Celso Luiz Silva Soares Filho

Classificação de Exames PET de Corpo Inteiro usando Representações MIP e Aprendizado Profundo

São Luís - MA

Março de 2026

Celso Luiz Silva Soares Filho

Universidade Federal do Maranhão – UFMA

Centro de Ciências Exatas e Tecnologias

Departamento de Engenharia Elétrica

Programa de pós-graduação em Engenharia Elétrica

Classificação de Exames PET de Corpo Inteiro usando Representações MIP e Aprendizado Profundo

Dissertação de mestrado apresentada ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal do Maranhão na área de Ciência da Computação, como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

São Luís - MA

Março de 2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Soares Filho, Celso Luiz Silva.
Classificação de Exames PET de Corpo Inteiro usando Representações MIP e Aprendizado Profundo / Celso Luiz Silva Soares Filho. - 2026.
70 f.

Coorientador(a) 1: Darlan Bruno Pontes Quintanilha.
Orientador(a): Anselmo Cardoso de Paiva.
Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, São Luís, 2026.

1. Tomografia Por Emissão de Pósitrons (pet). 2. Projeção de Intensidade Máxima (mip). 3. Classificação Automática. 4. Aprendizado Profundo. I. de Paiva, Anselmo Cardoso. II. Quintanilha, Darlan Bruno Pontes. III. Título.

Celso Luiz Silva Soares Filho

Classificação de Exames PET de Corpo Inteiro usando Representações MIP e Aprendizado Profundo

Dissertação de mestrado apresentada ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal do Maranhão na área de Ciência da Computação, como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

Trabalho apresentado em: São Luís - MA, 13 de Mraço de 2026.

Prof. Dr. Anselmo Cardoso de Paiva
Orientador

Prof. Dr. Darlan Bruno Pontes Quintanilha
Coorientador

Prof. Dr. Aristófanes Corrêa Silva
Universidade Federal do Maranhão
Banca Examinadora

Prof. Dr. António Manuel Trigueiros da Silva Cunha
Universidade de Trás-os-Montes e Alto Douro
Banca Examinadora

São Luís - MA
Março de 2026

Este trabalho é dedicado à minha mãe.

Agradecimentos

Agradeço primeiramente à Deus por tudo na vida, pela sabedoria, pela saúde. Gostaria de agradecer em seguida à minha mãe, Selma Sousa, pela educação e pelos ensinamentos que ela me proporcionou. Agradeço à ela também pelo incentivo que sempre me deu em continuar tentando, nunca desistir por causa de obstáculos e por ser exemplo que sempre me ajudou a continuar caminhando. Não posso esquecer de agradecer também a minha irmã Brenda, que sempre me auxiliou na resolução de problemas e nunca recusou ajuda. Por fim, à minha namorada Amanda por todo apoio e incentivo.

Agradeço também aos professores da Universidade Federal do Maranhão (UFMA) pelos conhecimentos concebidos por eles. Um agradecimento especial para os Professores Anselmo, Aristófanes e Darlan pela oportunidade e pela confiança dadas em mim para que eu pudesse participar de seu laboratório e pudesse adquirir experiências de valores inestimáveis. E à Giovanni Lucca pelo incentivo e ter sido inspiração para que eu pudesse iniciar o mestrado. Também gostaria de agradecer a Ramsey D. Badawi e Vivek Swarnakar, ambos da Universidade de Davis na Califórnia. Além de Cláudio de Souza Baptista e Mateus Queiroz Cunha da Universidade Federal de Campina Grande pelo suporte e ensinamentos.

Agradeço à FAPEMA, CAPES e CNPQ pelos recursos cedidos que foram utilizados para o desenvolvimento desta dissertação. Bem como agradeço ao Núcleo de Computação Aplicada pelo espaço e materiais que foram disponibilizados para me auxiliar.

No laboratório conheci e agradeço a pessoas que foram essenciais para meu desenvolvimento como aluno. São eles: Neilson Ribeiro, Jhones Soares, Matheus Raposo, Luís Felipe Rocha, Felipe Teles e Ricardo Marques. Foram responsáveis por um conhecimento que não poderia adquirir de outra forma.

Quando você quer alguma coisa,
todo o universo consipira para que
você realize o seu desejo.
(Paulo Coelho)

Resumo

O câncer é um dos maiores desafios de saúde pública global, com estimativas de 35 milhões de novos casos até 2035. Nesse cenário, o exame de Tomografia por Emissão de Pósitrons (PET) é essencial para o diagnóstico e monitoramento. No entanto, a interpretação clínica desses exames é uma tarefa exaustiva, sujeita à subjetividade do especialista e limitada pela alta complexidade dos dados volumétricos 3D. Este trabalho propõe um método de classificação automática de exames PET de corpo inteiro de pacientes com câncer de pulmão, linfoma, melanoma e saudáveis, utilizando técnicas de aprendizado profundo aplicadas a representações de Projeção de Intensidade Máxima (MIP). O método é estruturado em quatro etapas: geração de imagens MIP nos eixos coronal e sagital, pré-processamento, extração de características e classificação. Foram avaliados seis arquiteturas para extração de atributos (ConvNeXt, EfficientNet-B0, Swin e VGG19) e três classificadores (MLP, SVM e XGBoost). O método alcançou resultados de 96,45% para a métrica de AUC, 91,98% de Acurácia, 91,63% de *F1-Score*, 91,18% de sensibilidade e uma precisão de 92,08%. Esses resultados mostram que a utilização de representações MIP, combinada a um conjunto de arquiteturas especializadas por perspectiva, permite atingir um desempenho satisfatório, aproximando-se de abordagens que utilizam volumes 3D e exames híbridos (PET/TC).

Palavras-chave: Tomografia por Emissão de Pósitrons (PET). Projeção de Intensidade Máxima (MIP). Classificação Automática. Aprendizado Profundo.

Abstract

Cancer is one of the greatest global public health challenges, with an estimated 35 million new cases by 2035. In this context, Positron Emission Tomography (PET) is essential for diagnosis and monitoring. However, the clinical interpretation of these exams is an exhaustive task, subject to the specialist's subjectivity and limited by the high complexity of 3D volumetric data. This work proposes a method for the automatic classification of whole-body PET scans of patients with lung cancer, lymphoma, melanoma, and healthy individuals, using deep learning techniques applied to Maximum Intensity Projection (MIP) representations. The method is structured in four stages: generation of MIP images in the coronal and sagittal axes, preprocessing, feature extraction, and classification. Six architectures for feature extraction (ConvNeXt, EfficientNet-B0, Swin, and VGG19) and three classifiers (MLP, SVM, and XGBoost) were evaluated. The method achieved results of 96.45% for the AUC metric, 91.98% for the accuracy, 91.63% for the F1-Score, 91.18% for the sensitivity, and a precision of 92.08%. These results show that the use of MIP representations, combined with a set of perspective-specific specialized architectures, allows for satisfactory performance, approaching approaches that use 3D volumes and hybrid examinations (PET/CT).

Keywords: Positron Emission Tomography (PET). Maximum Intensity Projection (MIP). Automatic Classification. Deep Learning.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

AUC	Área Sob a Curva (do inglês, *Area Under the Curve*)
BCE	Entropia Cruzada Binária (do inglês, *Binary Cross Entropy*)
BW	Peso Corporal (do inglês, *Body Weight*)
CAD	Diagnóstico Assistido por Computador (do inglês, *Computer-Aided Diagnosis*)
CLAHE	Equalização Adaptativa de Histograma com Contraste Limitado (do inglês, *Contrast Limited Adaptive Histogram Equalization*)
CNN	Rede Neural Convolucional (do inglês, *Convolutional Neural Network*)
CT	Tomografia Computadorizada (do inglês, *Computed Tomography*)
DP	Desvio Padrão
FDG	Fluodesoxiglicose marcada com Flúor-18 (do inglês, *Fluorodeoxyglucose*)
FN	Falso Negativo
FP	Falso Positivo
FPR	Taxa de Falsos Positivos (do inglês, *False Positive Rate*)
F-Res	Modelos Fracionários-Residuais (do inglês, *Fractional-Residual Models*)
GPU	Unidade de Processamento Gráfico (do inglês, *Graphics Processing Unit*)
Grad-CAM	Mapeamento de Ativação de Classe Ponderado pelo Gradiente (do inglês, *Gradient-weighted Class Activation Mapping*)
IA	Inteligência Artificial
LN	Normalização de Camada (do inglês, *Layer Norm*)
LSTMs	Memória de Longo e Curto Prazo (do inglês, *Long Short-Term Memory*)
MIP	Projeção de Intensidade Máxima (do inglês, *Maximum Intensity Projection*)
ML	Aprendizado de Máquina (do inglês, *Machine Learning*)
MLP	Perceptron Multicamadas (do inglês, *Multilayer Perceptron*)

PET	Tomografia por Emissão de Pósitrons (do inglês, *Positron Emission Tomography*)
PSMA	Antígeno de Membrana Prostático Específico (do inglês, *Prostate-Specific Membrane Antigen*)
RBF	Função de Base Radial (do inglês, *Radial Basis Function*)
ReLU	Unidade Linear Retificada (do inglês, *Rectified Linear Unit*)
RGB	Modelo de cores Vermelho, Verde e Azul (do inglês, *Red, Green, Blue*)
RM	Ressonância Magnética
RNNs	Redes Neurais Recorrentes (do inglês, *Recurrent Neural Networks*)
ROC	Característica de Operação do Receptor (do inglês, *Receiver Operating Characteristic*)
sCT	Tomografia Computadorizada Sintética (do inglês, *Synthetic Computed Tomography*)
SGD	Gradiente Descendente Estocástico (do inglês, *Stochastic Gradient Descent*)
SPECT	Tomografia Computadorizada por Emissão de Fóton Único (do inglês, *Single Photon Emission Computed Tomography*)
SUV	Valor de Captação Padronizado (do inglês, *Standardized Uptake Value*)
SUVmax	Valor de Captação Padronizado Máximo
SVM	Máquinas de Vetores de Suporte (do inglês, *Support Vector Machine*)
SW-MSA	Autoatenção de Múltiplas Cabeças Baseada em Janela Deslocada (do inglês, *Shifted Window-based Multi-head Self-Attention*)
TC	Tomografia Computadorizada
TPE	Estimador Parzen de Estrutura em Árvore (do inglês, *Tree-structured Parzen Estimator*)
TPR	Taxa de Verdadeiros Positivos (do inglês, *True Positive Rate*)
VGG	Grupo de Geometria Visual (do inglês, *Visual Geometry Group*)
ViT	Transformadores de Visão (do inglês, *Vision Transformers*)
VN	Verdadeiro Negativo

VP	Verdadeiro Positivo
W-MSA	Autoatenção de Múltiplas Cabeças Baseada em Janela (do inglês, *Window-based Multi-head Self-Attention*)
XGBoost	*Extreme Gradient Boosting*

Sumário

1 Introdução

O câncer é uma doença que atinge milhões de pessoas anualmente. Segundo a GLOBOCAN, houve cerca de 20 milhões de novos casos de câncer apenas no ano de 2022, com o câncer de pulmão sendo o mais frequente e o com maior taxa de mortalidade, com 2,5 milhões de mortes, representando 12,4% de todos os casos naquele ano (BRAY et al., 2024). Além disso, as previsões baseadas em dados demográficos indicam que o número pode chegar a 35 milhões até 2035 (BRAY et al., 2024). Essa tendência de crescimento reflete uma mudança global nos perfis de risco. Observa-se uma transição em que o tabagismo prevalece em regiões como a América do Sul, Ásia e África, ao mesmo tempo em que a industrialização altera a tipologia da doença, visto que cânceres antes associados a agentes infecciosos dão espaço para aqueles ligados ao estilo de vida e à poluição ambiental (WILD; WEIDERPASS; STEWART, 2020).

Com o crescimento de novos casos, o diagnóstico e monitoramento durante os estágios iniciais da doença são muito importantes, por ser silenciosa e agressiva (World Health Organization, 2024). Logo, diversas formas de acompanhamento e diagnóstico foram desenvolvidas ao longo dos anos, como Tomografia Computadorizada (TC), a Ressonância Magnética (RM), a Tomografia por Emissão de Pósitrons (PET), biópsia histopatológica, etc. (ALI et al., 2024).

Os exames PET são frequentemente utilizados no acompanhamento e diagnóstico de diferentes tipos de câncer. Para a realização do exame, um radiotraçador é aplicado no paciente e, após um período, o exame é feito na máquina PET. Em muitos casos, esse exame é feito junto da TC, que traz informações morfológicas do corpo, enquanto o PET traz informações relacionadas à fisiologia do paciente (BUSHBERG et al., 2012). O resultado do exame é uma imagem 3D do corpo do paciente, que pode ser de partes específicas, como cabeça e pescoço, ou do corpo inteiro, dependendo do diagnóstico ou da suspeita do tipo de câncer. A capacidade do exame PET de detectar diferentes tipos de câncer se dá pelo uso do radiotraçador, que leva o composto radioativo para partes específicas do corpo (BUSHBERG et al., 2012).

Dessa forma, a rotina clínica dos especialistas para a interpretação dos exames exige a realização da inspeção visual combinada à análise semiquantitativa, utilizando métricas como o Valor de Captação Padronizado (SUVmax), sendo a forma de medir a atividade metabólica em regiões do corpo (KINAHAN; FLETCHER, 2010a). Um dos problemas para essa tarefa é a alta quantidade de falsos positivos gerada por processos inflamatórios ou infecciosos, mimetizando o comportamento metabólico tumoral. Somada à fadiga decorrente do alto volume de dados 3D e da complexidade, pode ocorrer uma

variabilidade na análise dos resultados, já que o diagnóstico manual se torna uma tarefa exaustiva e suscetível a inconsistências (FALLAHPOOR et al., 2024).

Além disso, a imagem gerada pelo exame depende do tipo de radiotraçador utilizado. O mais comum é o 18F-FDG, formado por um análogo à glicose e responsável por levar o composto radioativo flúor-18 para as áreas com maior atividade metabólica, como cérebro, coração, rins, inflamações, infecções e lesões cancerígenas (BUSHBERG et al., 2012). No entanto, há um desafio ao lidar com esse tipo de imagem pela escassez de radiologistas especializados nela, existindo uma demanda muito grande para poucos especialistas (FROOD et al., 2024).

Visando o suporte aos especialistas, o diagnóstico assistido por computador (CAD - do inglês *Computer-aided diagnosis*) vem se popularizando nos últimos anos como ferramenta essencial na prática clínica (CHAN; SAMALA; HADJIISKI, 2019). Dessa forma, a Inteligência Artificial (IA) vem se destacando por oferecer avaliações objetivas e reprodutíveis, fundamentadas na extração quantitativa de características das imagens médicas (ACHARYA et al., 2014). Além disso, os algoritmos de aprendizado profundo avançaram essa capacidade, permitindo a identificação automática de padrões complexos para facilitar o diagnóstico, detectando sintomas e aprimorando imagens (FALLAHPOOR et al., 2024).

No contexto de exames PET, algoritmos de IA são capazes de realizar classificação automática de exames, identificando se há presença de malignidade ou o subtipo da lesão. Além disso, essas técnicas são aplicadas em tarefas de detecção, segmentação, reconstrução e aprimoramento da qualidade das imagens (FALLAHPOOR et al., 2024). Frequentemente, a arquitetura desses métodos se dá em um extrator de características, otimizado durante o treinamento para codificar e extrair as representações visuais mais discriminantes e relevantes de cada exame (LITJENS et al., 2017). Tradicionalmente, essa extração tem sido dominada pelas Redes Neurais Convolucionais (CNNs), que operam através de convoluções locais para identificar padrões visuais (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Contudo, mais recentemente, arquiteturas baseadas em *Vision Transformers* (ViT) emergiram como uma alternativa robusta, destacando-se pela capacidade de capturar dependências globais de longo alcance e modelar relações complexas entre regiões distantes da imagem (DOSOVITSKIY, 2020).

Outro fator determinante é o ambiente de execução dos algoritmos. Por se tratar de exames 3D, há uma demanda computacional significativamente superior à observada em imagens 2D (GIL et al., 2023). Estratégias comuns para mitigar esse custo incluem o uso de fatias isoladas do volume ou Projeções de Intensidade Máxima (MIP — *Maximum Intensity Projection*) (RANA et al., 2020). Embora frequente em exames PET, a geração dessas representações exige cautela para minimizar a perda de informações clínicas relevantes. Identifica-se, contudo, uma lacuna no estado da arte, sendo a carência de métodos que

aliem baixo custo computacional à alta precisão, especialmente em cenários de classificação com mais de um tipo de lesão. Este trabalho contribui para o preenchimento dessa lacuna ao propor o uso de representações MIP em uma abordagem multivista, estabelecendo um comparativo rigoroso entre o desempenho de CNNs e Transformers e abordando mais de um tipo de lesão.

Portanto, nota-se que a análise manual de exames PET constitui uma tarefa demorada e trabalhosa, muitas vezes limitada pela subjetividade e pela fadiga visual do especialista. Neste contexto, uma avaliação automática pode ser benéfica para otimizar a rotina clínica. Ao equilibrar estratégias de eficiência computacional com arquiteturas robustas de aprendizado profundo, essas ferramentas podem oferecer benefícios tangíveis tanto para o médico quanto para o paciente, atuando não como substitutas, mas como suporte para acelerar o processo diagnóstico e elevar a precisão dos laudos médicos.

1.1 Objetivos

Nesta seção serão descritos o objetivo geral e os objetivos específicos.

1.1.1 Objetivo Geral

O objetivo geral deste trabalho é propor um método computacional para classificação automática de exames PET de corpo inteiro em saudáveis ou com lesão, utilizando representações MIP e arquiteturas de aprendizado profundo.

1.1.2 Objetivos Específicos

Para atingir o objetivo principal deste trabalho, foi necessário atingir os seguintes objetivos específicos:

- Analisar quantitativamente os níveis de SUV nos exames PET e nas regiões de lesão, de forma a auxiliar a geração de representações MIP;
- Criar representações MIP nos eixos coronal e sagital do exame;
- Implementar e adaptar arquiteturas CNNs e Transformers para usar como extratores de características;
- Analisar o impacto da utilização combinada das visões coronal e sagital no desempenho da classificação, comparando-a com abordagens de visão única;
- Efetuar uma análise comparativa entre o método proposto e métodos utilizados na literatura; e

- Utilizar técnicas de explicabilidade para validar se as áreas de atenção dos modelos coincidem com as regiões anatômicas com lesão.

1.2 Organização do trabalho

O presente trabalho está dividido em mais cinco capítulos, sendo eles:

Capítulo 2 - Trabalhos Relacionados - Os trabalhos relacionados ao tema, descrevendo seus métodos e resultados obtidos.

Capítulo 3 - Fundamentação Teórica - São explicados e abordados trabalhos relacionados ao tema, além dos conceitos necessários para a compreensão dos resultados e da metodologia. Nele são explicados os conceitos sobre exames PET, os tipos de lesões presentes nos exames utilizados no trabalho, inteligência artificial, aprendizado de máquina e aprendizado profundo.

Capítulo 4 - Materiais e Método - São descritos os passos necessários para desenvolver e implementar o método para geração de imagens MIP dos exames PET e classificação das mesmas utilizando dois tipos de visão do mesmo exame, além de explicar o funcionamento e o desenvolvimento de cada passo.

Capítulo 5 - Resultados e Discussão - São apresentados os resultados obtidos com a implementação do método, além de comparações com outras arquiteturas e outros métodos.

Capítulo 6 - Conclusão - Apresenta as considerações finais sobre o método proposto, além de descrever possíveis trabalhos futuros.

2 Trabalhos Relacionados

Este capítulo apresenta uma revisão dos trabalhos relacionados pertinentes ao escopo desta pesquisa. Os estudos selecionados abordam tarefas de classificação e detecção em exames PET, PET/CT ou PET/RM, bem como abordagens que empregam a classificação como etapa preliminar à segmentação de lesões.

Em Dirks et al. (2022), exploram a combinação da alta sensibilidade e especificidade do exame PET com 18F-FDG e a alta resolução da Tomografia Computadorizada (TC, do inglês Computed Tomography - CT), possibilitando uma avaliação precisa do estado da doença e da resposta ao tratamento. Nesse contexto, é proposto um sistema de detecção e segmentação automática de lesões de melanoma em exames PET/CT, empregando limiarização baseada na intensidade, seguida da classificação da região (normal ou maligna) utilizando CNN. O método alcançou um F1-Score de 0,7500 para a detecção e 0,8493 de Dice para a tarefa de segmentação.

Em Ren et al. (2025) desenvolvem um estudo que avalia a capacidade discriminativa do perfil metabólico PET/CT integrado com aprendizado de máquina para classificação da patologia. Foram utilizados três modelos de aprendizado de máquina diferentes, sendo eles *Random Forest* (RF), Support Vector Machines (SVM) e Redes Neurais Artificiais (RNA). Todos foram treinados utilizando os valores de SUV do corpo em pacientes com linfoma. Assim, foi possível alcançar o melhor resultado com RF, atingindo as métricas de AUC de 0,942, precisão de 0,9388 e especificidade de 1,00, superando o SVM e RNN.

Os autores Häggström et al. (2024) propuseram um método de classificação de exames 18F FDG PET/CT de pacientes com linfoma, realizando diversos tipos de experimentos, focados em diferentes tipos de otimização de métricas, como acurácia e sensibilidade. Além disso, foram utilizados os volumes 3D dos exames e representações MIP nas visões coronal e sagital com adaptações da ResNet 34 de acordo com a entrada. O melhor resultado obtido alcançou uma AUC de 0,949, acurácia de 0,890, sensibilidade de 0,868 e especificidade de 0,913.

Por outro lado, Sibille et al. (2020) fizeram um estudo buscando avaliar as configurações de CNNs para localizar e classificar padrões de captação de imagens de PET/CT de corpo inteiro com 18F FDG em pacientes com câncer de pulmão e linfoma. O método proposto utilizou-se das anotações feitas por especialistas na base de dados para desenvolver uma CNN capaz de detectar focos positivos para a captação do radiofármaco utilizado nos exames. Os experimentos foram realizados apenas com exames CT, apenas com PET, com a combinação entre os dois e com representações MIP dos exames PET. O melhor resultado foi utilizando combinações entre exames PET e CT, alcançando (0,98 e 0,95) de

AUC, sensibilidade (0,871 e 0,754) e especificidade (0,99 e 0,958) para a classificação de câncer de pulmão e linfoma, respectivamente.

Zhang et al. (2024) tiveram como objetivo desenvolver e avaliar o desempenho de um modelo preditivo baseado em exames PET/CT que seja capaz de distinguir os subtipos histológicos do adenocarcinoma pulmonar e de células escamosas com dados de variáveis clínicas, metabólicas e radiológicas. Utilizando um classificador de regressão logística, obteve a melhor performance, alcançando uma AUC de 0,870.

No trabalho de Park et al. (2021), o objetivo foi avaliar e desenvolver um sistema baseado em aprendizado profundo para diagnóstico de câncer de pulmão com TC e PET/CT com FDG. A metodologia empregou transferência de aprendizado (transfer learning) em uma ResNet-18, que foi treinada com dados de imagens de TC e metadados dos valores de SUVmáx dos exames PET, e os tamanhos das lesões advindas de exames PET/CT. Dessa forma, foi possível alcançar 0,877 de AUC, 0,825 de Acurácia, 0,856 de Precisão, Sensibilidade de 0,912 e F1-Score de 0,882.

O trabalho desenvolvido por Alves, Cardoso e Gama (2024) explorou a capacidade de aprendizado das CNNs de aprenderem características automaticamente, aplicando-as em imagens de 18F FDG PET/CT para reconhecimento de padrões de nódulos pulmonares. Foi realizada uma validação cruzada utilizando *Stacked* 3D CNNs, com as arquiteturas VGG, *Inception* e ResNet-50. Por fim, o método proposto alcançou 0,8385 de AUC, 0,8000 de sensibilidade, 0,6923 de especificidade, e uma acurácia de 0,7391.

O estudo de Heiliger et al. (2022) realiza uma segmentação automática em exames PET/CT de três tipos de lesões no conjunto de dados AutoPET, sendo câncer de pulmão, linfoma e melanoma. Antes da segmentação, os autores fazem uso de uma classificação automática utilizando projeções MIP. A classificação é realizada utilizando *ensemble*, com as projeções coronal e sagital normais e com a remoção do cérebro, reduzindo a falsos positivos naturais do exame, além de utilizar o *backbone* das ResNet 18 e 50. Para a classificação, foi utilizado o OU lógico, ou seja, um exame só era classificado como negativo, apenas se todos os modelos de classificação o classificassem assim. Reduzindo a quantidade de falsos negativos. Logo, alcançou-se uma acurácia média de 0,743 na validação cruzada utilizando 5 folds.

Pang et al. (2024) fizeram uso do conjunto de dados AutoPET e de exames de PET/RM, desenvolvendo um método capaz de detectar lesões. Para isso, o modelo de detecção foi baseado em modelos fracionários-residuais 2D e 3D (F-Res). Para estender o método aos exames de ressônância magnética, uma arquitetura foi treinada para transformar exames de RM em CT (sCT). Os resultados obtidos para os exames PET/RM e PET/CT foram de (0,81 vs 0,78) de acurácia, (0,81 vs 0,84) de precisão, (0,99 vs 0,98) de sensibilidade e (0,88 vs 0,89) de Dice, respectivamente.

Diante da análise dos trabalhos correlatos, identifica-se uma lacuna referente a métodos focados exclusivamente na classificação eficiente de múltiplas lesões. Diferente das abordagens revisadas, este trabalho propõe a classificação como tarefa central, apresentando as seguintes contribuições principais em relação ao estado da arte:

- **Classificação simultânea de múltiplos tipos de lesão:** O método abrange câncer de pulmão, linfoma e melanoma de forma integrada. A maioria dos trabalhos na literatura foca em apenas uma patologia específica, como (DIRKS et al., 2022) para melanoma, ou (ZHANG et al., 2024) e (ALVES; CARDOSO; GAMA, 2024), focados exclusivamente em câncer de pulmão.
- **Uso exclusivo de imagem PET:** Ao contrário de métodos que dependem do par anatômico (CT ou RM) predominantes na literatura revisada, como em (PANG et al., 2024). Logo, esta abordagem utiliza apenas o canal funcional, reduzindo a complexidade dos dados e o custo de armazenamento.
- **Comparativo entre CNNs e Transformers:** Realiza-se um embate direto entre arquiteturas convolucionais clássicas e modelos baseados em *Vision Transformers*, avaliando o compromisso entre custo computacional e precisão diagnóstica.
- **Interpretabilidade para análise clínica:** Implementação de técnicas de visualização (como mapas de calor) que permitem identificar as regiões da imagem determinantes para a classificação.

Ressalta-se que o uso da mesma base de dados (AutoPET) utilizada por (HEILIGER et al., 2022) e (PANG et al., 2024) viabiliza uma comparação direta da eficácia destas contribuições frente a métodos de segmentação e detecção recentes.

Assim, o método proposto pelo presente trabalho consiste na geração de imagens MIP coronal e sagital baseada em uma análise quantitativa dos níveis de SUV dos exames. Em seguida, as representações 2D geradas nos planos coronal e sagital do mesmo paciente são processadas em conjunto simultaneamente por backbones de diferentes arquiteturas, baseados em CNNs e Transformers. As features extraídas são concatenadas e testadas em diversos classificadores diferentes, com a intenção de comparar os resultados entre eles. Por fim, o diferencial do trabalho se dá na geração de imagens MIP válidas, capazes de diferenciar as áreas de lesão das áreas saudáveis, além de um método de classificação automática treinado com três tipos de lesões diferentes, sendo câncer de pulmão, linfoma e melanoma.

Na Tabela 2.1, pode-se ver um resumo dos trabalhos relacionados mencionados neste capítulo, apontando os autores, o objetivo, as técnicas utilizadas e a base de imagens empregada.

Tabela 2.1 – Visão geral dos trabalhos relacionados

Trabalhos	**Objetivo**	**Tipo de Lesão**	**Métodos utilizados**	**Base de imagens**	**Exame utilizado**
(DIRKS et al., 2022)	Detecção Segmentação	Melanoma	Limiarização CNN 2D	Privada	PET/CT
(REN et al., 2025)	Classificação	Linfoma	Dados tabulares (SUV) *Random Forest*	Privada	PET/CT
(HÄGGSTRÖM et al., 2024)	Classificação	Linfoma	ResNet-34 3D + MIP	Privada	PET/CT
(SIBILLE et al., 2020)	Detecção Classificação	Câncer de Pulmão Linfoma	CNN 2D	Privada	PET/CT
(ZHANG et al., 2024)	Classificação	Câncer de Pulmão	Dados Clínicos Regressão Logística	Privada	PET/CT
(PARK et al., 2021)	Classificação	Câncer de Pulmão	Metadados ResNet-18 2D	Privada	PET/CT
(ALVES; CARDOSO; GAMA, 2024)	Classificação de Nódulos	Câncer de Pulmão	Stacked 3D CNNs	Privada	PET/CT
(HEILIGER et al., 2022)	Classificação Segmentação	Câncer de Pulmão Linfoma Melanoma	Ensemble ResNet-18/50 MIP	Pública - AutoPET	PET PET/CT
(PANG et al., 2024)	Detecção	Câncer de Pulmão Linfoma Melanoma	F-ResNet 2D/3D	Pública - AutoPET Privada - PET/RM	PET/CT/RM
Método Proposto	Classificação	Câncer de Pulmão Linfoma Melanoma	MIP + Deep Features + Fusão Multivista	Pública - AutoPET	PET

Fonte: Autoral.

3 Fundamentação Teórica

Esta seção apresenta os fundamentos teóricos necessários para o entendimento do método proposto. São abordados os exames PET, aprendizado de máquina e aprendizado profundo, conceitos que são fundamentais para o bom entendimento do trabalho desenvolvido.

3.1 Tomografia por Emissão de Pósitrons

A Tomografia por Emissão de Pósitrons (PET, do inglês *Positron Emission Tomography*) é um exame da medicina nuclear, área que vem estudando o uso de isótopos radioativos para fins médicos desde 1920, e é usada para a geração de imagens a partir da concentração de radionuclídeos desde 1940 (SUETENS, 2009). Pesquisas na área para geração de imagens foram desenvolvidas até os anos 1970, mas o poder computacional da época impedia a implementação dos métodos criados. Assim, na década de 1980, a Tomografia Computadorizada por Emissão de Fóton Único (SPECT, do inglês *Single Photon Emission Computed Tomography*) surgiu e serviu de inspiração para a criação do exame PET, ao ser observado que duas câmeras de cintilação poderiam ser combinadas para detectar pares de fótons originados após a emissão de pósitrons (SUETENS, 2009).

As imagens geradas em exames PET representam a distribuição de radionuclídeos emissores de pósitrons no paciente (BUSHBERG et al., 2012). Para tal, moléculas traçadoras marcadas com um isótopo instável (radionuclídeo) são administradas, geralmente por via intravenosa (SUETENS, 2009). Uma vez no organismo, essas moléculas participam dos processos metabólicos, enquanto os isótopos instáveis emitem radiação gama. Isso permite mensurar a concentração da molécula traçadora e, consequentemente, avaliar a atividade metabólica do paciente (SUETENS, 2009).

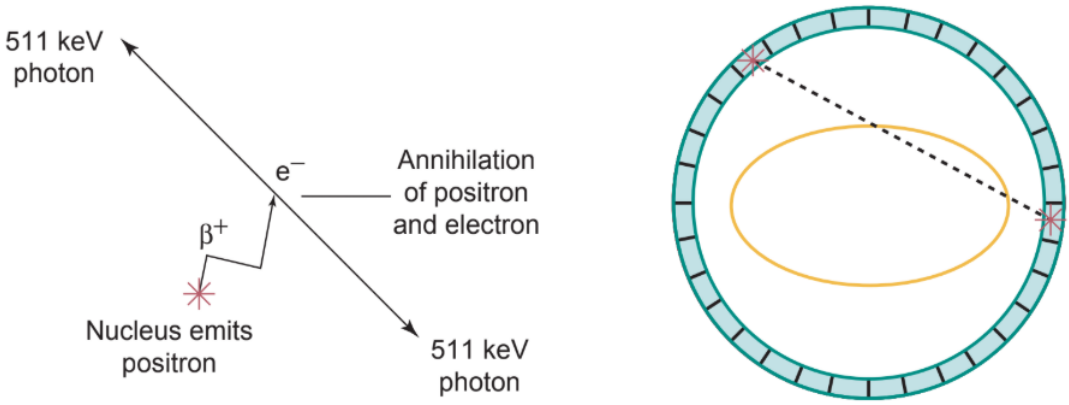
Por fim, com o passar dos anos e o desenvolvimento tecnológico aliado à necessidade clínica, sistemas de dupla modalidade foram desenvolvidos, tais como PET/CT e PET/RM. Ambos combinam a capacidade dos exames PET de realizar a análise fisiológica do paciente com o detalhamento morfológico fornecido pela Tomografia Computadorizada (TC) ou Ressonância Magnética (RM) (BUSHBERG et al., 2012). Além disso, a instrumentação PET evoluiu dos scanners convencionais de corpo inteiro, nos quais a mesa com o paciente se desloca através de um campo de visão axial limitado, onde os exames são realizados em partes específicas, como cabeça e pescoço, ou do topo da cabeça à parte superior das coxas (TOWNSEND, 2008), para os sistemas de corpo completo. Neste último, a captação da radiação é realizada simultaneamente em todo o organismo, devido ao longo campo de visão axial, o que aumenta drasticamente a sensibilidade e mitiga erros e ruídos associados

aos modelos anteriores (CHERRY et al., 2018; BADAWI et al., 2019).

Para a formação da imagem nesses sistemas, independentemente da arquitetura, é fundamental o uso de radiofármacos específicos. Nesse contexto, o exame PET é amplamente utilizado na oncologia para o diagnóstico e estadiamento de diversas neoplasias. Embora existam diferentes radiotraçadores, o mais comum é o 18F FDG (Fluodesoxiglicose marcada com flúor-18), um análogo da glicose marcado com o isótopo emissor de pósitrons flúor-18 (FLETCHER et al., 2008). Após a administração, o composto distribui-se pelo organismo, acumulando-se em áreas de alta atividade glicolítica, como cérebro, coração, rins, além de lesões tumorais, inflamações e infecções (SUETENS, 2009). Nota-se também intensa atividade na bexiga, decorrente não do metabolismo, mas da excreção renal do material radioativo (SUETENS, 2009).

Após a injeção intravenosa do radiotraçador 18F FDG, aguarda-se um período de captação de 30 a 60 minutos antes que o paciente seja posicionado no scanner para a aquisição das imagens (BOELLAARD et al., 2015). Com o decaimento do isótopo radioativo, ocorre a emissão de pósitrons (a antipartícula do elétron). Quando um pósitron colide com um elétron no tecido, ocorre o fenômeno da aniquilação, gerando dois fótons de radiação gama emitidos em direções opostas. Esses fótons são capturados simultaneamente pelos detectores da máquina, permitindo a reconstrução da distribuição do traçador no corpo a partir da interseção das linhas de resposta dos eventos detectados (BUSHBERG et al., 2012). A Figura 3.1 mostra, à esquerda, o raio de aniquilação e a direção dos fótons de radiação gama (511 keV), e, à direita, a representação desse mesmo raio sendo detectado pelos sensores da máquina PET.

Figura 3.1 – Representação do Raio de Aniquilação



Fonte: (BUSHBERG et al., 2012).

Após a detecção desses raios, a imagem 3D é gerada, mostrando em evidência as regiões de alta atividade metabólica, calculadas com base na absorção padronizada (SUV, do inglês - *Standardized Uptake Values*). As lesões cancerígenas possuem uma maior

absorção do radiotraçador, o que facilita a análise visual dos especialistas, possuindo altos níveis de SUV, deixando as lesões em evidência (BUSHBERG et al., 2012; SUETENS, 2009).

A Equação 3.1 apresenta a fórmula para o cálculo do SUV:

$$SUV = \frac{C_{img}}{D_{inj}/BW} \tag{3.1}$$

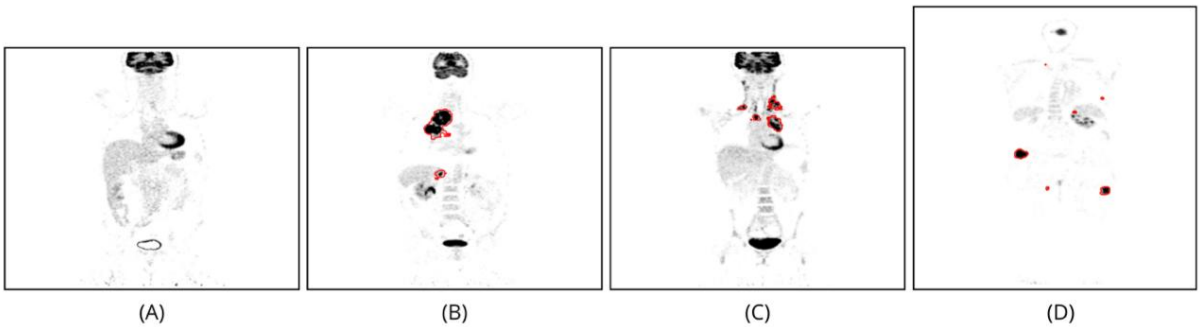
Onde C_{img} representa a concentração de radioatividade medida na imagem (kBq/mL), D_{inj} é a dose de radiofármaco injetada (MBq) e BW corresponde ao peso corporal do paciente (kg).

O câncer de pulmão geralmente tem os valores de absorção padronizados (SUV, do inglês - *Standardized Uptake Values*) mais baixos, onde muitas vezes é considerado o limiar de SUV $>$ 2,5 como lesão (KINAHAN; FLETCHER, 2010b). Além disso, apenas em casos raros há metástase para partes inferiores do corpo, fazendo com que não seja necessário realizar o exame PET de corpo completo (PUTRO et al., 2024).

Em casos de melanoma, a lesão pode estar presente em qualquer parte do corpo, sendo determinante a realização do exame no corpo de forma completa (TAS, 2012). Já em casos de linfoma, as lesões possuem grande evidência na região do tórax, frequentemente acometendo o mediastino e estruturas hilares. Contudo, devido à disseminação sistêmica característica das doenças linfoproliferativas, o mapeamento por imagem não deve se restringir ao tórax, sendo essencial a varredura corporal para a correta identificação de linfonodos afetados e o estadiamento preciso da doença (ANSELL, 2020).

A Figura 3.2 mostra exemplos extraídos de volumes PET no eixo coronal, sendo a fatia central de um paciente saudável (A), seguida das fatias com maior quantidade de voxels com lesão, marcadas em vermelho, sendo câncer de pulmão (B), linfoma (C) e melanoma (D), respectivamente.

Figura 3.2 – Exemplos de Fatias Extraídas de Exames PET.



Fonte: (GATIDIS et al., 2022) ADAPTADA.

Além disso, a alta captação fisiológica natural em algumas regiões torna a interpreta-

ção do especialista mais complexa, o que reforça a necessidade de métodos computacionais para o auxílio.

3.2 Aprendizado de Máquina

O aprendizado de máquina (ML, do inglês *Machine Learning*) é um campo multidisciplinar que abrange uma ampla gama de domínios de pesquisa, além de ser uma subárea da Inteligência Artificial (IA). O ML refere-se à capacidade de computadores aprenderem sem serem explicitamente programados (SAMUEL, 1959). Ao contrário dos seres humanos que aprendem com a experiência, as máquinas necessitam de dados (ALZUBI; NAYYAR; KUMAR, 2018).

Comumente, o ML tem sido utilizado em várias áreas da computação para projetar e programar algoritmos em diferentes tipos de mercados, como filtragem de spam de e-mails, detecção de fraudes, negociações online (ALZUBI; NAYYAR; KUMAR, 2018). Antes de resolver um problema, ele deve ser categorizado de forma adequada para que o algoritmo de aprendizado de máquina mais apropriado possa ser aplicado a ele, podendo ser tarefas de classificação, detecção de anomalias, regressão, agrupamento e aprendizado por reforço (ALZUBI; NAYYAR; KUMAR, 2018).

Independentemente do domínio de aplicação, os dados assumem um papel central no aprendizado de máquina. Uma vez que a eficácia dos algoritmos está intrinsecamente ligada aos dados de entrada, torna-se imprescindível a utilização de amostras representativas e de alta qualidade (BUDACH et al., 2022). Adicionalmente, a presença, ausência ou combinação parcial de rótulos nos dados é o fator determinante para a escolha do paradigma de aprendizado de máquina a ser adotado para a resolução do problema (ALNUAIMI; ALBALDAWI, 2024).

A diferença fundamental entre os paradigmas de aprendizado baseia-se na existência de uma variável-alvo nos dados. Dados rotulados são constituídos por pares de entrada e saída, provendo ao modelo o gabarito para o treinamento, ao passo que dados não rotulados carecem dessa informação prévia (SARKER, 2021). Nesse contexto, o aprendizado supervisionado emprega dados rotulados para aproximar uma função de mapeamento apta a predizer saídas para novos dados. Em contrapartida, o aprendizado não supervisionado explora dados não rotulados visando identificar padrões latentes ou estruturas intrínsecas, sem a orientação de um alvo predefinido (SARKER, 2021).

No contexto do aprendizado supervisionado, a classificação destaca-se como uma tarefa fundamental em que o objetivo é prever rótulos de classe discretos para novos dados. Nesse processo, os algoritmos utilizam um conjunto de treinamento composto por exemplos previamente rotulados para aprender uma função de mapeamento, associando as características de entrada a uma categoria específica. Diferentemente da regressão, que

estima valores contínuos, a classificação foca na distinção entre diferentes grupos, buscando estabelecer fronteiras de decisão que permitam ao modelo generalizar o conhecimento adquirido (ALNUAIMI; ALBALDAWI, 2024). Para a execução dessas tarefas, a literatura apresenta uma variedade de algoritmos robustos. Dentre eles, destacam-se as Máquinas de Vetores de Suporte (SVM, do inglês *Support Vector Machine*) e o XGBoost (do inglês *Extreme Gradient Boosting*).

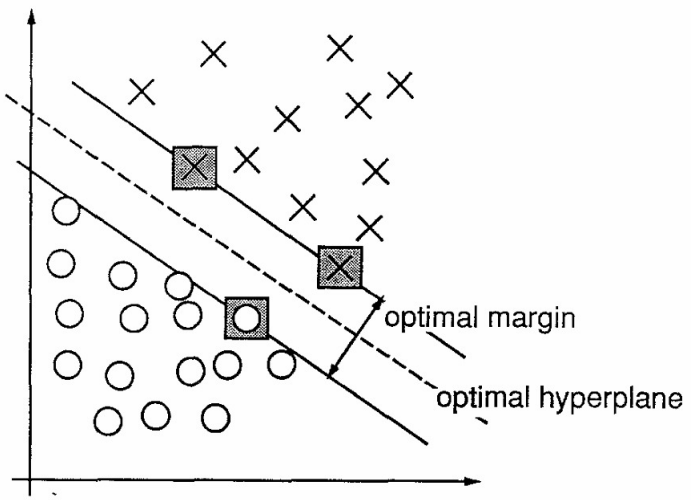
3.2.1 Maquinas de Vetores de Suporte

A Máquina de Vetores de Suporte (SVM) é um algoritmo de aprendizado supervisionado fundamentado na teoria de aprendizagem estatística (CORTES; VAPNIK, 1995). O objetivo central do SVM é encontrar uma fronteira de decisão que não apenas separe as diferentes classes de dados, mas que o faça maximizando a capacidade de generalização do modelo.

Diferente de classificadores lineares simples que buscam qualquer linha capaz de separar as classes, o SVM busca o hiperplano ótimo. Este é definido como a superfície de decisão que maximiza a margem, ou seja, a distância entre o hiperplano separador e os pontos de dados mais próximos de cada classe. A premissa teórica é que quanto maior a margem, menor será o erro esperado em dados futuros (CORTES; VAPNIK, 1995).

Os pontos de dados que residem exatamente no limite dessa margem são denominados vetores de suporte. Eles são os elementos mais críticos do conjunto de dados, pois são os únicos que definem a posição e a orientação do hiperplano ótimo. Todos os outros pontos além da margem são irrelevantes para a definição da fronteira de decisão, o que confere ao SVM uma eficiência notável em espaços de alta dimensão (CORTES; VAPNIK, 1995).

Figura 3.3 – Representação da Classificação com SVM



Fonte: (CORTES; VAPNIK, 1995).

A Figura 3.3 ilustra os conceitos centrais em um problema de classificação binária. A linha central tracejada representa o hiperplano ótimo, que atua como a fronteira de decisão final. O espaço entre as duas linhas sólidas paralelas é denominado margem ótima, cuja largura o algoritmo busca maximizar durante o treinamento. Os pontos de dados destacados com quadrados cinzas são os vetores de suporte.

3.2.2 XGBoost

O XGBoost (*Extreme Gradient Boosting*) é um algoritmo de aprendizado de máquina baseado em árvores de decisão que opera sob o paradigma de *ensemble*. Diferente de métodos que constroem árvores independentes (como o *Random Forest*), o XGBoost utiliza a técnica de *Gradient Boosting*, onde novos modelos são adicionados sequencialmente para corrigir os erros residuais dos modelos anteriores, transformando aprendizes fracos em um preditor forte (CHEN, 2016).

A arquitetura do XGBoost deriva diretamente do paradigma de *Ensemble Learning*. Métodos de ensemble operam sob a premissa de que a combinação estratégica de múltiplos classificadores tende a superar a performance de qualquer membro individual do conjunto (DIETTERICH, 2000). Essa superioridade estatística ocorre porque a agregação de modelos permite mitigar o risco de escolher um classificador localmente ótimo, mas globalmente impreciso, além de expandir o espaço de hipóteses representáveis. No contexto específico do XGBoost, esse comitê não é formado por votação simples, mas por uma construção aditiva onde cada novo classificador é especializado em corrigir as deficiências dos anteriores (DIETTERICH, 2000; CHEN, 2016).

A principal inovação do XGBoost em relação ao *Gradient Boosting* tradicional reside na sua função objetivo otimizada. A função de perda do algoritmo incorpora um termo de regularização formal que penaliza a complexidade do modelo, que por sua vez é baseado no número de folhas e na magnitude dos pesos das folhas (CHEN, 2016). Essa regularização é fundamental para evitar o *overfitting*, garantindo que o modelo generalize bem para dados não vistos.

Além da robustez matemática, o algoritmo foi projetado para alta eficiência computacional. Ele implementa processamento paralelo na construção das árvores, algoritmos de aproximação para encontrar as melhores divisões em dados densos e tratamento automático de valores ausentes (dados esparsos), tornando-o eficaz em grandes conjuntos de dados estruturados (CHEN, 2016).

3.3 Redes Neurais Convolucionais (CNNs)

O Aprendizado Profundo (*Deep Learning*) é uma subárea do aprendizado de máquina que utiliza modelos computacionais compostos por múltiplas camadas de pro-

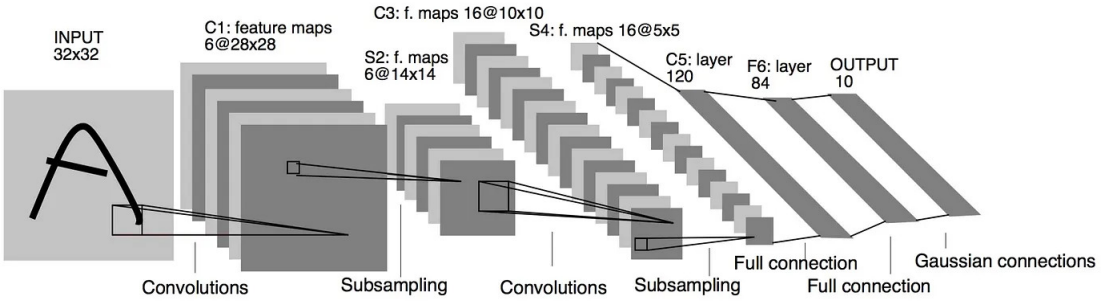
cessamento para aprender representações de dados com múltiplos níveis de abstração (LECUN; BENGIO; HINTON, 2015). Embora o termo *Deep Learning* tenha ganhado popularidade recentemente, seus fundamentos matemáticos remontam à década de 1940. A primeira tentativa formal de simular o funcionamento de um neurônio biológico ocorreu em 1943, quando foi proposto um modelo lógico de processamento neural (MCCULLOCH; PITTS, 1943). Posteriormente, foi desenvolvido o *Perceptron* em 1958, o primeiro algoritmo capaz de aprender pesos a partir de dados, embora limitado a problemas linearmente separáveis (ROSENBLATT, 1958). A evolução para as redes profundas modernas só se tornou viável décadas depois, com o aumento do poder computacional e o refinamento do algoritmo de *Backpropagation* (TERVEN et al., 2023).

Com a evolução de algoritmos baseados nesse tipo de tecnologia, surgiram arquiteturas especializadas para lidar com dados complexos. As Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) representam uma dessas classes, projetadas especificamente para processar dados matriciais, como imagens. Diferentes das Perceptrons de Multicamadas (MLPs) tradicionais que exigem que a entrada seja transformada em um vetor unidimensional, as CNNs preservam a estrutura espacial dos dados e utilizam o compartilhamento de pesos para reduzir drasticamente a complexidade computacional (LECUN; BENGIO; HINTON, 2015).

O marco inicial dessa arquitetura foi a LeNet, projetada para o reconhecimento de dígitos manuscritos, estabelecendo a alternância entre camadas de convolução, usadas para extração de características, e camadas de *pooling*, finalizando com camadas densas (MLP) para a classificação com fluxo padrão de processamento profundo (LECUN et al., 2002a).

A Figura 3.4 mostra a arquitetura da LeNet-5, onde os dados de entrada (*input*) percorrem uma sequência alternada de camadas convolucionais (C1, C3 e C5) e camadas de subamostragem (*pooling*) (S2 e S4). Em seguida, uma camada completamente conectada (F6) é utilizada para unir as características extraídas, que é utilizada como entrada em uma MLP, responsável por fazer a classificação final (LECUN et al., 2002a).

Figura 3.4 – Representação da LeNet - 5



Fonte: (LECUN et al., 2002a).

O funcionamento das CNNs baseia-se primordialmente na extração de caracte-

rísticas, *pooling* e classificação. As camadas convolucionais são utilizadas para extração de características. Nesta etapa, um *kernel* desliza sobre a imagem de entrada realizando produtos escalares locais. Esse processo gera mapas de características que detectam padrões visuais específicos, como bordas, texturas ou formas complexas, dependendo da profundidade da camada (LECUN et al., 2002a).

Em seguida, a camada de *Pooling* é aplicada após a convolução, com a função de reduzir a dimensionalidade espacial dos mapas de características. O método mais comum, o *Max Pooling*, seleciona apenas o valor máximo dentro de uma janela deslizante, reduzindo o custo computacional para as camadas subsequentes, além de conferir ao modelo robustez contra pequenas distorções e variações na posição dos objetos na imagem (LECUN et al., 2002a).

Após as etapas de extração e redução de características, é necessário classificar os padrões identificados. Nas arquiteturas clássicas de CNNs, essa tarefa é realizada por uma estrutura densamente conectada acoplada ao final da rede. Para isso, os mapas de características tridimensionais resultantes são linearizados na camada *flattening*, transformando-se em um vetor unidimensional que serve de entrada para uma rede neural artificial tradicional.

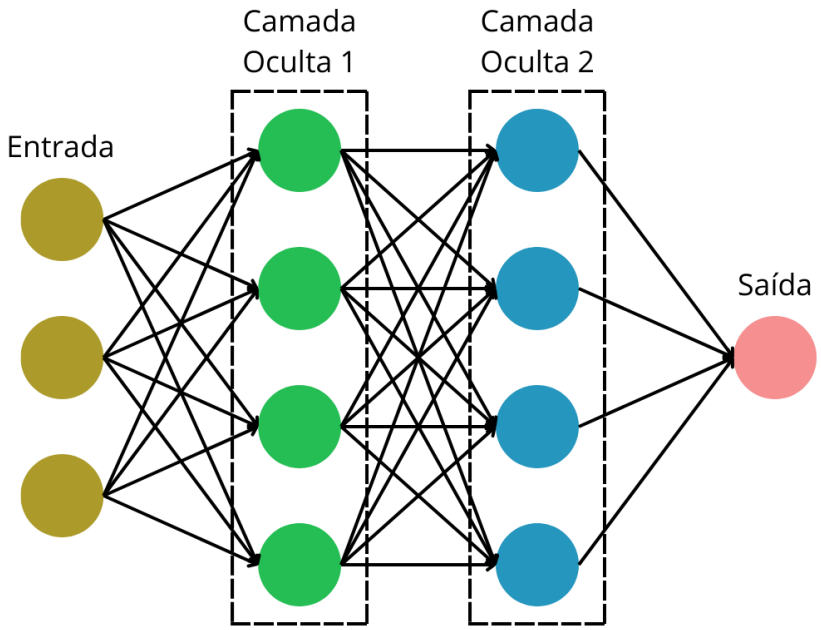


Figura 3.5 – Rede Neural Artificial Multicamadas (MLP, do inglês *Multilayer Perceptron*).

Uma Rede Neural Artificial (RNA) típica, como a utilizada no estágio final de uma CNN, é organizada em camadas sequenciais de neurônios interconectados. A Figura 3.5 ilustra essa estrutura. O processo inicia-se na camada de entrada que recebe o vetor linearizado. Em seguida, a informação passa pelas camadas ocultas, situadas entre a entrada e a saída, onde ocorre o processamento principal. O termo profundo em *Deep Learning* refere-se justamente à presença de múltiplas camadas ocultas empilhadas nessa

etapa (LECUN; BENGIO; HINTON, 2015). Ao final do processamento, a camada de saída produz a predição final do modelo.

O funcionamento matemático dessas camadas densas depende da interação entre três componentes críticos. Os neurônios artificiais são unidades básicas de processamento. Cada neurônio recebe sinais de entrada x, multiplica-os por pesos w, soma um termo de viés (*bias*) b e passa o resultado adiante (LECUN; BENGIO; HINTON, 2015). A Equação 3.2 refere-se ao cálculo realizado pelos neurônios:

$$z = \sum w_i x_i + b \tag{3.2}$$

Para que a rede seja capaz de resolver problemas não triviais, utiliza-se a função de ativação sobre o resultado do neurônio (APICELLA et al., 2021). A função de ativação introduz não-linearidade ao modelo, sendo a sigmoide e a ReLU exemplos comuns que permitem à rede aprender fronteiras de decisão complexas. Por fim, a validação do aprendizado é regida pela função de perda (*Loss Function*), métrica que avalia o desempenho do modelo durante o treinamento e guia o ajuste iterativo dos pesos para reduzir o erro global (TERVEN et al., 2023).

Com o aumento da capacidade de processamento via GPUs, arquiteturas mais profundas e eficientes superaram a LeNet. A VGG (*Visual Geometry Group*) destacou-se por demonstrar que o aumento da profundidade da rede utilizando filtros pequenos (3×3) melhorava significativamente a precisão (SIMONYAN; ZISSERMAN, 2014). Mais recentemente, a busca por eficiência em dispositivos com recursos limitados levou ao desenvolvimento da EfficientNet. Esta arquitetura utiliza um método de escalonamento composto para otimizar simultaneamente a profundidade, a largura e a resolução da rede (TAN; LE, 2019).

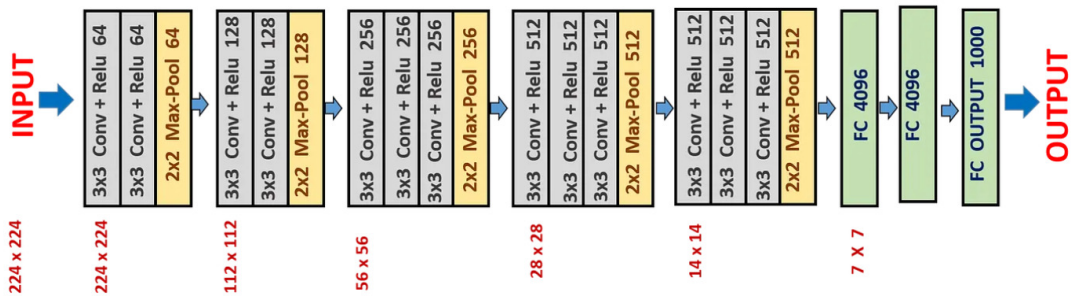
Por fim, houve uma mudança significativa na área de visão computacional com a chegada dos Transformers. Diferentes das CNNs, que analisam partes limitadas da imagem por vez através de convoluções, os Transformers utilizam mecanismos de atenção para considerar toda a imagem simultaneamente. Isso permite que o modelo entenda o contexto global dos dados de forma muito mais completa, superando as abordagens baseadas puramente em CNNs em diversas tarefas complexas de classificação (DOSOVITSKIY, 2020).

3.3.1 Visual Geometry Group (VGG)

A arquitetura VGG representou um avanço significativo no design de redes neurais convolucionais ao investigar o efeito da profundidade na precisão do reconhecimento de imagens (SIMONYAN; ZISSERMAN, 2014). O principal diferencial desta proposta foi a substituição de filtros de convolução de grandes dimensões por uma pilha exclusiva de filtros

muito pequenos, de tamanho (3×3). Essa mudança permitiu aumentar a não-linearidade da rede sem elevar excessivamente o número de parâmetros, possibilitando a construção de modelos mais profundos e com maior capacidade de generalização.

Figura 3.6 – Arquitetura VGG 16



Fonte: (BANGAR, 2022).

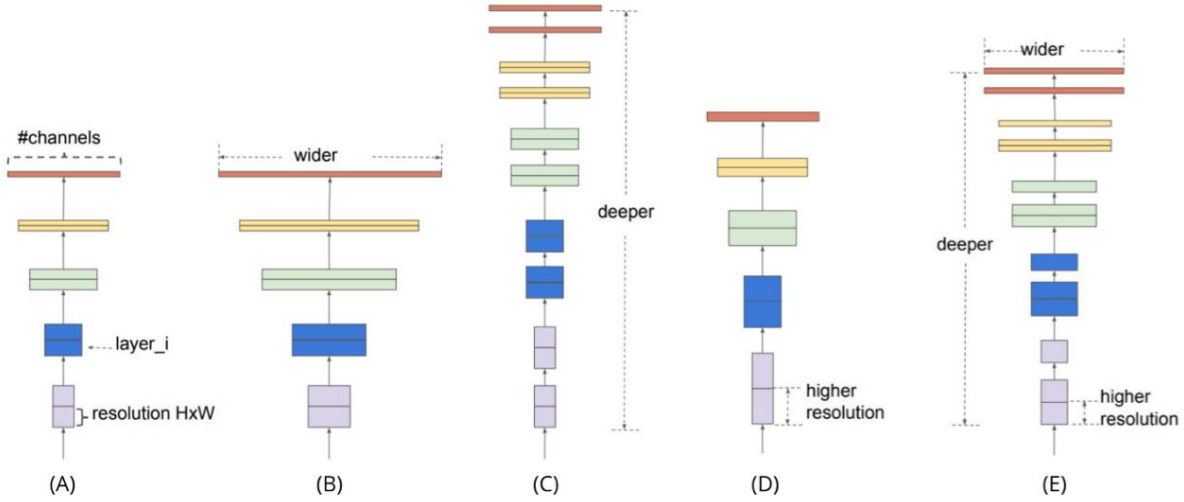
Na Figura 3.6 é possível ver a arquitetura da VGG 16. A entrada consiste em uma imagem de dimensões fixas que atravessa uma sequência de cinco blocos convolucionais. Cada bloco é composto por uma sucessão de camadas de convolução com os filtros 3×3 mencionados, seguidas por uma camada de *max-pooling* responsável pela redução espacial. À medida que os dados avançam na rede, a profundidade dos mapas de características aumenta progressivamente, dobrando de 64 canais no primeiro bloco até alcançar 512 canais nos últimos. O processamento é finalizado por três camadas totalmente conectadas, sendo as duas primeiras com 4096 canais cada e a última correspondendo às classes da tarefa de classificação (SIMONYAN; ZISSERMAN, 2014).

Além disso, existem variantes da mesma arquitetura, sendo as VGG-16 e VGG-19 as mais notáveis. A diferença entre elas está no número de camadas com pesos treináveis. A VGG-16 é composta por 13 camadas convolucionais e 3 camadas densas, totalizando 16 camadas de peso. Já a VGG-19 aprofunda essa arquitetura inserindo três camadas convolucionais adicionais nos últimos blocos de processamento, totalizando 19 camadas (SIMONYAN; ZISSERMAN, 2014).

3.3.2 EfficientNet

Buscando superar as limitações de escalonamento manual, a arquitetura EfficientNet, marcou uma mudança de paradigma ao introduzir o método de escalonamento composto (TAN; LE, 2019). Diferente das abordagens anteriores que aumentavam apenas uma dimensão da rede, a EfficientNet propõe uma mudança de escala em todas as três dimensões de forma uniforme e equilibrada. A justificativa para a mudança se deu pelo fato de que, embora o aumento isolado de uma dimensão melhore a performance até certo ponto, o ganho de acurácia satura se as outras dimensões não forem ajustadas proporcionalmente (TAN; LE, 2019).

Figura 3.7 – Comparação entre métodos de escalonamento e o Escalonamento Composto



Fonte: (TAN; LE, 2019).

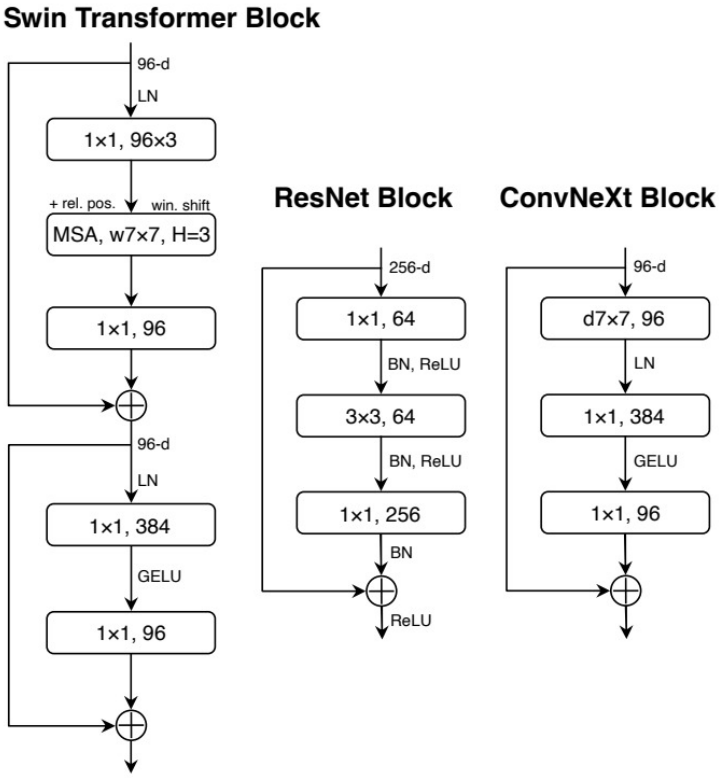
A Figura 3.7 mostra as diferentes estratégias para aumentar a capacidade de uma rede neural. Ela compara o modelo base (A) à esquerda com métodos que expandem apenas uma dimensão isoladamente, tornando a rede mais larga (B), mais profunda (C) ou com maior resolução (D). Já a imagem (E) representa o escalonamento composto da EfficientNet, mostra que a medida que aumenta a largura, a profundidade e a resolução também mudam de forma equilibrada para maximizar o desempenho (TAN; LE, 2019).

3.3.3 ConvNeXt

Com o avanço das arquiteturas baseadas em CNN e o surgimento da Transformers, a ConvNeXt foi criada para desafiar o domínio dos Vision Transformers (ViTs). A intenção do estudo foi modernizar uma ResNet padrão, aplicando características de Transformers, como estratégias de treinamento, mas mantendo a estrutura simples e eficiente de uma ConvNet sem mecanismos de atenção (LIU et al., 2022).

A Figura 3.8 apresenta o design do bloco da ConvNeXt, que adota uma estrutura de gargalo invertido, onde a camada interna é quatro vezes mais larga que a entrada, similar aos blocos MLP dos Transformers. A arquitetura utiliza uma convolução *depthwise* com núcleo grande (7×7) movida para o topo do bloco para simular a mistura de informações espaciais, substitui a normalização *Batch Norm* por *Layer Norm* e troca de funções de ativação, reduzindo a frequência de uso dessas funções para simplificar o processamento (LIU et al., 2022).

Figura 3.8 – Representação da Arquitetura da ConvNeXt



Fonte: (LIU et al., 2022).

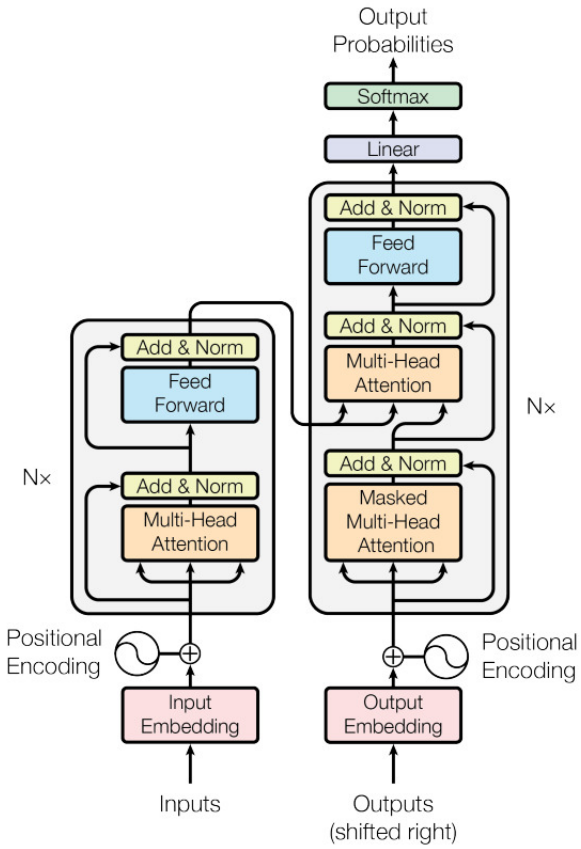
3.4 Transformers

Essas escolhas de design observadas na ConvNeXt derivam diretamente do sucesso da arquitetura Transformer, originalmente proposta para tarefas de processamento de linguagem natural visando superar as restrições computacionais das redes recorrentes (VASWANI et al., 2017). O principal motivador para sua criação foi a ineficiência dos modelos sequenciais (como RNNs e LSTMs), que processam dados passo a passo e impedem a paralelização durante o treinamento. Ao dispensar a recorrência em favor de mecanismos de atenção, o Transformer permitiu a modelagem de dependências globais e o processamento paralelo de toda a sequência, estabelecendo um novo padrão de eficiência para tarefas de tradução e compreensão de texto (VASWANI et al., 2017).

A Figura 3.9, mostra a arquitetura do Transformers, que segue uma lógica codificador-decodificador. O componente à esquerda (Codificador) é composto por uma pilha de N camadas idênticas, cada uma contendo duas subcamadas principais, sendo um mecanismo de autoatenção com múltiplas cabeças (*Multi-Head Attention*) e uma rede *feed-forward*. Já o componente à direita (Decodificador) insere uma terceira subcamada de atenção mascarada (*Masked Multi-Head Attention*), garantindo que a previsão para uma posição atual dependa apenas das posições anteriores conhecidas (VASWANI et al., 2017).

Outro componente importante para o funcionamento das arquiteturas *transformers*

Figura 3.9 – Arquitetura Transformers



Fonte: (VASWANI et al., 2017).

é o mecanismo deautoatenção (*Self-Attention*), responsável por relacionar diferentes posições de uma única sequência para computar uma representação da mesma. Para isso, a entrada é projetada em três vetores diferentes que são aprendidos durante o treinamento, sendo *Query* (Q), *Key* (K) e *Value* (V). A atenção é calculada como uma função que mapeia uma *query* e um conjunto de pares *key-value* para uma saída, utilizando o produto escalar entre Q e K para determinar a relevância entre os tokens. Para evitar gradientes instáveis, é submetido a uma função *Softmax* para obter os pesos normalizados. Esses pesos ponderam os vetores V, gerando a saída final conforme descrito na Equação 3.3 (VASWANI et al., 2017).

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.3}$$

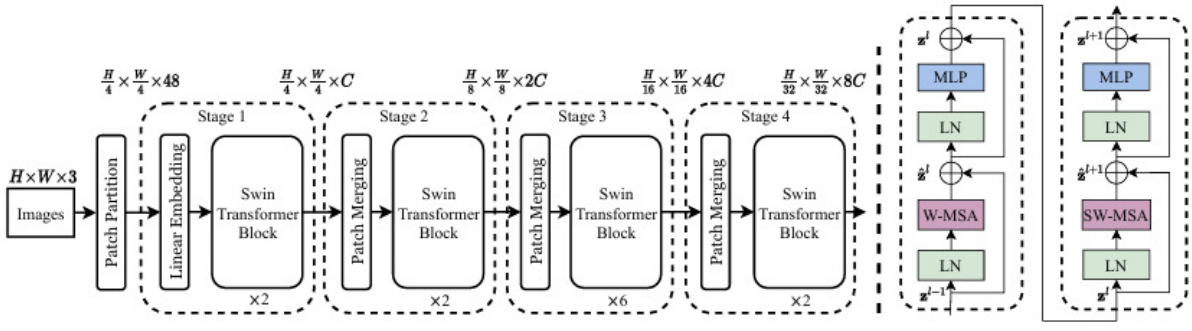
Um elemento crucial é a injeção de *Positional Encodings*. Como o Transformer não processa os dados em ordem sequencial temporal, esses vetores são somados aos *embeddings* de entrada para fornecer à rede informações sobre a posição relativa ou absoluta dos tokens na sequência. Além disso, conexões residuais (*Add*) seguidas de normalização de camada (*Norm*) são aplicadas ao redor de cada subcamada, facilitando o fluxo de gradientes e o treinamento (VASWANI et al., 2017).

3.4.1 Swin Transformers

O sucesso paradigmático dos Transformers no processamento de linguagem natural motivou sua adaptação para o domínio da visão computacional. A arquitetura Vision Transformer (ViT) foi pioneira nessa transição, demonstrando que uma rede baseada puramente em atenção poderia tratar partes de imagens (*patches*) como tokens de texto e atingir desempenho superior às CNNs em grandes volumes de dados (DOSOVITSKIY, 2020). Contudo, o ViT aplica a autoatenção globalmente, o que gera um custo computacional quadrático em relação à resolução da imagem, tornando-o ineficiente para processar imagens de alta resolução. Para superar essa limitação de escalabilidade e reintroduzir a hierarquia de características típica das convoluções, foi proposto o Swin Transformer, que restringe o cálculo da atenção a janelas locais e introduz um mecanismo de deslocamento (LIU et al., 2021).

O Swin Transformer (*Hierarchical Vision Transformer using Shifted Windows*) foi proposto para solucionar os desafios de adaptação do Transformer para visão computacional, especificamente as variações de escala dos objetos e a alta resolução das imagens (LIU et al., 2021). O principal diferencial desta arquitetura é a substituição da atenção global (usada no ViT) por um mecanismo de atenção computado apenas dentro de janelas locais não sobrepostas. Para permitir a comunicação entre essas janelas vizinhas, a rede introduz uma operação de deslocamento (*shifted window*), o que torna a complexidade computacional linear em relação ao tamanho da imagem, ao contrário da complexidade quadrática dos Transformers tradicionais.

Figura 3.10 – Arquitetura Swin Transformers



Fonte: (LIU et al., 2021).

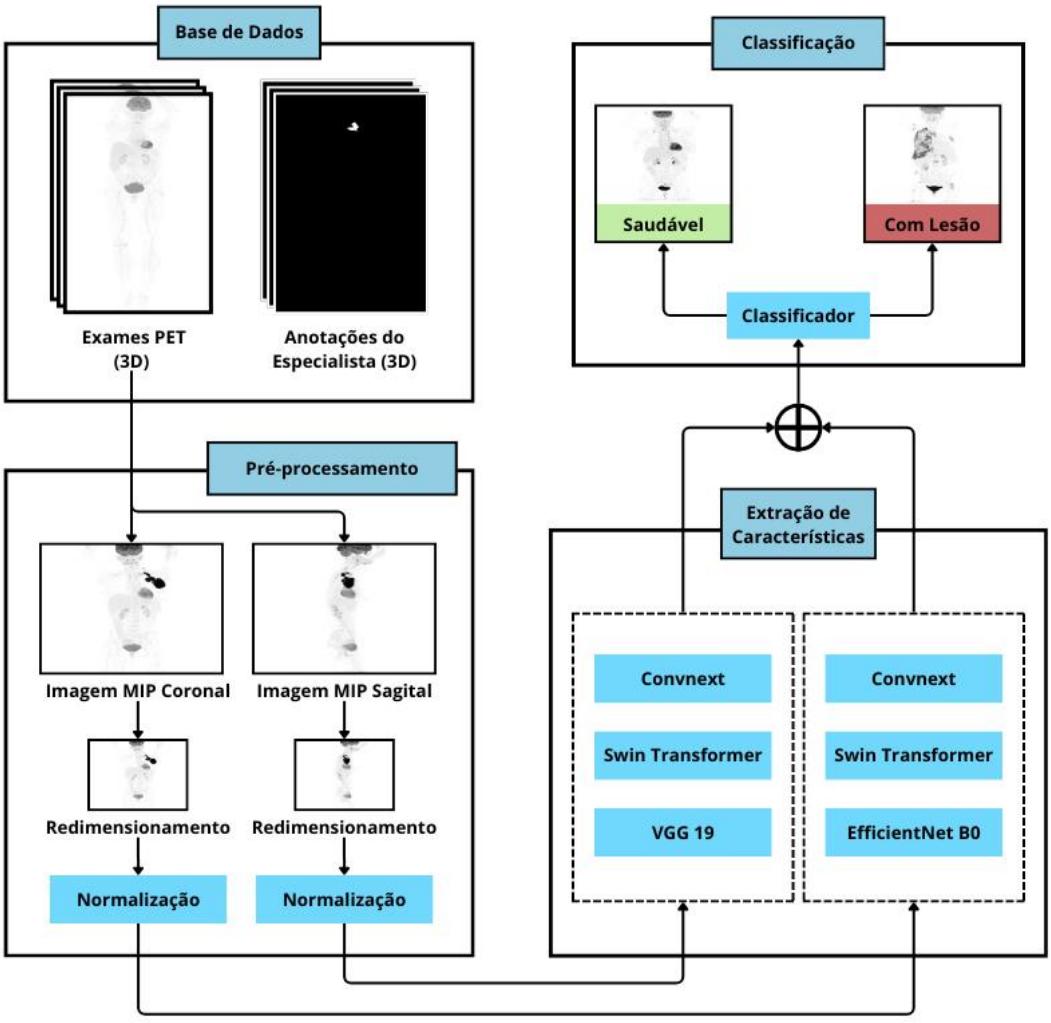
A Figura 3.10 detalha a estrutura hierárquica da rede. O processamento inicia no módulo *Patch Partition*, que divide a imagem de entrada em fragmentos de 4×4 pixels, seguido por um *Linear Embedding* que projeta esses fragmentos para uma dimensão arbitrada C. A arquitetura é organizada em quatro estágios. Para construir a hierarquia, camadas de *Patch Merging* são inseridas entre os estágios para reduzir a resolução espacial pela metade e dobrar o número de canais (de C para 2C, 4C e 8C) (LIU et al., 2021).

O bloco de construção fundamental, expandido à direita na Figura 3.10, opera sempre em pares consecutivos. O primeiro bloco utiliza o mecanismo de atenção padrão baseado em janelas (*W-MSA*), enquanto o segundo bloco aplica a atenção com janelas deslocadas (*SW-MSA*). Esse deslocamento cíclico dos dados antes do cálculo da atenção é o que permite o cruzamento de informações entre as fronteiras das janelas anteriores. Cada subcamada de atenção e de MLP é precedida por uma normalização (*Layer Norm - LN*) e seguida por uma conexão residual, garantindo a estabilidade do treinamento (LIU et al., 2021).

4 Materiais e Método

Este capítulo descreve detalhadamente os procedimentos e técnicas utilizados na condução deste trabalho, que tem como objetivo a classificação binária automática de exames PET em saudáveis ou com lesão. O método proposto é dividido em três etapas principais, sendo o pré-processamento, incluindo a geração de representações 2D por meio de projeções de intensidade máxima (MIP), redimensionamento e normalização das mesmas, a extração de características e a subsequente classificação. A Figura 4.1 apresenta o fluxo principal do método proposto. No decorrer deste capítulo, será descrita em detalhes cada etapa que compõe o método.

Figura 4.1 – Método Proposto



Fonte: Autoral.

4.1 Aquisição da Base

Neste trabalho, foi utilizado o conjunto de dados do desafio AutoPet III, que possui dois diferentes tipos de imagens. O primeiro usa dados de exames PET/TC com radiofármaco 18F FDG, enquanto o segundo usa traçador PSMA (GATIDIS et al., 2022; JEBLICK et al., 2024). No entanto, foi utilizado apenas o primeiro subconjunto, com o objetivo de classificar apenas os exames PET com 18F FDG, devido à sua maior prevalência na prática clínica.

Os dados utilizados possuem 1.014 exames diferentes de 900 pacientes com quatro diagnósticos diferentes, sendo câncer de pulmão, linfoma e melanoma e saudável. A Tabela 4.1 mostra a distribuição do conjunto de dados de acordo com o diagnóstico em relação ao sexo, número de exames e idade dos pacientes.

Tabela 4.1 – Distribuição dos Exames

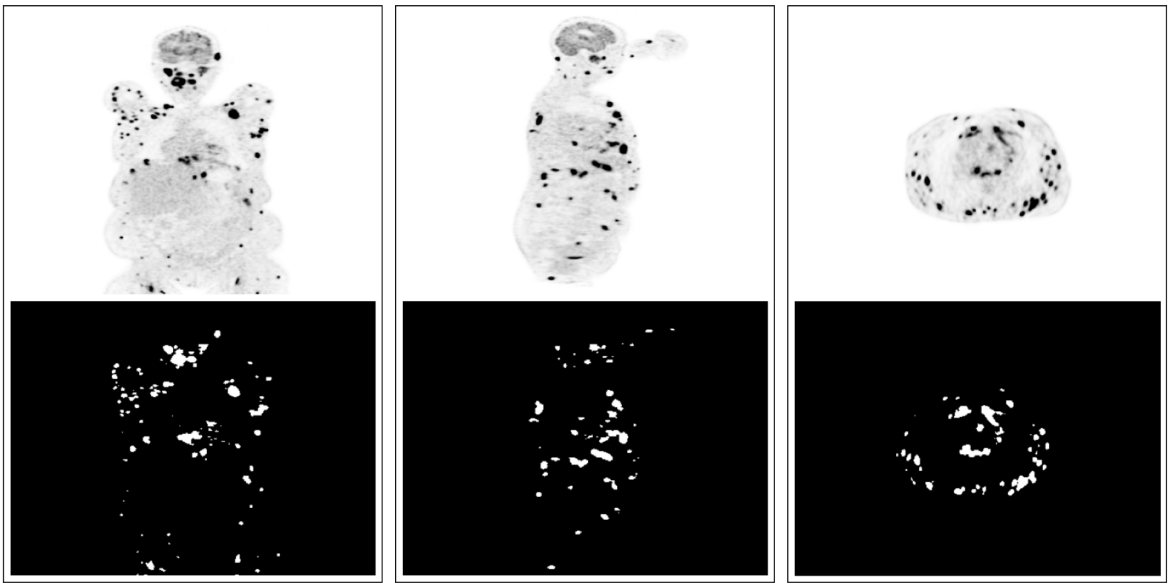
Diagnóstico	Gênero	Número de Estudos	Idade (Média ± DP)
Câncer de Pulmão	Feminino	65	64,2 ± 8,7
	Masculino	103	67,0 ± 9,0
Linfoma	Feminino	69	45,1 ± 19,7
	Masculino	76	47,3 ± 17,9
Melanoma	Feminino	77	65,0 ± 12,8
	Masculino	111	65,7 ± 13,7
Negativo	Feminino	233	59,1 ± 14,7
	Masculino	280	58,7 ± 15,1
Total	Feminino	444	58,5 ± 16,1
	Masculino	570	60,1 ± 15,9

Fonte: (GATIDIS et al., 2022).

A aquisição dos exames PET do conjunto de dados usado foi realizada após um período de jejum de pelo menos seis horas, seguido pela injeção intravenosa de aproximadamente 350 MBq de 18F FDG. A varredura de corpo inteiro iniciou-se cerca de 60 minutos após a administração do radiofármaco, utilizando um scanner Biograph mCT. Além disso, os dados foram anotados por radiologistas com experiência em imagens híbridas. O protocolo de anotação consistiu em duas etapas, na primeira as lesões tumorais foram identificadas por meio da análise visual das imagens em conjunto com os laudos clínicos, na segunda realizou-se a segmentação manual das lesões identificadas nos cortes axiais.

A Figura 4.2 ilustra um exemplo da fatia com maior quantidade de voxels com lesão de um paciente com melanoma nos cortes coronais, sagitais e axiais com suas respectivas anotações.

Figura 4.2 – Exemplo de Fatia Central de Exame Diagnosticado com Melanoma com suas respectivas anotações



Fonte: (GATIDIS et al., 2022).

4.2 Geração e Normalização das Imagens MIP

A utilização de representações 2D justifica-se pelo menor custo computacional no processamento em comparação aos volumes 3D. Em exames de corpo inteiro, as vistas coronal e sagital são as mais representativas da anatomia e distribuição metabólica, sendo, portanto, as escolhidas para os experimentos.

A Projeção de Intensidade Máxima (MIP) é uma técnica comum em PET para gerar essas representações. O processo consiste em selecionar um eixo de projeção, sendo axial, coronal ou sagital (x,y ou z) e, ao longo de cada linha de visão paralela a esse eixo, seleciona-se apenas o voxel de maior intensidade, que é então projetado no plano de saída para compor a imagem final (RANA et al., 2020).

Para o processamento por redes neurais, é necessário realizar a quantização dos valores de SUV (Standardized Uptake Value) para o formato de 8 bits (0 a 255) (GONZALEZ; WOODS, 2000). A abordagem mais frequente é a normalização Min-Max, que redimensiona os valores dentro dos limites mínimos e máximos de SUV encontrados. Entretanto, esta técnica pode se tornar sensível a ruídos e outliers, o que pode gerar inconsistências visuais no conjunto de dados.

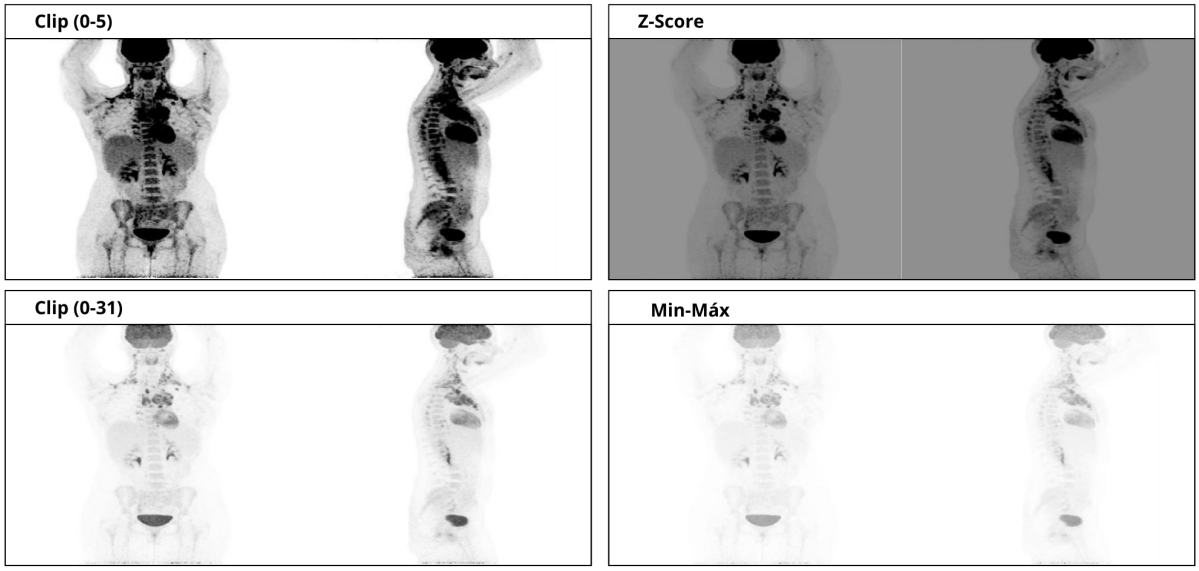
Como alternativa, a padronização via Z-score utiliza a média e o desvio padrão da distribuição de intensidades. Esta abordagem é mais robusta contra ruídos e valores discrepantes, pois a escala resultante baseia-se em propriedades estatísticas globais (média zero e variância unitária), promovendo maior consistência entre diferentes exames (LECUN

et al., 2002b).

Outra estratégia aplicada é o truncamento (clipping) dos níveis de SUV. O processo consiste em definir um limiar superior de SUV; todos os voxels com valores acima deste teto são fixados no limite estabelecido, e o intervalo resultante é remapeado para a escala de intensidades de 0 a 255. Essa técnica é menos sensível a ruídos, porém, é necessário um estudo para a escolha do limiar, visto que cada tipo de lesão possui suas características. Ren et al. (2021) utilizaram o valor máximo de SUV como cinco, em que o autor afirma que aumentar esse valor não melhora os valores de métricas para a tarefa de segmentação.

A Figura 4.3 mostra os diferentes resultados de normalização ao gerar imagens MIP nos eixos coronal (esquerda) e sagital (sagital) do mesmo exame em todos os exemplos que foram avaliados no método.

Figura 4.3 – Imagens MIP Geradas de Diferentes Formas

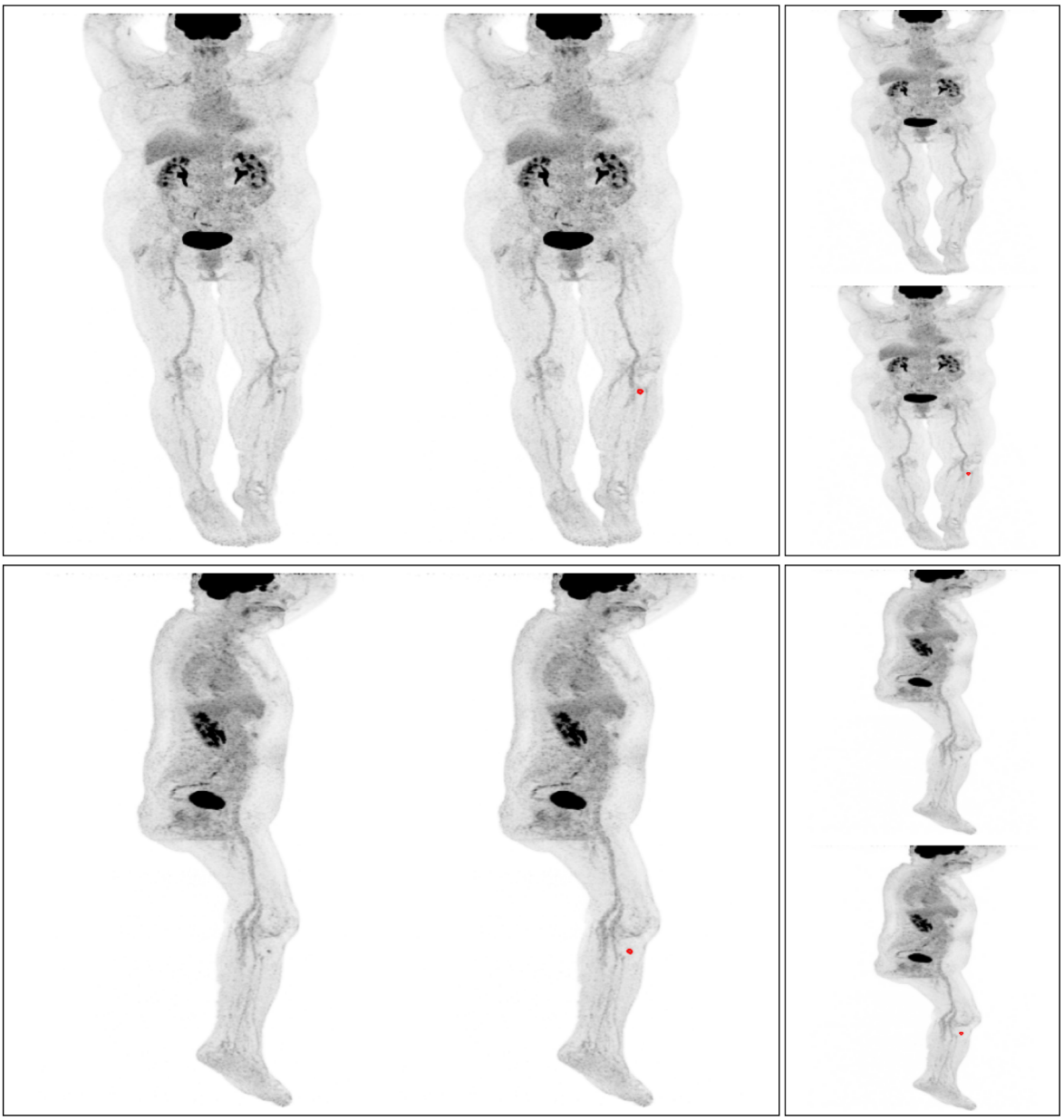


Fonte: Autoral.

Para viabilizar o uso de modelos pré-treinados, as imagens foram redimensionadas para a resolução de 224 × 224 pixels. Originalmente, as vistas coronais e sagitais possuem largura fixa de 400 pixels, enquanto a altura varia entre 200 e 661 pixels. Consequentemente, o redimensionamento por interpolação alterou a proporção original, introduzindo deformações morfológicas. Contudo, verificou-se que a integridade das informações patológicas foi preservada, sem a supressão de lesões (GONZALEZ; WOODS, 2000). A Figura 4.4 exemplifica um exame de melanoma nos cortes coronais e sagitais, comparando a lesão segmentada antes e após o redimensionamento.

Posteriormente, procedeu-se à normalização dos dados. Dado que as arquiteturas foram originalmente treinadas em bases de dados de imagens coloridas (RGB), adaptou-se o canal único de intensidade do PET replicando-o em três canais idênticos. Em seguida,

Figura 4.4 – Antes e Depois do Redimensionamento para 224 x 224



Fonte: Autoral.

os valores de intensidade foram padronizados com base na média e no desvio padrão do conjunto de dados *ImageNet*. Esta etapa é fundamental para assegurar a congruência estatística entre os dados de entrada e os parâmetros aprendidos pelas redes, otimizando a convergência e o desempenho do modelo.

4.3 Extração de Características

Para a etapa de extração de características, foi conduzida uma análise comparativa abrangente utilizando diversas arquiteturas do estado da arte. Adotou-se uma abordagem

de múltiplas vistas (*multiview*), na qual as imagens correspondentes aos cortes coronais e sagitais de um mesmo exame são processadas simultaneamente. Após a propagação pelas camadas da rede neural, os mapas de características resultantes de cada corte são convertidos em vetores unidimensionais e, em seguida, concatenados. De modo que a etapa de classificação receba um vetor de características conjunto, composto pela fusão das informações espaciais de ambos os planos.

A seleção das arquiteturas visou testar as redes clássicas, baseadas em convoluções (CNNs) e baseadas em *Transformers* com a intenção de comparar a eficiência das CNNs na extração de características locais com a capacidade dos Transformers de compreender o contexto global da imagem. No grupo das CNNs, foram avaliados os modelos ConvNeXt (LIU et al., 2022), EfficientNet (variantes B0, B4 e V2 Small) (TAN; LE, 2019; TAN; LE, 2021), ResNet (18 e 50) (HE et al., 2016) e VGG (16 e 19) (BANGAR, 2022). Já representando as arquiteturas baseadas em atenção, foram utilizados o Vision Transformer Base (ViT) (DOSOVITSKIY, 2020) e o Swin Transformer Base (LIU et al., 2021).

A inclusão de modelos baseados em CNNs justifica-se pela sua eficácia comprovada na detecção de padrões locais e na invariância à translação. Em exames PET, onde as lesões tumorais se manifestam frequentemente como regiões focais de alta captação de radiofármaco (SUV elevado), as convoluções são particularmente eficientes em extrair características de baixo nível, como bordas, texturas e gradientes de intensidade, essenciais para delinear a morfologia da lesão em relação aos tecidos adjacentes (RAGHU et al., 2021).

Por outro lado, a utilização de Transformers visa explorar o mecanismo de autoatenção (*self-attention*), que permite modelar dependências de longo alcance na imagem. Diferentemente das CNNs, que possuem um campo receptivo local limitado, modelos como o ViT e o Swin Transformer conseguem capturar o contexto global do exame. Essa capacidade é benéfica para a análise de imagens PET, pois auxilia na distinção entre ruídos de fundo e captações fisiológicas e patológicas, correlacionando a intensidade da lesão com a estrutura anatômica global do paciente (SHAMSHAD et al., 2023).

4.4 Classificação

Nesta fase final do método, as representações vetoriais extraídas pelas arquiteturas são unidas por concatenação e são usadas de entrada para a tarefa de classificação binária. Os classificadores devem ser capazes de mapear o vetor de características para um rótulo de classe, entre saudável e com lesão. Para garantir uma avaliação robusta, optou-se pela comparação entre três paradigmas de classificação amplamente utilizados na literatura.

O primeiro é a Máquina de Vetores de Suporte (SVM), escolhida por sua eficácia em espaços de alta dimensão ao buscar um hiperplano ótimo de separação. O segundo é o

XGBoost (*Extreme Gradient Boosting*), um algoritmo baseado em árvores de decisão que se destaca pelo alto desempenho em dados estruturados. Esses dois classificadores também podem exigir menos dados do que treinar uma rede neural. Por fim, utilizou-se uma Rede Neural Artificial do tipo MLP (*Multilayer Perceptron*) para capturar não-linearidades complexas.

O treinamento dos classificadores foi realizado utilizando o conjunto de treinamento, enquanto o conjunto de validação foi empregado para monitorar o desempenho durante o ajuste dos modelos. A análise dos resultados priorizou as métricas de F1-Score e Recall, fundamentais para mensurar a capacidade do método de identificar corretamente as lesões e minimizar a ocorrência de falsos negativos, garantindo maior segurança diagnóstica.

4.5 Métricas de Avaliação

O desempenho do método na classificação das imagens PET é avaliado por meio de métricas quantitativas baseadas nos resultados da matriz de confusão. Essa classificação é então comparada com o diagnóstico real. Existem quatro resultados possíveis para cada classificação ao ser confrontada com o diagnóstico do especialista (THARWAT, 2021):

- Verdadeiro Positivo (VP): Ocorre quando o modelo classifica corretamente a presença de uma lesão. Ou seja, a predição indica a classe positiva (com doença) e o diagnóstico real confirma a existência da condição;
- Verdadeiro Negativo (VN): Ocorre quando o modelo classifica corretamente o exame como saudável. Neste caso, a predição indica a classe negativa (sem lesão) e o diagnóstico real confirma que o paciente não possui a lesão;
- Falso Positivo (FP): Acontece quando o modelo prediz incorretamente a presença de uma lesão em um paciente saudável; e
- Falso Negativo (FN): Acontece quando o modelo prediz incorretamente que o exame é saudável, falhando em detectar uma lesão existente.

Com base nos quatro resultados fundamentais descritos anteriormente (VP, VN, FP, FN), diversas métricas quantitativas foram calculadas para avaliar o desempenho do método. Neste trabalho, adotou-se o uso da Acurácia, Precisão, Sensibilidade (*Recall*), F1-Score e da Curva ROC.

A acurácia representa a proporção global de acertos do modelo em relação ao número total de amostras classificadas. Essa métrica pode ser descrita pela equação 4.1 (THARWAT, 2021):

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \tag{4.1}$$

A precisão avalia a confiabilidade das predições positivas. Ela indica a porcentagem de amostras classificadas como contendo lesão que realmente correspondem a casos patológicos. Essa métrica pode ser descrita pela equação 4.2 (THARWAT, 2021):

$$Precisão = \frac{VP}{VP + FP} \tag{4.2}$$

A sensibilidade (*Recall*) mede a capacidade do modelo de detectar corretamente os casos positivos (Equação 4.3). Em contextos médicos, o *Recall* é uma métrica crítica, pois falhas na detecção (falsos negativos) podem comprometer o início do tratamento do paciente.

$$Sensibilidade = \frac{VP}{VP + FN} \tag{4.3}$$

O *F1-Score* consiste na média harmônica entre a precisão e a sensibilidade. Essa métrica é particularmente útil para buscar um equilíbrio entre a qualidade da predição e a capacidade de detecção, especialmente em conjuntos de dados desbalanceados. Essa métrica pode ser descrita pela equação 4.4 (THARWAT, 2021):

$$F1\text{-}Score = 2 \cdot \frac{Precisão \cdot Recall}{Precisão + Recall} \tag{4.4}$$

Por fim, a Curva ROC (*Receiver Operating Characteristic*) é uma ferramenta gráfica utilizada para avaliar a capacidade de discriminação do classificador sob diferentes limiares de decisão. A curva é gerada plotando-se a Taxa de Verdadeiros Positivos (TVP, equivalente à Sensibilidade) no eixo vertical contra a Taxa de Falsos Positivos (FPR) no eixo horizontal. Essa métrica pode ser descrita pela equação 4.5 (THARWAT, 2021):

$$FPR = \frac{FP}{FP + VN} \tag{4.5}$$

5 Resultados e Discussão

Neste capítulo, serão apresentados os resultados obtidos em cada um dos experimentos realizados neste trabalho. Todos os experimentos foram conduzidos utilizando a linguagem de programação Python (ROSSUM; JR, 1995), versão 3.8. A plataforma CUDA (NVIDIA; VINGELMANN; FITZEK, 2020) foi adotada para dar suporte à execução do treinamento em uma GPU NVIDIA GeForce GTX 1080 Ti equipada com aproximadamente 11 GB de VRAM.

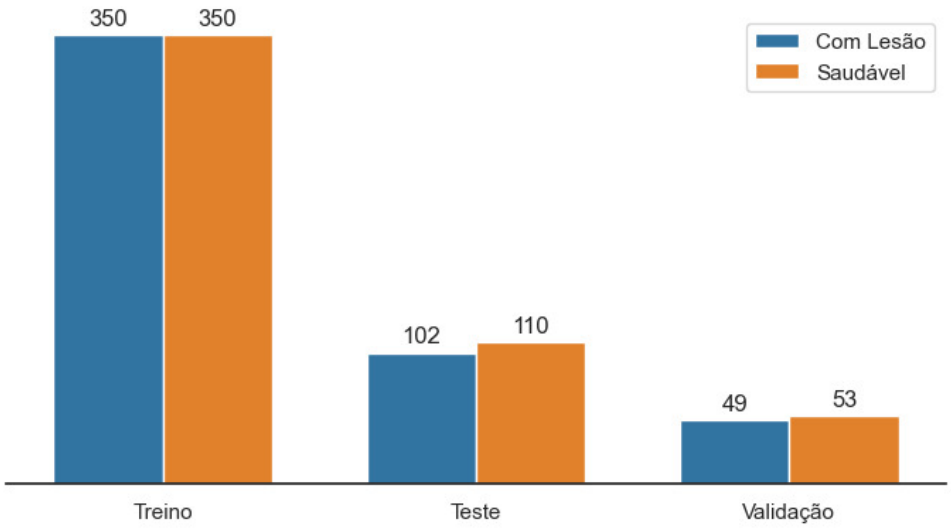
5.1 Divisão da Base de Imagens

A divisão da base de imagens utilizada é uma etapa muito importante para a realização dos experimentos a seguir. Por existir diferentes tipos de diagnósticos, e cada um com suas características individuais, é necessário ter uma boa distribuição para que cada subconjunto seja representativo. Dessa forma, foi realizada uma divisão com 80% das imagens para treino, 20% para teste e 10% do treino para a validação. Além disso, foi levado em consideração o paciente, mantendo exames diferentes do mesmo paciente no mesmo subconjunto. A divisão dos exames foi realizada por meio de amostragem aleatória estratificada, usando como extratos o tamanho das lesões e diagnóstico, de modo a garantir que estivessem presentes de forma balanceada nos subconjuntos. A Figura 5.1 mostra a distribuição de exames da classe positiva, sendo os pacientes que possuem algum tipo de lesão, e da classe negativa, contendo os exames saudáveis. A Figura 5.2 mostra a distribuição dos subconjuntos em relação aos tipos diferentes de diagnóstico.

Visto que no conjunto de dados não há informações sobre o tamanho das lesões, essa informação foi obtida a partir das anotações dos especialistas. Para cada exame foi realizada uma contagem da quantidade de voxels ativos que pertencem a alguma lesão. Em seguida, todos os valores de quantidade de voxels das regiões de lesão são ordenados em ordem crescente e divididos em três grupos, com aproximadamente 33% dos dados para cada, sendo chamados de pequeno, médio e grande. A Figura 5.3 mostra a distribuição em relação ao diagnóstico e ao tamanho da lesão, respectivamente.

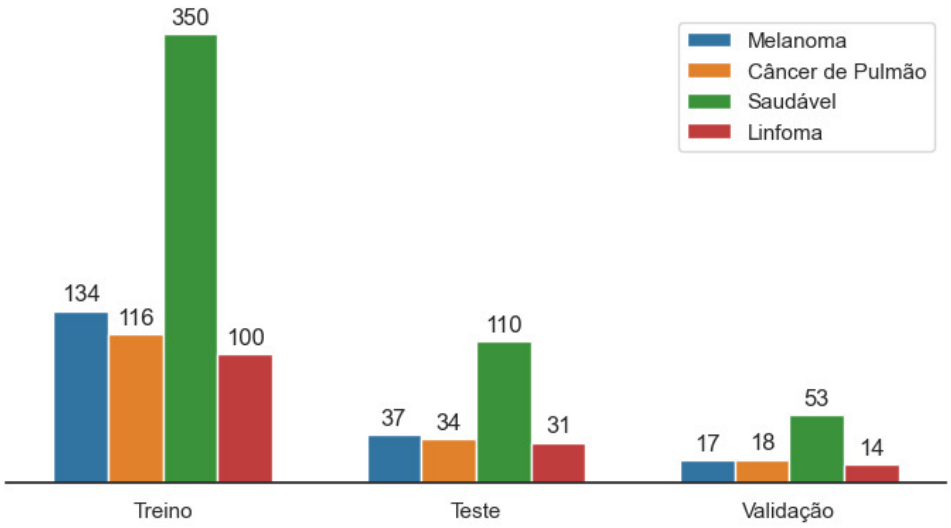
Por fim, a Tabela 5.1 mostra o resultado da divisão dos 1.014 exames dos 900 pacientes em treino com 700 exames, validação com 102 e teste com 212.

Figura 5.1 – Distribuição de exames entre os subconjuntos por classe



Fonte: Autoral.

Figura 5.2 – Distribuição dos tipos de diagnóstico por subconjunto



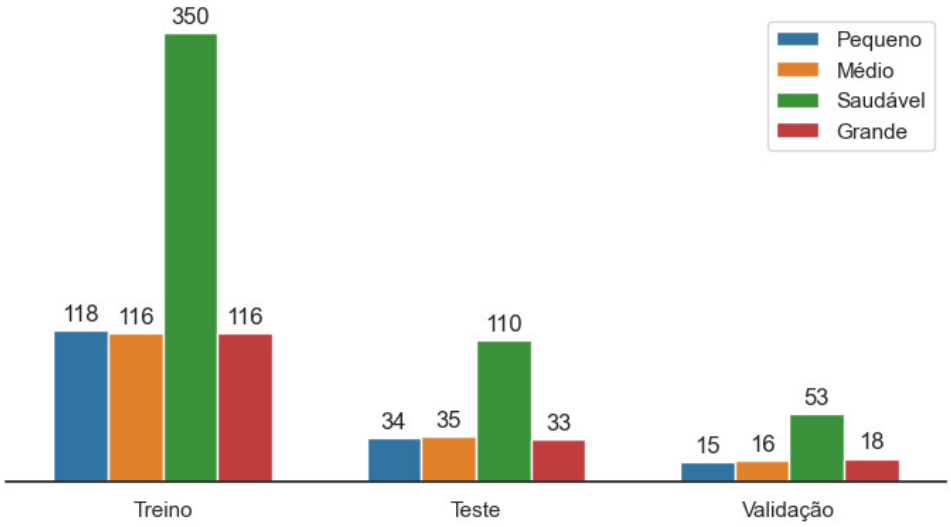
Fonte: Autoral.

5.2 Avaliação da Normalização das Imagens MIP

Após a divisão do conjunto de dados em três subconjuntos, foi realizada a geração de representações 2D dos exames. Para isso, foi utilizado um método comum ao lidar com imagens de exames PET, sendo as representações MIP. Como descrito na seção 4.2, foram geradas representações a partir da análise dos níveis de SUV dos exames presentes no subconjunto de treino e de outras maneiras propostas na literatura.

Os métodos mais comuns são Min-Max e Z-score, amplamente utilizados na literatura. Porém, em (REN et al., 2021) foi proposto que definir um limiar entre 0 e 5 para redistribuir os valores de voxels em intensidades de pixels não afetaria o resultado.

Figura 5.3 – Distribuição dos tamanhos das lesões por subconjunto



Fonte: Autoral.

Tabela 5.1 – Distribuição Geral do Conjunto de Dados.

Grupo	**Característica**	**Teste**	**Treino**	**Validação**
Diagnóstico	Câncer de Pulmão	34	116	18
	Linfoma	31	100	14
	Melanoma	37	134	17
	Saudável	110	350	53
Tamanho da Lesão	Grande	33	116	18
	Médio	35	116	16
	Pequeno	34	118	15
	Saudável	110	350	53
Condição	Com Lesão	102	350	49
	Saudável	110	350	53

Fonte: Autoral.

Em contrapartida, no presente trabalho foi realizado um estudo nos exames com base na intensidade dos níveis de SUV das áreas de lesão. A importância da maneira com que as imagens são geradas influencia diretamente o contraste delas. Em seguida, uma série de arquiteturas foi utilizada para realizar a tarefa de classificação binária, e a maneira que mais vezes se destacou foi escolhida para os futuros experimentos.

Para a análise dos níveis de SUV, foi utilizada apenas a região de lesões dos exames do subconjunto de treino, fazendo uso das anotações dos especialistas. Realizaram-se contagens dos valores mínimos e máximos dos níveis de SUV nessas regiões. A Tabela 5.2 apresenta os valores dos níveis de SUV das lesões nos exames presentes no treino.

Diante dessa análise, observou-se que o maior valor de SUV encontrado foi 30,03

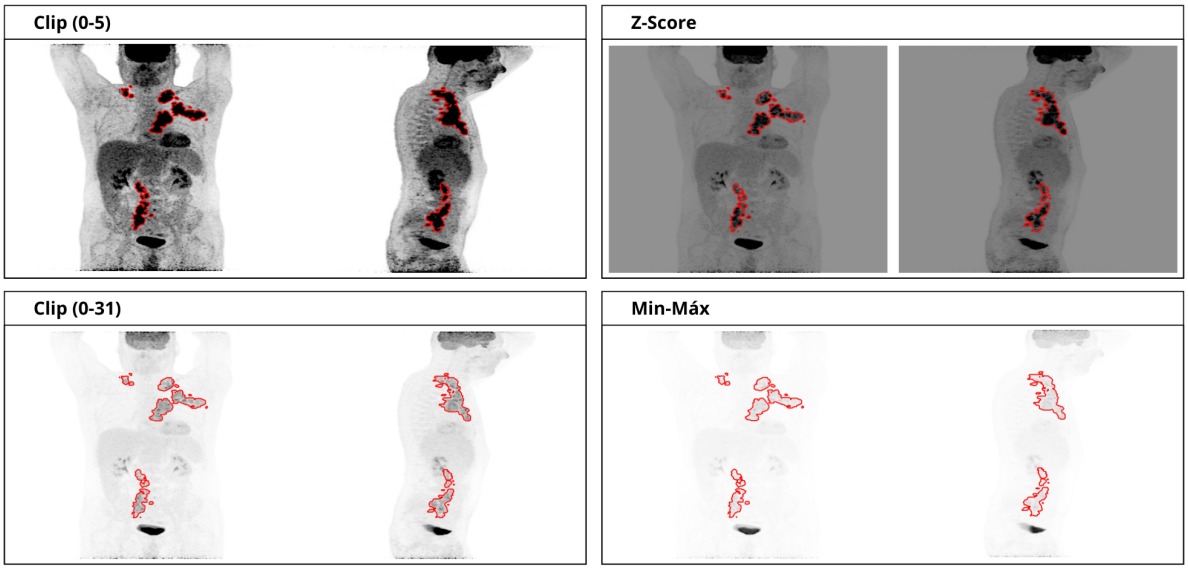
Tabela 5.2 – Estatísticas descritivas dos valores de SUV por diagnóstico no conjunto de treino.

Diagnóstico	**SUV (média ± DP)**	**Mínimo**	**Máximo**	**Nº de Casos**
Câncer de Pulmão	4,28 ± 1,37	1,17	11,91	116
Linfoma	6,23 ± 2,38	2,31	15,46	100
Melanoma	6,13 ± 3,07	1,44	30,03	134

para um caso de melanoma, sendo o valor máximo escolhido para o *clipping* na geração de representações MIP o intervalo entre 0 e 31.

Para escolher qual tipo de imagem seria utilizado para os experimentos futuros, optou-se por realizar a classificação binária utilizando diferentes arquiteturas e representações min-máx, *Z-score*, *clipping* (0 - 5) e *clipping* (0 - 31) nos eixos coronal e sagital para cada exame. A Figura 5.4 mostra a comparação entre os quatro diferentes métodos e a região da lesão segmentada em vermelho em um exame com linfoma.

Figura 5.4 – Imagens MIP Geradas de Diferentes Formas



Fonte: Autoral.

Os resultados dessa comparação, apresentados nas Tabelas 5.3 e 5.4 para os cortes coronais e sagitais, respectivamente. Na parte inferior de ambas as tabelas há a informação de quantas vezes cada tipo de representação foi melhor. Apesar de que o *clipping* entre 0 e 5 tenha se sobressaido no corte coronal, no sagital ele não foi melhor em nenhum momento, sendo superado pelo *clipping* entre 0 e 31, que possuiu as melhores métricas em nove casos, enquanto o anterior foi apenas cinco. Portanto, as representações geradas a partir do clipping entre 0 e 31 foram escolhidas para a realização de todos os seguintes experimentos.

Tabela 5.3 – Resultados de F1-Score e Sensibilidade para cada Arquitetura

Arquitetura	*Clipping* (0-5)	*Clipping* (0-31)	Min_Máx	*Z-Score*
ConvNext	79,59%/79,59%	**86,32%/83,67%**	72,73%/73,47%	71,43%/71,43%
EfficientNet B0	**77,55%/77,55%**	77,08%/75,51%	75,51%/75,51%	68,13%/63,27%
EfficientNet B4	**72,55%/72,51%**	70,97%/67,35%	70,21%/67,35%	72,34%/69,39%
EfficientNet V2 S	78,65%/71,43%	**81,72%/77,55%**	73,56%/65,31%	80,46%/71,43%
ResNet 18	**79,21%/81,63%**	79,17%/77,55%	77,55%/77,55%	76,40%/69,39%
ResNet 50	71,26%/63,27%	**77,89%/75,51%**	67,50%/55,10%	77,42%/73,47%
Swin B	77,55%/77,55%	78,10%/83,67%	**81,19%/83,67%**	81,25%/79,59%
VGG 16	80,41%/79,59%	83,52%/77,55%	78,79%/79,59%	**85,42%/83,67%**
VGG 19	**82,00%/83,67%**	80,81%/81,63%	79,61%/83,67%	77,89%/75,51%
ViT	**76,60%/73,47%**	65,26%/63,27%	64,65%/65,31%	66,67%/61,22%
Métricas	F1/Sensibilidade	F1/Sensibilidade	F1/Sensibilidade	F1/Sensibilidade
Qtd. melhor	5	3	1	1

Fonte: Autoral.

Tabela 5.4 – Resultados de F1-Score e Sensibilidade para cada Arquitetura

Arquitetura	*Clipping* (0-5)	*Clipping* (0-31)	Min_Máx	*Z-Score*
ConvNext	81,13%/87,76%	**82,69%/87,76%**	71,74%/67,35%	74,51%/77,55%
EfficientNet B0	68,82%/65,31%	69,39%/69,39%	**76,29%/75,51%**	69,39%/69,39%
EfficientNet B4	66,67%/65,31%	**71,11%/65,31%**	67,35%/67,35%	65,26%/63,27%
EfficientNet V2 S	69,72%/77,55%	**78,16%/69,39%**	74,75%/75,51%	74,42%/65,31%
ResNet 18	64,20%/53,06%	**80,43%/75,51%**	78,72%/75,51%	76,92%/81,63%
ResNet 50	66,67%/63,27%	80,37%/87,76%	74,23%/73,47%	**82,11%/79,59%**
Swin B	80,00%/81,63%	81,82%/91,84%	79,28%/89,80%	**82,88%/93,88%**
VGG 16	76,77%/77,55%	**81,19%/83,67%**	76,77%/77,55%	76,47%/79,59%
VGG 19	76,92%/81,63%	**84,21%/81,63%**	75,51%/75,51%	79,61%/83,67%
ViT	60,47%/53,06%	73,79%/77,55%	66,67%/65,31%	**75,25%/77,55%**
Métricas	F1/Sensibilidade	F1/Sensibilidade	F1/Sensibilidade	F1/Sensibilidade
Qtd. melhor	0	6	1	3

Fonte: Autoral.

Ao analisar visualmente as representações geradas (Figura 5.4), observa-se que as diferentes técnicas de processamento impactam diretamente a percepção de contraste e a preservação de detalhes das lesões. O método de *clipping* (0-5) resulta em imagens com contraste acentuado, destacando focos de alta captação de SUV, o que justifica seu bom desempenho no corte coronal para diversas arquiteturas. No entanto, essa saturação pode ocultar nuances volumétricas essenciais no corte sagital. Em contrapartida, o *clipping* (0-31) oferece um equilíbrio superior entre brilho e contraste, preservando a morfologia das lesões sem perder a intensidade do sinal. Essa maior riqueza de informações anatômicas permitiu que modelos como ConvNext e VGG 19 atingissem métricas de sensibilidade mais robustas no plano sagital. Por outro lado, as técnicas de *Z-Score* e Min-Máx geraram imagens com fundos acinzentados ou com baixo contraste, sendo excessivamente claras, dificultando a distinção entre a lesão e o tecido saudável, o que se refletiu na baixa frequência dessas

representações como o melhor caso nas tabelas comparativas. Portanto, a escolha do *clipping* (0-31) baseia-se na sua capacidade de fornecer características visuais consistentes que potencializam a extração de características pelas redes neurais em ambos os eixos de análise.

5.3 Extração de Características

Definida a forma de geração das imagens MIP, selecionaram-se as arquiteturas que deverão compor o método para cada plano de corte. Foram testadas redes baseadas em CNNs e *Transformers* para os eixos coronal e sagital, de forma independente, visando identificar o melhor desempenho para cada visão. A escolha fundamentou-se nas métricas de *F1-Score* e sensibilidade. Todos os experimentos utilizaram hiperparâmetros padronizados: taxa de aprendizado de 5 $\times 10^{-5}$, *batch size* de 32, otimizador AdamW e função de perda *Focal Loss*, com treinamento de 300 épocas e paciência de 30 para o *early stopping*.

A Tabela 5.5 detalha as arquiteturas testadas e os resultados obtidos para a imagem MIP do plano coronal. Nesse corte, as arquiteturas ConvNeXt e EfficientNet B0 foram selecionadas por apresentarem o melhor desempenho combinado, com 86,32% de F1-Score e 83,67% de sensibilidade. A Swin Base também foi incluída devido à sua sensibilidade de 83,67%, garantindo robustez na detecção. A priorização de modelos com alta sensibilidade é crucial em contextos médicos, pois assegura a redução de falsos negativos, garantindo que o maior número possível de lesões seja identificado para um diagnóstico seguro.

Tabela 5.5 – Resultados para Clipping entre 0 e 31 para imagens Coronais

Arquiteturas	**Acurácia**	**F1-Score**	**Sensibilidade**	**Precisão**
Convnext	**87,25%**	**86,32%**	**83,67%**	89,13%
EfficientNet B0	**87,25%**	**86,32%**	**83,67%**	89,13%
EfficientNet B4	73,53%	70,97%	67,35%	75,00%
EfficientNet v2 Small	83,33%	81,72%	77,55%	86,36%
ResNet 18	80,39%	79,17%	77,55%	80,85%
ResNet 50	79,41%	77,89%	75,51%	80,43%
Swin Base	77,45%	78,10%	**83,67%**	73,21%
VGG 16	85,29%	83,52%	77,55%	**90,48%**
VGG 19	81,37%	80,81%	81,63%	80,00%
ViT	67,65%	65,26%	63,27%	67,39%

Fonte: Autoral.

Já a Tabela 5.6 apresenta os dados para o eixo sagital. Nela, a VGG 19 obteve o maior F1-Score (84,21%), enquanto a Swin Base destacou-se com a maior sensibilidade absoluta dos experimentos, atingindo 91,84%. A ConvNeXt complementou a seleção com

métricas de 82,69% de F1-Score e 87,76% de sensibilidade. Em aplicações clínicas, a manutenção de uma sensibilidade elevada é o fator determinante, visto que a detecção precoce e precisa de todos os focos da doença é vital para o sucesso do tratamento e a sobrevida do paciente.

Tabela 5.6 – Resultados para Clipping entre 0 e 31 para imagens Sagitais

Arquiteturas	**Acurácia**	**F1-Score**	**Sensibilidade**	**Precisão**
Convnext	82,35%	82,69%	87,76%	78,18%
EfficientNet B0	70,59%	69,39%	69,39%	69,39%
EfficientNet B4	74,51%	71,11%	65,31%	78,05%
EfficientNet v2 Small	81,37%	78,16%	69,39%	**89,47%**
ResNet 18	82,35%	80,43%	75,51%	86,05%
ResNet 50	79,41%	80,37%	87,76%	74,14%
Swin Base	80,39%	81,82%	**91,84%**	73,77%
VGG 16	81,37%	81,19%	83,67%	78,85%
VGG 19	**85,29%**	**84,21%**	81,63%	86,96%
ViT	73,53%	73,79%	77,55%	70,37%

Fonte: Autoral.

Após a definição das arquiteturas, os modelos selecionados para cada visão, incluindo ConvNeXt, EfficientNet B0 e Swin para o eixo coronal, além de ConvNeXt, Swin e VGG 19 para o plano sagital, foram submetidos a uma etapa de ajuste fino. Para realizar essa otimização, utilizou-se o *framework* Optuna (AKIBA et al., 2019), que opera com base no algoritmo *Tree-structured Parzen Estimator* (TPE). O TPE é um método de Otimização Bayesiana que analisa o histórico de execuções para criar modelos probabilísticos das regiões promissoras do espaço de busca. O algoritmo funciona dividindo as observações passadas em dois grupos, sendo um composto pelos melhores hiperparâmetros e outro pelos demais, buscando maximizar a razão entre essas distribuições para selecionar os próximos candidatos com maior potencial de sucesso.

A Tabela 5.7 apresenta os hiperparâmetros testados nas arquiteturas selecionadas. O objetivo da otimização foi a maximização da métrica de *F1-Score*. Através da exploração sistemática de variáveis contínuas e categóricas, buscou-se refinar o aprendizado dos modelos nas representações MIP previamente escolhidas, permitindo que o algoritmo TPE identificasse as configurações de hiperparâmetros mais eficientes para a detecção das lesões.

Além disso, as arquiteturas foram treinadas utilizando a técnica de *augmentation online*, com o objetivo de expandir a variabilidade do conjunto de dados de treinamento. O *augmentation* consiste na aplicação de transformações geométricas e de intensidade em imagens reais para gerar amostras sintéticas, sendo uma estratégia fundamental para melhorar a generalização de modelos em contextos de dados médicos limitados (SHORTEN;

Tabela 5.7 – Espaço de busca definido para a otimização com Optuna.

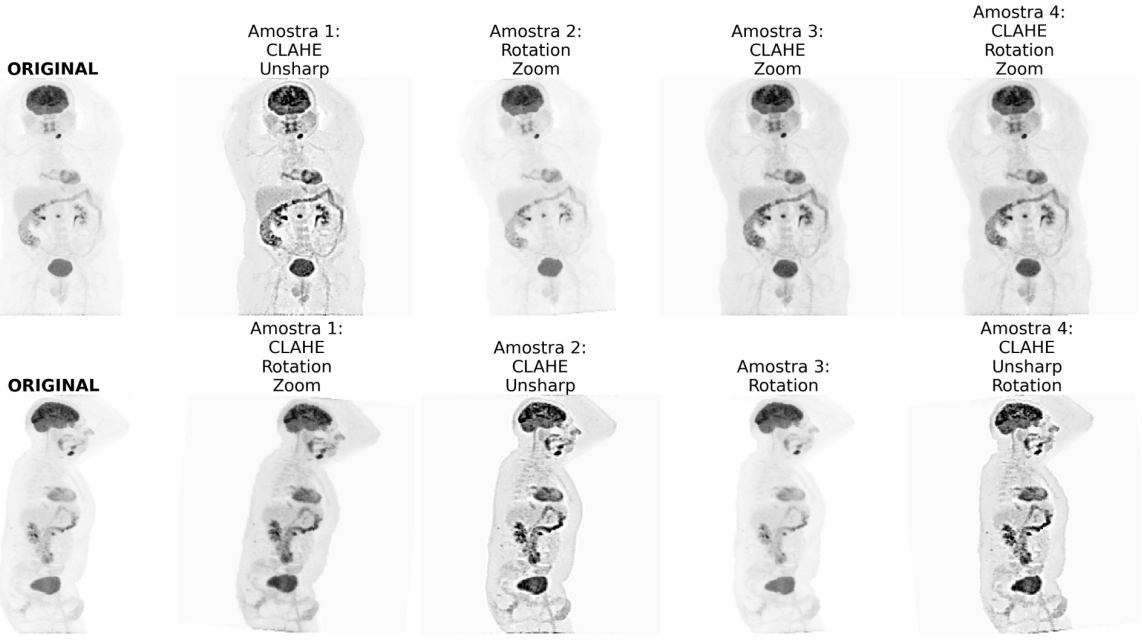
Hiperparâmetro	**Tipo**	**Espaço de Busca**
Taxa de Aprendizado (lr)	Contínuo (Log)	$[1 \times 10^{-5}, 5 \times 10^{-5}]$
Otimizador	Categórico	Adam AdamW SGD
Função de Perda	Categórico	Focal Loss Binary Cross-Entropy (BCE)

Fonte: Autoral.

KHOSHGOFTAAR, 2019).

As variações implementadas compreenderam o ajuste de contraste por meio de CLAHE (*Contrast Limited Adaptive Histogram Equalization*) e Unsharp Mask, além de rotações aleatórias de até 5º (em ambos os sentidos) e zoom de até 2%. Cada transformação possuía uma probabilidade de 50% de aplicação por amostra, assegurando uma pequena diversidade visual tanto na fase de treinamento quanto na etapa de otimização de hiperparâmetros. A Figura 5.5 apresenta exemplos comparativos de imagens originais e processadas com a aplicação do *augmentation*.

Figura 5.5 – Exemplos de Imagens com augmentation



Fonte: Autoral.

Por fim, a Tabela 5.8 mostra os resultados das arquiteturas no conjunto de validação após a otimização com e sem uso de augmentation. Nota-se que o modelo VGG19 exigiu consistentemente o menor *batch size* devido à sua densa estrutura de parâmetros e ao alto

consumo de memória. Adicionalmente, a alternância entre as funções de perda BCE e Focal Loss sugere que a inserção de augmentation alterou a percepção de dificuldade dos exemplos durante o treino, levando o otimizador a adaptar a penalização dos erros para maximizar o F1-Score em cada cenário.

Tabela 5.8 – Hiperparâmetros otimizados.

Cenário	**Plano**	**Modelo**	**LR**	**Otim.**	**Loss**	**Batch**	**F1**
Sem Aug.	Coronal	ConvNeXt	4,68e-05	Adam	Focal	16	92,63%
		Eff. B0	2,35e-05	AdamW	BCE	8	80,00%
		Swin Base	4,11e-05	AdamW	BCE	16	91,09%
	Sagital	ConvNeXt	2,67e-05	Adam	Focal	8	90,00%
		Swin Base	2,98e-05	AdamW	BCE	32	91,09%
		VGG19	3,71e-05	Adam	Focal	4	90,32%
Com Aug.	Coronal	ConvNeXt	1,94e-05	Adam	BCE	8	91,84%
		Eff. B0	2,94e-05	AdamW	BCE	16	78,43%
		Swin Base	2,81e-05	AdamW	BCE	16	90,20%
	Sagital	ConvNeXt	2,54e-05	AdamW	BCE	8	89,58%
		Swin Base	3,04e-05	Adam	Focal	8	90,91%
		VGG19	2,59e-05	Adam	BCE	4	91,30%

Fonte: Autoral.

5.4 Classificação

Após o treinamento das arquiteturas selecionadas, as mesmas são utilizadas apenas para a extração de características dos pares de imagens. Após isso, todas as características dos seis modelos são concatenadas (ConvNeXt, EfficientNet-B0, Swin Base e VGG19) e utilizadas como entrada em três classificadores diferentes: MLP, SVM e XGBoost. Estes, por sua vez, foram treinados utilizando os conjuntos de treino e validação.

O classificador *Multi-Layer Perceptron* (MLP) foi configurado com duas camadas ocultas contendo 128 e 64 neurônios, respectivamente. Utilizou-se a função de ativação *ReLU* para introduzir não linearidade e o otimizador *Adam* para o ajuste dos pesos. O treinamento foi limitado a um máximo de 500 iterações, garantindo tempo suficiente para a convergência da rede neural.

Para o *Support Vector Machine* (SVM), adotou-se o *kernel* RBF (*Radial Basis Function*), que permite ao modelo mapear os dados em dimensões superiores para encontrar fronteiras de decisão não lineares. O parâmetro *probability* foi ativado, permitindo que o classificador gere estimativas probabilísticas para cada predição, o que é essencial para o cálculo de métricas como a área sob a curva ROC (AUC).

O modelo *XGBoost* foi treinado utilizando o algoritmo de *gradient boosting* com a função de custo *binary:logistic*. Foram executadas 100 rodadas de reforço (*boosting rounds*) com uma taxa de aprendizado (*eta*) de 0,1. Para evitar o *overfitting*, a profundidade máxima das árvores foi limitada a seis, e utilizaram-se apenas 80% das amostras e colunas em cada iteração de treinamento.

Com as seis arquiteturas otimizadas com o algoritmo TPE, foi realizado um experimento testando todas as combinações possíveis entre as arquiteturas coronal e sagital. O experimento seguiu de forma que sempre houvesse ao menos uma arquitetura para um corte ou para as três, totalizando 49 combinações possíveis.

A Tabela 5.9 mostra o melhor desempenho no conjunto de testes e a melhor combinação entre arquiteturas e os classificadores que foram treinados utilizando os subconjuntos de treino e validação. Nota-se que, ainda que próximas, dessa vez a utilização do *augmentation* melhorou o desempenho dos classificadores. Destaca-se principalmente o cenário com o conjunto de dados expandido, onde a MLP alcançou os índices mais elevados de F1-Score (91,62%) e Sensibilidade (91,17%), apenas deixando de utilizar a arquitetura ConvNeXt para imagens do plano sagital. Considerando que a combinação utilizando MLP com *augmentation* obteve os melhores resultados, as discussões a seguir serão nesta configuração de arquiteturas e classificador.

Tabela 5.9 – Resultados de Classificação Final

Cenário	Modelo	Acurácia	F1-Score	Sensibilidade	Precisão	AUC	Arq. Coronal	Arq. Sagital
Sem Augmentation	MLP	90,56%	90,00%	88,23%	91,83%	95,90%	Convnext Base Swin Base	Swin Base VGG 19
	SVM	91,03%	90,45%	88,23%	92,78%	95,92%	Convnext Base Swin Base	Swin Base VGG 19
	XGBoost	89,62%	89,42%	**91,17%**	87,73%	95,16%	Convnext Base EfficientNet B0	Convnext Base Swin Base VGG 19
Com Augmentation	MLP	**91,98%**	**91,63%**	**91,18%**	92,08%	**96,45%**	Convnext Base Swin Base EfficientNet B0	Swin Base VGG 19
	SVM	91,03%	90,54%	89,21%	91,91%	95,49%	Swin Base	Convnext Base Swin Base VGG 19
	XGBoost	**91,98%**	91,45%	89,21%	**93,81%**	96,22%	Convnext Base Swin Base	Convnext Base Swin Base

Fonte: Autoral.

O desempenho superior no cenário com *augmentation* indica que o aumento da base de dados permitiu aos classificadores explorar melhor a diversidade das características extraídas. Ao unir Transformers, que possuem uma visão global da imagem, com CNNs, que focam em detalhes e texturas locais, as limitações individuais de cada técnica são compensadas. Essa integração é fundamental, pois permite que a deficiência de uma arquitetura ou plano anatômico em determinada classe seja suprida pelas virtudes do outro.

Em seguida, foi levantado o tipo de diagnóstico para o resultado da classificação. Obtiveram-se 93 exames classificados como VP, 102 como VN, 8 como FP e 9 como FN. A Tabela 5.10 apresenta quais diagnósticos para a melhor combinação citada anteriormente.

Tabela 5.10 – Resultados da Classificação por Diagnóstico.

Classificação	**Câncer de Pulmão**	**Linfoma**	**Melanoma**	**Saudável**
FN	0	2	7	0
FP	0	0	0	8
VN	0	0	0	102
VP	34	29	30	0

Fonte: Autoral.

A análise da Tabela 5.10 revela que o melanoma apresenta maior dificuldade classificatória; isso acontece devido às grandes dimensões dos exames originais. O redimensionamento para o padrão de entrada da rede neural degrada a resolução espacial, prejudicando a morfologia, o que justifica a ocorrência dos falsos negativos para esta lesão. Em oposição, o câncer de pulmão demonstra maior simplicidade por não apresentar metástases nas extremidades inferiores do corpo. Essa concentração regional dos achados patológicos permitiu a identificação correta de todos os 34 casos positivos sem a incidência de erros de classificação.

Além disso, a etapa de geração das imagens MIP combinada ao redimensionamento causa perda de informações estruturais dos exames originais. Esse processamento afeta a representação das lesões nos exames de linfoma e melanoma, resultando diretamente na maior quantidade de erros de classificação observada para esses diagnósticos.

5.5 Estudos de Caso

Esta seção apresenta estudos de caso para avaliar qualitativamente o método de classificação automática em representações MIP de exames PET. A análise qualitativa utiliza mapas de calor gerados pelo algoritmo *Grad-CAM* (SELVARAJU et al., 2017) para visualizar o desempenho dos modelos em casos específicos para as arquiteturas baseadas em CNNs e com uma adaptação para a *Swin*, visto que ela é baseada em *Transformers*. A adaptação consiste no rearranjo espacial dos tensores para converter os *tokens* de atenção em mapas bidimensionais, permitindo o cálculo dos gradientes sobre a estrutura hierárquica de *patches*.

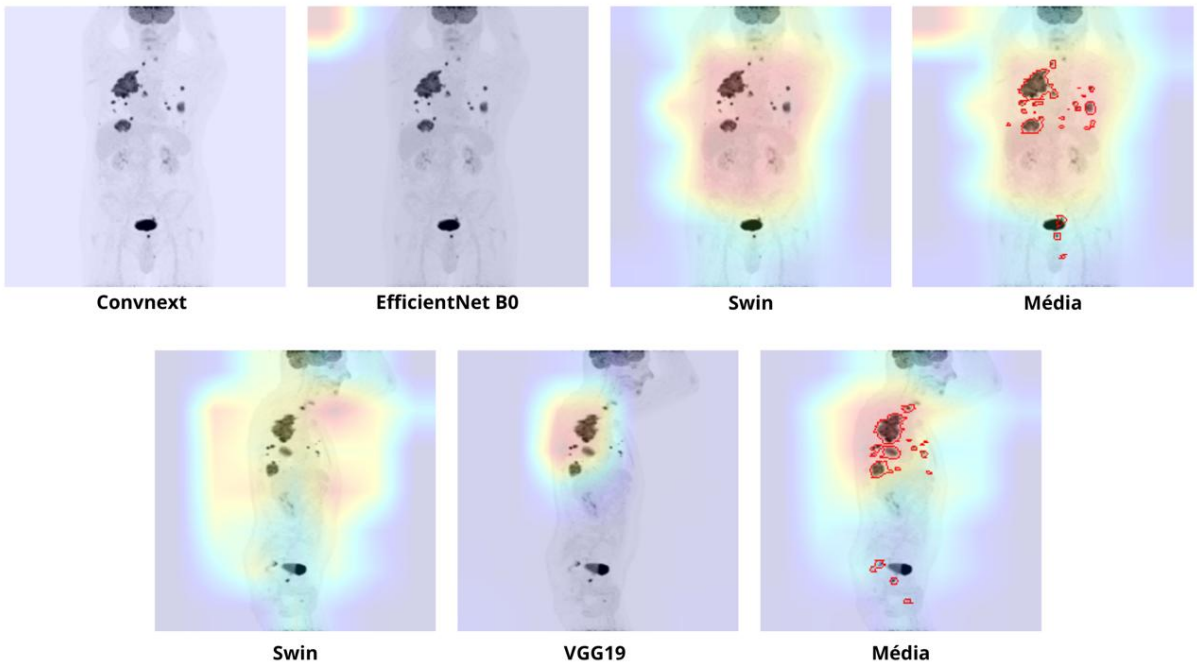
5.5.1 Estudos Qualitativos

Os estudos de caso são apresentados para avaliar, de forma qualitativa, o método de classificação automática nas representações MIP de exames PET. Esses estudos ofereceram

uma análise visual do desempenho dos modelos utilizados, destacando casos específicos de classificações corretas, falsos positivos e falsos negativos, utilizando os mapas de calor fornecidos pelo algoritmo *Grad-CAM* (SELVARAJU et al., 2017).

Um exemplo em que a classificação de um verdadeiro positivo ocorre com a probabilidade de 100% para um exame de câncer de pulmão é mostrado na Figura 5.6. Observa-se que a arquitetura ConvNeXt não possui atenção para nenhuma área específica da imagem. Além disso, a EfficientNet-B0 não observa áreas relevantes para a classificação. Em contrapartida, a Swin e a VGG19 tiveram seu foco nas áreas com maior ativação e no centro do corpo do paciente, sendo essenciais para a classificação como verdadeiro positivo. Este caso foi bem-sucedido provavelmente pela combinação das projeções coronal e sagital, uma vez que a projeção sagital apresentou resultados mais consistentes, complementando o resultado na projeção coronal.

Figura 5.6 – Caso de exame com lesão classificado corretamente

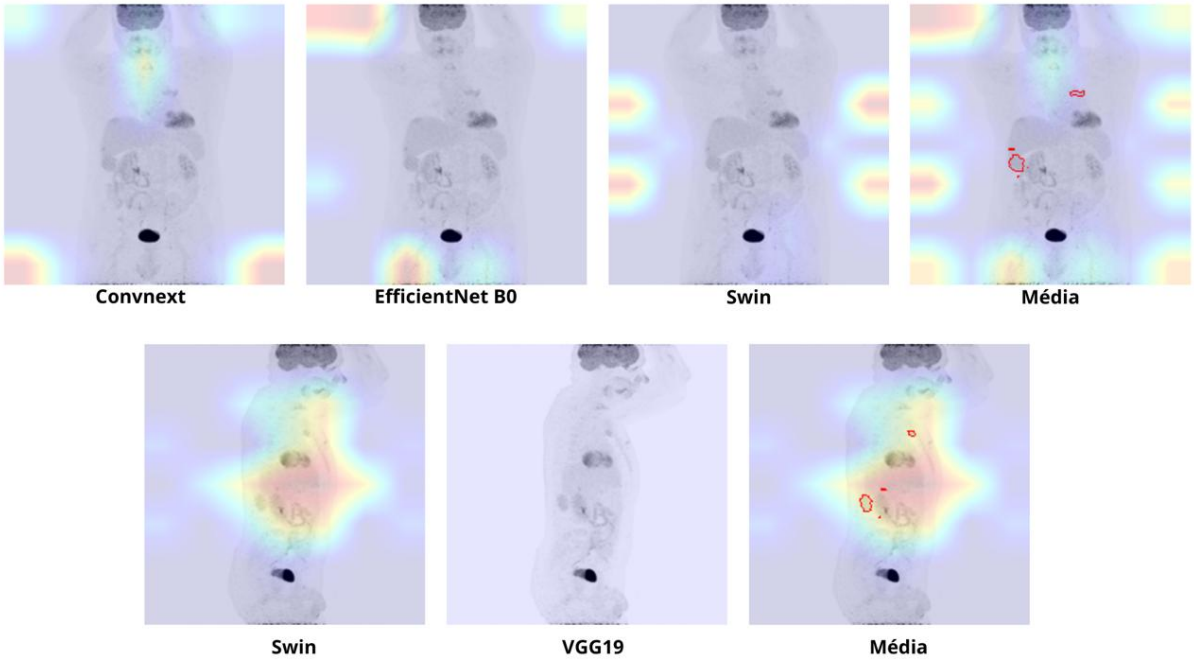


Fonte: Autoral.

Por outro lado, à medida que o grau de confiança de um verdadeiro positivo diminui, as arquiteturas alteram suas regiões de ativação. Na Figura 5.7, observa-se um exemplo de verdadeiro positivo para linfoma classificado com menor certeza, apresentando uma probabilidade de 0,540. É possível analisar que a ConvNeXt passa a exibir uma ativação no mapa de calor. Em contrapartida, a VGG19, que no caso anterior apresentava grande foco na lesão, agora não possui nenhuma ativação significativa. Dessa forma, nota-se que a Swin foi a arquitetura que melhor contribuiu para o acerto da classificação.

A VGG19 continuou apresentando esse mesmo comportamento, não apresentando

Figura 5.7 – Exemplos de Verdadeiro Positivo com menor grau de de confiança para a classe positiva



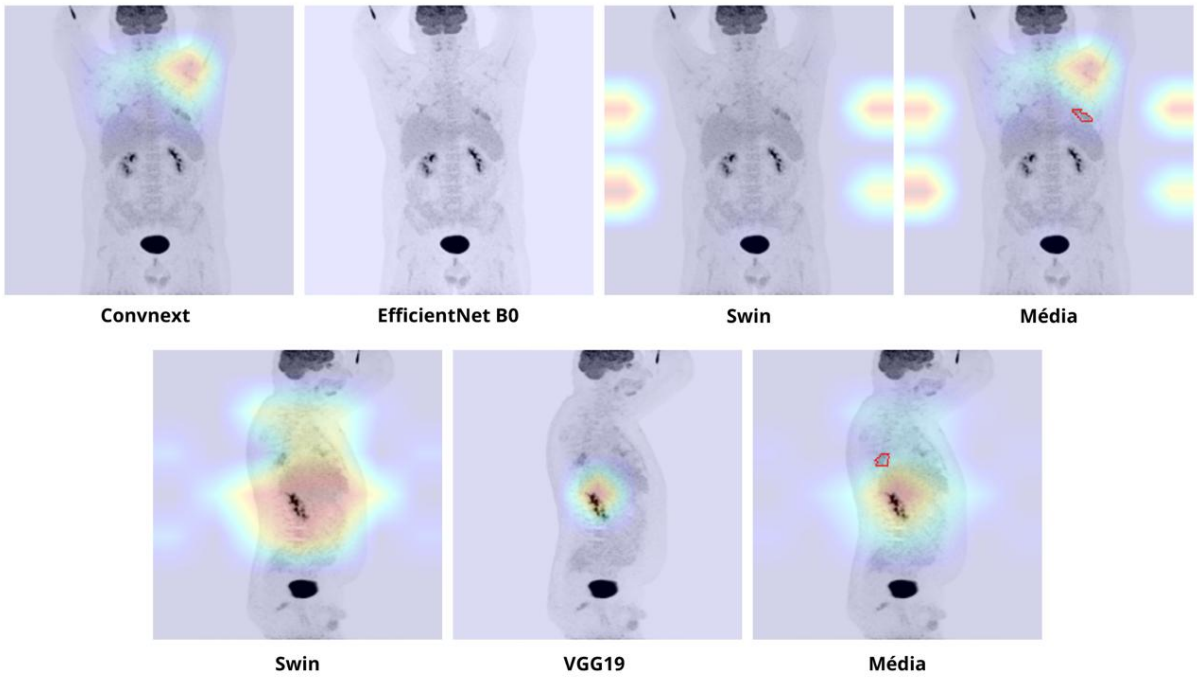
Fonte: Autoral.

ativação. As únicas exceções foram os casos de verdadeiro positivo com alto grau de confiança e o de falso negativo. A Figura 5.8 mostra um exemplo de falso negativo de linfoma, onde a lesão possui uma baixa ativação metabólica,e as arquiteturas no eixo coronal focaram em áreas distintas, nenhuma sendo realmente a região da lesão. Já para a imagem no corte sagital, a Swin manteve o comportamento de observar uma grande parte do corpo, em especial a região do tronco. A VGG19 manteve o foco apenas em uma região com alto grau de ativação, mas que não correspondia a uma lesão. As setas vermelhas na Figura 5.8 indicam a região da lesão.

Ainda nos casos de falso negativo, o exame de melanoma que apresentou a maior confiança para a classe negativa obteve uma saída de $2{,}2244868 \times 10^{-7}$. Nesse caso, a lesão era reduzida e possuía uma intensidade muito baixa, sendo de difícil visualização. A Figura 5.9 ilustra esse caso, e as setas apontam para a região da lesão.

Para os casos de falso positivo, a EfficientNet-B0 manteve o mesmo comportamento observado no verdadeiro positivo. Entretanto, observou-se que arquiteturas como a ConvNeXt e a Swin (nos planos coronal e sagital) utilizaram a região da bexiga urinária para a classificação. Esta é uma região com alta ativação natural ao realizar o exame, configurando um falso positivo natural em virtude da elevada concentração do radiofármaco no órgão responsável pela excreção do traçador. A Figura 5.10 ilustra esse exemplo de falso positivo com o grau de confiança de 0,540.

Figura 5.8 – Caso de exame com lesão classificado incorretamente sem lesão



Fonte: Autoral.

5.6 Comparação com Trabalhos Relacionados

Ao comparar o método proposto com trabalhos relacionados, é possível notar que, inicialmente, os resultados obtidos superam significativamente os demais nas métricas de acurácia e F1-Score, conforme a Tabela 5.11.

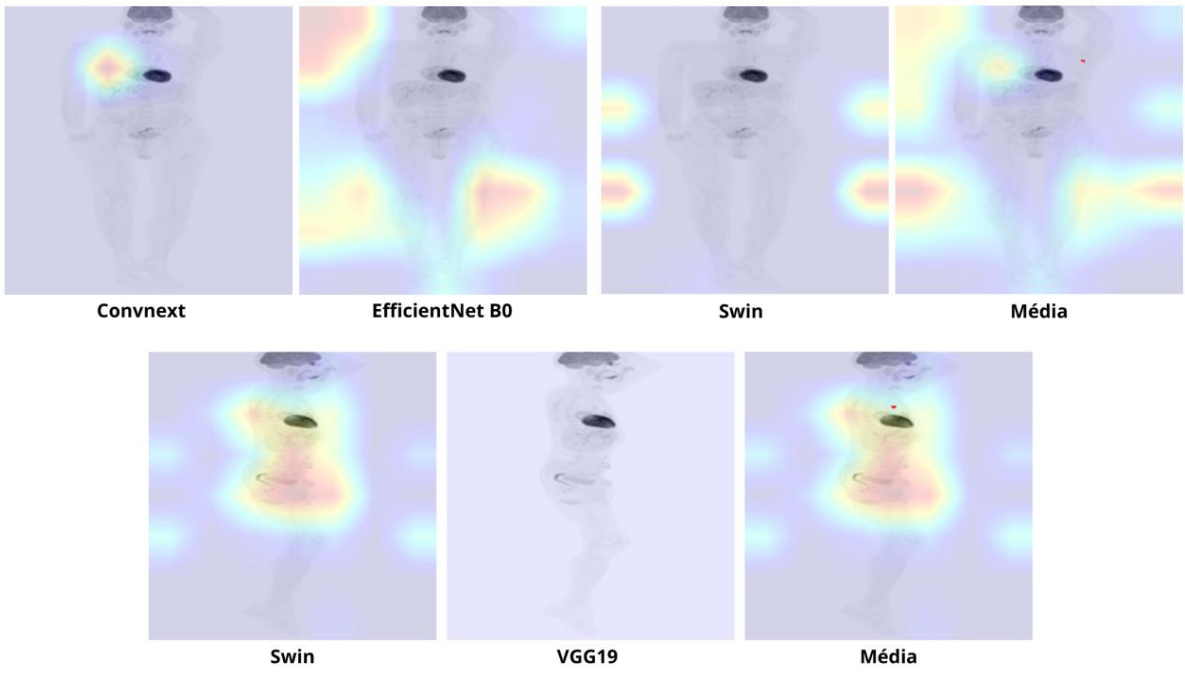
Tabela 5.11 – Comparação com Trabalhos Relacionados.

Trabalho	Aplicação	AUC	Acurácia	F1	Sens.	Precisão	Conjunto de Dados	Exame Utilizado
(DIRKS et al., 2022)*	Melanoma	-	-	75,00%	-	-	Privado	PET/CT
(REN et al., 2025)	Linfoma	94,20%	-	-	-	**93,88%**	Privado	PET/CT
(HÄGGSTRÖM et al., 2024)	Linfoma	94,90%	89,00%	-	86,80%	-	Privado	PET/CT
(SIBILLE et al., 2020)	Linfoma	95,00%	-	-	75,40%	-	Privado	PET/CT
	Câncer de Pulmão	**98,00%**	-	-	87,10%	-	Privado	PET/CT
(ZHANG et al., 2024)	Adenocarcinoma Pulmonar	87,00%	-	-	-	-	Privado	PET/CT
(PARK et al., 2021)	Câncer de Pulmão	87,70%	82,50%	88,20%	91,20%	85,60%	Privado	PET/CT
(ALVES; CARDOSO; GAMA, 2024)	Nódulos Pulmonares	83,85%	73,91%	-	80,00%	-	Privado	PET/CT
(HEILIGER et al., 2022)	Pulmão, Linfoma e Melanoma	-	74,30%	-	-	-	AutoPET	PET PET/CT
(PANG et al., 2024)*	Pulmão, Linfoma e Melanoma	-	78,00%	89,00%	**98,00%**	84,00%	AutoPET	PET/CT/RM
Método Proposto	Pulmão, Linfoma e Melanoma	96,45%	**91,98%**	**91,63%**	91,18%	92,08%	AutoPET	PET

* Trabalhos que fazem Detecção
Fonte: Autoral.

O trabalho de Heiliger et al.(2022) realizou uma classificação binária como uma etapa preliminar para o seu método de segmentação utilizando a combinação entre exames PET e TC. Já Pang et al. (2024) fizeram uso de combinações de imagens 3D com 2D, exames

Figura 5.9 – Exemplo de Falso Negativo com Menor grau de de Confiança



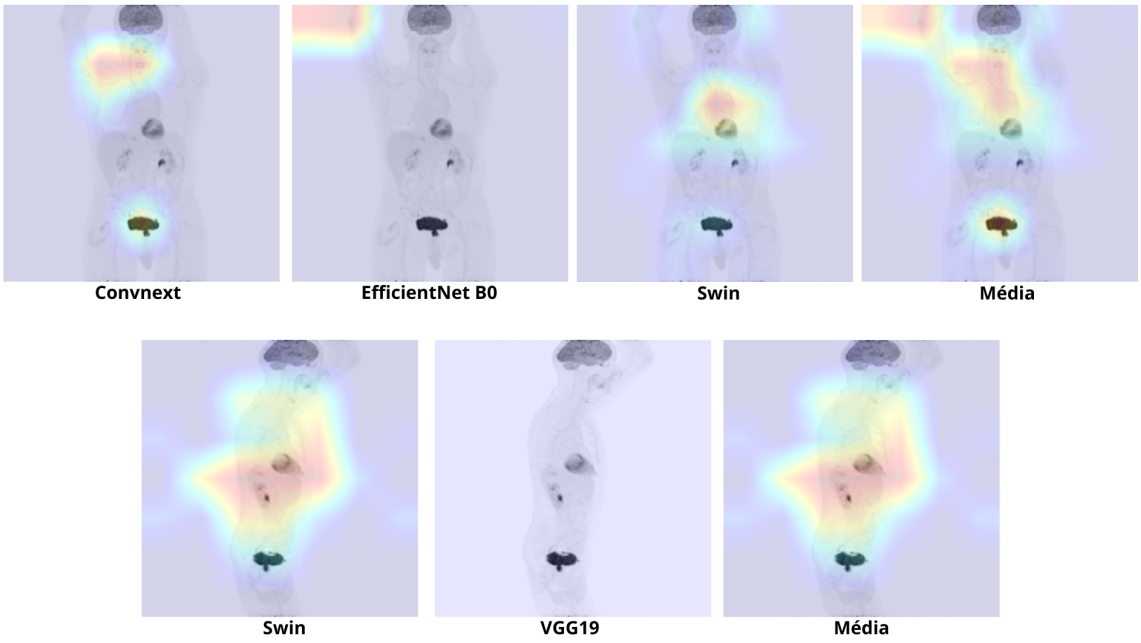
Fonte: Autoral.

PET e CT para detecção. Embora o método desses autores apresente alta sensibilidade (98%) e F1-Score (89%) na tarefa de detecção, a acurácia global (78%) permanece abaixo da obtida pelo presente trabalho, que utilizou uma abordagem com representações 2D dos exames nos cortes coronais e sagitais, junto com técnicas de *deep features*.

A superioridade do método proposto também se manifesta na sua capacidade de generalização. Enquanto grande parte da literatura foca em lesões específicas, o método proposto demonstra eficácia equilibrada na classificação simultânea de câncer de pulmão, linfoma e melanoma. Ao atingir uma AUC de 96,45%, o sistema supera propostas consolidadas, como as de (REN et al., 2025) e (HÄGGSTRÖM et al., 2024), que, embora apresentem bons resultados para linfoma, não possuem a mesma versatilidade multilesão.

Por fim, a geração de representações MIP nos eixos coronal e sagital se demonstrou eficiente ao alcançar resultados próximos de abordagens que usam técnicas 3D e a combinação de imagens PET e CT. Isso é visto ao comparar a acurácia de (ALVES; CARDOSO; GAMA, 2024), que atingiu 73,91% classificando nódulos pulmonares, e (HÄGGSTRÖM et al., 2024), ao classificar exames com linfoma e saudáveis. Nesse contexto, o método proposto se consolida como uma alternativa mais eficiente, usando apenas imagens PET em 2D e, mesmo assim, oferece desempenho superior ou equivalente a métodos que utilizam volumes 3D ou combinações com TC.

Figura 5.10 – Caso de exame sem lesão classificado incorretamente com lesão



Fonte: Autoral.

6 Conclusão

A identificação de lesões cancerígenas em exames PET é uma tarefa que requer muito tempo e esforço do profissional na sua execução. Métodos de classificação automática para identificar se esses exames possuem ou não algum tipo de lesão têm sido desenvolvidos de forma a auxiliar o profissional da saúde no diagnóstico mais preciso e rápido do paciente.

Este trabalho propôs e validou um método de classificação automática de exames PET utilizando aprendizado profundo. A abordagem utilizou uma geração de imagens MIP baseada em um estudo do conjunto de dados que não foi explorado em trabalhos anteriores para auxiliar no problema proposto. Utilizam-se, ainda, seis arquiteturas para extração de características e três classificadores diferentes para o objetivo final da tarefa, sendo elas ConvNeXt, EfficientNet-B0 e Swin para representações no corte coronal e ConvNeXt, VGG19 e Swin para as imagens geradas a partir do eixo sagital. Já os classificadores testados foram MLP, SVM e XGBoost. O método proposto conseguiu desempenho superior ao de trabalhos relacionados na mesma base de dados AutoPET.

Os resultados indicam que o método proposto consegue classificar as representações MIP dos volumes de exames PET do conjunto de dados Auto Pet III de forma automática utilizando Convnext Base, Swin Base e EfficientNet B0 para projeções coronais e Swin Base e VGG19 para projeções sagitais para extração de características e uma MLP para classificação. Assim, alcançando as métricas de 91,98% de acurácia, 91,62% de F1-Score, 91,17% de Sensibilidade, 92,07% de precisão e AUC de 96,45%. Nos experimentos realizados, técnicas como a geração de imagens MIP contribuíram significativamente para o problema, aproximando-se de trabalhos que utilizam combinações de imagens PET e TC e o processamento de exames 3D. Outra contribuição foi a utilização de um conjunto de técnicas e arquiteturas específicas para a extração de características para cada tipo de perspectiva do mesmo exame, onde cada arquitetura extrai informações diferentes e cada uma possui sua especialidade. Notou-se também que testar as combinações entre as arquiteturas auxiliou na melhoria das métricas.

Considerando os resultados dos experimentos e a comparação com outros trabalhos, pode-se afirmar que o método consegue realizar a tarefa. Compreende-se, ainda, que os resultados levantam indícios de que o método proposto pode, de fato, auxiliar o profissional da saúde na tarefa de classificação dos exames PET. Portanto, entende-se que os objetivos deste trabalho foram atingidos.

Como trabalhos futuros, sugere-se utilizar exames com o radiotraçador PSMA em pacientes com diagnóstico de carcinoma de próstata e casos saudáveis, visando validar a versatilidade do modelo em diferentes perfis metabólicos. Além disso, geração das imagens

MIP com maior profundidade de bits, visando perder menos informações durante a geração das mesmas e preservar detalhes sutis do contraste radiofarmacêutico. Outra modificação que pode trazer benefícios é a utilização de um bloco de atenção após a concatenação das características e na MLP responsável pela classificação, dessa forma, é possível realizar o cálculo de atenção entre as características extraídas para a redução de dimensionalidade, mantendo apenas aquelas que são úteis para a classificação. Por fim, sugere-se a realização da validação cruzada para permitir uma avaliação da capacidade de generalização do método, garantindo que as métricas alcançadas não sejam dependentes de uma divisão específica entre treino e teste.

Referências

ACHARYA, U. R. et al. Linear and nonlinear analysis of normal and cad-affected heart rate signals. *Computer methods and programs in biomedicine*, Elsevier, v. 113, n. 1, p. 55–68, 2014. Citado na página 17.

AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. [S.l.: s.n.], 2019. p. 2623–2631. Citado na página 54.

ALI, Z. et al. Advances in radiological techniques for cancer diagnosis: A narrative review of current technologies. Health Nexus, 2024. Citado na página 16.

ALNUAIMI, A. F.; ALBALDAWI, T. H. Concepts of statistical learning and classification in machine learning: An overview. In: EDP SCIENCES. *BIO Web of Conferences*. [S.l.], 2024. v. 97, p. 00129. Citado 2 vezes nas páginas 28 e 29.

ALVES, V. M.; CARDOSO, J. dos S.; GAMA, J. Classification of pulmonary nodules in 2-[18f] fdg pet/ct images with a 3d convolutional neural network. *Nuclear Medicine and Molecular Imaging*, Springer, v. 58, n. 1, p. 9–24, 2024. Citado 4 vezes nas páginas 23, 24, 61 e 62.

ALZUBI, J.; NAYYAR, A.; KUMAR, A. Machine learning from theory to algorithms: an overview. In: IOP PUBLISHING. *Journal of physics: conference series*. [S.l.], 2018. v. 1142, p. 012012. Citado na página 28.

ANSELL, S. M. Hodgkin lymphoma: A 2020 update on diagnosis, risk-stratification, and management. *American journal of hematology*, Wiley Online Library, v. 95, n. 8, p. 978–989, 2020. Citado na página 27.

APICELLA, A. et al. A survey on modern trainable activation functions. *Neural Networks*, Elsevier, v. 138, p. 14–32, 2021. Citado na página 33.

BADAWI, R. D. et al. First human imaging studies with the explorer total-body pet scanner. *Journal of Nuclear Medicine*, Society of Nuclear Medicine, v. 60, n. 3, p. 299–303, 2019. Citado na página 26.

BANGAR, S. *VGG-Net Architecture Explained*. Medium, 2022. Disponível em: <https://medium.com/@siddheshb008/vgg-net-architecture-explained-71179310050f>. Acesso em: 19 dez. 2025. Citado 2 vezes nas páginas 34 e 45.

BOELLAARD, R. et al. Fdg pet/ct: Eanm procedure guidelines for tumour imaging: version 2.0. *European journal of nuclear medicine and molecular imaging*, Springer, v. 42, n. 2, p. 328–354, 2015. Citado na página 26.

BRAY, F. et al. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, Wiley Online Library, v. 74, n. 3, p. 229–263, 2024. Citado na página 16.

BUDACH, L. et al. The effects of data quality on machine learning performance. *arXiv preprint arXiv:2207.14529*, 2022. Citado na página 28.

BUSHBERG, J. et al. *The Essential Physics of Medical Imaging*. 3rd. ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2012. Citado 5 vezes nas páginas 16, 17, 25, 26 e 27.

CHAN, H.-P.; SAMALA, R. K.; HADJIISKI, L. M. Cad and ai for breast cancer—recent development and challenges. *The British journal of radiology*, The British Institute of Radiology., v. 93, n. 1108, p. 20190580, 2019. Citado na página 17.

CHEN, T. Xgboost: A scalable tree boosting system. *Cornell University*, 2016. Citado na página 30.

CHERRY, S. R. et al. Total-body pet: maximizing sensitivity to create new opportunities for clinical research and patient care. *Journal of Nuclear Medicine*, Society of Nuclear Medicine, v. 59, n. 1, p. 3–12, 2018. Citado na página 26.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 29.

DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *International workshop on multiple classifier systems*. [S.l.], 2000. p. 1–15. Citado na página 30.

DIRKS, I. et al. Computer-aided detection and segmentation of malignant melanoma lesions on whole-body 18f-fdg pet/ct using an interpretable deep learning approach. *Comput. Methods Programs Biomed.*, v. 221, p. 106902, 2022. Citado 3 vezes nas páginas 23, 24 e 61.

DOSOVITSKIY, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. Citado 4 vezes nas páginas 17, 33, 38 e 45.

FALLAHPOOR, M. et al. Deep learning techniques in pet/ct imaging: A comprehensive review from sinogram to image space. *Computer methods and programs in biomedicine*, Elsevier, v. 243, p. 107880, 2024. Citado na página 17.

FLETCHER, J. W. et al. Recommendations on the use of 18f-fdg pet in oncology. *Journal of Nuclear Medicine*, Society of Nuclear Medicine, v. 49, n. 3, p. 480–508, 2008. Citado na página 26.

FROOD, R. et al. Comparative effectiveness of standard vs. ai-assisted pet/ct reading workflow for pre-treatment lymphoma staging: a multi-institutional reader study evaluation. *Frontiers in Nuclear Medicine*, Frontiers Media SA, v. 3, p. 1327186, 2024. Citado na página 17.

GATIDIS, S. et al. A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. *Scientific Data*, Nature Publishing Group UK London, v. 9, n. 1, p. 601, 2022. Citado 3 vezes nas páginas 27, 41 e 42.

GIL, J. et al. Deep learning-based feature extraction from whole-body pet/ct employing maximum intensity projection images: preliminary results of lung cancer data. *Nuclear Medicine and Molecular Imaging*, Springer, v. 57, n. 5, p. 216–222, 2023. Citado na página 17.

GONZALEZ, R.; WOODS, R. *Processamento de imagens digitais*. Edgard Blucher, 2000. ISBN 9788521202646. Disponível em: <https://books.google.com.br/books?id=d3MnAgAACAAJ>. Citado 2 vezes nas páginas 42 e 43.

HÄGGSTRÖM, I. et al. Deep learning for [18f] fluorodeoxyglucose-pet-ct classification in patients with lymphoma: a dual-centre retrospective analysis. *The Lancet Digital Health*, Elsevier, v. 6, n. 2, p. e114–e125, 2024. Citado 3 vezes nas páginas 24, 61 e 62.

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 45.

HEILIGER, L. et al. Autopet challenge: combining nn-unet with swin unetr augmented by maximum intensity projection classifier. *arXiv preprint arXiv:2209.01112*, 2022. Citado 3 vezes nas páginas 23, 24 e 61.

JEBLICK, K. et al. Dataset, *A whole-body PSMA-PET/CT dataset with manually annotated tumor lesions (PSMA-PET-CT-Lesions) (Version 1)*. [S.l.]: The Cancer Imaging Archive, 2024. Version 1. Citado na página 41.

KINAHAN, P. E.; FLETCHER, J. W. PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy. *Seminars in Ultrasound, CT and MRI*, Elsevier, v. 31, n. 6, p. 496–505, 2010. Citado na página 16.

KINAHAN, P. E.; FLETCHER, J. W. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Seminars in Ultrasound, CT and MRI*, v. 31, n. 6, p. 496–505, 2010. ISSN 0887-2171. PET-CT in RT Planning and Assessment. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0887217110000880>. Citado na página 27.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012. Citado na página 17.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado 2 vezes nas páginas 31 e 33.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 2002. Citado 2 vezes nas páginas 31 e 32.

LECUN, Y. et al. Efficient backprop. In: *Neural networks: Tricks of the trade*. [S.l.]: Springer, 2002. p. 9–50. Citado na página 43.

LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017. Citado na página 17.

LIU, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 10012–10022. Citado 3 vezes nas páginas 38, 39 e 45.

LIU, Z. et al. A convnet for the 2020s. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2022. p. 11976–11986. Citado 3 vezes nas páginas 35, 36 e 45.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 31.

NVIDIA; VINGELMANN, P.; FITZEK, F. H. *CUDA*. [S.l.], 2020. Release: 10.2.89. Disponível em: <https://developer.nvidia.com/cuda-toolkit>. Acesso em: 7 jan. 2026. Citado na página 48.

PANG, L. et al. Comparison of the accuracy of a deep learning method for lesion detection in pet/ct and pet/mri images. *Molecular Imaging and Biology*, Springer, v. 26, n. 5, p. 802–811, 2024. Citado 3 vezes nas páginas 23, 24 e 61.

PARK, Y.-J. et al. Performance evaluation of a deep learning system for differential diagnosis of lung cancer with conventional ct and fdg pet/ct using transfer learning and metadata. *Clinical Nuclear Medicine*, LWW, v. 46, n. 8, p. 635–640, 2021. Citado 2 vezes nas páginas 24 e 61.

PUTRO, Y. A. P. et al. Right thigh mass metastasis from lung cancer mimicking primary soft tissue sarcoma: A case report. *The American Journal of Case Reports*, v. 25, p. e942416–1, 2024. Citado na página 27.

RAGHU, M. et al. Do vision transformers see like convolutional neural networks? *Advances in neural information processing systems*, v. 34, p. 12116–12128, 2021. Citado na página 45.

RANA, N. et al. Correction of positron emission tomography maximum intensity projection image artifact using retro reconstruction method. *Indian Journal of Nuclear Medicine*, Medknow, v. 35, n. 3, p. 235–237, 2020. Citado 2 vezes nas páginas 17 e 42.

REN, J. et al. Pet normalizations to improve deep learning auto-segmentation of head and neck tumors in 3d pet/ct. In: *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. [S.l.]: Springer, 2021. p. 83–91. Citado na página 49.

REN, J. et al. The significance of pet/ct combined with machine learning models for the classification of lymphoma involvement and metastases in enlarged lymph nodes. *Frontiers in Oncology*, v. 15, p. 1643924, 2025. Citado 3 vezes nas páginas 24, 61 e 62.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 31.

ROSSUM, G. van; JR, F. L. D. *Python reference manual*. Amsterdam, NLD, 1995. CWI Report CS-R9525. Citado na página 48.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, IBM, v. 3, n. 3, p. 210–229, 1959. Citado na página 28.

SARKER, I. H. Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, Springer, v. 2, n. 3, p. 160, 2021. Citado na página 28.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 618–626. Citado 2 vezes nas páginas 58 e 59.

SHAMSHAD, F. et al. Transformers in medical imaging: A survey. *Medical image analysis*, Elsevier, v. 88, p. 102802, 2023. Citado na página 45.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of big data*, Springer, v. 6, n. 1, p. 1–48, 2019. Citado na página 55.

SIBILLE, L. et al. 18f-fdg pet/ct uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. *Radiology*, Radiological Society of North America, v. 294, n. 2, p. 445–452, 2020. Citado 2 vezes nas páginas 24 e 61.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado 2 vezes nas páginas 33 e 34.

SUETENS, P. *Fundamentals of Medical Imaging*. 2. ed. Cambridge: Cambridge University Press, 2009. ISBN 978-0-521-51915-1. Citado 3 vezes nas páginas 25, 26 e 27.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International conference on machine learning*. [S.l.], 2019. p. 6105–6114. Citado 4 vezes nas páginas 33, 34, 35 e 45.

TAN, M.; LE, Q. Efficientnetv2: Smaller models and faster training. In: PMLR. *International conference on machine learning*. [S.l.], 2021. p. 10096–10106. Citado na página 45.

TAS, F. Metastatic behavior in melanoma: timing, pattern, survival, and influencing factors. *Journal of oncology*, Wiley Online Library, v. 2012, n. 1, p. 647684, 2012. Citado na página 27.

TERVEN, J. et al. Loss functions and metrics in deep learning. *arXiv preprint arXiv:2307.02694*, 2023. Citado 2 vezes nas páginas 31 e 33.

THARWAT, A. Classification assessment methods. *Applied computing and informatics*, Emerald Publishing Limited, v. 17, n. 1, p. 168–192, 2021. Citado 2 vezes nas páginas 46 e 47.

TOWNSEND, D. Multimodality imaging of structure and function. *Physics in Medicine & Biology*, IOP Publishing, v. 53, n. 4, p. R1, 2008. Citado na página 25.

VASWANI, A. et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017. Citado 2 vezes nas páginas 36 e 37.

WILD, C. P.; WEIDERPASS, E.; STEWART, B. W. World cancer report. International Agency for Research on Cancer, 2020. Citado na página 16.

World Health Organization. *Global cancer burden growing, amidst mounting need for services*. 2024. Acesso em: 03 dez. 2025. Disponível em: <https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services>. Citado na página 16.

ZHANG, Y. et al. Machine learning for differentiating lung squamous cell cancer from adenocarcinoma using clinical-metabolic characteristics and 18f-fdg pet/ct radiomics. *Plos one*, Public Library of Science San Francisco, CA USA, v. 19, n. 4, p. e0300170, 2024. Citado 3 vezes nas páginas 23, 24 e 61.