do treinamento, já que provém de diferentes equipamentos, protocolos de aquisição e populações de pacientes, o que introduz variabilidade não representada no conjunto de treinamento. A alta precisão para tuberculose (578/(578+4) = 99,31%) indica que, quando o modelo identifica um caso como tuberculose, essa classificação é altamente confiável.

### 5.6.2 Com segmentação

A Figura 48 apresenta a matriz de confusão do pipeline em cascata no *dataset* externo com segmentação, e a Tabela 22 apresenta as métricas correspondentes.

Figura 48 – Matriz de confusão no *dataset* externo com segmentação.




Fonte: Autor.

Tabela 22 – Métricas de desempenho no *dataset* externo com segmentação.

| **Métrica** | **Acurácia** | **Precisão** | ***Recall*** | ***F1-Score*** |
|---|---|---|---|---|
| Modelo | 0,9426 | 0,9183 | 0,8673 | 0,8901 |

Fonte: Autor.

Com segmentação, o modelo alcançou acurácia de 0,9426, precisão de 0,9183, *recall* de 0,8673 e *F1-score* de 0,8901 no bando de dados externo, ou seja, desempenho inferior ao obtido sem segmentação (*F1-score* = 0,9420). A matriz de confusão (Figura 48) mostra que a queda de desempenho se distribui entre as duas classes: 69 casos normais foram classificados como tuberculose (contra 4 sem segmentação) e 172 casos de tuberculose foram classificados como normal (contra 122 sem segmentação). O aumento expressivo de falsos positivos para a classe normal (69 vs 4) é particularmente notável e sugere que as máscaras geradas pelo *CMUNeXt* sobre imagens externas, não vistas durante o treinamento do segmentador, podem apresentar menor qualidade, introduzindo artefatos que confundem o classificador.

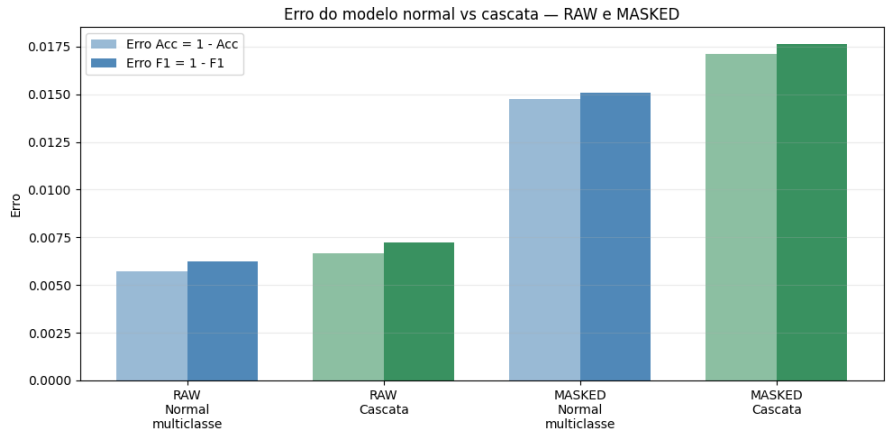
Esse resultado é o único, em toda a avaliação, em que a segmentação prejudicou o desempenho no dataset externo de forma mais acentuada do que no dataset interno. A diferença de *F1-score* entre as duas configurações no *dataset* externo (0,9420 vs

0,8901 = $\Delta 0,0519$) é maior do que a observada no *dataset* interno (0,9928 vs 0,9824 = $\Delta 0,0104$), indicando que a segmentação de imagens externas representa um ponto de fragilidade do *workflow* proposto. Esse achado reforça a importância de avaliar sistemas de segmentação e classificação em conjuntos de dados completamente independentes, uma vez que a qualidade das máscaras impacta diretamente o desempenho da etapa de classificação subsequente.

## 5.7 COMPARAÇÃO ENTRE ARQUITETURAS

A Figura 49 apresenta um comparativo visual do erro de acurácia ($1 - \text{Acurácia}$) e do erro de *F1-score* ($1 - \textit{F1-Score}$) entre as quatro configurações avaliadas no *dataset* interno: arquitetura paralela sem segmentação (*RAW* Normal/Multiclasse), arquitetura em cascata sem segmentação (*RAW* Cascata), arquitetura paralela com segmentação (*MASKED* Normal/Multiclasse) e arquitetura em cascata com segmentação (*MASKED* Cascata).

Figura 49 – Comparação de desempenho entre as arquiteturas propostas sem (*raw*) e com segmentação (*masked*).




Fonte: Autor.

O gráfico evidencia que as configurações sem segmentação apresentam erros consistentemente menores do que as configurações com segmentação, confirmando o padrão observado ao longo de toda a análise ablativa e dos resultados por arquitetura. Entre as configurações *RAW*, a diferença de erro entre a arquitetura paralela e a cascata é

pequena (aproximadamente 0,0012 pontos em acurácia e *F1-score*), indicando que as duas abordagens são equivalentes no *dataset* interno sem segmentação.

Dessa forma, para as configurações com segmentação, o erro é maior em ambas as arquiteturas, com a cascata apresentando erro ligeiramente superior à arquitetura paralela. Esse resultado contrasta com a expectativa teórica de que a cascata, ao filtrar casos normais na primeira etapa, deveria reduzir erros no estágio multiclasse. A explicação mais provável é que, com segmentação, o estágio binário comete mais erros de triagem (como observado na Seção 5.3), e esses erros se propagam irreversivelmente para o resultado final da cascata. Esse fenômeno é conhecido como propagação de erro em sistemas em cascata.

Em síntese, a arquitetura paralela sem segmentação apresentou o melhor desempenho global no *dataset* interno, com *F1-score* de 0,9940, seguida pela cascata sem segmentação (*F1-score* = 0,9928). Para o bando de dados externo, a configuração sem segmentação também superou a com segmentação em ambas as arquiteturas, com a cascata *raw* atingindo *F1-score* de 0,9420. Esses resultados indicam que, no contexto do presente trabalho, o pipeline sem segmentação apresenta maior robustez e generalização, enquanto a segmentação, apesar de conceitualmente motivada, não resultou em ganho de desempenho mensurável nas condições avaliadas.

## 5.8 COMPARAÇÃO COM A LITERATURA

A Tabela 23 apresenta uma comparação entre o sistema proposto e trabalhos relacionados da literatura que abordam a classificação de patologias pulmonares em imagens de raio-X torácico. De maneira geral, observa-se que os trabalhos revisados adotam abordagens variadas em termos de arquitetura, tamanho do dataset e número de classes, o que torna a comparação direta entre os resultados limitada, uma vez que diferenças no conjunto de dados, no protocolo de avaliação e nas classes consideradas influenciam diretamente as métricas reportadas.

Entre os trabalhos analisados, a maioria aborda a classificação binária, distinguindo imagens normais de patológicas, enquanto o presente trabalho propõe uma classificação multiclasse de quatro categorias clínicas relevantes (normal, pneumonia, COVID-19 e tuberculose), tarefa intrinsecamente mais complexa. Apesar dessa maior exigência, o sistema proposto na configuração sem segmentação atingiu acurácia de

Tabela 23 – Comparação com trabalhos relacionados da literatura. Seg. = Segmentação; Ens. = *Ensemble*; FT = *Fine-Tuning*; — = não reportado.

| **Trabalho** | **Dataset(s)** | **Nº Imagens** | **Classes** | **Seg.** | **Ens.** | **FT** | **Acurácia** | **AUC** | **F1-Score** |
|---|---|---|---|---|---|---|---|---|---|
| (COLAÇO, 2022) | Chest X-ray14 + Centro de Telemedicina (Privado) | 11.510 (base) + 200 (teste) | 2 classes: Normal vs. Patologia | Sim (ResNet50 U-Net) | Sim | Sim | 79,5% (ens.) / 81,5% (ResNet18) | 89,10% | 80,80% |
| (HASSAN, 2026) | ChestX6 | 18.036 | 6 classes: Normal, COVID-19, TB, Pneumonia, Enfisema | Não (*Grad-CAM*) | Não | Não | 93,38%–96,52% | 99,00% | 96,37% |
| (CHICHANOSKI, 2022) | MC, Shenzhen e *dataset* misto COVID | 800 (seg.) + 7.103 (class.) | 3 classes: Normal, COVID-19, Pneumonia | Sim (U-Net) | Não | Não | 99,47% (COVID) / 96,13% (Normal) | — | 92,70% |
| (SHARMA et al., 2024) | MC, Shenzhen e NIAID TB Portal | 704 (seg.) + 1.400 (class.) | 2 classes: Normal vs. Tuberculose | Sim (U-Net) | Não | Não | 99,29% | 99,90% | 99,29% |
| (BHARATI; PODDER; MONDAL, 2020) | NIH Chest X-ray | $>$ 112.000 | 2 classes: Normal vs. Doença | Sim (STN integrada) | Sim (VDSNet) | Não | 73,00% | — | — |
| **Proposto (*RAW*)** | ChestX6 + *Dataset* externo | $\approx$ 13.900 (treino) + 4.200 (ext.) | 4 classes: Normal, COVID-19, Pneumonia, Tuberculose | Não | Sim (5 modelos) | Sim | **99,33%** | **99,98%** | **99,28%** |
| **Proposto (*Masked*)** | ChestX6 + *Dataset* externo | $\approx$ 13.900 (treino) + 4.200 (ext.) | 4 classes: Normal, COVID-19, Pneumonia, Tuberculose | Sim (*CMUNeXt*) | Sim (5 modelos) | Sim | **98,29%** | **99,83%** | **98,24%** |

Fonte: Autor.

99,33% e *F1-score* de 99,28%, resultados competitivos com os melhores valores reportados na literatura para tarefas de menor complexidade. Adicionalmente, o uso de ensemble com cinco arquiteturas heterogêneas e *fine-tuning* em duas fases, combinado com a avaliação em um *dataset* externo independente, confere ao sistema proposto um grau de robustez e validação superior ao de parte dos trabalhos comparados, que se restringem à avaliação em conjuntos de teste provenientes da mesma distribuição do treinamento.

# 6 CONCLUSÃO

Este último capítulo apresenta as principais conclusões do trabalho, destacando os resultados mais relevantes obtidos a partir da aplicação dos métodos propostos. Além disso, são discutidas as limitações do estudo e propostas direções para trabalhos futuros que possam ampliar e aprimorar os achados aqui apresentados.

## 6.1 PRINCIPAIS CONCLUSÕES

As doenças pulmonares representam um desafio crescente para os sistemas de saúde no Brasil e no mundo, com impacto direto na morbimortalidade da população. A pandemia de COVID-19 agravou esse cenário ao sobrecarregar os recursos de saúde, reduzir o acesso a diagnósticos e tratamentos de outras patologias pulmonares (como a tuberculose, cujos óbitos no Brasil cresceram 85,71% entre 2018 e 2023) e evidenciar a urgência de soluções tecnológicas capazes de ampliar a capacidade diagnóstica sem depender exclusivamente da disponibilidade de profissionais especializados.

Motivado por essa problemática, este trabalho propôs o desenvolvimento de um sistema de apoio ao diagnóstico de patologias pulmonares em imagens de raio-X torácico, baseado em um *pipeline* de *deep learning* composto por três módulos integrados: segmentação pulmonar, classificação binária em *ensemble* e classificação multiclasse em *ensemble*, avaliados em duas arquiteturas distintas: a paralela e a em cascata.

O módulo de segmentação, implementado com a arquitetura CMUNeXt, demonstrou desempenho expressivo na delimitação da região pulmonar, atingindo *IoU* de 95,16% e *Dice* de 97,52%, com *recall* de 97,00% e especificidade de 99,00%. Esses resultados confirmam a viabilidade do CMUNeXt como ferramenta de pré-processamento para imagens de raio-X, gerando máscaras de qualidade suficiente para o *pipeline* de classificação subsequente.

O módulo de classificação foi composto por cinco arquiteturas CNN treinadas individualmente, compreendendo Híbrido (ResNet152 + EfficientNet-B5), ResNet50, DenseNet121, VGG16 e EfficientNet-B0, com as predições combinadas por três métodos de *ensemble learning*: *hard voting*, *soft voting* e média ponderada pela *AUC*. A análise ablativa conduzida ao longo do desenvolvimento evidenciou a contribuição positiva de cada componente incorporado ao *pipeline*. A aplicação de *data augmentation* resultou

em ganho de 16,79 pontos percentuais no *F1-score* no *dataset* externo (de 59,35% para 76,14%), demonstrando que a diversificação artificial das amostras de treinamento é fundamental para a generalização do modelo. O *fine-tuning* em duas fases promoveu incrementos consistentes em todas as métricas, tanto no *dataset* interno (*F1-score* de 98,85% para 99,16%) quanto no externo (*F1-score* de 79,67% para 82,87%), confirmando que o ajuste progressivo das camadas superiores do *backbone* ao domínio médico melhora a capacidade discriminativa sem comprometer a generalização. O *Test Time Augmentation* contribuiu adicionalmente com ganhos de precisão no *dataset* externo, ao produzir estimativas mais robustas por meio da média de múltiplas predições aumentadas.

Os resultados finais dos *ensembles* demonstraram desempenho altamente competitivo. No estágio binário sem segmentação, o *ensemble* alcançou acurácia e *F1-score* de 0,9943, com *AUC* de 0,9998, um resultado que evidencia a elevada capacidade discriminativa do conjunto de modelos para a triagem entre imagens normais e patológicas. No estágio multiclasse sem segmentação, o desempenho foi equivalente, com acurácia de 0,9943 e *F1-score* de 0,9940, classificando corretamente a totalidade dos casos de tuberculose e com erros residuais concentrados na distinção entre COVID-19 e normal, par de classes historicamente desafiador na literatura de classificação de raio-X.

A comparação entre as arquiteturas paralela e em cascata revelou que, no *dataset* interno, o desempenho é equivalente para a configuração sem segmentação (*F1-score* de 0,9940 e 0,9928, respectivamente), enquanto a arquitetura paralela apresenta leve vantagem com segmentação. No *dataset* externo, o *pipeline* em cascata sem segmentação atingiu *F1-score* de 0,9420, demonstrando boa capacidade de generalização para imagens fora da distribuição de treinamento. A configuração com segmentação apresentou queda mais acentuada no *dataset* externo (*F1-score* = 0,8901), evidenciando que a qualidade das máscaras geradas sobre imagens externas representa um ponto de fragilidade do *pipeline* que merece atenção em trabalhos futuros.

Um resultado transversal a todas as configurações avaliadas foi a superioridade do *pipeline* sem segmentação sobre o *pipeline* com segmentação, tanto no *dataset* interno quanto no externo. Embora a segmentação seja conceitualmente motivada, ao restringir a análise à região pulmonar e eliminar informações de fundo irrelevantes, os resultados indicam que, nas condições avaliadas, as informações contextuais preservadas nas imagens originais contribuem positivamente para a classificação. Esse achado está

alinhado com trabalhos recentes da literatura que questionam a contribuição universal da segmentação em *pipelines* de classificação de imagens médicas (TEIXEIRA et al., 2021), e reforça a necessidade de avaliar empiricamente cada componente do *pipeline* antes de incorporá-lo ao sistema final.

Em síntese, o sistema proposto neste trabalho demonstrou ser uma solução promissora para o auxílio ao diagnóstico de patologias pulmonares em raio-X torácico, capaz de classificar corretamente a grande maioria das imagens em quatro categorias clínicas relevantes (normal, pneumonia, COVID-19 e tuberculose), com desempenho superior ao de modelos individuais e com evidências de generalização para dados externos. A abordagem multi-modelo com *ensemble learning* comprovou que a combinação de arquiteturas heterogêneas é capaz de mitigar os erros individuais de cada modelo, resultando em predições mais robustas e confiáveis, o que representa um requisito fundamental para qualquer sistema de apoio ao diagnóstico em contexto clínico real.

## 6.2 TRABALHOS FUTUROS

Apesar dos resultados promissores obtidos, o sistema desenvolvido apresenta oportunidades de melhoria que podem ser exploradas em trabalhos futuros. As principais direções identificadas são descritas a seguir.

### 6.2.1 Ampliação e Diversificação do *Dataset*

O desempenho de modelos de *deep learning* é fortemente dependente da quantidade e diversidade dos dados de treinamento. A incorporação de novos conjuntos de dados públicos e, especialmente, de imagens provenientes de instituições clínicas brasileiras (com suas particularidades de equipamento, protocolo de aquisição e população de pacientes) poderia melhorar significativamente a capacidade de generalização do sistema. Além disso, a inclusão de classes adicionais de patologias pulmonares, como derrame pleural, atelectasia e nódulos pulmonares, ampliaria o escopo clínico do sistema e aumentaria seu potencial de aplicação prática.

### 6.2.2 Aprimoramento do Módulo de Segmentação

O módulo de segmentação, embora tenha apresentado resultados expressivos na tarefa de delimitação pulmonar isoladamente, não resultou em ganho de desempenho no *pipeline* de classificação em nenhuma das configurações avaliadas, nem na base de dados interna nem na externa. Esse resultado sugere que o sistema de classificação é capaz de extrair informações relevantes das imagens originais sem necessitar da restrição à região pulmonar, e que a segmentação, nas condições do presente trabalho, introduz variabilidade suficiente para superar seus potenciais benefícios. Uma direção de trabalho futuro é o aprimoramento desse módulo, tanto por meio do treinamento prolongado do CMUNeXt, já que o modelo atual foi treinado com apenas 3 épocas, deixando margem para investigação de convergência a longo prazo, comum em outros trabalhos, como por exemplo em (COLAÇO, 2022) com 30 épocas.

Além disso, outro caminho é treinar com conjuntos de dados mais diversos e representativos, quanto pela avaliação de arquiteturas mais recentes de segmentação médica. Adicionalmente, a aplicação de pós-processamento morfológico sobre as máscaras geradas, como operações de abertura e fechamento, e um recorte (*crop*) automático da região pulmonar segmentada, eliminando ao máximo o fundo preto residual, poderia reduzir a variabilidade introduzida pelas máscaras e fornecer regiões de interesse mais precisas e consistentes para a etapa de classificação.

### 6.2.3 Representação por *Patches* Internos ao Pulmão

Uma abordagem alternativa e inovadora para a representação das imagens segmentadas consiste na extração de pequenos *patches* de conteúdo (janelas de $8 \times 8$ *pixels*) a partir das regiões internas ao pulmão segmentado, combinando-os para formar uma nova imagem composta exclusivamente por fragmentos de tecido pulmonar. Formalmente, a imagem resultante teria dimensão $k \cdot 8 \times k \cdot 8$ *pixels*, onde $k$ representa o número de *patches* na direção horizontal (e vertical, assumindo disposição quadrada). Essa representação, não necessariamente interpretável visualmente por humanos, concentraria apenas as informações texturais e estruturais internas ao pulmão, eliminando completamente o fundo e as bordas, que podem introduzir ruído no processo de classificação. Essa técnica poderia ser aplicada tanto para a geração dos dados de

treinamento quanto para as imagens de inferência, e sua viabilidade como estratégia de pré-processamento para redes CNN merece investigação experimental.

### 6.2.4 Otimização de Hiperparâmetros e Arquiteturas

O presente trabalho utilizou configurações de hiperparâmetros definidas com base na literatura e em experimentos preliminares. A realização de uma busca sistemática de hiperparâmetros, por meio de técnicas como *grid search*, *random search* ou otimização bayesiana, para cada modelo do *ensemble* poderia resultar em ganhos adicionais de desempenho. Adicionalmente, a avaliação de arquiteturas mais recentes, como os modelos baseados em *Vision Transformers* (ViT) e suas variantes híbridas com CNNs, representa uma direção promissora, dado o crescente desempenho dessas arquiteturas em tarefas de classificação de imagens médicas (HUSSAIN et al., 2025).

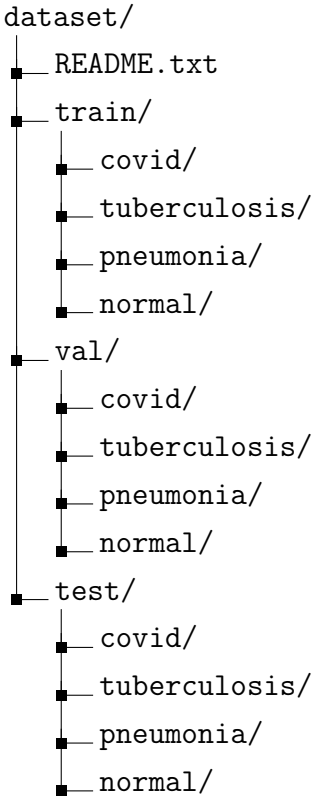
### 6.2.5 Validação Clínica

O sistema proposto foi avaliado exclusivamente em *datasets* públicos e em um *dataset* externo de validação. A realização de um estudo de validação clínica prospectiva, com imagens coletadas em ambiente hospitalar real e avaliação comparativa com o diagnóstico de radiologistas experientes, seria o passo necessário para avaliar o real potencial do sistema como ferramenta de apoio ao diagnóstico. Esse tipo de validação é essencial para identificar limitações não capturadas por métricas de *benchmark* e para adaptar o sistema às necessidades e fluxos de trabalho do ambiente clínico brasileiro.

# APÊNDICE A – ESTRUTURA DE DIRETÓRIOS DO *DATASET*

A seguir é apresentada a estrutura de diretórios do *dataset* composto neste trabalho, disponibilizado publicamente em (OLIVEIRA; SILVA, 2026).

Figura 50 – Estrutura de diretórios do *dataset*.




Fonte: Autor.

# REFERÊNCIAS

ALINSAIF, S. Covid-19 image classification: A comparative performance analysis of hand-crafted vs. deep features. **Computation**, v. 12, n. 1, p. 7, 2024.

ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, n. 1, p. 53, 2021.

BBC News Brasil. **Como a Covid afetou os pulmões e deixou sequelas em milhares de pacientes**. 2020. Acessado em: 1 dez. 2024. Disponível em: https://www.bbc.com/portuguese/geral-53398537.

BEHZADI-KHORMOUJI, H. et al. Deep learning, reusable and problem-based architectures for detection of consolidation on chest x-ray images. **Computer Methods and Programs in Biomedicine**, v. 185, p. 105162, 2020.

BELLMAN, R. **An Introduction to Artificial Intelligence: Can Computers Think?** Boston, MA: Boyd & Fraser Publishing Company, 1978.

BERGMANN, D.; STRYKER, C. **What is a Loss Function?** 2025. Acesso em: 19 maio 2025. Disponível em: https://www.ibm.com/think/topics/loss-function.

BHARATI, S.; PODDER, P.; MONDAL, M. R. H. Hybrid deep learning for detecting lung diseases from x-ray images. **Informatics in Medicine Unlocked**, Elsevier, v. 20, p. 100391, 2020. Disponível em: https://doi.org/10.1016/j.imu.2020.100391.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York, NY: Springer, 2006. ISBN 978-0387310732.

BUSLAEV, A. et al. Albumentations: Fast and flexible image augmentations. **Information**, v. 11, n. 2, p. 125, 2020. Disponível em: https://www.mdpi.com/2078-2489/11/2/125.

CARIN, L. et al. **Introduction to Machine Learning**. 2019. Course. Disponível em: https://www.coursera.org/learn/machine-learning-duke.

CATAL, C. Performance evaluation metrics for software fault prediction studies. **Acta Polytechnica Hungarica**, v. 9, n. 4, p. 193–209, 2012. Disponível em: https://acta.uni-obuda.hu/Catal_36.pdf.

CHARNIAK, E.; MCDERMOTT, D. **Introduction to Artificial Intelligence**. Boston, MA: Addison-Wesley, 1985. ISBN 978-0201119459.

CHICHANOSKI, G. **Algoritmos para Auxílio no Diagnóstico de Doenças Pulmonares Utilizando Inteligência Artificial**. Dissertação (Mestrado) — Universidade Estadual de Londrina, 2022. Acessado em: 2 junho 2025. Disponível em: https://repositorio.uel.br/items/b051272d-4865-488a-8940-4f9b5dcc63ee.

CHOWDHURY, M. E. H. et al. **COVID-19 Radiography Database**. 2020. Kaggle Dataset. Dataset contendo radiografias de tórax para COVID-19, pneumonia viral e casos normais. Disponível em: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

COLAÇO, R. S. C. **Identificação Semiautomática de Patologia em Imagens de Raio X do Tórax com Técnicas de Inteligência Artificial**. Dissertação (Mestrado) — Universidade NOVA de Lisboa, 2022. Disponível em: https://run.unl.pt/handle/10362/157942.

DASHA, A. **What is Self-Supervised Machine Learning and How Does It Work?** 2022. Blog Post. Disponível em: https://dasha.ai/blog/self-supervised-machine-learning.

DEMATOS, G. et al. Feedforward versus recurrent neural networks for forecasting monthly japanese yen exchange rates. **Financial Engineering and the Japanese Markets**, v. 3, n. 1, p. 59–75, 1996.

EL-GAMAL, E.-Z. A.; ELMOGY, M.; ATWAN, A. Current trends in medical image registration and fusion. **Egyptian Informatics Journal**, v. 17, n. 1, p. 99–124, 2016. Disponível em: https://www.sciencedirect.com/science/article/pii/S111086651500047X.

ELIWA, E. H. I. Enhancing skin cancer diagnosis through fine-tuning of pretrained models: A two-phase transfer learning approach. **International Journal of Breast Cancer**, John Wiley & Sons, v. 2025, p. 4362941, 2025. Disponível em: https://doi.org/10.1155/ijbc/4362941.

FACELI, K. et al. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. Rio de Janeiro: LTC, 2011. ISBN 978-85-216-1836-8.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA: MIT Press, 2016. ISBN 978-0262035613. Disponível em: http://www.deeplearningbook.org.

HAAS, W. **Aprendizado de Máquina Aplicado à Classificação de Imagens de Melanoma**. Dissertação (Mestrado) — Instituto Federal de Santa Catarina (IFSC), 2020. Trabalho de Conclusão de Curso. Disponível em: https://repositorio.ifsc.edu.br/bitstream/handle/123456789/2841/TCC_Welinton_Haas.pdf.

HASSAN, A. **ChestX6-hybrid-model**. Kaggle, 2026. Acessado em: 5 de junho de 2026. Disponível em: https://www.kaggle.com/code/ahmedhassan9/chestx6-hybrid-model.

HAYASHI, H. A hybrid of generative and discriminative models based on the gaussian-coupled softmax layer. **arXiv preprint arXiv:2305.05912**, 2023. Disponível em: https://arxiv.org/abs/2305.05912.

HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2016. p. 770–778. Disponível em: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

HOBBS, D. L. Chest radiography for radiologic technologists. **Radiologic Technology**, v. 78, n. 6, p. 494–516, Jul-Aug 2007. Disponível em: https://www.isu.edu/media/libraries/radiographic-science/pdf/chest.pdf.

HUANG, G. et al. Densely connected convolutional networks. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2017. p. 4700–4708. Disponível em: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_201

HUSSAIN, S. et al. Modern diagnostic imaging technique applications and risk factors in the medical field: A review. **BioMed Research International**, v. 2022, p. 1–19, 2022. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9192206/.

HUSSAIN, T. et al. Effresnet-vit: A fusion-based convolutional and vision transformer model for explainable medical image classification. **IEEE Access**, IEEE, v. 13, p. 54040–54068, 2025.

IBM. **Types of Machine Learning**. 2021. Acessado em: 10 out. 2023. Disponível em: https://www.ibm.com/think/topics/machine-learning-types.

ISLAM, S. K. M. S. et al. Introduction of medical imaging modalities. **arXiv preprint arXiv:2306.01022**, 2023. Disponível em: https://arxiv.org/abs/2306.01022.

KAMILA, I. P. et al. A good evaluation based on confusion matrix for lung diseases classification using convolutional neural networks. **Advance Sustainable Science, Engineering and Technology**, v. 6, n. 1, 2024. Disponível em: https://journal.upgris.ac.id/index.php/asset/article/view/17330/pdf.

KERMANY, D.; ZHANG, K.; GOLDBAUM, M. **Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification**. Mendeley Data, 2018. Disponível em: https://data.mendeley.com/datasets/rscbjbr9sj/2.

KHAN, S. et al. Impact of ct intensity and contrast variability on deep-learning-based lung-nodule detection: A systematic review of preprocessing and harmonization strategies (2020–2025). **Diagnostics**, MDPI, v. 16, n. 2, p. 201, 2026. Disponível em: https://doi.org/10.3390/diagnostics16020201.

KIM, H. E. et al. Transfer learning for medical image classification: a literature review. **BMC Medical Imaging**, BioMed Central (Springer Nature), v. 22, n. 1, p. 69, 2022. Disponível em: https://doi.org/10.1186/s12880-022-00793-7.

KIRAN, S.; JABEEN, I. **Dataset of Tuberculosis Chest X-rays Images**. Mendeley Data, 2024. Disponível em: https://data.mendeley.com/datasets/8j2g3csprk/2.

KURZWEIL, R. **The Age of Intelligent Machines**. Cambridge, MA: MIT Press, 1990. ISBN 978-0262111215.

MCMILLIAN, E.; BANERJEE, A.; BUENO-OROVIO, A. From 2D to 3D, deep learning-based shape reconstruction in magnetic resonance imaging: A review. **arXiv preprint arXiv:2510.01296**, 2025. Disponível em: https://arxiv.org/abs/2510.01296.

MEHTA, M. et al. Large language models: a primer and gastroenterology applications. **Frontiers in Digital Health**, v. 6, p. 1364379, 2024.

Ministério da Saúde do Brasil. **Brasil recupera índice de detecção de tuberculose e aperfeiçoa tratamento de pacientes**. 2023. Acessado em: 1 dez. 2024. Disponível em: https://www.gov.br/saude/pt-br/assuntos/noticias/2023/dezembro/brasil-recupera-indice-de-deteccao-de-tuberculose-e-aperfeicoa-tratamento-de-pacientes.

MOLCHANOV, D. et al. Greedy policy search: A simple baseline for learnable test-time augmentation. In: **Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)**. PMLR, 2020. (Proceedings of Machine Learning Research, v. 124), p. 1308–1317. Disponível em: https://arxiv.org/abs/2002.09103.

MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. **Array**, v. 16, p. 100258, 2022. Disponível em: https://www.sciencedirect.com/science/article/pii/S2590005622000911.

NASSER, A. A.; AKHLOUFI, M. A. Deep learning methods for chest disease detection using radiography images. **SN Computer Science**, v. 4, n. 4, p. 388, 2023.

National Cancer Institute. **Radiology**. 2025. Acesso em: 19 maio 2025. Disponível em: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/radiology.

National Institute of Biomedical Imaging and Bioengineering (NIBIB). **Nuclear Medicine**. 2025. Acesso em: 19 maio 2025. Disponível em: https://www.nibib.nih.gov/science-education/science-topics/nuclear-medicine.

NGUYEN, S. T.; TULABANDHULA, T.; WATSON-MANHEIM, M. B. User friendly and adaptable discriminative ai: Using the lessons from the success of llms and image generation models. **arXiv preprint arXiv:2312.06826**, 2023. Disponível em: https://arxiv.org/abs/2312.06826.

OLIVEIRA, L. L. G. A. d.; SILVA, A. G. **Chest X-Ray Dataset for Pulmonary Disease Classification (COVID-19, Tuberculosis, Pneumonia, Normal)**. Zenodo, 2026. Disponível em: https://doi.org/10.5281/zenodo.20312106.

PANESAR, A. et al. Machine learning for health care: Transformative, collaborative, and clinically relevant. **Artificial Intelligence in Medicine**, v. 3, p. 100048, 2021. Disponível em: https://www.sciencedirect.com/science/article/pii/S2666764921000485.

PETERSILGE, C. A. et al. Visible light imaging: Clinical aspects with an emphasis on medical photography - a himss-siim enterprise imaging community whitepaper. **Journal of Digital Imaging**, v. 35, n. 3, p. 385–395, 2022. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC9156602/.

PONTI, M. A.; COSTA, G. B. P. da. Como funciona o deep learning. In: **Tópicos em Gerenciamento de Dados e Informações**. São Carlos, SP, Brasil: Sociedade Brasileira de Computação, 2017. cap. 3, p. 63–93. ISBN 978-85-7669-400-7. Disponível em: https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/25/49/206-1.

POOLE, D.; MACKWORTH, A.; GOEBEL, R. **Computational Intelligence: A Logical Approach**. New York, NY: Oxford University Press, 1998. ISBN 978-0195102703.

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.

RAHMAN, T. et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. **Sensors**, v. 21, n. 2, p. 572, 2021.

RAVI, V.; ACHARYA, V.; ALAZAB, M. A multichannel efficientnet deep learning-based stacking ensemble approach for lung disease detection using chest x-ray images. **Cluster Computing**, v. 26, n. 2, p. 1181–1203, 2023.

REMESEIRO, B.; BOLóN-CANEDO, V. A review of feature selection methods in medical applications. **Computers in Biology and Medicine**, v. 112, p. 103375, 2019. Disponível em: https://www.sciencedirect.com/science/article/pii/S0010482519302525.

ROCHA, J.; MENDONÇA, A. M.; CAMPILHO, A. Review on deep learning methods for chest x-ray based abnormality detection and thoracic pathology classification. **U.Porto Journal of Engineering**, v. 7, n. 4, p. 16–32, 2021.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**. [S.l.]: Springer, Cham, 2015. (Lecture Notes in Computer Science, v. 9351), p. 234–241.

ROUND, M. E. Chest x-ray. In: WILLERSON, J. T. et al. (Ed.). **Cardiovascular Medicine**. London: Springer, 2007. p. 79–91. Disponível em: https://doi.org/10.1007/978-1-84628-715-2_4.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4th. ed. Harlow, England: Pearson, 2020. ISBN 978-0134610993. Disponível em: https://aima.cs.berkeley.edu/.

SARKER, I. H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. **SN Computer Science**, v. 2, n. 5, p. 420, 2021.

SHARMA, S.; MITTAL, R.; GOYAL, N. An assessment of machine learning and deep learning techniques with applications. In: **ECS Transactions**. [S.l.]: The Electrochemical Society, 2022. v. 107, n. 1, p. 8979–8991.

SHARMA, V. et al. Deep learning models for tuberculosis detection and infected region visualization in chest x-ray images. **Intelligent Medicine**, Elsevier, v. 4, n. 2, 2024. Disponível em: https://doi.org/10.1016/j.imed.2023.06.001.

SHERKATGHANAD, Z. et al. Baytta: Uncertainty-aware medical image classification with optimized test-time augmentation using bayesian model averaging. **arXiv preprint arXiv:2406.17640**, 2024. Disponível em: https://arxiv.org/abs/2406.17640.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 2019. Disponível em: https://doi.org/10.1186/s40537-019-0197-0.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: **International Conference on Learning Representations (ICLR)**. [s.n.], 2015. Disponível em: https://arxiv.org/abs/1409.1556.

STEHMAN, S. V. Selecting and interpreting measures of thematic classification accuracy. **Remote Sensing of Environment**, v. 62, n. 1, p. 77–89, 1997. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425797000837.

TAMAAZOUSTI, Y.; BORGNE, H. L.; HUDELOT, C. Recent advances in self-supervised learning: A survey and future directions. **SN Computer Science**, v. 2, n. 3, p. 204, 2021. Disponível em: https://link.springer.com/article/10.1007/s42979-021-00592-x.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **Proceedings of the 36th International Conference on Machine Learning (ICML)**. PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. Disponível em: https://proceedings.mlr.press/v97/tan19a.html.

TANG, F. et al. Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. **arXiv preprint arXiv:2308.01239**, 2023. Disponível em: https://arxiv.org/abs/2308.01239.

TEIXEIRA, L. O. et al. Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images. **Sensors**, MDPI, v. 21, n. 21, p. 7116, 2021.

Telemedicina Morsch. **Doenças do Pulmão**. 2024. Acessado em: 1 dez. 2024. Disponível em: https://telemedicinamorsch.com.br/blog/doencas-do-pulmao.

Tesla Imagem. **Ultrassonografia Obstétrica: Como é Realizada?** 2025. Acesso em: 19 maio 2025. Disponível em: https://teslaimagem.com.br/ultrassonografia-obstetrica-como-e-realizada/.

THOMAS, A. M. K.; BANERJEE, A. K. **The History of Radiology**. Oxford, United Kingdom: Oxford University Press, 2013. (Oxford Medical Histories). ISBN 9780199639977. Disponível em: https://academic.oup.com/book/24971.

University of Edinburgh – Edinburgh Imaging. **What is microscopy**. 2025. Acesso em: 19 maio 2025. Disponível em: https://clinical-sciences.ed.ac.uk/edinburgh-imaging/for-patients-study-participants/tell-me-more-about-my-scan/what-is-microscopy.

WADA, D. T.; RODRIGUES, J. A. H.; SANTOS, M. K. Aspectos técnicos e roteiro de análise da radiografia de tórax. **Medicina (Ribeirão Preto)**, v. 52, n. Supl. 1, p. 5–16, 2019. Disponível em: https://pdfs.semanticscholar.org/f10d/b2f8a66cccc673210efe218916bbd3e63382.pdf.

World Health Organization. **Tuberculosis Profile – Brazil**. 2023. Acessado em: 1 dez. 2024. Disponível em: https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_tab="charts"lan="EN"iso2="BR"entity_type="country

.

YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, Springer, v. 9, p. 611–629, 2018.

ZUKIć, D. et al. **2D to 3D Reconstruction in Medical Imaging: Challenges and Solutions**. 2025. Kitware Blog. Acedido em: 14 de fevereiro de 2026. Disponível em: https://www.kitware.com/2d-to-3d-reconstruction-in-medical-imaging-challenges-and-solutions/.

# Modelo híbrido de *Deep Learning* com *ensemble* de CNNs para a detecção de doenças pulmonares em radiografias de tórax

**Leonardo Luiz Gambalonga Alves de Oliveira**[1], **Alexandre Gonçalves Silva**[1]

[1]Departamento de Informática e Estatística (INE) – Centro Tecnológico (CTC)
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brasil

leonardoluizg.oliveira@gmail.com, alexandre.goncalves@ufsc.br

***Abstract.*** *Lung diseases are a major public-health burden, aggravated by the COVID-19 pandemic. Chest X-ray is the most accessible screening exam, but its interpretation depends on specialists and is error-prone. This work developed a computer-aided diagnosis system for lung pathologies in chest radiographs, based on a deep-learning pipeline with three modules: lung segmentation (CMU-NeXt), binary classification (normal/pathological) and multiclass classification (normal, pneumonia, COVID-19, tuberculosis), the last two using ensembles of five CNNs, evaluated in cascade and parallel architectures. On the internal test set the multiclass ensemble reached 99.43% accuracy and 99.40% F1-score, and 94.20% F1-score on an unseen external dataset. Configurations without segmentation consistently outperformed those with it, and ensemble learning of heterogeneous architectures produced robust predictions.*

***Resumo.*** *As doenças pulmonares representam um grave problema de saúde pública, agravado pela pandemia de COVID-19. A radiografia de tórax é o exame de triagem mais acessível, mas sua interpretação depende de especialistas e está sujeita a erros. Este trabalho desenvolveu um sistema de apoio ao diagnóstico de patologias pulmonares em radiografias torácicas, baseado em um pipeline de aprendizado profundo com três módulos: segmentação pulmonar (CMUNeXt), classificação binária (normal/patológico) e classificação multiclasse (normal, pneumonia, COVID-19 e tuberculose), estas por meio de ensembles de cinco CNNs, avaliados em arquiteturas em cascata e paralela. No conjunto de teste interno, o ensemble multiclasse atingiu 99,43% de acurácia e 99,40% de F1-score, e 94,20% de F1-score em uma base externa não vista. As configurações sem segmentação superaram consistentemente as com segmentação.*

## 1. Introdução

No Brasil, as doenças respiratórias estão entre as principais causas de morbidade e de demanda por atenção à saúde [Brasil, Ministério da Saúde 2024]. A pandemia iniciada em dezembro de 2019, causada pelo vírus SARS-CoV-2, agravou esse cenário ao direcionar recursos de saúde para o combate ao novo vírus, reduzindo diagnósticos e tratamentos de outras patologias. Como consequência, enquanto em 2018 a tuberculose foi responsável por cerca de 7 mil mortes no país, em 2023 causou aproximadamente 13 mil óbitos, um aumento de 85,71% [World Health Organization 2023].

Entre as formas de diagnóstico, a radiografia de tórax é a mais requisitada por seu baixo custo e natureza não invasiva [Rocha et al. 2021]. Entretanto, o tempo de análise e a

suscetibilidade a interpretações dúbias, mesmo por profissionais experientes, evidenciam a necessidade de soluções complementares que agilizem e aprimorem a precisão diagnóstica. Nesse contexto, sistemas de apoio ao diagnóstico baseados em *Deep Learning* (DL) tornaram-se uma alternativa de grande interesse, capazes de aprender automaticamente representações hierárquicas a partir dos *pixels* da imagem [Colaço 2022].

Este trabalho tem como objetivo desenvolver um sistema de detecção de patologias pulmonares em radiografias de tórax que combine diferentes redes neurais convolucionais (CNNs) por meio de *ensemble learning*, organizado em um *pipeline* hierárquico de segmentação, triagem binária e classificação multiclasse. As principais contribuições são: (i) a composição e a disponibilização pública de uma base de dados balanceada com quatro classes (normal, pneumonia, COVID-19 e tuberculose); (ii) a proposta e a comparação de duas arquiteturas de integração (em cascata e paralela); e (iii) uma análise ablativa que quantifica a contribuição de cada componente do *pipeline*, incluindo a avaliação empírica do impacto da segmentação pulmonar sobre a classificação, em uma base interna e em uma base externa independente.

## 2. Trabalhos Relacionados

A análise de radiografias de tórax (*Chest X-Ray*, CXR) é uma área de grande potencial para sistemas de apoio ao diagnóstico (*Computer-Aided Diagnosis*, CAD), embora seja uma tarefa complexa: a natureza bidimensional da imagem sobrepõe estruturas anatômicas e pode obscurecer achados patológicos sutis, além de estar sujeita à variabilidade interobservador [Nasser and Akhloufi 2023]. A capacidade do DL de aprender características diretamente dos dados, dispensando a extração manual (*handcrafted features*) típica do aprendizado de máquina clássico, redefiniu os paradigmas da imagiologia médica.

Uma etapa de pré-processamento frequentemente citada é a segmentação pulmonar, que isola os campos pulmonares do restante da imagem. Arquiteturas codificador-decodificador, como a *U-Net*, tornaram-se o padrão para essa tarefa [Chichanoski 2022]. Para a classificação, a evolução foi das primeiras abordagens com *Support Vector Machines* e *Random Forests* sobre descritores manuais, até as CNNs pré-treinadas em *ImageNet* e ajustadas por *fine-tuning* ao domínio médico, como *ResNet*, *DenseNet* e a família *EfficientNet*, esta última focada em eficiência por escalonamento composto [Ravi et al. 2023]. Para lidar com a sobreposição de características entre patologias, técnicas de *ensemble learning* (por *voting* ou *stacking*) aumentam a robustez e a precisão em relação a modelos individuais [Sharma et al. 2022].

Apesar desse progresso, observa-se uma lacuna na exploração de modelos híbridos e hierárquicos que combinem essas tarefas de forma explícita e sequencial, simulando um fluxo de raciocínio diagnóstico: um primeiro estágio que atua como triador de anormalidade e, apenas para os casos positivos, um comitê de especialistas que determina a patologia específica. É nessa lacuna que o presente trabalho se insere. Cabe ainda destacar que a contribuição universal da segmentação em *pipelines* de classificação tem sido questionada por trabalhos recentes [Teixeira et al. 2021], o que motiva a avaliação empírica conduzida neste estudo.

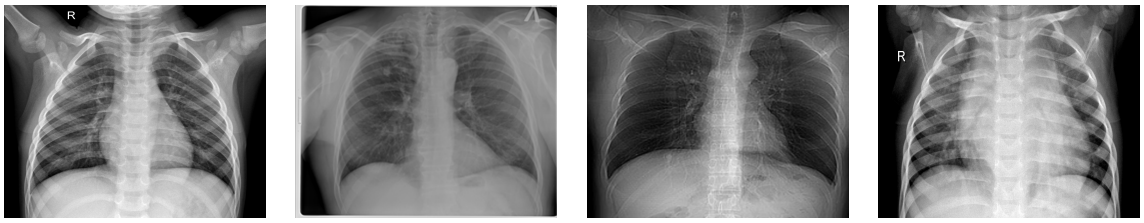
## 3. Materiais e Métodos

### 3.1. Base de dados

A base de dados utilizada foi composta a partir de conjuntos públicos, reunindo imagens de quatro classes. As imagens de pneumonia provêm do *Guangzhou Women and Children's Medical Center* [Kermany et al. 2018]; as de COVID-19 e parte das imagens normais, do *COVID-19 Radiography Database* [Chowdhury et al. 2020]; e as de tuberculose, de um conjunto público [Kiran and Jabeen 2024]. As imagens foram redimensionadas para $224 \times 224$ *pixels* e divididas de forma estratificada em treino, validação e teste (70%, 15% e 15%). A distribuição final é apresentada na Tabela 1, e o conjunto resultante foi publicado em repositório público para fins de reprodutibilidade [Oliveira and Silva 2026]. A Figura 1 ilustra uma radiografia de cada classe.

**Tabela 1. Distribuição dos casos na base de dados composta.**

| Classe | Total | Treino | Validação | Teste |
|---|---|---|---|---|
| COVID-19 | 2.300 | 1.610 | 345 | 345 |
| Tuberculose | 2.300 | 1.610 | 345 | 345 |
| Pneumonia | 2.300 | 1.610 | 345 | 345 |
| Normal | 7.100 | 4.970 | 1.065 | 1.065 |
| **Total** | **13.900** | **9.800** | **2.100** | **2.100** |



(a) Normal (b) Tuberculose (c) COVID-19 (d) Pneumonia

**Figura 1. Exemplos de radiografias para cada uma das quatro classes.**

Para avaliar a capacidade de generalização, utilizou-se ainda uma base externa independente, composta por 4.200 radiografias (700 de tuberculose e 3.500 normais) provenientes da Universidade do Qatar em conjunto com a Universidade de Dhaka [Rahman et al. 2020], não vista durante o treinamento.

### 3.2. Pré-processamento e aumento de dados

Além da normalização e do redimensionamento, aplicou-se *data augmentation* exclusivamente durante o treinamento, por meio da biblioteca *Albumentations* [Buslaev et al. 2020]. As transformações (espelhamento horizontal, variação de brilho e contraste, rotação de até $15^{\circ}$, ruído gaussiano e CLAHE) foram escolhidas para simular variações realistas de aquisição, evitando distorções anatomicamente implausíveis.

### 3.3. Segmentação pulmonar

A segmentação pulmonar foi realizada com a arquitetura *CMUNeXt*, uma rede codificador-decodificador leve e eficiente, adequada a conjuntos de dados de tamanho moderado. As máscaras geradas foram utilizadas para produzir uma versão segmentada da base (*masked*), permitindo comparar o desempenho da classificação com e sem a restrição da análise à região pulmonar.

### 3.4. Arquitetura do *pipeline* de classificação

O *pipeline* de classificação é composto por dois estágios apoiados em um comitê de cinco CNNs: um modelo Híbrido (ResNet152 + EfficientNet-B5), ResNet50, DenseNet121, VGG16 e EfficientNet-B0. O primeiro estágio (binário) realiza a triagem entre *normal* e *patológico*; o segundo (multiclasse) identifica a patologia específica entre normal, pneumonia, COVID-19 e tuberculose. As predições dos cinco modelos são combinadas por três métodos de agregação: *hard voting*, *soft voting* e média ponderada pela AUC (*weighted*).

Duas estratégias de integração foram propostas e comparadas (Figura 2). Na arquitetura em cascata, apenas as imagens identificadas como patológicas no estágio binário são encaminhadas ao estágio multiclasse, reduzindo falsos positivos por dupla verificação, ao custo de erros irreversíveis do estágio binário. Na arquitetura paralela, os dois estágios operam de forma independente sobre as imagens segmentadas, eliminando a propagação de erros entre estágios, mas impondo maior complexidade ao módulo multiclasse.

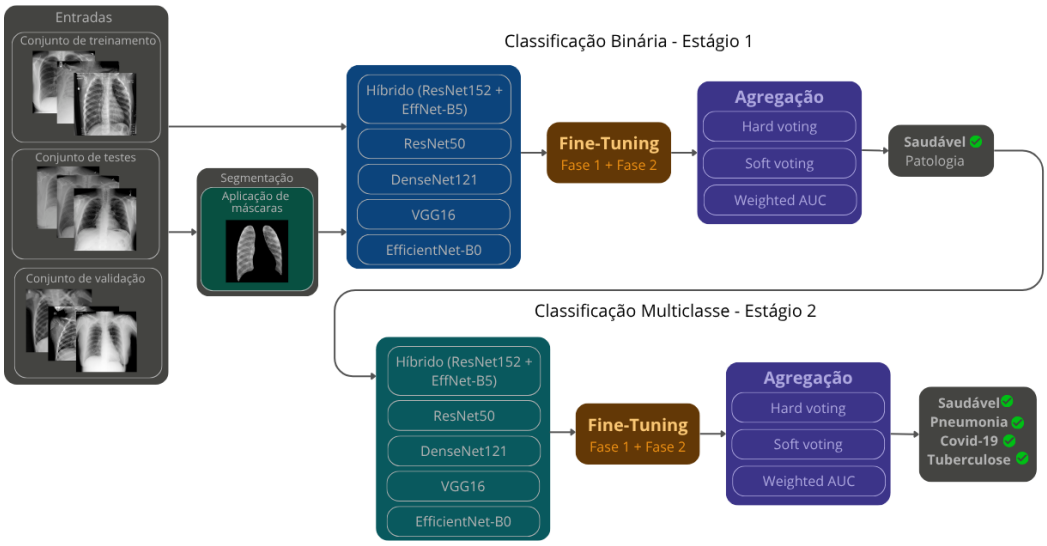
### 3.5. Estratégias de refinamento

Cada modelo do comitê passou por *fine-tuning* em duas fases: na primeira, os *backbones* pré-treinados são congelados e apenas o classificador é ajustado (15 épocas, taxa de aprendizado $5 \times 10^{-4}$); na segunda, os blocos convolucionais mais profundos são descongelados seletivamente e ajustados com taxa de aprendizado reduzida ($1 \times 10^{-5}$ para o *backbone* e $5 \times 10^{-5}$ para o classificador, 20 épocas). Aplicou-se também *Test Time Augmentation* (TTA) e harmonização de histograma no espaço LAB para aproximar as imagens de teste externas do padrão de treinamento. O desenvolvimento foi realizado em *Python* com *PyTorch* e *scikit-learn*, com treinamento em GPU (NVIDIA GeForce RTX 3050 e ambientes *Google Colab*/*Vlab*).

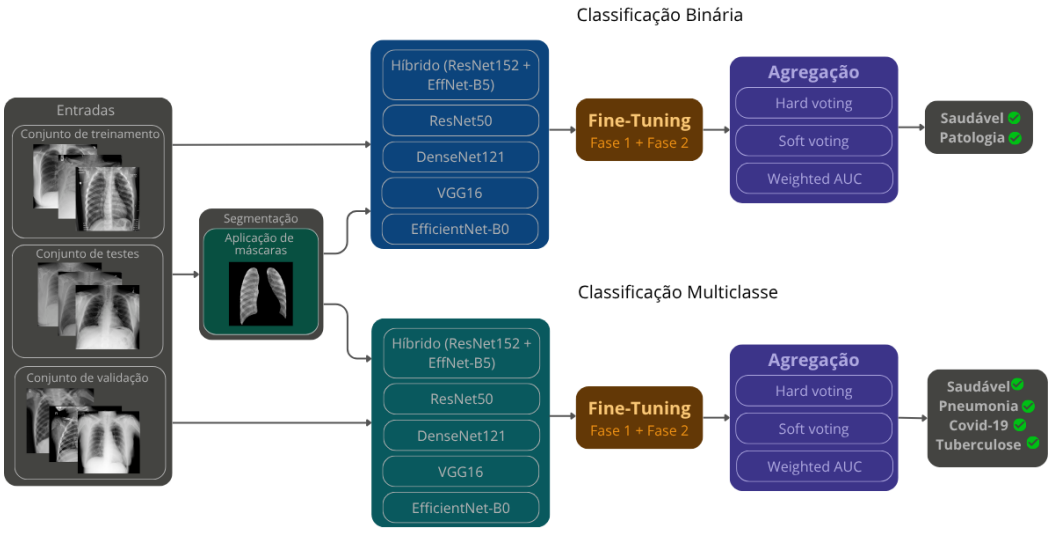
## 4. Resultados e Discussão

### 4.1. Segmentação

Cabe ressaltar que o *CMUNeXt* foi treinado por apenas 3 épocas, uma vez que já nesse ponto atingiu *IoU* de validação de 95,16% e coeficiente *Dice* (F1) de 97,52%, com *recall* de 97,00% e especificidade de 99,00% (Tabela 2), desempenho considerado satisfatório para o propósito de gerar máscaras a serem utilizadas no *pipeline* de classificação, que constitui o foco principal deste trabalho. Os erros residuais concentram-se na fronteira entre a região pulmonar e o mediastino, onde a transição entre as classes é anatomicamente gradual, sem erros grosseiros, o que valida a qualidade das máscaras geradas. Ainda assim, a *IoU* permanecia em tendência de crescimento ao fim da terceira época (de 93,26% para 95,16%), indicando que um treinamento mais prolongado poderia elevar ainda mais o desempenho da segmentação, o que é apontado como trabalho futuro.




(a) Cascata




(b) Paralela

**Figura 2. Arquiteturas de integração propostas para o *pipeline*.**

**Tabela 2. Desempenho do *CMUNeXt* na segmentação pulmonar.**

| Métrica | IoU | Acurácia | Precisão | *Recall* | F1 | Especif. |
|---|---|---|---|---|---|---|
| Valor | 95,16% | 98,60% | 98,49% | 97,00% | 97,52% | 99,00% |

### 4.2. Análise ablativa

A análise ablativa foi conduzida com o modelo multiclasse individual (arquitetura Híbrida), quantificando o efeito de cada componente. Como resumido na Tabela 3, o *data augmentation* teve impacto marginal na base interna, porém decisivo na base externa, elevando o F1-score de 59,35% para 76,14% ($+16,79$ pontos percentuais). O TTA com harmonização elevou o F1-score externo para 79,67%, e o *fine-tuning* em duas fases promoveu ganhos consistentes em ambas as bases (F1 interno de 98,85% para 99,16%; externo de 79,67% para 82,87%). A aplicação da segmentação, por outro lado, reduziu o desempenho na base interna (F1 de 99,16% para 94,90%), antecipando um achado recorrente em toda a avaliação.

**Tabela 3. Análise ablativa (F1-score) do modelo multiclasse individual.**

| Configuração | F1 interno | F1 externo |
|---|---|---|
| *Baseline* (só normalização) | 98,75% | 59,35% |
| + *Data Augmentation* | 98,85% | 76,14% |
| + TTA e *Harmonization* | – | 79,67% |
| + *Fine-Tuning* (2 fases) | 99,16% | 82,87% |
| + Segmentação | 94,90% | 42,43% |

### 4.3. Estágio binário

O *ensemble* binário, compartilhado pelas duas arquiteturas, alcançou acurácia e F1-score de 99,43% (*hard voting*) e AUC de 99,98% sem segmentação, com apenas 12 erros em 2.100 amostras (Tabela 4). Desses, 9 são falsos negativos (imagens patológicas classificadas como normais), o erro mais crítico na cascata, pois tais casos não prosseguem ao estágio multiclasse. Com segmentação, o desempenho caiu para 98,29%.

**Tabela 4. Desempenho do *ensemble* binário (*hard voting*).**

| Configuração | Acurácia | F1-Score | AUC |
|---|---|---|---|
| Sem segmentação | 0,9943 | 0,9943 | 0,9998 |
| Com segmentação | 0,9829 | 0,9829 | 0,9983 |

### 4.4. Classificação multiclasse: cascata *versus* paralela

A Tabela 5 compara as duas arquiteturas na base interna. A arquitetura paralela sem segmentação obteve o melhor F1-score (99,40%), classificando corretamente a totalidade dos 345 casos de tuberculose e com erros residuais concentrados entre COVID-19 e normal (Figura 3), par de classes reconhecidamente desafiador na literatura. A cascata sem

segmentação apresentou desempenho equivalente (99,28%), com *recall* mais equilibrado. Em ambas as arquiteturas, a segmentação reduziu o desempenho.

**Tabela 5. Desempenho multiclasse na base interna (melhor método de agregação).**

| Arquitetura | Config. | Acurácia | Precisão | *Recall* | F1 |
|---|---|---|---|---|---|
| Paralela | Sem seg. | 0,9943 | 0,9938 | 0,9942 | 0,9940 |
| | Com seg. | 0,9852 | 0,9835 | 0,9864 | 0,9849 |
| Cascata | Sem seg. | 0,9933 | 0,9947 | 0,9908 | 0,9928 |
| | Com seg. | 0,9829 | 0,9851 | 0,9798 | 0,9824 |

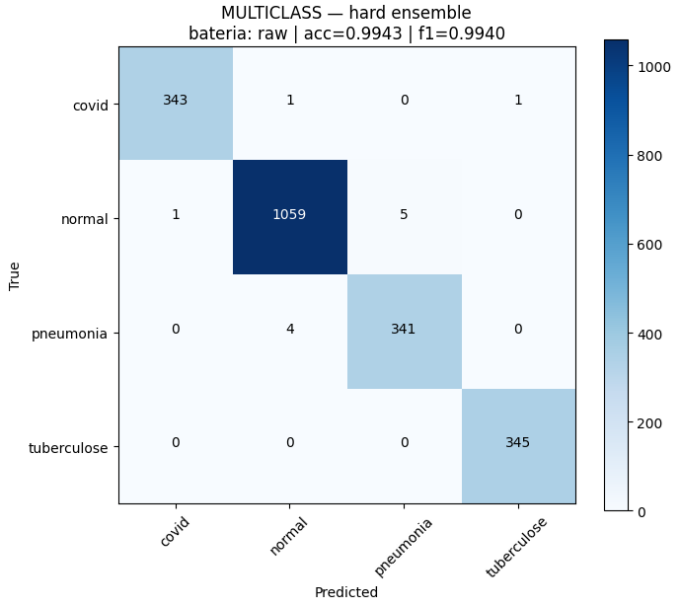



**Figura 3. Matriz de confusão do *ensemble* multiclasse (paralela, sem segmentação) na base interna.**

### 4.5. Avaliação na base externa

Na base externa não vista, o *pipeline* em cascata sem segmentação atingiu acurácia de 97,00% e F1-score de 94,20% (Tabela 6), demonstrando boa generalização para dados fora da distribuição de treinamento. Dos 3.500 casos normais, 3.496 foram corretamente identificados; para a tuberculose, 578 dos 700 casos foram detectados (Figura 4), com precisão de 99,31%. Com segmentação, o F1-score caiu para 89,01%, e a diferença entre as configurações foi maior na base externa ($\Delta = 0,0519$) do que na interna ($\Delta = 0,0104$), indicando que a segmentação de imagens externas, não vistas pelo segmentador, é um ponto de fragilidade do *workflow*.

**Tabela 6. Desempenho do *pipeline* em cascata na base externa.**

| Configuração | Acurácia | Precisão | *Recall* | F1-Score |
|---|---|---|---|---|
| Sem segmentação | 0,9700 | 0,9797 | 0,9123 | 0,9420 |
| Com segmentação | 0,9426 | 0,9183 | 0,8673 | 0,8901 |

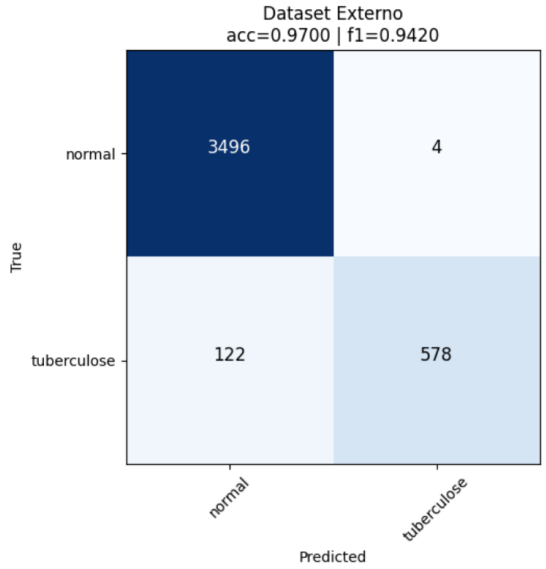



**Figura 4. Matriz de confusão do *pipeline* em cascata na base externa (sem segmentação).**

Um resultado transversal a todas as configurações foi a superioridade do *pipeline* sem segmentação sobre o com segmentação, tanto na base interna quanto na externa. Embora conceitualmente motivada, a segmentação, nas condições avaliadas, elimina informações contextuais úteis à classificação e introduz variabilidade por meio de imprecisões nas máscaras, achado alinhado com a literatura que questiona sua contribuição universal [Teixeira et al. 2021].

### 4.6. Comparação com a literatura

A Tabela 7 posiciona o sistema proposto frente a trabalhos relacionados. Diferentemente da maioria, que aborda 2 ou 3 classes, o sistema proposto trata 4 classes clínicas com desempenho competitivo, combinando *ensemble* e *fine-tuning* e reportando avaliação em base externa independente.

**Tabela 7. Comparação com trabalhos relacionados (Seg. = segmentação; Ens. = *ensemble*).**

| Trabalho | Classes | Seg. | Ens. | Acurácia | AUC | F1 |
|---|---|---|---|---|---|---|
| [Colaço 2022] | 2 | Sim | Sim | 79,5% | 89,1% | 80,8% |
| [Hassan 2026] | 6 | Não | Não | 93,4–96,5% | 99,0% | 96,4% |
| [Chichanoski 2022] | 3 | Sim | Não | 96,1–99,5% | — | 92,7% |
| [Sharma et al. 2024] | 2 | Sim | Não | 99,3% | 99,9% | 99,3% |
| [Bharati et al. 2020] | 2 | Sim | Sim | 73,0% | — | — |
| **Proposto (RAW)** | 4 | Não | Sim | **99,3%** | **99,98%** | **99,3%** |
| **Proposto (Masked)** | 4 | Sim | Sim | 98,3% | 99,8% | 98,2% |

## 5. Conclusão

Este trabalho propôs um sistema de apoio ao diagnóstico de patologias pulmonares em radiografias de tórax, baseado em um *pipeline* de *deep learning* com segmentação (CMUNeXt) e classificação binária e multiclasse por *ensemble* de cinco CNNs, avaliado em arquiteturas em cascata e paralela. O sistema classificou corretamente a grande maioria das imagens em quatro classes clínicas, atingindo F1-score de 99,40% na base interna e 94,20% em uma base externa independente, com desempenho superior ao de modelos individuais. A combinação de arquiteturas heterogêneas por *ensemble learning* mostrou-se eficaz para mitigar erros individuais e produzir predições mais robustas. Um achado relevante foi que o *pipeline* sem segmentação superou consistentemente o com segmentação, reforçando a necessidade de avaliar empiricamente cada componente antes de incorporá-lo ao sistema final.

Como trabalhos futuros, destacam-se a ampliação e diversificação da base de dados, o aprimoramento do módulo de segmentação (treinamento mais prolongado do CMUNeXt, atualmente limitado a poucas épocas, e pós-processamento morfológico das máscaras), a representação de *patches* internos ao pulmão para detecção de opacidades em regiões internas, a otimização sistemática de hiperparâmetros e a avaliação de arquiteturas baseadas em *Vision Transformers*, além de uma validação clínica prospectiva em ambiente hospitalar real.

## Referências


Bharati, S., Podder, P., and Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from x-ray images. *Informatics in Medicine Unlocked*, 20:100391.

Brasil, Ministério da Saúde (2024). Doenças respiratórias. Biblioteca Virtual em Saúde. Acessado em: 12 jul. 2026.

Buslaev, A., Parinov, A., Khvedchenya, E., Iglovikov, V. I., and Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2):125.

Chichanoski, G. (2022). Algoritmos para auxílio no diagnóstico de doenças pulmonares utilizando inteligência artificial. Master's thesis, Universidade Estadual de Londrina. Acessado em: 2 junho 2025.

Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al-Emadi, N., and Reaz, M. B. I. (2020). Covid-19 radiography database. Kaggle Dataset. Dataset contendo radiografias de tórax para COVID-19, pneumonia viral e casos normais.

Colaço, R. S. C. (2022). Identificação semiautomática de patologia em imagens de raio x do tórax com técnicas de inteligência artificial. Master's thesis, Universidade NOVA de Lisboa.

Hassan, A. (2026). Chestx6-hybrid-model. Acessado em: 5 de junho de 2026.

Kermany, D., Zhang, K., and Goldbaum, M. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification.

Kiran, S. and Jabeen, I. (2024). Dataset of tuberculosis chest x-rays images.

Nasser, A. A. and Akhloufi, M. A. (2023). Deep learning methods for chest disease detection using radiography images. *SN Computer Science*, 4(4):388.

Oliveira, L. L. G. A. d. and Silva, A. G. (2026). Chest x-ray dataset for pulmonary disease classification (covid-19, tuberculosis, pneumonia, normal).

Rahman, T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Ayari, M. A., and Chowdhury, M. E. H. (2020). Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization. *IEEE Access*, 8:191586–191601.

Ravi, V., Acharya, V., and Alazab, M. (2023). A multichannel efficientnet deep learning-based stacking ensemble approach for lung disease detection using chest x-ray images. *Cluster Computing*, 26(2):1181–1203.

Rocha, J., Mendonça, A. M., and Campilho, A. (2021). Review on deep learning methods for chest x-ray based abnormality detection and thoracic pathology classification. *U.Porto Journal of Engineering*, 7(4):16–32.

Sharma, S., Mittal, R., and Goyal, N. (2022). An assessment of machine learning and deep learning techniques with applications. In *ECS Transactions*, volume 107, pages 8979–8991. The Electrochemical Society.

Sharma, V., Nillmani, Gupta, S. K., and Shukla, K. K. (2024). Deep learning models for tuberculosis detection and infected region visualization in chest x-ray images. *Intelligent Medicine*, 4(2).

Teixeira, L. O., Pereira, R. M., Bertolini, D., Oliveira, L. S., Nanni, L., Cavalcanti, G. D. C., and Costa, Y. M. G. (2021). Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images. *Sensors*, 21(21):7116.

World Health Organization (2023). Tuberculosis profile – brazil. Acessado em: 1 dez. 2024.