

Detecção de Tumores Mamários por CNNs em Termografias

Arthur Santos de Oliveira
Disciplina de Inteligência Artificial
Universidade Federal de São Paulo
São José dos Campos, Brasil
oliveira.arthur@unifesp.com

Resumo—A detecção precoce de doenças mamárias, como o câncer de mama, é crucial para o sucesso do tratamento e a melhoria das taxas de sobrevivência. Este trabalho propõe o desenvolvimento de um sistema automatizado para a detecção de tumores mamários em imagens de termografia, utilizando as arquiteturas de redes neurais convolucionais ResNet50, EfficientNet e VGG19. Técnicas de aumento de dados e transferência de aprendizado foram aplicadas para melhorar o desempenho dos modelos. Além disso, uma análise comparativa dos resultados dessas três redes foi realizada, empregando diferentes tipos de pré-processamento e validação cruzada. A técnica CAM foi utilizada para gerar mapas de ativação que destacam as áreas das imagens mais relevantes para as decisões do modelo. Os resultados incluem gráficos de acurácia com desvio padrão, curvas de aprendizagem e matrizes de confusão para avaliar o desempenho dos modelos.

Palavras-chave—Detecção de doenças mamárias, Redes neurais convolucionais, ResNet50, Efficientnet, Vgg19, Data augmentation, Mixup, PyTorch, CAM, validação cruzada.

I. INTRODUÇÃO E MOTIVAÇÃO

A. Contextualização do Problema

A detecção precoce de doenças mamárias, como o câncer de mama, é crucial para o sucesso do tratamento e a melhoria das taxas de sobrevivência. Tradicionalmente, a detecção dessas doenças é realizada por meio de mamografias analisadas por especialistas médicos. No entanto, a análise manual pode ser demorada e sujeita a erros humanos.

A termografia infravermelha médica é uma tecnologia promissora para o diagnóstico precoce de anormalidades mamárias, devido às suas vantagens como ser livre de radiação, não invasiva e indolor. As câmeras térmicas usadas são altamente sensíveis e de alta resolução, capazes de detectar pequenas variações de temperatura na pele que podem indicar a presença de tumores [1]. Um termograma mamário saudável apresenta um padrão de temperatura simétrico entre ambas as mamas; qualquer assimetria pode indicar uma anormalidade [1].

Além disso, a termografia é um exame emergente no diagnóstico do câncer de mama, que obtém uma imagem infravermelha representando a temperatura da superfície da pele [2]. Esse método requer equipamentos mais baratos e pode detectar tumores milimétricos devido à natureza bioquímica dos tumores [3].

Portanto, há uma crescente necessidade de desenvolver métodos automatizados que possam auxiliar na detecção precoce e precisa dessas doenças.

B. Motivação

Estudar e implementar técnicas de aprendizado profundo para a detecção de doenças mamárias é interessante porque:

- **Impacto Social:** Pode salvar vidas ao permitir a detecção precoce e precisa de doenças graves, melhorando significativamente as taxas de sobrevivência. A detecção eficaz do câncer de mama, em particular, pode reduzir a mortalidade e proporcionar tratamentos menos invasivos.
- **Desafios Técnicos:** Envolve a aplicação de redes neurais convolucionais (CNNs), um dos tópicos mais avançados e promissores em inteligência artificial. As CNNs são especialmente eficazes em tarefas de classificação e análise de imagens, tornando-as ideais para a detecção de tumores em termogramas mamários. Além disso, o trabalho aborda desafios específicos como o pré-processamento de imagens usando técnicas como corte (crop) e filtro laplaciano para verificar como essas operações impactam na acurácia e no mapa de ativação de classe, observando assim quais características impactam para se classificar doenças mamárias.
- **Inovação:** Proporciona a oportunidade de contribuir para a pesquisa médica com novas soluções tecnológicas, como a integração de termografia e aprendizado profundo para a detecção precoce do câncer de mama. Este trabalho também explora o uso de técnicas avançadas de data augmentation, como Mixup e random flip horizontal, que são essenciais para aumentar a diversidade do conjunto de dados e melhorar a generalização dos modelos.

II. CONCEITOS FUNDAMENTAIS

A. Termografia Digital Infravermelha (DITI)

A termografia digital infravermelha é uma técnica de imagem que detecta variações de temperatura na superfície das mamas. Embora não seja aprovada pelo FDA como uma ferramenta de triagem independente, essa técnica oferece uma abordagem inovadora para a detecção de anomalias no tecido mamário, especialmente quando usada em conjunto com métodos tradicionais, como a mamografia.

A premissa básica da termografia é que, à medida que as células cancerosas se multiplicam, elas necessitam de um suprimento aumentado de sangue rico em oxigênio para sustentar seu crescimento. Esse aumento no fluxo sanguíneo em torno de um tumor gera elevações de temperatura na área afetada, permitindo que a termografia identifique essas variações térmicas. Em um experimento [9], Folkman observou a dependência do crescimento de um tumor na angiogênese ao transplantar células tumorais em camundongos, constatando que uma grande quantidade de sangue flui para os vasos conectados ao câncer, tornando-os mais quentes e provocando sua expansão e alongamento. Gautherie [10], por sua vez, afirmou que essa elevação de temperatura ocorre devido ao aumento da atividade metabólica nas células tumorais, o que faz com que a área apareça mais brilhante e quente nas imagens térmicas.

A termografia apresenta como um dos seus principais benefícios o fato de não emitir radiação, ao contrário da mamografia, que utiliza raios-X de baixa dose para gerar imagens internas das mamas. Isso torna a termografia uma alternativa atraente para pacientes que desejam evitar a exposição à radiação. No entanto, é fundamental enfatizar que a termografia não substitui a mamografia na detecção do câncer de mama. Estudos [11] indicam que as câmeras infravermelhas modernas possuem uma sensibilidade tão alta que são capazes de detectar diferenças de temperatura de até 0,025 °C, o que permite gerar gráficos de calor extremamente precisos da distribuição de temperatura sobre a superfície da pele.

Os dispositivos de termografia foram aprovados pelo FDA apenas como uma ferramenta "adjuvante", o que significa que devem ser utilizados em conjunto com um teste de triagem primário, como a mamografia. Pacientes que realizam o teste de termografia isoladamente não devem se sentir seguros com os resultados, pois essa técnica não foi validada para uso sem um método complementar de diagnóstico.

Portanto, embora a termografia possa ser uma ferramenta útil para complementar a triagem do câncer de mama, ela não deve substituir os métodos tradicionais, como a mamografia, que ainda são considerados o padrão-ouro na detecção precoce da doença. A teoria por trás da termografia mamária pressupõe que o tecido mamário está livre de processos anormais que emitem padrões de calor previsíveis na superfície da pele. Quando processos fisiológicos, como distúrbios vasculares ou inflamação, alteram esse padrão, a termografia pode capturar essas mudanças, fornecendo informações valiosas sobre o estado fisiológico funcional da área corporal examinada.

B. Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais (CNNs) são modelos especializados em processar dados estruturados em grades, como imagens. Elas são compostas por camadas convolucionais que aplicam filtros para extrair características locais das imagens, capturando padrões como bordas, texturas e formas específicas. Cada camada convolucional gera mapas de características que são combinados em camadas subsequentes para aprender representações cada vez mais abstratas das imagens.

As CNNs revolucionaram a visão computacional ao possibilitar o aprendizado de características diretamente dos dados, eliminando a necessidade de extração manual de características. Este paradigma permitiu avanços significativos em aplicações como reconhecimento de objetos, diagnóstico médico e análise de imagens. Foram utilizados três modelos de rede neural: ResNet50, EfficientNet, e VGG19:

ResNet50 é uma rede neural convolucional profunda com 50 camadas, conhecida por sua arquitetura de "residual connections" ou conexões residuais. Essas conexões ajudam a mitigar o problema do gradiente desaparecido em redes muito profundas, permitindo que a ResNet50 atinja alto desempenho em tarefas de classificação de imagens.

EfficientNet é uma arquitetura otimizada que equilibra de forma eficiente a profundidade, largura e resolução da rede. Ela utiliza um método de escalonamento que ajusta essas três dimensões simultaneamente, o que resulta em um modelo altamente eficiente e preciso, mesmo com menor número de parâmetros comparado a outras redes profundas.

VGG19 é uma rede profunda com 19 camadas, caracterizada por sua simplicidade e uso de filtros de tamanho 3x3 em todas as camadas convolucionais. Embora tenha mais parâmetros devido à sua profundidade, o VGG19 é eficaz em capturar características de alto nível em imagens, sendo amplamente utilizado em tarefas de classificação e extração de características.

C. Mapeamento de Ativação de Classe (CAM)

Modelos de deep learning frequentemente superam os métodos tradicionais de machine learning em tarefas complexas, especialmente em visão computacional. No entanto, uma limitação significativa dos modelos de deep learning é a dificuldade de interpretá-los. As previsões desses modelos são muitas vezes vistas como caixas-pretas, o que significa que, embora possamos ver as entradas e as saídas, entender o processo de tomada de decisão por trás dessas previsões pode ser extremamente desafiador devido às interações complexas entre as várias camadas de uma rede neural convolucional (CNN).

A interpretabilidade é crucial para garantir a confiabilidade dos modelos de machine learning. Em particular, nas CNNs, que são amplamente utilizadas para tarefas como classificação de imagens, segmentação, detecção de objetos e estimativa de pose, é essencial entender se o modelo está focando nas regiões corretas de uma imagem ao fazer previsões.

Para abordar esse desafio de interpretabilidade, foram propostos vários métodos, sendo um dos mais antigos e amplamente utilizados o Mapeamento de Ativação de Classe (CAM). O CAM permite visualizar as regiões de uma imagem que influenciam as decisões do modelo, ajudando os pesquisadores a compreender melhor como o modelo toma suas decisões e, assim, aumentar sua confiabilidade. Ao aplicar o CAM, podemos verificar se o modelo está realmente utilizando as características corretas da imagem para fazer suas previsões, garantindo uma maior transparência e confiança nas decisões do modelo.

Neste trabalho, o CAM foi utilizado para compreender como a rede neural decide se uma imagem de termografia mamária deve ser classificada como doente ou saudável. Com isso, foi possível destacar as regiões que a rede considera importantes para a classificação, além de identificar e ocultar áreas que a rede leva em conta, mas que não deveriam influenciar a decisão.

D. Data Augmentation

Data augmentation é uma técnica crucial para aumentar a diversidade e a quantidade dos dados de treinamento, melhorando a generalização e robustez dos modelos de CNN. Ela envolve a aplicação de transformações aleatórias nas imagens de treinamento, como rotação, escala, espelhamento, alterações de brilho e contraste. Essas transformações aumentam a variabilidade dos dados sem a necessidade de coletar novas amostras, mitigando problemas de overfitting e melhorando a capacidade do modelo de generalizar para novos dados. Nesse trabalho foi utilizado um giro (flip) aleatório nas imagens do dataset.

Além dessas técnicas mais simples, existe o Mixup que é uma técnica de aumento de dados que cria novas amostras ao combinar dois pontos de dados existentes. A ideia principal é sintetizar novos exemplos a partir de combinações lineares de pares de amostras de entrada e seus rótulos correspondentes. A fórmula básica, conforme descrita no artigo original, é:

$$\hat{x} = \lambda x_i + (1 - \lambda)x_j \quad (1)$$

$$\hat{y} = \lambda y_i + (1 - \lambda)y_j \quad (2)$$

Em (1), \hat{x} representa a nova amostra gerada a partir de duas amostras originais x_i e x_j , enquanto em (2) \hat{y} é o rótulo correspondente, também gerado como uma combinação dos rótulos originais y_i e y_j . O parâmetro λ é um valor que controla o peso de cada amostra na combinação, normalmente amostrado de uma distribuição beta. Mixup tem se mostrado particularmente eficaz no treinamento de redes neurais, especialmente quando a distribuição dos dados de treinamento é linear. Ao incorporar Mixup no processo de treinamento, as lacunas entre os pontos de dados originais no espaço de características são preenchidas, o que permite que o modelo aprenda de uma maneira mais robusta e tenha maiores chances de fazer previsões precisas. Na visão computacional, essa técnica é frequentemente referida como "mesclagem de imagens", e compartilha o mesmo conceito.

No presente trabalho, a técnica Mixup foi utilizada porque o conjunto de dados disponível para o treinamento dos modelos era relativamente pequeno. Ao gerar novas amostras sintéticas, Mixup teve por objetivo ajudar a aumentar a diversidade dos dados, tentando permitir que os modelos de deep learning treinados—ResNet50, EfficientNet e VGG19—se tornassem mais generalizáveis e menos propensos ao overfitting.

III. TRABALHOS RELACIONADOS

A. Diagnóstico de câncer de mama usando termografia e redes neurais convolucionais

O estudo de Ekici e Jawzal [4] se concentra no uso de termografia e CNNs para melhorar a detecção precoce do câncer de mama. Eles destacam a termografia como uma técnica de imagem que, embora não seja ainda o padrão ouro, possui um risco baixo e pode ser aprimorada com avanços tecnológicos. A proposta do trabalho é utilizar CNNs para processar imagens termográficas, superando as limitações de precisão e validade que historicamente têm sido obstáculos para a adoção ampla da termografia.

O objetivo do estudo foi revisar a literatura relacionada à imagem térmica e à triagem termográfica do câncer de mama, identificando lacunas na pesquisa existente. Além de desenvolver um analisador termográfico de alta performance usando CNNs no ambiente MATLAB e testar o sistema desenvolvido com imagens reais para validar a eficácia do trabalho. Por fim, comparar o desempenho do sistema desenvolvido com sistemas anteriores encontrados na literatura. Este trabalho é um exemplo de esforços para aprimorar a detecção de câncer de mama através de métodos menos invasivos, contribuindo para a área de diagnóstico precoce do câncer de mama.

Os resultados obtidos por outros pesquisadores nesta área são promissores e mostram a eficácia das CNNs na detecção de câncer de mama através da termografia:

- **[5]:** Utilizou seleção de características e redes Bayesianas, analisando imagens de 80 pacientes, com 20 imagens para cada um (40 saudáveis e 40 não saudáveis), alcançando uma precisão impressionante de 100%.
- **[6]:** Focou em características extraídas por CNNs, trabalhando com um total de 137 pacientes e 20 imagens por paciente (95 saudáveis e 95 não saudáveis), e obteve uma precisão de 95%.
- **[4]:** Utilizou características de CNNs otimizadas por algoritmo Bayesiano, analisando um total de 140 pacientes, sendo 98 saudáveis e 32 não saudáveis. O sistema desenvolvido alcançou uma precisão de 98,95%.

B. Detecção de câncer de mama a partir da termografia baseada em redes neurais profundas

O uso de redes neurais profundas para a detecção de câncer de mama a partir de termografia é uma área ativa de pesquisa. Um outro estudo importante nesta área é o de Seyfullah Kiyimet, Muhammet Yavuz Aslankaya, Murat Taskiran e Bulent Bolat, apresentado na conferência IEEE em Izmir, Turquia, de 31 de outubro a 2 de novembro de 2019 [7].

Este estudo aborda a detecção de câncer de mama utilizando termografia e redes neurais profundas. A pesquisa destaca as limitações dos métodos tradicionais de imagem, como a mamografia, que podem ser invasivos e menos eficazes em mulheres jovens com tecido mamário denso. A termografia é apresentada como uma alternativa promissora devido à sua natureza não invasiva e à capacidade de detectar alterações

térmicas causadas por atividades químicas e vasculares anormais associadas ao câncer.

Os objetivos deste estudo incluem explorar o conjunto de dados DMR (Database For Mastology Research), que contém 4446 imagens de 287 voluntários obtidas com a câmera termográfica FLIR SC-620, incluindo mapas de cores e matrizes térmicas. Também envolve revisar a literatura existente, analisando estudos anteriores que utilizaram diferentes algoritmos de extração de características e classificadores, como KNN, SVM e técnicas de segmentação baseadas em K-means. Outro objetivo é desenvolver e implementar uma Rede Neural Convolucional (CNN) para a extração de características convolucionais e a detecção de câncer de mama a partir de imagens térmicas, abordando uma lacuna observada na literatura.

Além disso, o estudo visa descrever as etapas metodológicas, incluindo a aquisição de dados, o pré-processamento de imagens e o treinamento e validação do modelo de CNN. Como resultado de estudos experimentais, a rede ResNet50 alcançou o maior desempenho de teste na detecção de câncer de mama em 88,89%.

C. Redes neurais convolucionais profundas para classificação do câncer de mama usando termografia infravermelha

Um importante estudo [8] utilizou uma rede neural convolucional profunda, especificamente a ResNet-101 com aprendizado por transferência, para analisar termogramas mamários. Este método capitaliza as vantagens da termografia, como sua natureza não invasiva e confortável, enquanto enfrenta desafios como a falta de procedimentos padronizados de interpretação. Os dados foram coletados de múltiplos centros, incluindo um banco de dados público e termogramas de instituições mexicanas, utilizando o Protocolo Glamorgan para aquisição de imagens. Após a aplicação de critérios de exclusão, o estudo finalizou com 311 termogramas, rotulados utilizando o sistema BI-RADS para categorizar lesões como normais ou anormais.

O modelo de aprendizado profundo foi treinado e validado em dois conjuntos de dados distintos: um desbalanceado, refletindo uma coorte típica de triagem com maior número de casos normais, e outro balanceado, equilibrando os casos normais e anormais. Técnicas de aumento de dados foram aplicadas para melhorar a generalização do modelo, incluindo transformações geométricas aleatórias nos termogramas. Os resultados experimentais mostraram que a CNN ResNet-101 alcançou sensibilidade de 92,3% no conjunto de dados balanceado, e sensibilidade de 84,6% no conjunto desbalanceado.

IV. OBJETIVO

O presente trabalho tem como objetivo principal a comparação dos modelos ResNet50, EfficientNet e VGG19 na classificação de doenças mamárias utilizando imagens de termografia, focando na distinção entre mamas saudáveis e doentes. Para isso, será realizada uma análise aprofundada do impacto de diferentes técnicas de pré-processamento nas imagens, com o intuito de compreender como essas abordagens influenciam a acurácia dos modelos e a curva de aprendizagem

durante o treinamento. Além disso, o estudo buscará avaliar a eficácia da técnica de data augmentation Mixup na melhoria da classificação das imagens termográficas, investigando seu efeito na robustez e generalização dos modelos. Para complementar a análise, será utilizada a técnica de Class Activation Mapping (CAM) para identificar quais características das imagens estão sendo consideradas pelas redes neurais em suas classificações. Essa abordagem permitirá um entendimento mais claro sobre os fatores que influenciam o desempenho do modelo e proporcionará insights para aprimorar suas capacidades. Esses objetivos visam contribuir para a detecção do câncer de mama por meio da combinação de técnicas de aprendizado de máquina e imagem térmica, promovendo um entendimento mais profundo sobre o funcionamento dos modelos utilizados e suas aplicações na área da saúde.

V. METODOLOGIA EXPERIMENTAL

A. Descrição

Nesta seção, apresentamos uma visão geral da metodologia experimental adotada no nosso estudo, que envolve um rigoroso processo de pré-processamento e validação dos dados. O conjunto de dados será submetido a seis diferentes etapas de pré-processamento, com o objetivo de verificar como cada modelo se comporta com distintas abordagens de preparação dos dados.

Cada uma dessas etapas de pré-processamento será validada por meio de uma validação cruzada de 10 folds. Em cada fold, os modelos ResNet50, EfficientNet e VGG19 serão treinados por 10 épocas, permitindo uma avaliação abrangente do desempenho de cada abordagem. As métricas coletadas durante esse processo incluem a curva de aprendizagem, gráficos de acurácia, matriz de confusão e Class Activation Mapping (CAM).

A curva de aprendizagem foi calculada a partir da média dos valores de acurácia e perda de cada época nos folds. Em cada época i , somamos os 10 valores de acurácia e perda, e ao final, esses valores são divididos para fornecer a curva de aprendizagem baseada na média dos folds.

O gráfico de acurácia foi gerado utilizando as médias das validações de cada fold, proporcionando uma visualização clara do desempenho dos modelos ao longo do treinamento. Para a matriz de confusão, os valores de cada fold foram somados, uma vez que cada imagem foi utilizada para treinamento apenas uma vez durante o processo de validação cruzada.

Por fim, o modelo do último fold foi salvo para realizar inferências, permitindo a geração do CAM para 50 imagens aleatórias do conjunto de dados. Essa abordagem nos proporciona uma compreensão mais profunda das decisões dos modelos e facilita a interpretação dos resultados obtidos.

Todos esse processo é demonstrado a partir da "Fig. 1".

B. Domínio da aplicação

Neste estudo, utilizamos a base de dados DMR-IR (Database for Mastology Research - Infrared), que contém imagens termográficas da região das mamas de mulheres. As imagens foram coletadas pelo Laboratório Visual para Pesquisa em

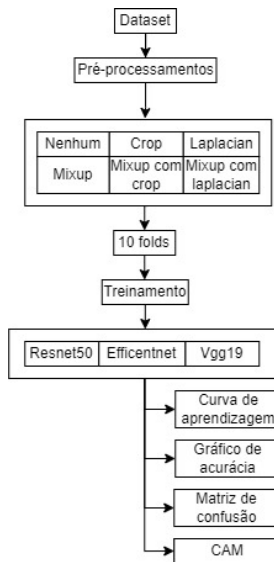


Figura 1. Diagrama do processo metodológico.

Mastologia da Universidade Federal Fluminense (UFF), localizado em Niterói, Brasil. As capturas foram realizadas com uma câmera térmica FLIR SC620, gerando imagens com resolução de 480 x 640 pixels.

A DMR-IR inclui três classes distintas de imagens: saudáveis (healthy), doentes (sick), e desconhecidas (unknown). Para o propósito deste projeto, focamos exclusivamente nas classes "saudável" e "doente", descartando a classe "desconhecida". Esta última representa casos sem diagnóstico ou com classificação incerta, o que poderia introduzir ruído nos dados de treinamento e dificultar tanto o processo de aprendizagem do modelo quanto a interpretação dos resultados.

A base de dados utilizada contém um total de 200 imagens, sendo 100 classificadas como saudáveis e 100 como doentes. As imagens foram convertidas para escala de cinza, uma vez que a análise termográfica tem como foco principal a variação de temperatura, que é mais claramente representada em imagens em tons de cinza. A remoção de cores facilita a interpretação dos dados térmicos, eliminando complexidades adicionais que poderiam surgir com o uso de imagens coloridas.

Além disso, optamos por concentrar o estudo em imagens capturadas de ângulo frontal. Essa escolha se baseia na premissa de que, para análise térmica, as imagens em escala de cinza são suficientes para capturar variações de temperatura de maneira eficaz, impactando diretamente a classificação entre "saudável" e "doente".

A classificação correta entre essas duas categorias é de extrema importância para o treinamento do modelo, pois permitirá que ele aprenda a distinguir padrões térmicos associados à presença ou ausência de anomalias que possam indicar câncer de mama. Este trabalho visa contribuir para o avanço de métodos de diagnóstico menos invasivos e mais acessíveis, usando redes neurais convolucionais para análise termográfica.

C. Pré-processamento

O pré-processamento dos dados é uma etapa crucial para garantir que as imagens estejam em um formato adequado para o treinamento das redes neurais convolucionais. Neste estudo, adotamos uma série de transformações baseadas nos parâmetros otimizados para cada uma das redes utilizadas: ResNet50, EfficientNet e VGG19.

As imagens foram redimensionadas (resize) para 224 x 224 pixels, um tamanho que foi recomendado pela documentação do Pytorch. O redimensionamento (resize) ajusta a resolução das imagens para garantir consistência nos dados de entrada e compatibilidade com a arquitetura das redes, que geralmente requerem um tamanho de entrada fixo.

Além disso, as imagens passaram por uma normalização usando os valores [0.485, 0.456, 0.406] para o desvio padrão e [0.229, 0.224, 0.225] para a média. Esse processo de normalização é fundamental para ajustar a distribuição de pixel values (valores de pixels) das imagens, garantindo que elas estejam dentro de um intervalo que facilita o treinamento das redes neurais. A normalização ajuda a estabilizar e acelerar o processo de aprendizagem, evitando problemas como gradientes explodidos ou desaparecidos, comuns em redes profundas.

As imagens foram, então, embaralhadas e armazenadas em uma classe que herda do Dataset do PyTorch. O PyTorch, uma das principais bibliotecas de deep learning, oferece suporte a operações de pré-processamento de dados, como o RandomFlip, que foi utilizado para aplicar uma transformação aleatória de flip horizontal nas imagens. Esta técnica aumenta a variabilidade dos dados de treinamento, contribuindo para a robustez do modelo ao evitar que ele fique excessivamente adaptado (overfitting) a padrões específicos das imagens.

Para cada um dos modelos (ResNet50, EfficientNet e VGG19), foram avaliados seis diferentes cenários de pré-processamento:

- **Sem pré-processamento:** As imagens são utilizadas em seu estado original após o redimensionamento e normalização.
- **Crop na região inferior das axilas:** Esta técnica consiste em remover a parte inferior das imagens, onde as axilas estão localizadas. A decisão de aplicar este crop foi baseada em análises com Class Activation Mapping (CAM), que mostraram que as axilas, por serem áreas de alta temperatura, estavam influenciando indevidamente as previsões da rede. Na "Fig. 2" é possível ver um exemplo dessa operação. Com o crop a imagem da termografia ficou conforme a "Fig. 3".
- **Filtro Laplaciano:** O filtro Laplaciano foi aplicado para destacar as mudanças bruscas de temperatura nas imagens. Este filtro realça bordas e contornos, potencialmente melhorando a detecção de regiões de alta concentração de temperatura, que podem indicar a presença de doenças. Essa configuração pode ser visualizada na "Fig. 4".
- **Mixup:** O mixup é uma técnica de data augmentation que combina aleatoriamente pares de imagens e suas respectivas labels. Esta abordagem ajuda a generalizar o

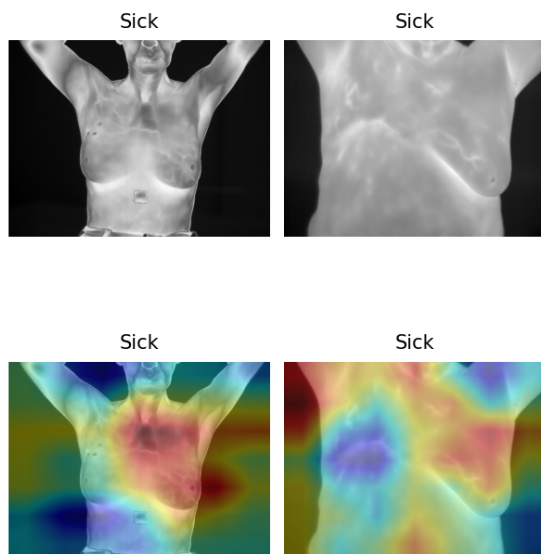


Figura 2. Exemplo de uso do CAM.

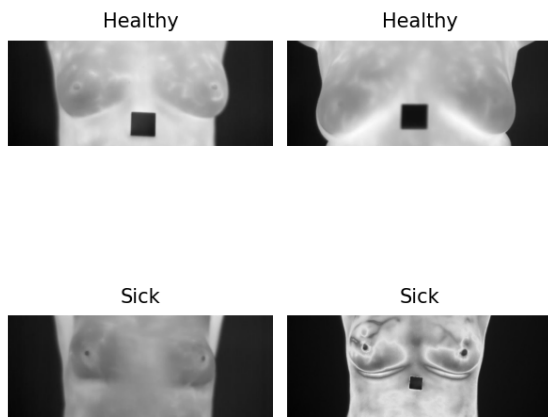


Figura 3. Crop das axilas.

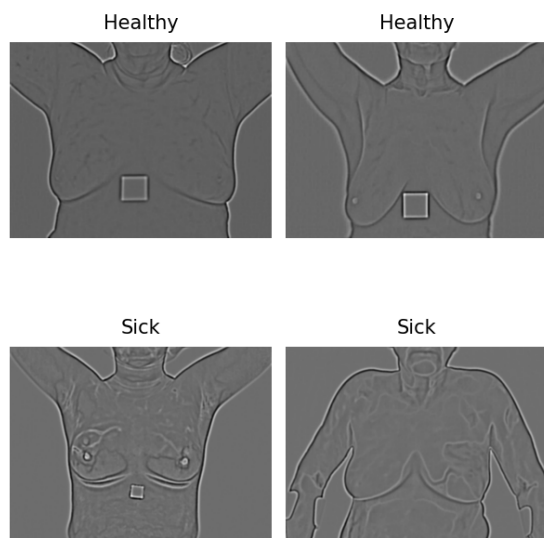


Figura 4. Termografia com filtro laplaciano.

modelo, reduzindo o risco de overfitting, especialmente importante quando se trabalha com um conjunto de dados limitado.

- **Crop na região inferior das axilas com mixup:** Combina o corte das axilas com a técnica de mixup, tentando eliminar o ruído enquanto generaliza melhor o modelo.
- **Filtro Laplaciano com mixup:** Integra o realce das bordas e contornos proporcionado pelo filtro Laplaciano com a robustez adicional proporcionada pelo mixup.

O uso do mixup é particularmente interessante para este projeto devido ao número limitado de imagens disponíveis, o que pode levar ao overfitting, então seu uso pode se apresentar eficaz.

Essas diferentes abordagens de pré-processamento foram testadas para avaliar o impacto de cada uma na performance dos modelos, com o objetivo de identificar a combinação mais eficaz para a detecção precisa de anomalias nas imagens termográficas.

D. Passos e Técnicas

Conforme dito este estudo tem como objetivo classificar imagens de termografia mamária em duas categorias distintas: doentes e saudáveis, utilizando abordagens avançadas de aprendizado profundo. Para essa tarefa, selecionamos três modelos de rede neural de alto desempenho: ResNet50, VGG19 e EfficientNet, conhecidos por sua eficácia em tarefas de classificação de imagens.

Para adaptar essas arquiteturas à classificação de imagens de termografia mamária, realizamos um processo metódico que envolveu a inicialização dos modelos com pesos pré-treinados. Essa abordagem permite que as redes capturem características visuais relevantes, acelerando o processo de treinamento e melhorando o desempenho. Ajustamos as camadas de saída dos modelos para corresponder ao número de classes da tarefa. Especificamente, substituímos a última camada por uma nova camada totalmente conectada configurada para duas classes: 'saudável' e 'doente'.

O treinamento dos modelos foi realizado através de um processo de fine-tuning, onde ajustamos os pesos da camada de saída com base nas novas imagens de termografia, mantendo as camadas intermediárias congeladas. Essa modificação possibilita que o modelo aprenda características específicas dos dados, adaptando as representações aprendidas às particularidades do novo conjunto.

- **K-Fold Cross-Validation** A validação cruzada com k-fold é uma técnica essencial para garantir a robustez e a generalização dos modelos de aprendizado profundo. Neste estudo, implementamos uma validação cruzada com 10 folds, dividindo o conjunto de dados em dez partes iguais. Em cada iteração, uma dessas partes é utilizada como conjunto de validação, enquanto as outras nove são empregadas para o treinamento do modelo. Essa estratégia assegura que cada dado disponível seja utilizado tanto para treinamento quanto para validação, resultando em uma avaliação mais precisa do desempenho do modelo. Além disso, a média dos resultados obtidos

em cada iteração fornece uma medida confiável de como o modelo deve se comportar em novos dados não vistos, aspecto crucial em contextos médicos onde a precisão é vital. Essa validação pode ser observada na "Fig. 5".

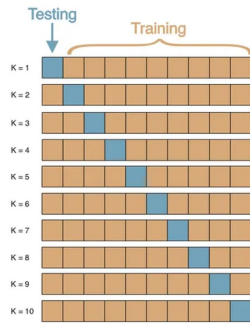


Figura 5. Exemplo de validação cruzada com 10 folds.

Durante o processo de k-fold cross-validation, monitoramos diversas métricas de desempenho, incluindo acurácia, precisão, sensibilidade e especificidade. Essa análise nos permitiu identificar variações no desempenho do modelo e ajustar hiperparâmetros, como a taxa de aprendizado e o número de épocas, visando otimizar os resultados finais.

- **Class Activation Mapping (CAM)** A interpretabilidade dos modelos de aprendizado profundo é uma consideração crítica em aplicações médicas, uma vez que entender como e por que um modelo realiza determinada predição é essencial. Para atender a essa necessidade, empregamos a técnica de Class Activation Mapping (CAM), que possibilita visualizar as regiões das imagens que o modelo considera mais relevantes para suas decisões. O CAM gera mapas de ativação que destacam as áreas da imagem que mais contribuíram para a classificação em uma determinada categoria. Esses mapas são particularmente úteis para validar o comportamento do modelo, assegurando que ele esteja focando em características clínicas relevantes, como áreas suspeitas de anomalias térmicas, ao invés de ruídos ou artefatos não relacionados. Em nosso estudo, aplicamos o CAM aos modelos treinados para verificar se as decisões de classificação estavam fundamentadas em padrões térmicos coerentes com diagnósticos médicos. Essa análise garantiu que o modelo não estivesse apenas aprendendo a distinguir as imagens com base em características irrelevantes, mas sim em aspectos críticos que podem ser aplicados na prática clínica. A capacidade de interpretar e validar as predições do modelo é fundamental para a eventual implementação dessa tecnologia em ambientes de saúde.
- **Scheduler de Taxa de Aprendizado** Para garantir o treinamento eficaz de redes neurais profundas, é necessária uma gestão cuidadosa da taxa de aprendizado, que controla a magnitude dos passos dados na direção da minimização da função de perda. Para otimizar o desempenho de nossos modelos, implementamos um scheduler

de taxa de aprendizado que ajusta dinamicamente esse parâmetro ao longo do treinamento. Essa técnica é particularmente importante em tarefas complexas, como a classificação de imagens médicas, onde o modelo pode se beneficiar de uma alta taxa de aprendizado nas etapas iniciais, permitindo uma rápida convergência, e de uma taxa mais baixa nas etapas posteriores para refinar os ajustes e evitar oscilações ao redor de mínimos locais.

Em nosso experimento, utilizamos um scheduler que reduzia a taxa de aprendizado em momentos estratégicos, com base na observação da função de perda. Essa abordagem permitiu que os modelos superassem platôs de desempenho e continuassem a melhorar mesmo após várias épocas de treinamento, resultando em um modelo final mais refinado e preciso.

E. Bibliotecas e Tecnologias

No presente estudo, diversas tecnologias e bibliotecas foram empregadas para a implementação e análise de modelos de classificação de imagens por meio de redes neurais convolucionais (CNNs). A seguir, é detalhada as funções específicas de cada uma dessas ferramentas.

- **glob:** A biblioteca glob foi utilizada para a exploração sistemática dos caminhos das imagens dentro das subpastas do dataset. Através de expressões glob, foi possível localizar e listar eficientemente todos os arquivos de imagem, garantindo que o pipeline de dados fosse robusto e abrangente, sem a necessidade de intervenção manual na especificação dos caminhos.
- **OpenCV:** O OpenCV desempenhou um papel fundamental no pré-processamento das imagens, realizando várias operações essenciais:
 - **Filtro Laplaciano:** Implementado para a detecção de bordas, o filtro Laplaciano foi utilizado para realçar contornos, o que pode contribuir para a extração de características relevantes durante o treinamento do modelo.
 - **Normalização das Imagens:** As imagens foram normalizadas para garantir que os valores de pixel estivessem dentro de uma faixa padronizada, o que é crucial aplicar após o filtro Laplaciano, uma vez que ao realçar as bordas pode gerar valores de pixel que são negativos ou que excedem 255. Esses valores fora do intervalo precisam ser ajustados para serem visualizados corretamente como uma imagem.
 - **Conversão de Canais:** A conversão entre diferentes espaços de cor foi realizada para adaptar as imagens às necessidades específicas do modelo, como a conversão para escala de cinza ou outros formatos necessários durante a exibição dos resultados do CAM.
 - **Leitura e Escrita de Imagens:** O OpenCV foi empregado tanto para carregar as imagens no ambiente de trabalho quanto para salvá-las após a aplicação de transformações.

- **Redimensionamento:** As imagens foram redimensionadas para as dimensões requeridas pelo modelo, assegurando uma entrada consistente durante o treinamento e a inferência.
- **Color Map para CAM:** Durante a visualização das regiões de interesse destacadas pelos CAMs, foi aplicado um color map para melhorar a interpretabilidade dos resultados.
- **Matplotlib:** A biblioteca matplotlib foi utilizada para visualização dos dados e análise dos resultados experimentais:
 - **Exibição das Imagens e Labels:** As imagens, juntamente com suas respectivas classes, foram exibidas para inspeção visual, permitindo a validação do pré-processamento e pós-processamento para exibição dos resultados da inferência do mapa de ativação de classe.
 - **Matriz de Confusão:** A matriz de confusão foi plotada para avaliar o desempenho do modelo em termos de classificação correta e incorreta das classes, oferecendo uma visão detalhada das áreas de confusão entre as categorias.
 - **Curva de Aprendizagem:** Foram plotadas curvas de aprendizagem para monitorar a evolução da acurácia e da função de perda durante o treinamento, permitindo a detecção de possíveis problemas como overfitting.
 - **Gráficos de Barras:** Gráficos de barras foram utilizados para comparar a acurácia e o desvio padrão dos modelos sob diferentes configurações de pré-processamento, facilitando a análise comparativa dos resultados.
- **PyTorch:** O PyTorch foi a estrutura central para a implementação e treinamento dos modelos de deep learning no projeto:
 - **Criação e Manipulação de Datasets:** Foi usado para o carregamento personalizado das imagens e seus rótulos utilizando `torch.utils.data.Dataset`, integrando transformações necessárias antes de alimentar o modelo. Além disso, foram criados iteradores eficientes sobre os datasets, otimizando o processamento em lotes e o uso de múltiplos núcleos para acelerar o carregamento dos dados.
 - **Transformação de Dados:** As transformações foram aplicadas utilizando o módulo `transforms` da `torchvision`, incluindo redimensionamento, normalização e flips horizontais aleatórios, que serviram como técnicas de aumento de dados para melhorar a generalização do modelo.
 - **Treinamento e Validação do Modelo:** Foi feito treinamento do modelo, utilizando backpropagation e otimizadores do PyTorch, como `torch.optim`, além de um scheduler para ajustar dinamicamente a taxa de aprendizado. Durante o treinamento, foi aplicada a técnica de Mixup, para gerar combinações interpoladas de exemplos de treinamento.
- **Inferência e Avaliação do Modelo:** Desempenhou o papel de avaliação do modelo nos dados de teste, ademais fornecer dados para as métricas como acurácia e para a matriz de confusão para análise de desempenho.
- **Utilização de Modelos Pré-Treinados:** Modelos pré-treinados, ResNet50, EfficientNet e Vgg19 da `torchvision.models`, foram fine-tunados para se adequar ao número de classes do problema em estudo, aproveitando o conhecimento prévio embutido nos pesos do modelo.
- **Treinamento em GPU:** O código foi otimizado para execução em GPUs, permitindo uma aceleração significativa tanto no treinamento quanto na inferência, essencial para lidar com a complexidade das CNNs.
- **Scikit-learn:** O scikit-learn foi empregado para tarefas de validação e avaliação dos modelos:
 - **Validação Cruzada com 10 Folds:** O scikit-learn foi utilizado para implementar a validação cruzada com 10 folds, dividindo o dataset em subconjuntos para assegurar uma avaliação robusta do desempenho do modelo.
 - **Cálculo de Métricas:** Foram calculadas métricas como acurácia, além de gerar a matriz de confusão para analisar o desempenho das previsões do modelo.

VI. RESULTADOS E DISCUSSÕES

A. Acurácia

Nesta seção, apresentamos e analisamos os resultados obtidos a partir da aplicação de diferentes técnicas de pré-processamento em três arquiteturas de redes neurais convolucionais: ResNet50, EfficientNet e VGG19. O objetivo foi investigar o impacto dessas técnicas na precisão da detecção de doenças mamárias utilizando imagens de termografia.

Cada técnica de pré-processamento foi selecionada com base em sua capacidade de melhorar a qualidade das imagens ou de reduzir o ruído que poderia afetar as previsões dos modelos. Em particular, exploramos a remoção de áreas específicas, como as axilas, a aplicação de filtros para realce de contornos e a combinação de imagens com a técnica de data augmentation Mixup. Além disso, também analisamos o efeito combinado dessas técnicas para verificar se a integração de múltiplas abordagens poderia oferecer um ganho adicional de desempenho.

A seguir, apresentamos os resultados de cada modelo, comparando as médias de precisão e os desvios padrão obtidos para cada configuração de pré-processamento. Essa análise é essencial para entender como cada técnica contribui para a robustez e generalização dos modelos, bem como para identificar as abordagens mais eficazes para a tarefa em questão.

Conforme apresentado na "Fig. 6", os resultados obtidos utilizando o modelo ResNet50 mostram que a média de

precisão sem qualquer pré-processamento foi de 92,5%, com um desvio padrão de 6,80%. Este resultado sólido estabelece uma linha de base robusta para o desempenho do modelo em imagens termográficas. A aplicação do Crop na região inferior das axilas reduziu a precisão média para 90,0%, com um desvio padrão de 5,92%, sugerindo que a remoção dessa área não beneficiou o modelo, embora tenha reduzido a variabilidade nos resultados.

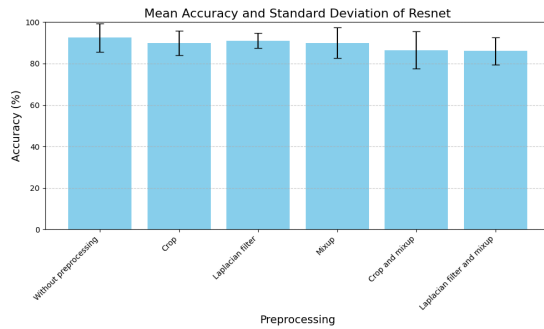


Figura 6. Acurácias do modelo Resnet50.

Ainda na "Fig. 6", a aplicação do Filtro Laplaciano melhorou a estabilidade das previsões, resultando em um desvio padrão de 3,63%, o menor entre todas as técnicas testadas para o ResNet50, com uma precisão média de 91,07%. Por outro lado, a utilização do Mixup resultou em uma leve diminuição na precisão para 90,0%, com um desvio padrão mais elevado de 7,42%, indicando maior variabilidade nos resultados e sugerindo que a técnica pode introduzir complexidades adicionais.

Quando combinados, Crop e Mixup resultaram na maior redução de precisão para 86,5%, com um desvio padrão elevado de 8,96%, o que sugere que essa combinação de técnicas pode não ser eficaz para o ResNet50 em termos de precisão e consistência. A combinação do Filtro Laplaciano com Mixup também resultou em uma redução significativa, com uma precisão média de 86,02% e um desvio padrão de 6,67%, reafirmando que a combinação dessas técnicas pode não ser ideal para este modelo.

No caso do modelo EfficientNet, conforme ilustrado na "Fig. 7", o desempenho inicial sem pré-processamento foi de 85,5%, com um desvio padrão de 8,79%. A aplicação do Crop na região inferior das axilas não melhorou os resultados, resultando em uma queda na precisão para 83,5% e um desvio padrão de 8,67%. A aplicação do Filtro Laplaciano também não proporcionou melhorias, com uma precisão média de 82,55% e um desvio padrão de 6,85%.

A técnica Mixup sozinha resultou em uma precisão de 82,0%, com um desvio padrão reduzido para 6,00%, o que indica que, embora a precisão média tenha diminuído, houve uma melhora na consistência das previsões. As combinações Crop com Mixup e Filtro Laplaciano com Mixup resultaram em precisões médias de 82,0% e 82,55%, respectivamente, com desvios padrão mais elevados (9,27% e 9,59%), sugerindo que essas combinações não foram eficazes em melhorar o desempenho do EfficientNet.

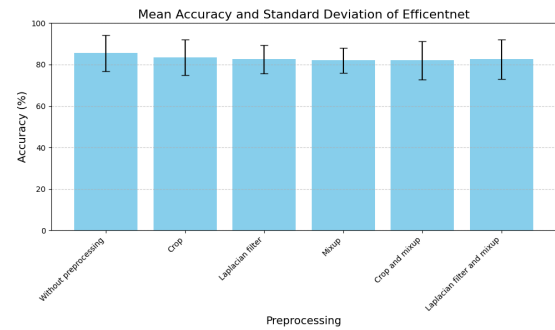


Figura 7. Acurácias do modelo Efficientnet.

Já o modelo VGG19, conforme demonstrado na "Fig. 8", apresentou um desempenho inicial forte, com uma precisão de 92,0% sem pré-processamento e um desvio padrão de 7,48%. A aplicação do Crop na região inferior das axilas manteve a precisão em 92,0%, mas reduziu significativamente o desvio padrão para 3,32%, indicando uma maior consistência nas previsões. A aplicação do Filtro Laplaciano resultou em uma ligeira redução na precisão para 89,1%, com um desvio padrão de 6,11%.

A técnica Mixup manteve uma precisão alta de 91,0%, mas com um desvio padrão elevado de 8,60%, sugerindo uma maior variabilidade nos resultados. A combinação Crop com Mixup resultou em uma precisão de 87,5%, com um desvio padrão de 6,02%, enquanto a combinação Filtro Laplaciano com Mixup resultou em uma precisão de 88,05%, com um desvio padrão de 7,82%. Esses resultados sugerem que, embora essas combinações tenham gerado alguma redução na precisão, a variabilidade dos resultados permaneceu alta.

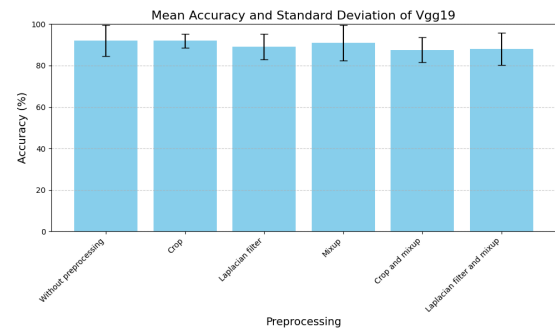


Figura 8. Acurácias do modelo Vgg19.

Os dados apresentados indicam que as técnicas de pré-processamento têm impactos variados dependendo da arquitetura do modelo. Para o ResNet50, o Filtro Laplaciano sozinho melhorou a estabilidade das previsões, enquanto combinações como Crop com Mixup e Filtro Laplaciano com Mixup não trouxeram benefícios claros. No caso do EfficientNet, nenhuma das técnicas ou combinações aplicadas resultou em melhorias significativas, destacando a sensibilidade do modelo às mudanças introduzidas pelo pré-processamento. Já o VGG19 se beneficiou da aplicação do Crop, que manteve a

alta precisão e reduziu a variabilidade das previsões. Esses resultados reforçam a importância de ajustar as técnicas de pré-processamento às características específicas do modelo e do conjunto de dados.

B. Matriz de confusão

Para uma compreensão mais aprofundada da performance dos modelos, foram geradas e analisadas as matrizes de confusão sem qualquer pré-processamento. As matrizes de confusão selecionadas, correspondentes às redes sem pré-processamento, são discutidas abaixo.

A matriz de confusão para o modelo ResNet50 sem pré-processamento, conforme apresentado na "Fig. 9", revelou um desempenho robusto em ambas as classes, com 89 amostras corretamente classificadas como "healthy" e 96 como "sick". Contudo, o modelo apresentou algumas falhas, classificando erroneamente 11 amostras que eram "healthy" como "sick" e 4 amostras que eram "sick" como "healthy".

Estes resultados indicam que, embora o ResNet50 tenha demonstrado uma alta precisão geral, a classificação de amostras "healthy" parece ser um pouco mais desafiadora, como evidenciado pelo maior número amostras "healthy" classificadas como "sick". A baixa ocorrência de amostras "sick" classificadas como "healthy" sugere uma elevada sensibilidade do modelo em identificar corretamente amostras doentes, o que é crucial em cenários médicos para evitar diagnósticos falsos de saúde.

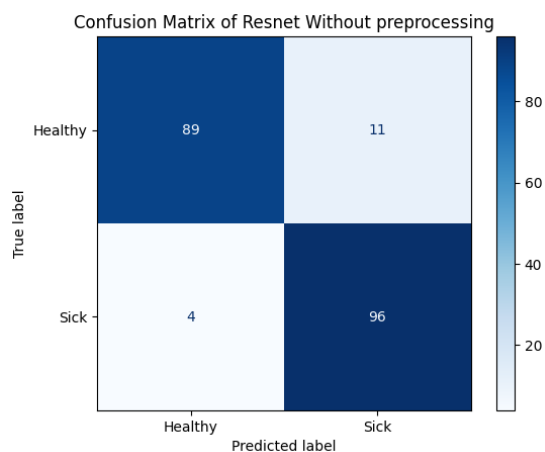


Figura 9. Matriz de confusão Resnet50.

Para o modelo EfficientNet, a matriz de confusão apresentada na "Fig. ??", mostrou uma performance distinta. O modelo corretamente classificou 71 amostras como "healthy" e 100 como "sick". No entanto, houve uma quantidade relativamente alta de 29 amostras "healthy" que foram classificadas erroneamente como "sick", enquanto todas as amostras "sick" foram corretamente identificadas.

Esse resultado indica que o EfficientNet tem uma tendência mais acentuada a classificar amostras "healthy" como "sick" em comparação ao ResNet50. A ausência de nenhuma amostra

"sick" foi classificada como "healthy" destaca a alta sensibilidade do modelo para a detecção de amostras doentes. Embora a precisão geral do EfficientNet seja menor em comparação ao ResNet50, sua capacidade de evitar diagnósticos falsos de saúde pode ser uma característica valiosa em determinados contextos clínicos.

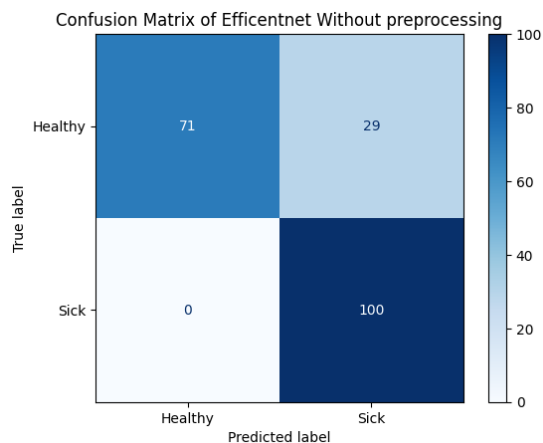


Figura 10. Matriz de confusão Efficientnet.

Por fim, o modelo VGG19, conforme mostrado na "Fig. 11", apresentou uma matriz de confusão equilibrada, com 92 amostras corretamente classificadas tanto como "healthy" quanto como "sick". Entretanto, o modelo cometeu 8 erros em ambas as direções: 8 amostras "healthy" foram classificadas como "sick" e 8 amostras "sick" foram classificadas como "healthy".

Esse equilíbrio nas classificações corretas e nos erros de classificação sugere que o VGG19 possui uma abordagem relativamente balanceada, com uma menor tendência a favorecer uma classe sobre a outra. No entanto, a presença de erros em ambas as direções aponta para uma necessidade de melhorias adicionais para aumentar a precisão e a consistência das previsões, especialmente em um contexto onde tanto falsos positivos quanto falsos negativos podem ter implicações significativas.

As análises das matrizes de confusão para os três modelos revelam características distintas em suas capacidades de classificação. O ResNet50 e o EfficientNet mostram uma alta sensibilidade para detectar amostras doentes, com o EfficientNet evitando completamente classificações de "sick" como "healthy", mas ao custo de um maior erro na classificação de "healthy". Já o VGG19 apresenta um comportamento mais equilibrado, embora com uma necessidade de redução de erros para melhorar sua confiabilidade em aplicações práticas.

C. Curva de aprendizagem

As curvas de aprendizagem dos modelos ResNet50, EfficientNet e VGG19 foram geradas a partir da média dos valores de acurácia e perda de cada época nos folds da validação cruzada, serão discutidas apenas as curvas sem pré-processamento. Cada curva reflete a evolução do desempenho dos modelos durante o processo de treinamento e fornece insights sobre a sua capacidade de generalização.

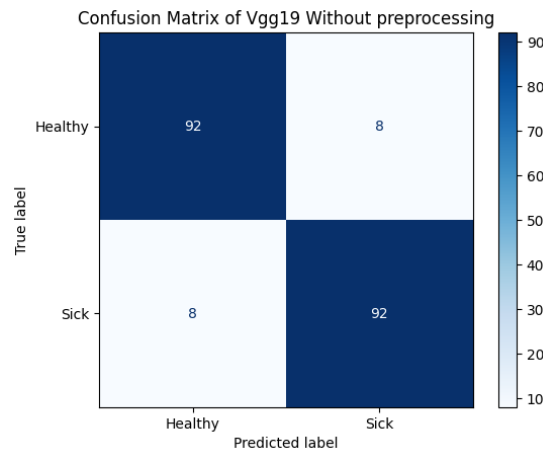


Figura 11. Matriz de confusão Vgg19.

A curva de aprendizagem, visualizada na "Fig. 12" da ResNet50 mostra uma tendência de redução contínua na perda ao longo das épocas, com uma acurácia crescente que se aproxima de um platô após a quinta época. A perda decresce de forma estável, indicando que o modelo está convergindo bem, e a acurácia atinge valores próximos de 100%, sugerindo que o modelo está aprendendo as características do conjunto de dados de maneira eficaz. No entanto, a alta acurácia pode também ser um indicativo de overfitting, especialmente se a acurácia no conjunto de validação fosse menor que no conjunto de treinamento.

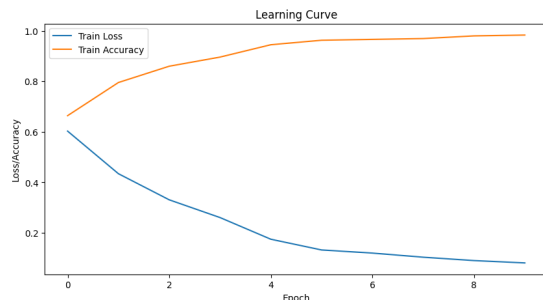


Figura 12. Curva de aprendizagem Resnet50.

A curva de aprendizagem da EfficientNet, "Fig. 13" revela um padrão semelhante ao da ResNet50, com uma redução constante na perda e um aumento na acurácia. No entanto, observa-se que a perda do EfficientNet tende a se estabilizar em um valor maior do que o da ResNet50, enquanto a acurácia é ligeiramente menor. Isso pode indicar que o modelo está tendo mais dificuldade em capturar as complexidades dos dados, ou que a arquitetura da EfficientNet, mesmo com sua eficiência computacional, não é tão eficaz quanto a da ResNet50 para este conjunto específico de dados.

A curva de aprendizagem da VGG19, vista na "Fig 14", também mostra uma redução contínua na perda e um aumento na acurácia, muito semelhante ao observado na ResNet50. A perda decresce rapidamente nas primeiras épocas,

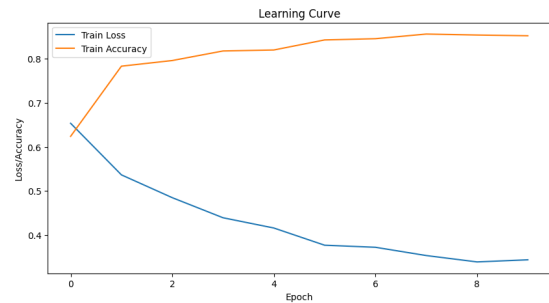


Figura 13. Curva de aprendizagem Efficientnet.

estabilizando-se em valores muito baixos, enquanto a acurácia aumenta de forma constante, atingindo valores próximos de 100%. Isso sugere que o VGG19, apesar de sua arquitetura mais antiga e profunda, ainda é altamente eficaz para a tarefa em questão, possivelmente devido à sua capacidade de capturar detalhes finos nas imagens.

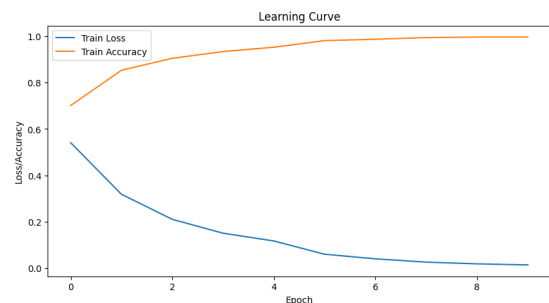


Figura 14. Curva de aprendizagem Vgg19.

As curvas de aprendizagem dos três modelos indicam que todos os modelos estão aprendendo bem os padrões dos dados de treinamento. A ResNet50 e a VGG19 apresentam desempenhos semelhantes, ambos convergindo rapidamente e alcançando altas acurácias, sugerindo que esses modelos estão particularmente bem ajustados para a tarefa de classificação proposta. A EfficientNet, por outro lado, apresenta uma curva de aprendizagem que converge mais lentamente e atinge uma acurácia ligeiramente menor, o que pode indicar uma adaptação menos eficaz ao conjunto de dados ou a necessidade de ajustes adicionais na arquitetura ou nos hiperparâmetros para melhorar seu desempenho.

VII. CONCLUSÕES E TRABALHOS FUTUROS

Neste estudo, exploramos a aplicação de três arquiteturas de redes neurais convolucionais — ResNet50, EfficientNet, e VGG19 — na tarefa de classificação de imagens termográficas para a detecção de tumores mamários. Os experimentos envolveram a avaliação dos modelos com diversas técnicas de pré-processamento, e foram realizadas análises detalhadas utilizando métricas como acurácia, matrizes de confusão, curvas de aprendizagem e CAM.

Os resultados das acurácias mostraram que a aplicação de pré-processamento impacta significativamente o desempenho

dos modelos. A ResNet50, sem pré-processamento, obteve a maior acurácia, demonstrando sua capacidade de generalizar bem para a tarefa, mesmo quando os dados não são processados previamente. Isso pode ser atribuído à sua arquitetura, que é particularmente eficaz em extrair características robustas de imagens complexas como as termográficas. Já a VGG19 apresentou uma acurácia comparável, reforçando sua eficácia na captura de detalhes finos, ademais com crop obteve um desvio padrão de apenas 3,32%.

Por outro lado, a EfficientNet, embora tenha sido projetada para ser uma arquitetura mais leve e eficiente, não atingiu o mesmo nível de desempenho. Isso sugere que, para conjuntos de dados como o utilizado neste estudo, com imagens de natureza tão específica, a otimização de pré-processamento ou ajustes na arquitetura da EfficientNet poderiam ser necessários para alcançar resultados comparáveis às outras arquiteturas.

As matrizes de confusão reforçaram as observações feitas com base nas acurácias. A ResNet50 e a VGG19 mostraram-se mais equilibradas em suas classificações, com uma menor taxa de falsos positivos e falsos negativos. A ResNet50, em particular, foi eficiente em identificar corretamente as imagens de pacientes doentes, o que é fundamental para aplicações médicas. A EfficientNet, embora tenha apresentado uma taxa de acerto alta para pacientes doentes, mostrou dificuldades em classificar corretamente as imagens de pacientes saudáveis, indicando uma necessidade de melhoria na capacidade de generalização do modelo.

As curvas de aprendizagem, que foram calculadas a partir da média das acurácias e perdas dos folds, mostraram que a ResNet50 e a VGG19 não só alcançaram alta acurácia, mas também mantiveram uma curva de perda decrescente constante, indicando uma boa capacidade de aprendizado. A EfficientNet, por outro lado, apresentou uma curva de aprendizagem mais instável, sugerindo que, sem pré-processamento, o modelo pode ter dificuldades em capturar as nuances dos dados, resultando em um desempenho inferior.

Os resultados deste estudo abrem várias oportunidades para pesquisas futuras. Primeiramente, seria interessante investigar mais exaustivamente impacto de diferentes técnicas de pré-processamento nas arquiteturas analisadas. Essa investigação pode ser apoiada ao CAM gerado dos modelos com os diferentes tipos de pré-processamento, visto que para esse trabalho ele serviu como base para testar o crop abaixo das axilas e aplicação do filtro laplaciano.

Além das análises realizadas, é interessante considerar a possibilidade de implementar uma segmentação mais precisa da região das mamas nas imagens termográficas. Embora o uso de técnicas de crop, como o corte abaixo das axilas, tenha sido testado, os resultados mostraram que esse tipo de pré-processamento não trouxe os benefícios esperados em termos de melhoria na acurácia dos modelos.

A segmentação específica das mamas pode contribuir para isolar com mais precisão a área de interesse, eliminando possíveis ruídos e informações irrelevantes que podem estar presentes nas imagens. Isso permitiria que os modelos focassem mais diretamente nas regiões onde os sinais de tumores

mamários são mais pronunciados, potencialmente melhorando o desempenho dos modelos em identificar corretamente as anomalias.

Em suma, este trabalho demonstrou que as arquiteturas ResNet50 e VGG19 são altamente eficazes para a detecção de tumores mamários em imagens termográficas, especialmente sem a necessidade de pré-processamento pesado. No entanto, há ainda um vasto campo de possibilidades a ser explorado para otimizar essa tarefa, com o objetivo de fornecer um suporte diagnóstico mais preciso e acessível na prática clínica.

REFERÊNCIAS

- [1] M. A. S. A. Husaini, M. H. Habaebi, S. A. Hameed, M. R. Islam and T. S. Gunawan, "A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks," in *IEEE Access*, vol. 8, pp. 208922–208937, 2020.
- [2] Zuluaga-Gomez, J., Zerhouni, N., Al Masry, Z., Devalland, C., and Varnier, C., "A survey of breast cancer screening techniques: thermography and electrical impedance tomography," *Journal of Medical Engineering & Technology*, vol. 43, no. 5, pp. 305–322, 2019.
- [3] W. Amalu, W. Hobbins, J. Head, and R. Elliot, "Infrared imaging of the breast— an overview," *The Biomedical Engineering Handbook*, 3rd ed., Medical Devices and Systems, CRC Press, Baton Rouge, 2006.
- [4] T. Ekici and H. Jawzal, "Breast cancer diagnosis using thermography and convolutional neural networks," *Medical Hypotheses*, vol. 137, 2020, 109542.
- [5] L. F. Silva, A. A. S. M. D. Santos, R. S. Bravo, A. C. Silva, D. C. Muchaluat-Saade, and A. Conci, "Hybrid analysis for indicating patients with breast cancer using temperature time series," *Comput Methods Programs Biomed*, vol. 130, pp. 142–153, 2016, doi: 10.1016/j.cmpb.2016.03.002.
- [6] B. de Freitas Oliveira Baffa and L. Grassano Lattari, "Convolutional neural networks for static and dynamic breast infrared imaging classification," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images, IEEE*, 2018, pp. 174–181, doi: 10.1109/SIBGRAPI.2018.00029.
- [7] S. Kiyomet, M. Y. Aslankaya, M. Taskiran and B. Bolat, "Breast Cancer Detection From Thermography Based on Deep Neural Networks," *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Izmir, Turkey, 2019, pp. 1–5.
- [8] J. C. Torres-Galván, E. Guevara, E. S. Kolosovas-Machuca, A. Ocegüera-Villanueva, J. L. Flores, and F. J. González, "Deep convolutional neural networks for classifying breast cancer using infrared thermography," *Quantitative InfraRed Thermography Journal*, vol. 19, no. 4, pp. 283–294, 2021.
- [9] J. Folkman, "What is the evidence that tumors are angiogenesis dependent?," *J Natl Cancer Inst*, vol. 82, pp. 4–6, 1990.
- [10] M. Gautherie, "Thermopathology of breast cancer: measurement and analysis of in vivo temperature and blood flow," *Ann N Y Acad Sci*, vol. 335, pp. 383–415, 1980.
- [11] J. D. Bronzino, "The Biomedical Engineering Handbook," Boca Raton, FL: CRC Press, 2000.