VICTOR DANIEL OLIVEIRA GAETE

ANÁLISE COMPARATIVA ENTRE TRANSFORMERS VISUAIS E CNNS NA CLASSIFICAÇÃO DE CALCIFICAÇÕES E MASSAS MAMÁRIAS

BELO HORIZONTE

2026

ANÁLISE COMPARATIVA ENTRE TRANSFORMERS VISUAIS E CNNS NA CLASSIFICAÇÃO DE CALCIFICAÇÕES E MASSAS MAMÁRIAS

Artigo científico apresentado ao Curso de Graduação em Engenharia Elétrica do Centro Federal de Educação Tecnológica de Minas Gerais como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Alexandre Rodrigues Farias

Coorientador: Prof. Dr. Túlio Charles de Oliveira Carvalho

BELO HORIZONTE
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
2026

SUMÁRIO

ANÁLISE COMPARATIVA ENTRE TRANSFORMERS VISUAIS E CNNS NA CLASSIFICAÇÃO DE CALCIFICAÇÕES E MASSAS MAMÁRIAS[1]

Victor Daniel Oliveira Gaete[2]
Prof. Dr. Alexandre Rodrigues Farias[3]
Prof. Dr. Túlio Charles de Oliveira Carvalho[4]

RESUMO

Este trabalho apresenta uma análise comparativa entre redes neurais convolucionais e Transformers visuais aplicados à classificação binária de lesões mamográficas, considerando massas e calcificações. Foram utilizados recortes de regiões de interesse, da base CBIS-DDSM como conjunto principal, com avaliação externa na base VinDr-Mammo. O desempenho dos modelos foi avaliado principalmente por meio da métrica ROC-AUC, complementada por acurácia, precisão, sensibilidade, especificidade e F1-score. Os resultados indicam diferenças de desempenho entre os tipos de achado, com melhor desempenho geral na classificação de massas em comparação às calcificações. Além disso, observou-se que a generalização externa varia entre arquiteturas e tipos de lesão, evidenciando a importância da validação em bases independentes.

Palavras-chave: mamografia; redes neurais convolucionais; Transformers visuais; classificação de imagens; aprendizado profundo.

ABSTRACT

This work presents a comparative analysis between convolutional neural networks and visual Transformers applied to the binary classification of mammographic lesions, considering masses and calcifications. Region-of-interest patches from the CBIS-DDSM dataset were used as the main dataset, with external evaluation on the VinDr-Mammo dataset. Model performance was mainly assessed using the ROC-AUC metric, complemented by accuracy, precision, sensitivity, specificity, and F1-score. The results indicate performance differences between lesion types, with better overall performance for mass classification compared to calcifications. In addition, external generalization varied across architectures and lesion types, highlighting the importance of validation on independent datasets.

[1] Artigo científico elaborado para conclusão do curso de graduação em Engenharia Elétrica do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG).

[2] Graduando em Engenharia Elétrica pelo Centro Federal de Educação Tecnológica de Minas Gerais. E-mail: vdgaete@gmail.com.

[3] Orientador. Professor Doutor do Centro Federal de Educação Tecnológica de Minas Gerais. E-mail: alexandrefarias@cefetmg.br.

[4] Coorientador. Professor Doutor do Centro Federal de Educação Tecnológica de Minas Gerais. E-mail: tulio@cefetmg.br.

Keywords: mammography; convolutional neural networks; Visual Transformers; image classification; deep learning.

Data de submissão: 16 de junho de 2026
Data de aprovação: 26 de Junho de 2026

1 INTRODUÇÃO

O câncer de mama representa um dos principais desafios globais em saúde pública. De acordo com a Organização Mundial da Saúde (OMS), a doença foi responsável por aproximadamente 670 mil mortes em 2022 e foi o tipo de câncer mais comum entre mulheres em 157 de 185 países avaliados (World Health Organization, 2026). Estimativas do GLOBOCAN 2022 indicam cerca de 2,3 milhões de novos casos e 666 mil mortes no mesmo ano, correspondendo a aproximadamente 11,6% de todos os novos casos de câncer no mundo (Bray *et al.*, 2024). Projeções da IARC indicam que os novos casos anuais poderão atingir 3,2 milhões até 2050, com cerca de 1,1 milhão de mortes por ano (International Agency for Research on Cancer, 2025). Nesse contexto, estratégias de rastreamento e detecção precoce tornam-se fundamentais.

Apesar de sua ampla utilização clínica, a interpretação de mamografias ainda apresenta desafios. A sobreposição de tecidos, a variação da densidade mamária, o baixo contraste de determinadas lesões e a semelhança visual entre achados benignos e malignos podem dificultar a análise, contribuindo para falsos positivos, falsos negativos e maior dependência da experiência do especialista (Mota *et al.*, 2025; Barazi; Gunduru, 2023). Além disso, o desconforto ou a dor associados ao exame podem atuar como barreiras à adesão ao rastreamento em parte das pacientes, embora permaneça uma ferramenta fundamental para a detecção precoce do câncer de mama (Whelehan *et al.*, 2013; Montoro; Alcaraz; Gálvez-Sánchez, 2023). Assim, métodos computacionais de apoio à decisão têm sido investigados para aumentar a consistência da avaliação mamográfica (Wang, 2024; Amin *et al.*, 2025).

Conforme discutido por Takahashi *et al.* (2024), redes neurais convolucionais (CNNs) são amplamente utilizadas por sua capacidade de aprender automaticamente características visuais locais, como bordas, texturas e padrões morfológicos, enquanto Transformers visuais passaram a ser investigados mais recentemente.

Apesar dos avanços, ainda não há consenso claro sobre a superioridade de uma família de modelos em relação à outra. Parte dessa limitação decorre de comparações realizadas sob condições experimentais distintas, envolvendo diferenças nas bases de dados, no pré-processamento, nas divisões dos dados, nos critérios de seleção dos modelos e nas métricas de avaliação (Varoquaux; Cheplygina, 2022). Com isso, torna-se difícil atribuir diferenças de desempenho exclusivamente à arquitetura utilizada.

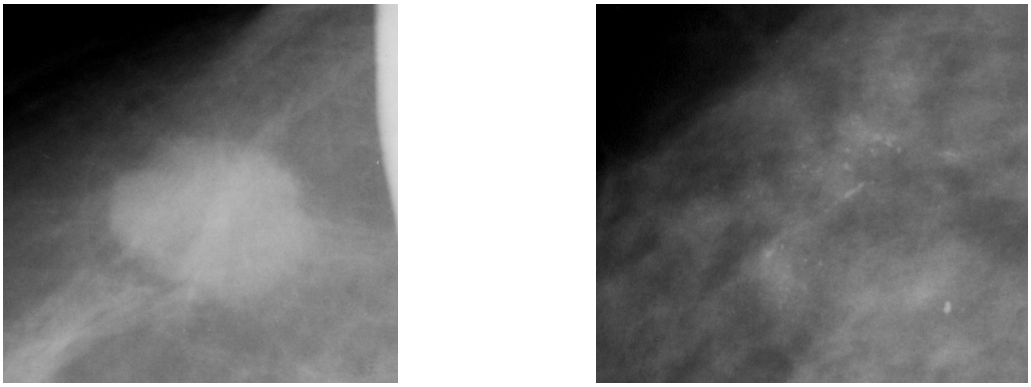
Diante disso, este trabalho apresenta uma análise comparativa entre CNNs e Transformers visuais para a classificação binária de lesões mamográficas benignas e malignas, considerando separadamente massas e calcificações. A comparação é conduzida por meio de um protocolo experimental padronizado, com avaliação interna na base CBIS-DDSM e validação externa no conjunto VinDr-Mammo (Nguyen *et al.*, 2022). Com isso, busca-se contribuir para uma comparação mais controlada e reprodutível entre diferentes famílias arquiteturais aplicadas à classificação de lesões mamográficas.

2 FUNDAMENTAÇÃO TEÓRICA

A classificação automática de lesões mamográficas depende da relação entre os padrões visuais presentes nas imagens e a capacidade dos modelos computacionais de distingui-los. Em mamografias, diferentes achados podem variar quanto à forma, textura, contraste e escala, fatores que influenciam diretamente a complexidade da tarefa.

Nesse contexto, esta seção apresenta os fundamentos necessários para compreender a comparação experimental proposta. Inicialmente, são discutidas as características visuais de massas e calcificações. Em seguida, são apresentados os princípios das redes neurais convolucionais e das arquiteturas baseadas em Transformers visuais. Por fim, são descritas as métricas utilizadas para avaliar o desempenho dos modelos em uma tarefa de classificação binária.

Figura 1 - Exemplos reais de recortes mamográficos contendo massa e calcificação.



(a) Massa (b) Calcificação

Fonte: Elaborado pelo autor a partir de imagens da base CBIS-DDSM (Lee *et al.*, 2017).

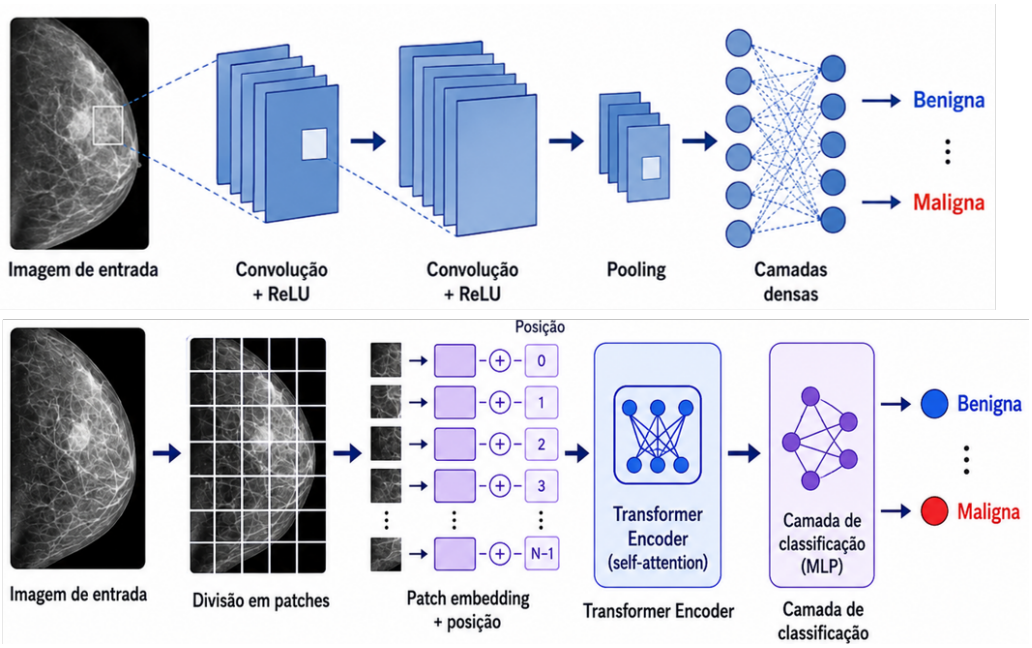
A mamografia é um dos principais exames utilizados no rastreamento e no apoio ao diagnóstico do câncer de mama, permitindo identificar achados como massas, calcificações, assimetrias e distorções arquiteturais (Heath *et al.*, 2000; Lee *et al.*, 2017; Wang, 2024). No contexto da classificação automática, massas e calcificações são particularmente relevantes por apresentarem padrões visuais distintos e por estarem amplamente representadas em bases públicas de referência, como a CBIS-DDSM (Lee *et al.*, 2017).

Massas geralmente aparecem como regiões de maior extensão espacial, com variações

de forma, margem e densidade. Sua avaliação envolve aspectos como contorno, delimitação e relação com o tecido mamário adjacente. Calcificações, por outro lado, tendem a se manifestar como pequenas estruturas de alta intensidade, frequentemente agrupadas, com variações de distribuição, morfologia e contraste (Heath *et al.*, 2000; Lee *et al.*, 2017; Qureshi *et al.*, 2025).

Essas diferenças justificam a análise separada de massas e calcificações neste trabalho. Ao tratar cada tipo de achado como um recorte experimental independente, busca-se avaliar o desempenho dos modelos em padrões mamográficos distintos, preservando a especificidade visual de cada tarefa de classificação (Litjens *et al.*, 2017; Manigrasso *et al.*, 2025; Wang, 2024).

Figura 2 - Representação esquemática do funcionamento de uma CNN e de um Vision Transformer (ViT).



Fonte: Elaborado pelo autor.

2.1 Redes neurais convolucionais

As redes neurais convolucionais (CNNs) constituem uma das principais famílias de modelos de aprendizado profundo utilizadas em visão computacional e imagens médicas. Consolidadas a partir do trabalho de LeCun *et al.* (1998), sua arquitetura é baseada em operações de convolução, nas quais filtros aprendidos durante o treinamento são aplicados sobre a imagem para extrair características visuais relevantes. Ao explorar conexões locais e compartilhamento de pesos, as CNNs reduzem o número de parâmetros e tornam o aprendizado mais eficiente para imagens.

A representação construída pelas CNNs ocorre de forma hierárquica. Nas camadas iniciais, os filtros tendem a responder a padrões simples, como bordas, orientações,

contrastes e texturas. Em camadas mais profundas, esses padrões são combinados em características mais abstratas e discriminativas para a tarefa de interesse.

Na área de imagens médicas, CNNs são amplamente empregadas em tarefas de classificação, detecção e segmentação, pois permitem aprender características diretamente dos dados, reduzindo a dependência de descritores manuais. Em mamografias, essa capacidade é relevante para identificar padrões como forma, margem, densidade, textura e contraste, associados à diferenciação entre achados benignos e malignos.

Diversas arquiteturas convolucionais foram propostas para aprimorar o treinamento e a representação desses modelos. A ResNet introduz conexões residuais, que facilitam o treinamento de redes profundas (He *et al.*, 2016). A DenseNet utiliza conexões densas, promovendo reutilização de características ao longo da rede (Huang *et al.*, 2017). Já a ConvNeXt reformula blocos convolucionais tradicionais a partir de princípios inspirados em arquiteturas modernas, buscando aumentar a competitividade das CNNs em relação aos Transformers visuais (Liu *et al.*, 2022; Woo *et al.*, 2023).

2.2 Arquiteturas baseadas em Transformers visuais

As arquiteturas baseadas em Transformers foram originalmente desenvolvidas para processamento de linguagem natural, a partir do trabalho de Vaswani *et al.* (2017), e posteriormente adaptadas para visão computacional. Nesse contexto, o Vision Transformer (ViT) é um tipo de Transformer visual, proposto para aplicar diretamente essa estrutura em tarefas de classificação de imagens. O ViT representa a imagem como uma sequência de *patches*, que são processados por mecanismos de autoatenção (Dosovitskiy *et al.*, 2021). Diferentemente das CNNs, que extraem características por meio de operações locais, os Transformers visuais permitem que diferentes regiões da imagem interajam diretamente entre si, favorecendo a modelagem de relações espaciais de longo alcance (Vaswani *et al.*, 2017; Dosovitskiy *et al.*, 2021).

Na área de imagens médicas, Transformers têm sido utilizados em tarefas de classificação, segmentação e detecção de lesões (He *et al.*, 2023; Azad *et al.*, 2023). Em mamografias, mecanismos de atenção podem auxiliar na modelagem de relações entre diferentes regiões da imagem, complementando a análise de características locais (Ayana *et al.*, 2023; Sarker *et al.*, 2024).

Por apresentarem menor viés indutivo para padrões visuais locais quando comparados às CNNs, Transformers visuais frequentemente dependem de grandes volumes de dados ou de estratégias específicas de treinamento para aprender representações robustas (Dosovitskiy *et al.*, 2021; Touvron *et al.*, 2021). Nesse contexto, variantes foram propostas para melhorar sua eficiência, estabilidade e desempenho. O DeiT (*Data-efficient Image Transformer*) emprega estratégias de treinamento e destilação para bases de menor escala (Touvron *et al.*, 2021). O Swin Transformer utiliza atenção em janelas locais deslocadas,

reduzindo o custo computacional e incorporando estrutura hierárquica à representação visual (Liu *et al.*, 2021). O SwinV2 propõe aprimoramentos voltados à estabilidade do treinamento e à escalabilidade de modelos baseados em atenção (Liu *et al.*, 2022).

2.3 Avaliação de classificadores em imagens médicas

A avaliação de classificadores em imagens médicas deve considerar não apenas a proporção de acertos, mas também a natureza dos erros e a capacidade de generalização para dados independentes. Na classificação de lesões mamográficas, falsos negativos podem estar associados à não identificação de casos malignos, enquanto falsos positivos podem levar a exames complementares e intervenções desnecessárias. Assim, a avaliação deve contemplar tanto a capacidade discriminativa global quanto o comportamento dos modelos em pontos operacionais específicos.

Neste trabalho, a comparação entre arquiteturas é conduzida por métricas independentes e dependentes de limiar. Entre as métricas independentes, destaca-se a ROC-AUC, que quantifica a capacidade discriminativa do modelo ao longo de diferentes limiares, medindo o quanto o classificador tende a atribuir escores mais altos às amostras positivas do que às negativas (Hanley; McNeil, 1982; Fawcett, 2006). Por não depender de um único limiar fixo, a ROC-AUC é adotada como métrica principal de comparação.

No entanto, a ROC-AUC não descreve completamente o comportamento operacional do classificador, especialmente em cenários com desbalanceamento entre classes ou custos diferentes para cada tipo de erro. Por isso, também são consideradas métricas complementares, como PR-AUC, acurácia, sensibilidade, especificidade e F1-score (Saito; Rehmsmeier, 2015).

Para definir um ponto operacional único, uma alternativa frequente é o índice de Youden, dado por

$$J = \text{sensibilidade} + \text{especificidade} - 1, \tag{1}$$

que busca maximizar simultaneamente sensibilidade e especificidade (Youden, 1950). Como as métricas dependentes de limiar variam conforme o ponto escolhido, a definição do limiar deve ser realizada de forma controlada, preferencialmente a partir do conjunto de validação, evitando ajuste direto sobre o conjunto de teste.

Além da avaliação interna, a validação externa é importante para investigar a generalização dos modelos. Em imagens médicas, diferenças entre bases podem surgir de variações nos equipamentos, protocolos clínicos, características populacionais, critérios de anotação e distribuição dos casos. Neste trabalho, essa análise é realizada com o conjunto VinDr-Mammo (Nguyen *et al.*, 2022), utilizado como base externa para avaliar a robustez dos modelos treinados.

Em conjunto, essas métricas e estratégias de avaliação permitem analisar a discriminação global dos modelos, seu comportamento em pontos de decisão definidos e sua

capacidade de generalizar para dados mamográficos independentes.

3 TRABALHOS RELACIONADOS

O aprendizado profundo tem sido amplamente utilizado em mamografias para tarefas de detecção, segmentação e classificação de lesões. Revisões recentes destacam que esses modelos permitem aprender representações visuais diretamente dos dados, reduzindo a dependência de descritores manuais e contribuindo para sistemas de apoio ao diagnóstico (Litjens *et al.*, 2017; Wang, 2024; Amin *et al.*, 2025).

Bases públicas têm papel importante na comparação de métodos para análise mamográfica. O DDSM e sua versão curada, o CBIS-DDSM, são frequentemente utilizados por disponibilizarem mamografias anotadas com massas, calcificações e informações patológicas associadas (Heath *et al.*, 2000; Lee *et al.*, 2017). Bases externas, como a VinDr-Mammo, permitem complementar essa avaliação ao testar a generalização dos modelos em domínios distintos daquele utilizado no treinamento (Nguyen *et al.*, 2022).

Grande parte dos estudos em mamografia utiliza modelos de aprendizado profundo baseados em arquiteturas convolucionais, devido à sua capacidade de representar padrões visuais relevantes para a caracterização de lesões. Mais recentemente, arquiteturas baseadas em atenção também passaram a ser investigadas em imagens médicas, motivadas por sua capacidade de modelar relações entre diferentes regiões da imagem (Takahashi *et al.*, 2024).

No domínio da mamografia, estudos recentes investigam estratégias como modelos multivisão, mecanismos de atenção e abordagens voltadas à representação de padrões sutis, incluindo microcalcificações (Manigrasso *et al.*, 2025; Mansour *et al.*, 2025; Qureshi *et al.*, 2025). Essas abordagens indicam que diferentes tipos de achado podem demandar formas distintas de representação, especialmente ao comparar massas e calcificações.

Apesar dos avanços, a comparação direta entre estudos ainda é limitada por diferenças nas bases utilizadas, nos recortes de entrada, no pré-processamento, nas divisões de treino e teste, na seleção de hiperparâmetros e nas métricas reportadas. Além disso, muitos trabalhos apresentam resultados agregados, sem distinguir massas e calcificações, ou restringem a avaliação ao mesmo domínio de desenvolvimento. Diante disso, observa-se a necessidade de comparações mais controladas entre diferentes famílias de arquiteturas, com avaliação separada por tipo de achado, protocolos padronizados e validação externa.

4 METODOLOGIA EXPERIMENTAL

Esta seção descreve o protocolo utilizado para comparar arquiteturas convolucionais e Transformers visuais na classificação de lesões mamográficas, mantendo separadas as etapas de ajuste, validação e avaliação final.

4.1 Bases de dados e recortes analisados

Foram utilizadas duas bases públicas de mamografia, compostas por exames previamente coletados, anonimizados e disponibilizados para fins de pesquisa. Dessa forma, este trabalho não envolveu coleta direta de dados clínicos, contato com pacientes ou identificação individual das participantes. As análises foram conduzidas exclusivamente sobre imagens e anotações públicas, respeitando as condições de uso definidas pelos respectivos repositórios.

O CBIS-DDSM foi adotado como base principal de desenvolvimento, por conter recortes de massas e calcificações com informações patológicas associadas (Heath *et al.*, 2000; Lee *et al.*, 2017). A divisão entre treino e teste utilizada neste trabalho seguiu a separação original disponibilizada pelo próprio CBIS-DDSM, preservando o teste fixo como subconjunto independente para avaliação final. O VinDr-Mammo foi empregado exclusivamente na validação externa, com o objetivo de avaliar a generalização dos modelos em outro domínio (Nguyen *et al.*, 2022).

As análises foram conduzidas separadamente para massas e calcificações, devido às diferenças morfológicas entre esses achados. A distribuição das amostras do CBIS-DDSM é apresentada na Tabela 1.

Tabela 1 - Distribuição das amostras por recorte, classe e divisão experimental no CBIS-DDSM.

Recorte	Divisão	Benigno	Maligno	Total
Massas	Treino	681	637	1318
Massas	Teste	231	147	378
Calcificações	Treino	1001	544	1545
Calcificações	Teste	197	129	326

Fonte: elaborado pelo autor.

O teste fixo do CBIS-DDSM, definido pela divisão original da base, não foi utilizado na otimização de hiperparâmetros, seleção de modelos ou definição de limiares, reduzindo o risco de viés na avaliação final (Varma; Simon, 2006; Roberts *et al.*, 2021).

Tabela 2 - Distribuição das classes no conjunto VinDr-Mammo utilizado na validação externa.

Recorte	Benigno	Maligno	Total
Massas	567	652	1219
Calcificações	63	444	507

Fonte: elaborado pelo autor.

A Tabela 2 resume os subconjuntos externos utilizados na validação em VinDr-Mammo. Foram considerados apenas exames com achado identificado como massa ou calcificação, região de interesse, do inglês *Region of Interest* (ROI), associada e dimensão

adequada para processamento. Como essa base foi usada apenas para validação externa, seus dados não participaram do treinamento, da otimização de hiperparâmetros ou da definição dos limiares.

4.2 Protocolo de validação

Figura 3 - Fluxo geral do protocolo experimental adotado neste trabalho.

Fonte: Elaborado pelo autor.

O protocolo foi organizado em três etapas: otimização de hiperparâmetros, validação cruzada interna e avaliação final, conforme a Figura 3. A otimização foi realizada com Optuna apenas sobre o treino do CBIS-DDSM, usando ROC-AUC de validação como métrica objetivo.

Após a definição da melhor configuração de cada arquitetura e recorte, os modelos foram avaliados por validação cruzada estratificada em 5 folds, com agrupamento por paciente quando disponível.

Figura 4 - Esquema da validação cruzada estratificada em 5 folds.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Iter. 1	Validação	Treino	Treino	Treino	Treino
Iter. 2	Treino	Validação	Treino	Treino	Treino
Iter. 3	Treino	Treino	Validação	Treino	Treino
Iter. 4	Treino	Treino	Treino	Validação	Treino
Iter. 5	Treino	Treino	Treino	Treino	Validação

Fonte: Elaborado pelo autor.
Legenda: em cada iteração, um fold é reservado para validação e os demais são utilizados para treinamento.

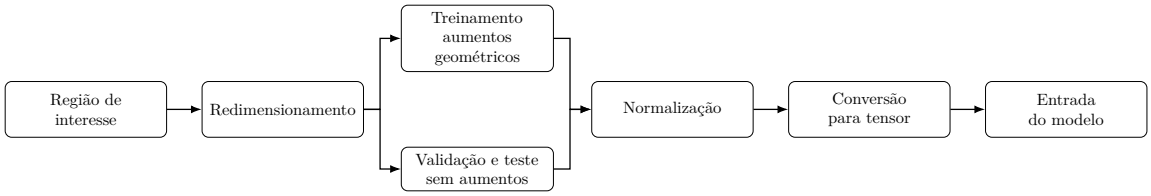
Ao final, o desempenho foi estimado no teste fixo do CBIS-DDSM e, complementarmente, no VinDr-Mammo, sem reotimização de hiperparâmetros ou limiares.

4.3 Preparação das imagens

Os experimentos foram conduzidos sobre regiões de interesse previamente extraídas. Todas as arquiteturas seguiram o mesmo pipeline, composto por redimensionamento, normalização e conversão para tensor, conforme a Figura 5.

Durante o treinamento, foram aplicados aumentos geométricos leves: espelhamento horizontal com probabilidade de 0,5, rotações de até $\pm 7°$ e transformações afins moderadas, com translações de até $\pm 5\%$ e escala entre 0,95 e 1,05, aplicadas com probabilidade de 0,3 (Shorten; Khoshgoftaar, 2019).

Figura 5 - Fluxo geral de preparação das imagens, com distinção entre treinamento e avaliação.



Fonte: Elaborado pelo autor.

Nas etapas de validação, teste interno e validação externa, foi utilizado pipeline determinístico, sem aumentos de dados.

4.4 Arquiteturas avaliadas

Foram avaliadas dez arquiteturas, distribuídas entre CNNs e Transformers visuais, conforme a Tabela 3.

Tabela 3 - Arquiteturas utilizadas na análise principal.

Família	Modelo	Entrada	Parâmetros
Transformer	SwinV2-Small	224×224	48,9
Transformer	SwinV2-Tiny	224×224	27,6
Transformer	SwinV2-Base	256×256	86,9
Transformer	ViT-Base	224×224	86,1
Transformer	DeiT-Small	224×224	21,7
CNN	ConvNeXtV2-Tiny	256×256	27,9
CNN	ConvNeXt-Tiny	256×256	28,1
CNN	ResNet-50	256×256	24,6
CNN	EfficientNetV2-S	256×256	20,8
CNN	DenseNet-169	256×256	13,3

Fonte: elaborado pelo autor.
Nota: os valores de parâmetros são apresentados em milhões.

Cada arquitetura foi avaliada separadamente para massas e calcificações, mantendo o mesmo protocolo experimental.

4.5 Otimização, treinamento e limiar de decisão

Os hiperparâmetros de cada combinação entre arquitetura e recorte foram otimizados com Optuna (Akiba *et al.*, 2019), usando apenas o treino do CBIS-DDSM e ROC-AUC de validação como métrica objetivo. Em seguida, a melhor configuração foi fixada e reavaliada por validação cruzada estratificada em 5 folds.

O treinamento utilizou AdamW, *label smoothing*, *gradient clipping*, ajuste dinâmico da taxa de aprendizado e parada antecipada baseada na ROC-AUC de validação (Loshchilov; Hutter, 2019; Szegedy *et al.*, 2015). Os checkpoints foram selecionados exclusivamente pelo desempenho de validação.

Para cada arquitetura, as predições dos cinco modelos foram agregadas pela média das probabilidades, usada como escore final de cada amostra (Dietterich, 2000). O limiar operacional foi definido nas predições de validação interna pelo índice de Youden:

$$J = \text{sensibilidade} + \text{especificidade} - 1 \tag{2}$$

O limiar que maximizou esse índice foi mantido fixo no teste interno e na validação externa (Youden, 1950; Varma; Simon, 2006).

4.6 Métricas de avaliação

A métrica principal foi a ROC-AUC, por avaliar a capacidade discriminativa dos modelos ao longo de diferentes limiares, sem depender de um ponto operacional fixo (Hanley; McNeil, 1982; Fawcett, 2006). Essa escolha é adequada ao estudo, pois os recortes avaliados apresentam diferentes proporções entre classes (Richardson *et al.*, 2024).

4.7 Comparação dos resultados

A comparação entre modelos foi conduzida de forma descritiva, usando a ROC-AUC como métrica principal. Para cada recorte, os modelos foram ordenados pelo desempenho no teste interno da CBIS-DDSM, enquanto a validação externa em VinDr-Mammo foi usada para analisar a generalização.

As análises foram realizadas separadamente para massas e calcificações e também agrupadas por família arquitetural. Métricas dependentes de limiar foram usadas como medidas complementares. Como não foram realizados testes estatísticos formais de hipótese, as conclusões foram baseadas em rankings, diferenças relativas de desempenho e consistência entre avaliação interna e externa.

4.8 Ambiente computacional

Todos os experimentos foram executados em um ambiente Python com suporte a GPU NVIDIA GeForce RTX 3060. Os modelos foram implementados em PyTorch (Paszke *et al.*, 2019), com apoio do PyTorch Lightning para organização do treinamento (Falcon; The PyTorch Lightning Team, 2019). A otimização de hiperparâmetros foi realizada com Optuna (Akiba *et al.*, 2019), e o rastreamento dos experimentos foi conduzido com MLflow (Zaharia *et al.*, 2018).

5 RESULTADOS

Esta seção apresenta os resultados obtidos na comparação entre arquiteturas convolucionais e modelos baseados em Transformers visuais para a classificação de massas e calcificações mamográficas. A ROC-AUC foi adotada como métrica principal de avaliação, por quantificar a capacidade discriminativa dos modelos independentemente de um limiar fixo de decisão. Métricas dependentes de limiar, como acurácia, precisão, sensibilidade, especificidade e F1-score, foram consideradas de forma complementar, utilizando o ponto operacional definido pelo índice de Youden na validação interna.

A apresentação dos resultados foi organizada em torno de dois cenários de avaliação: o desempenho interno no conjunto de teste fixo da CBIS-DDSM e o desempenho externo na base VinDr-Mammo. Em ambos os casos, os modelos são identificados pelo nome da arquitetura. Os identificadores completos dos experimentos, bem como as demais métricas calculadas, são mantidos nas tabelas complementares.

5.1 Desempenho interno em CBIS-DDSM

A Tabela 4 apresenta os resultados obtidos no conjunto de teste interno da CBIS-DDSM para os recortes de massas e calcificações. No recorte de massas, os valores de ROC-AUC variaram entre 0,7521 e 0,8317, com maior desempenho para o SwinV2-Base. Em seguida, apareceram o SwinV2-Tiny, com ROC-AUC de 0,8122, e o ConvNeXtV2-Tiny, com 0,8109. No mesmo recorte, os valores de acurácia variaram entre 67,7% e 73,5%, sendo o maior valor observado para o DeiT-Small.

No recorte de calcificações, a ROC-AUC variou entre 0,7302 e 0,7887. Assim como em massas, o maior valor foi obtido pelo SwinV2-Base, seguido pelo ConvNeXtV2-Tiny, com ROC-AUC de 0,7767, e pelo SwinV2-Small, com 0,7715. Para a acurácia, os valores ficaram entre 63,5% e 69,6%, com maior valor observado para o SwinV2-Small.

Tabela 4 - Desempenho interno em CBIS-DDSM para massas e calcificações.

Modelo	Massas		Calcificações	
	ROC-AUC	Acurácia	ROC-AUC	Acurácia
SwinV2-Base	0,8317	73,0%	0,7887	68,4%
SwinV2-Tiny	0,8122	71,2%	0,7652	66,3%
ConvNeXtV2-Tiny	0,8109	72,0%	0,7767	68,1%
SwinV2-Small	0,8018	70,6%	0,7715	69,6%
ConvNeXt-Tiny	0,7999	71,7%	0,7549	63,5%
DeiT-Small	0,7990	73,5%	0,7560	67,2%
ViT-Base	0,7956	67,7%	0,7539	68,1%
DenseNet-169	0,7788	71,2%	0,7395	68,1%
ResNet-50	0,7546	67,7%	0,7302	64,1%
EfficientNetV2-S	0,7521	70,9%	0,7436	65,3%

Fonte: elaborado pelo autor.

Considerando os dois recortes, o SwinV2-Base apresentou o maior valor de ROC-AUC na avaliação interna tanto para massas quanto para calcificações. Entre os modelos convolucionais, o ConvNeXtV2-Tiny obteve os maiores valores de ROC-AUC nos dois recortes, ocupando a terceira posição em massas e a segunda posição em calcificações.

5.2 Desempenho externo em VinDr-Mammo

A Tabela 5 apresenta os resultados da validação externa em VinDr-Mammo. No recorte de massas, os valores de ROC-AUC variaram entre 0,6951 e 0,7967. O maior valor foi obtido pelo ViT-Base, seguido pelo SwinV2-Small, com ROC-AUC de 0,7861, e pelo DeiT-Small, com 0,7808.

No recorte de calcificações, os valores de ROC-AUC variaram entre 0,3741 e 0,6382. O maior valor foi obtido pelo SwinV2-Small, seguido pelo SwinV2-Tiny, com ROC-AUC de 0,6242, e pelo ConvNeXt-Tiny, com 0,5944. Diferentemente da avaliação interna, o SwinV2-Base não ocupou a primeira posição na validação externa, aparecendo com ROC-AUC de 0,7700 em massas e 0,5417 em calcificações.

Em termos de primeira posição, a avaliação interna e a validação externa apresentaram ordenações distintas. Em massas, o SwinV2-Base liderou no teste interno da CBIS-DDSM, enquanto o ViT-Base obteve o maior valor externo em VinDr-Mammo. Em calcificações, o SwinV2-Base também liderou internamente, enquanto o SwinV2-Small apresentou o maior valor externo.

Tabela 5 - Desempenho externo em VinDr-Mammo para massas e calcificações, considerando a métrica ROC-AUC.

Modelo	Massas	Calcificações
ViT-Base	**0,7967**	0,4412
SwinV2-Small	0,7861	**0,6382**
DeiT-Small	0,7808	0,4676
ConvNeXtV2-Tiny	0,7743	0,5766
SwinV2-Base	0,7700	0,5417
EfficientNetV2-S	0,7655	0,5160
SwinV2-Tiny	0,7599	0,6242
DenseNet-169	0,7521	0,4674
ConvNeXt-Tiny	0,7456	0,5944
ResNet-50	0,6951	0,3741

Fonte: elaborado pelo autor.

5.3 Métricas dependentes de limiar

Além da ROC-AUC, foram calculadas métricas dependentes do ponto operacional definido pelo índice de Youden. Esse limiar foi estimado a partir das predições de validação interna e aplicado sem nova otimização ao teste interno da CBIS-DDSM e à validação externa em VinDr-Mammo.

Nas tabelas principais desta seção, a acurácia foi apresentada juntamente com a ROC-AUC para o cenário interno. As demais métricas dependentes de limiar, incluindo precisão, sensibilidade, especificidade e F1-score, são apresentadas nas tabelas complementares. Essa organização permite manter a ROC-AUC como métrica principal da comparação, ao mesmo tempo em que preserva a caracterização dos modelos no ponto operacional definido.

5.4 Comparação descritiva dos rankings

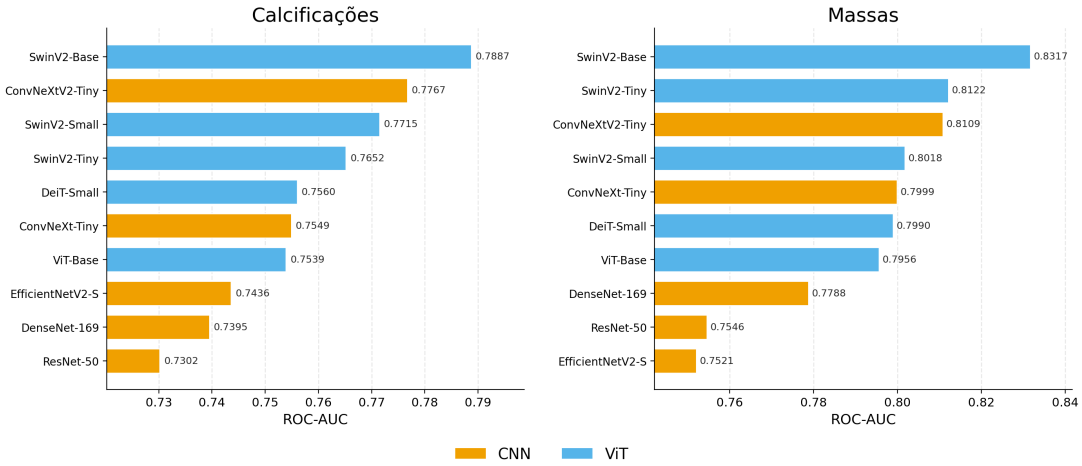
A comparação dos rankings foi realizada a partir da ROC-AUC. No cenário interno, o SwinV2-Base ocupou a primeira posição em massas e calcificações. Na validação externa, as primeiras posições foram ocupadas por diferentes arquiteturas: ViT-Base em massas e SwinV2-Small em calcificações.

Ao agrupar os modelos por família arquitetural, observa-se que modelos baseados em Transformers visuais ocuparam a primeira posição nos dois recortes da avaliação interna e externa. Também houve presença de arquiteturas convolucionais entre as primeiras posições, especialmente o ConvNeXtV2-Tiny na avaliação interna e o ConvNeXt-Tiny na validação externa de calcificações. A distribuição completa dos modelos por família é apresentada graficamente na Figura 11.

5.5 Visualizações complementares

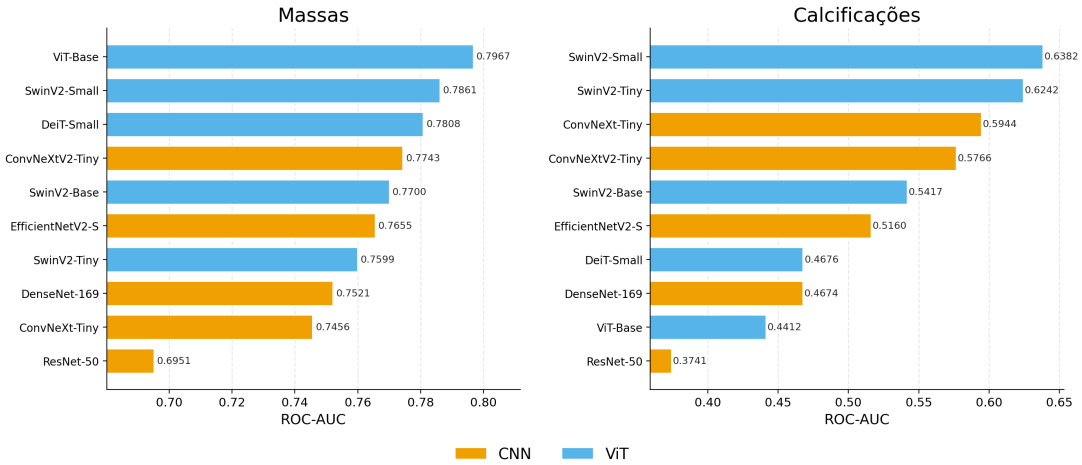
As figuras desta subseção complementam os resultados tabulares. As Figuras 6 e 7 apresentam os rankings de ROC-AUC nos cenários interno e externo, respectivamente. A Figura 8 apresenta a diferença de ROC-AUC entre massas e calcificações para cada arquitetura. As Figuras 9 e 10 mostram, respectivamente, as curvas ROC e as matrizes de confusão normalizadas dos modelos selecionados. Por fim, a Figura 11 apresenta a distribuição da ROC-AUC interna por família arquitetural.

Figura 6 - Ranking interno por ROC-AUC em CBIS-DDSM.



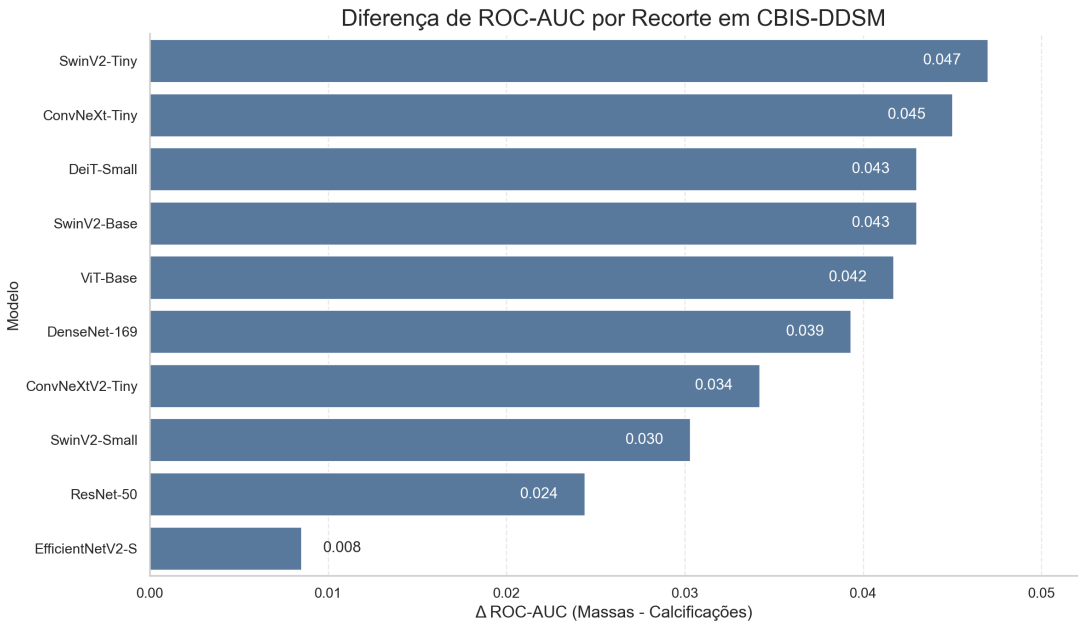
Fonte: elaborado pelo autor.

Figura 7 - Ranking externo por ROC-AUC em VinDr-Mammo.



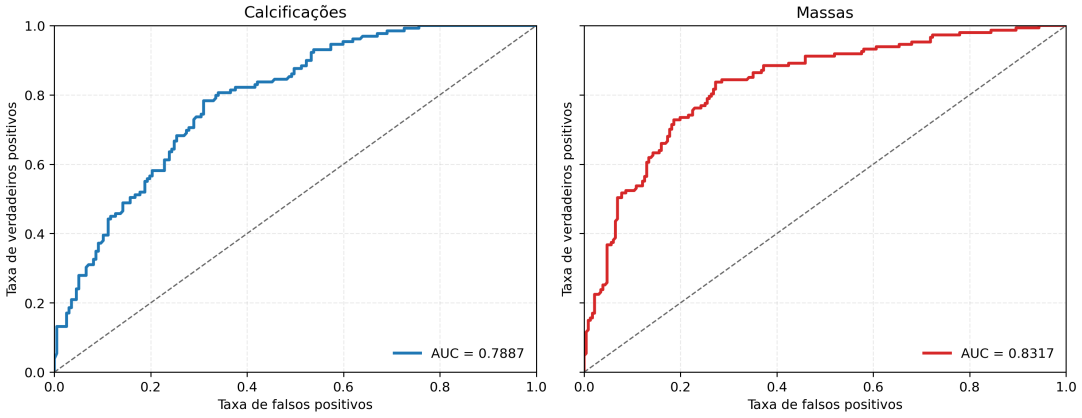
Fonte: elaborado pelo autor.

Figura 8 - Diferença de ROC-AUC entre massas e calcificações.



Fonte: elaborado pelo autor.

Figura 9 - Curvas ROC dos modelos selecionados.

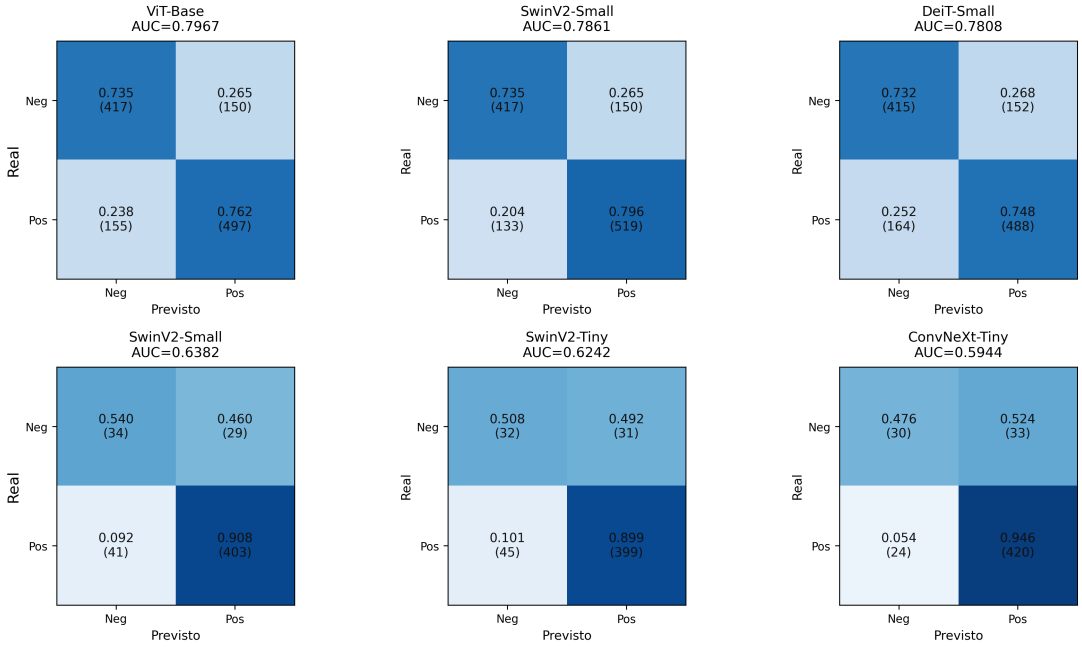


Fonte: elaborado pelo autor.

6 DISCUSSÃO

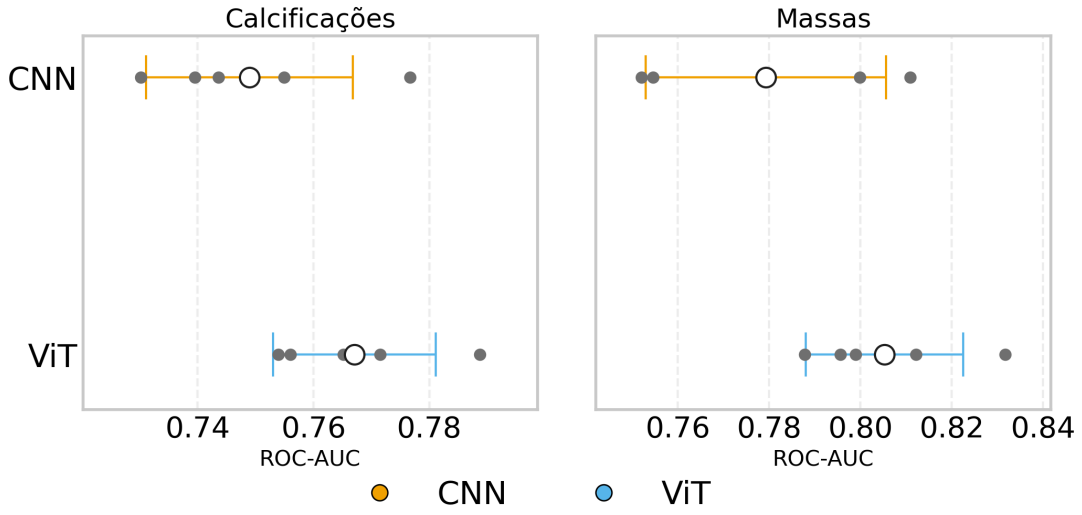
Esta seção discute os principais resultados obtidos a partir do protocolo experimental adotado, considerando o desempenho dos modelos em função do tipo de lesão, da família arquitetural e da capacidade de generalização externa. A análise é organizada de forma a relacionar os resultados quantitativos com as características visuais das lesões mamárias e com as limitações impostas pelas bases de dados utilizadas. Os resultados indicam que o desempenho dos modelos não depende apenas da arquitetura empregada, mas também do tipo de achado mamográfico e do domínio de avaliação. Assim, a comparação entre CNNs e Transformers visuais deve ser interpretada considerando simultaneamente o recorte clínico analisado, a base de teste e a estabilidade do desempenho em validação externa.

Figura 10 - Matrizes de confusão na validação externa.



Fonte: elaborado pelo autor.

Figura 11 - Distribuição da ROC-AUC por família arquitetural.



Fonte: elaborado pelo autor.

6.1 Efeito do tipo de lesão

Os resultados evidenciaram diferenças consistentes entre massas e calcificações. Tanto na avaliação interna em CBIS-DDSM quanto na validação externa em VinDr-Mammo, os modelos apresentaram maior ROC-AUC no recorte de massas, como também observado na Figura 8.

Esse comportamento é compatível com as características visuais de cada achado. Massas tendem a ocupar regiões mais extensas e são descritas por forma, margem e densidade, enquanto calcificações são achados menores, avaliados principalmente por morfologia e distribuição (American College of Radiology, 2013; Spak *et al.*, 2017). Assim, a classificação de calcificações tende a ser mais sensível à resolução de entrada, ao pré-processamento e à preservação de detalhes finos (Geras *et al.*, 2017).

6.2 Famílias arquiteturais e desempenho comparativo

A comparação entre famílias arquiteturais mostrou desempenho competitivo entre CNNs e Transformers visuais. Na avaliação interna em CBIS-DDSM, os Transformers apresentaram médias superiores nos dois recortes avaliados, conforme a Figura 11, com destaque para o SwinV2-Base na Figura 6.

Entretanto, houve sobreposição entre os resultados individuais das duas famílias. CNNs modernas, como ConvNeXtV2-Tiny e ConvNeXt-Tiny, também estiveram entre os modelos mais bem posicionados. Na validação externa, os maiores valores de ROC-AUC foram obtidos por Transformers, com ViT-Base em massas e SwinV2-Small em calcificações, conforme a Figura 7. Ainda assim, os resultados não indicam superioridade global de uma família arquitetural, pois o desempenho variou conforme o tipo de lesão, a base de avaliação e o modelo considerado.

6.3 Generalização externa em VinDr-Mammo

A validação externa em VinDr-Mammo permitiu avaliar os modelos em uma base independente, com características distintas do CBIS-DDSM (Nguyen *et al.*, 2022). Os resultados mostraram que o ranking interno não foi preservado integralmente na avaliação externa: em massas, o melhor modelo passou de SwinV2-Base para ViT-Base; em calcificações, de SwinV2-Base para SwinV2-Small.

Essa mudança sugere influência de diferenças entre as bases, como aquisição das imagens, resolução, critérios de anotação, composição das classes e distribuição dos achados. Em calcificações, o efeito pode ser mais evidente devido à dependência de detalhes locais e ao desbalanceamento do subconjunto externo, composto por 63 amostras benignas e 444 malignas. Esse comportamento também aparece nas matrizes de confusão da Figura 10,

com menor desempenho para a classe negativa.

6.4 Protocolo experimental

O protocolo experimental adotado buscou reduzir vieses na comparação entre modelos por meio de separação por paciente, validação cruzada estratificada em cinco folds, teste interno fixo, otimização de hiperparâmetros e validação externa.

Além disso, o ponto operacional foi definido a partir da validação interna, evitando ajuste direto dos limiares nos conjuntos de teste. Essa escolha torna mais adequada a interpretação de métricas dependentes de limiar, como acurácia, sensibilidade, especificidade e F1-score, em conjunto com a ROC-AUC.

6.5 Limitações do estudo

As principais limitações deste estudo estão relacionadas ao uso de uma única base principal para treinamento, à validação externa restrita ao VinDr-Mammo e à avaliação baseada em regiões de interesse previamente definidas. Embora os recortes de ROI permitam uma comparação controlada entre arquiteturas, eles não representam integralmente o fluxo clínico, no qual o sistema deve localizar e classificar achados em mamografias completas.

Além disso, os resultados dependem do conjunto de modelos, hiperparâmetros, resoluções de entrada e estratégias de pré-processamento avaliados. Portanto, não devem ser interpretados como conclusão definitiva sobre a superioridade de CNNs ou Transformers em mamografia, mas como evidência experimental obtida sob um protocolo padronizado e reprodutível.

7 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho comparou arquiteturas convolucionais e modelos baseados em Transformers visuais na classificação binária de lesões mamográficas em recortes de massas e calcificações do conjunto CBIS-DDSM. Os experimentos seguiram um protocolo unificado, com pré-processamento padronizado, otimização de hiperparâmetros, validação cruzada estratificada em 5 folds, separação por paciente, teste interno fixo, agregação das predições e validação externa em VinDr-Mammo.

Os resultados mostraram melhor desempenho no recorte de massas em comparação ao de calcificações, tanto na avaliação interna quanto na validação externa. Esse comportamento indica que o tipo de lesão influencia diretamente a dificuldade da tarefa, possivelmente devido a diferenças de escala, contraste e padrões morfológicos entre os achados.

Quanto às arquiteturas, os Transformers visuais apresentaram bons resultados em alguns cenários, com destaque para modelos da família SwinV2. No entanto, CNNs modernas, como ConvNeXtV2-Tiny e ConvNeXt-Tiny, também foram competitivas. Assim, não foi observada superioridade global de uma família arquitetural sobre a outra.

A validação externa em VinDr-Mammo mostrou que o desempenho interno em CBIS-DDSM não garante a mesma ordenação dos modelos em outro domínio, reforçando a importância de avaliar a generalização externa. Além disso, a análise pelo índice de Youden evidenciou a necessidade de reportar métricas dependentes de limiar juntamente com métricas globais, como a ROC-AUC.

Como trabalhos futuros, propõe-se ampliar a validação externa para outros bancos mamográficos, como MIAS e INbreast, aprofundar a análise de calcificações com maior resolução e estratégias multi-escala, avaliar modelos multivisão e investigar calibração probabilística, interpretabilidade visual e robustez entre sementes. Também se recomenda avançar para abordagens com mamografias completas, aproximando o protocolo de cenários clínicos reais.

REFERÊNCIAS

AKIBA, Takuya; SANO, Shotaro; YANASE, Toshihiko; OHTA, Takeru; KOYAMA, Masanori. Optuna: A next-generation hyperparameter optimization framework. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 2019. p. 2623–2631.

American College of Radiology. **ACR BI-RADS Atlas Fifth Edition: Reference Card**. 2013. Disponível em: https://msrads.web.unc.edu/wp-content/uploads/sites/15695/2018/04/BIRADS-Reference-Card_web_F.pdf. Acesso em: 25 jun. 2026.

AMIN, Ashwini; U, Dinesh Acharya; KOTESHWARA, Prakashini; C, Siddalingaswamy P; MATHEW, Stanley. A systematic literature review on mammography: deep learning techniques for breast cancer detection with global and asian perspectives. **BMC Cancer**, v. 25, p. 1627, 2025.

AYANA, Gelan; DESE, Kokeb; DEREJE, Yisak; KEBEDE, Yonas; BARKI, Hika; AMDISSA, Dechassa; HUSEN, Nahimiya; MULUGETA, Fikadu; HABTAMU, Bontu; CHOE, Se-Woon. Vision-transformer-based transfer learning for mammogram classification. **Diagnostics**, v. 13, n. 2, 2023. ISSN 2075-4418. Disponível em: <https://www.mdpi.com/2075-4418/13/2/178>.

AZAD, Reza; KAZEROUNI, Amirhossein; HEIDARI, Moein; AGHDAM, Ehsan Khodapanah; MOLAEI, Amir; JIA, Yiwei; JOSE, Abin; ROY, Rijo; MERHOF, Dorit. Advances in medical image analysis with vision transformers: A comprehensive review. **Medical image analysis**, v. 91, p. 103000, 2023. Disponível em: <https://api.semanticscholar.org/CorpusID:255545907>.

BARAZI, Hassana; GUNDURU, Mounika. Mammography bi rads grading. In: **StatPearls [Internet]**. [S.l.]: StatPearls Publishing, 2023.

BRAY, Freddie; LAVERSANNE, Mathieu; SUNG, Hyuna; FERLAY, Jacques; SIEGEL, Rebecca L.; SOERJOMATARAM, Isabelle; JEMAL, Ahmedin. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, v. 74, n. 3, p. 229–263, 2024.

DIETTERICH, Thomas G. Ensemble methods in machine learning. **Multiple Classifier Systems**, Springer, p. 1–15, 2000.

DOSOVITSKIY, Alexey; BEYER, Lucas; KOLESNIKOV, Alexander; WEISSENBORN, Dirk; ZHAI, Xiaohua; UNTERTHINER, Thomas; DEHGHANI, Mostafa; MINDERER, Matthias; HEIGOLD, Georg; GELLY, Sylvain; USZKOREIT, Jakob; HOULSBY, Neil. An image is worth 16x16 words: Transformers for image recognition at scale. In: **International Conference on Learning Representations**. [S.l.: s.n.], 2021.

FALCON, William; The PyTorch Lightning Team. **PyTorch Lightning**. 2019. Disponível em: https://github.com/Lightning-AI/lightning. Acesso em: 25 jun. 2026.

FAWCETT, Tom. An introduction to roc analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006.

GERAS, Krzysztof J.; WOLFSON, Stacey; SHEN, Yiqiu; WU, Nan; KIM, S. Gene; KIM, Eric; HEACOCK, Laura; PARIKH, Ujas; MOY, Linda; CHO, Kyunghyun. High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks. **arXiv preprint arXiv:1703.07047**, 2017. Disponível em: <https://arxiv.org/abs/1703.07047>.

HANLEY, James A.; MCNEIL, Barbara J. The meaning and use of the area under a receiver operating characteristic (roc) curve. **Radiology**, v. 143, n. 1, p. 29–36, 1982.

HE, Kelei; GAN, Chen; LI, Zhuoyuan; REKIK, Islem; YIN, Zihao; JI, Wen; GAO, Yang; WANG, Qian; ZHANG, Junfeng; SHEN, Dinggang. Transformers in medical image analysis. **Intelligent Medicine**, v. 3, n. 1, p. 59–78, 2023. ISSN 2667-1026. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2667102622000717>.

HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. 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.

HEATH, Michael; BOWYER, Kevin; KOPANS, Daniel; KEGELMEYER, Philip; MOORE, Richard; CHANG, Kevin; MUNISHKUMARAN, S. The digital database for screening mammography. In: **Proceedings of the 5th International Workshop on Digital Mammography**. [S.l.: s.n.], 2000. p. 212–218.

HUANG, Gao; LIU, Zhuang; MAATEN, Laurens van der; WEINBERGER, Kilian Q. Densely connected convolutional networks. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2017. p. 4700–4708.

International Agency for Research on Cancer. **Breast cancer cases and deaths are projected to rise globally**. fev. 2025. Disponível em: https://www.iarc.who.int/news-events/breast-cancer-cases-and-deaths-are-projected-to-rise-globally/. Acesso em: 25 jun. 2026.

LECUN, Yann; BOTTOU, Léon; BENGIO, Yoshua; HAFFNER, Patrick. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LEE, Rebecca Sawyer; GIMENEZ, Francisco; HOOGI, Assaf; MIYAKE, Kanae Kawai; GOROVOY, Mia; RUBIN, Daniel L. A curated mammography data set for use in computer-aided detection and diagnosis research. **Scientific Data**, Nature Publishing Group, v. 4, p. 170177, 2017.

LITJENS, Geert; KOOI, Thijs; BEJNORDI, Babak Ehteshami; SETIO, Arnaud Arindra Adiyoso; CIOMPI, Francesco; GHAFOORIAN, Mohsen; LAAK, Jeroen A. W. M. van der; GINNEKEN, Bram van; SÁNCHEZ, Clara I. A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60–88, 2017.

LIU, Ze; HU, Han; LIN, Yutong; YAO, Zhuliang; XIE, Zhenda; WEI, Yixuan; NING, Jia; CAO, Yue; ZHANG, Zheng; DONG, Li; WEI, Furu; GUO, Baining. Swin transformer v2: Scaling up capacity and resolution. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2022. p. 12009–12019.

LIU, Ze; LIN, Yutong; CAO, Yue; HU, Han; WEI, Yixuan; ZHANG, Zheng; LIN, Stephen; GUO, Baining. 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.

LIU, Zhuang; MAO, Hanzi; WU, Chao-Yuan; FEICHTENHOFER, Christoph; DARRELL, Trevor; XIE, Saining. 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.

LOSHCHILOV, Ilya; HUTTER, Frank. Decoupled weight decay regularization. In: **International Conference on Learning Representations**. [S.l.: s.n.], 2019.

MANIGRASSO, Francesco; MILAZZO, Rosario; RUSSO, Alessandro Sebastian; LAMBERTI, Fabrizio; STRAND, Fredrik; PAGNANI, Andrea; MORRA, Lia. Mammography classification with multi-view deep learning techniques: Investigating graph and transformer-based architectures. **Medical Image Analysis**, v. 99, p. 103320, 2025.

MANSOUR, Abdullah G. M. Al; ALSHOMRANI, Faisal; ALFAHAID, Abdullah; ALMUTAIRI, Abdulaziz T. M. Mammovit: A custom vision transformer architecture for accurate birads classification in mammogram analysis. **Diagnostics**, v. 15, n. 3, p. 285, 2025.

MONTORO, Clara Isabel; ALCARAZ, María del Carmen; GÁLVEZ-SÁNCHEZ, Carmen M. Experience of pain and unpleasantness during mammography screening: a cross-sectional study on the roles of emotional, cognitive, and personality factors. **Behavioral Sciences**, v. 13, n. 5, p. 377, 2023.

MOTA, Bruna Salani; SHIMIZU, Carlos; REIS, Yedda N.; GONçALVES, Rodrigo; Soares Junior, Jose Maria; BARACAT, Edmund Chada; FILASSI, José Roberto. Dense breasts and women's health: which screenings are essential? **Clinics**, v. 80, p. 100743, 2025. ISSN 1807-5932. Disponível em: <https://www.sciencedirect.com/science/article/pii-/S1807593225001620>.

NGUYEN, Hieu T.; NGUYEN, Ha Q.; PHAM, Hieu H.; LAM, Khanh; LE, Linh T.; DAO, Minh; VU, Van. Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. **medRxiv**, 2022. Disponível em: <https://www.medrxiv.org/content/early/2022/03/10/2022.03.07.22272009>.

PASZKE, Adam; GROSS, Sam; MASSA, Francisco; LERER, Adam; BRADBURY, James; CHANAN, Gregory; KILLEEN, Trevor; LIN, Zeming; GIMELSHEIN, Natalia; ANTIGA, Luca *et al.* Pytorch: An imperative style, high-performance deep learning library. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2019. v. 32.

QURESHI, Saad Ahmed; TARIQ, Abdul Samad; KHAN, M. T.; RAZA, A. Deep learning-based breast cancer detection and classification using microcalcifications in mammograms: A systematic review. **Healthcare**, v. 13, n. 5, p. 486, 2025.

RICHARDSON, Eve; TREVIZANI, Raphael; GREENBAUM, Jason A; CARTER, Hannah; NIELSEN, Morten; PETERS, Bjoern. The receiver operating characteristic curve accurately assesses imbalanced datasets. **Patterns (N. Y.)**, Elsevier BV, v. 5, n. 6, p. 100994, jun. 2024.

ROBERTS, Michael; DRIGGS, Derek; THORPE, Matthew; GILBEY, James; YEUNG, Michael; URSPRUNG, Stephen; AVILES-RIVERO, Angelica I.; ETMANN, Christian; MCCAGUE, Cathal; BEER, Lucian; WEIR-MCCALL, Jonathan R.; TENG, Zhongzhao; GKRANIA-KLOTSAS, Effrossyni; RUDD, James H. F.; SALA, Evis; SCHöNLIEB, Carola-Bibiane. Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. **Nature Machine Intelligence**, v. 3, p. 199–217, 2021.

SAITO, Takaya; REHMSMEIER, Marc. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. **PLOS ONE**, v. 10, n. 3, p. e0118432, 2015.

SARKER, Sushmita; SARKER, Prithul; BEBIS, George; TAVAKKOLI, Alireza. **MV-Swin-T: Mammogram Classification with Multi-view Swin Transformer**. 2024. Disponível em: https://arxiv.org/abs/2402.16298. Acesso em: 25 jun. 2026.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. **Journal of Big Data**, v. 6, p. 60, 2019.

SPAK, David A.; PLAXCO, Julia S.; SANTIAGO, Lourdes; DRYDEN, Megan J.; DOGAN, Basak E. BI-RADS Fifth Edition: A Summary of Changes. **Diagnostic and Interventional Imaging**, v. 98, n. 3, p. 179–190, 2017.

SZEGEDY, Christian; VANHOUCKE, Vincent; IOFFE, Sergey; SHLENS, Jonathon; WOJNA, Zbigniew. Rethinking the inception architecture for computer vision. **CoRR**, abs/1512.00567, 2015. Disponível em: <http://arxiv.org/abs/1512.00567>.

TAKAHASHI, Satoshi; SAKAGUCHI, Yusuke; KOUNO, Nobuji; TAKASAWA, Ken; ISHIZU, Kenichi; AKAGI, Yu; AOYAMA, Rina; TERAYA, Naoki; SHINKAI, Norio; MACHINO, Hidenori; KOBAYASHI, Kazuma; ASADA, Ken; KOMATSU, Masaaki; KANEKO, Syuzo; SUGIYAMA, Masashi; HAMAMOTO, Ryuji. Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. **Journal of Medical Systems**, v. 48, p. 84, 09 2024.

TOUVRON, Hugo; CORD, Matthieu; DOUZE, Matthijs; MASSA, Francisco; SABLAYROLLES, Alexandre; JÉGOU, Hervé. Training data-efficient image transformers and distillation through attention. In: **Proceedings of the 38th International Conference on Machine Learning**. [S.l.: s.n.], 2021. v. 139, p. 10347–10357.

VARMA, Sudhir; SIMON, Richard. Bias in error estimation when using cross-validation for model selection. **BMC Bioinformatics**, v. 7, p. 91, 2006.

VAROQUAUX, Gaël; CHEPLYGINA, Veronika. Machine learning for medical imaging: methodological failures and recommendations for the future. **npj Digital Medicine**, Nature Publishing Group, v. 5, n. 1, p. 48, 2022.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia. Attention is all you need. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2017. v. 30.

WANG, Lulu. Mammography with deep learning for breast cancer detection. **Frontiers in Oncology**, v. 14, 2024.

WHELEHAN, Patsy; EVANS, Andy; WELLS, Mary; MACGILLIVRAY, Steve. The effect of mammography pain on repeat participation in breast cancer screening: a systematic review. **The Breast**, v. 22, n. 4, p. 389–394, 2013.

WOO, Sanghyun; DEBNATH, Shoubhik; HU, Ronghang; CHEN, Xinlei; LIU, Zhuang; KWEON, In So; XIE, Saining. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2023. p. 16133–16142.

World Health Organization. **Breast cancer**. abr. 2026. Disponível em: https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Acesso em: 25 jun. 2026.

YOUDEN, William J. Index for rating diagnostic tests. **Cancer**, v. 3, n. 1, p. 32–35, 1950.

ZAHARIA, Matei; CHEN, Andrew; DAVIDSON, Aaron; GHODSI, Ali; HONG, Sue Ann; KONWINSKI, Andy; MURCHING, Siddharth; NYKODYM, Tomas; OGILVIE, Paul; PARKHE, Mani; XIE, Fen. Accelerating the machine learning lifecycle with mlflow. In: **IEEE Data Engineering Bulletin**. [S.l.: s.n.], 2018. v. 41, n. 4, p. 39–45.

MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA - NG

ATA Nº 17/2026 - DEE (11.56.08)

Nº do Protocolo: NÃO PROTOCOLADO

Belo Horizonte-MG, 16 de junho de 2026.

ATA DE DEFESA DO PROJETO FINAL DE CURSO

Aos 16 dias do mês de junho do ano de dois mil e vinte e seis, realizou-se a sessão pública de defesa do projeto final de curso intitulado "***Análise Comparativa entre Transformers Visuais e CNNs na Classificação de Calcificações e Massas Mamárias"***, apresentado por **Victor Daniel Oliviera Gate**, matrícula 20193020944 do **Curso de Engenharia Elétrica - CNG**. Os trabalhos foram iniciados às 14h10 pelo Professor (Orientador) Dr. Alexandre Rodrigues Farias, do Departamento de Eletrônica e Biomédica - DEEB, presidente da banca examinadora, constituída pelos seguintes membros do Departamento de Engenharia Elétrica campus Nova Gameleira: Prof. Dr. Túlio Charles de Oliveira Carvalho (Coorientador), Profa. Dra. Rosilene Nietzsch Dias e do Prof. Dr. José Hissa Ferreira. Após a exposição oral, o candidato foi arguido pelos componentes da banca que decidiram por **aprovar** o trabalho. Para constar, redigiu-se a presente ata que, após aprovada por todos os presentes, será assinada pelos membros da banca e pelo discente.

Belo Horizonte, 16 de junho de 2026

(Assinado digitalmente em 17/06/2026 10:31)
ALEXANDRE RODRIGUES FARIAS
PROFESSOR ENS BASICO TECN TECNOLOGICO
DEEB (11.56.12)
Matrícula: ###307#2

(Assinado digitalmente em 16/06/2026 22:57)
JOSE HISSA FERREIRA
PROFESSOR DO MAGISTERIO SUPERIOR
DEE (11.56.08)
Matrícula: ####49#2

(Assinado digitalmente em 16/06/2026 19:36)
ROSILENE NIETZSCH DIAS
PROFESSOR ENS BASICO TECN TECNOLOGICO
DEE (11.56.08)
Matrícula: ###044#9

(Assinado digitalmente em 22/06/2026 16:16)
TULIO CHARLES DE OLIVEIRA CARVALHO
PROFESSOR ENS BASICO TECN TECNOLOGICO
DEE (11.56.08)
Matrícula: ###653#3

(Assinado digitalmente em 23/06/2026 13:42)
VICTOR DANIEL OLIVEIRA GAETE
DISCENTE
Matrícula: 2019######4

Processo Associado: 23062.027369/2026-88