INSTITUTO FEDERAL
Pernambuco

CLASSIFICAÇÕES EXPLICÁVEIS PARA IMAGENS DE LESÕES DE PELE

EXPLAINABLE CLASSIFICATIONS FOR SKIN LESIONS IMAGES

Beatriz Freitas da Silva
bfs1@discente.ifpe.edu.br

Camila Maria da Silva
cms7@discente.ifpe.edu.br

Rodrigo Cesar Lira da Silva
rodrigo.lira@paulista.ifpe.edu.br

RESUMO

A pele é o maior órgão do corpo humano em termos de área e peso, desempenhando funções essenciais como a proteção contra agentes externos, a regulação da temperatura e a percepção sensorial. Lesões de pele comprometem esse organismo e, em casos graves, podem indicar doenças dermatológicas, como o câncer de pele, que é o tipo mais frequente no Brasil. Por isso, identificar sinais suspeitos precocemente é fundamental para garantir o diagnóstico e o tratamento eficaz. O uso de Inteligência Artificial (IA) tem se tornado cada vez mais frequente para ajudar no diagnóstico e prevenção das lesões de pele malignas. Porém, seus resultados têm sido acompanhados de intensas discussões éticas e morais, especialmente relacionadas à privacidade dos dados dos pacientes, à responsabilidade pelos diagnósticos realizados pela IA e ao risco de vieses que influenciam diretamente os resultados dos algoritmos. Nesse cenário, a Inteligência Artificial Explicável (XAI) ganha destaque por proporcionar a interpretabilidade dos modelos, utilizando técnicas como regras inferenciais, mapas de ativação e modelos explicativos locais, permitindo maior compreensão e validação dos diagnósticos. Este artigo, propõe mostrar a classificação de lesões de pele por meio de redes neurais artificiais utilizando os classificadores EfficientNetB0, VGG16 e ResNet50, aliados a técnicas de XAI para identificar as características mais relevantes na classificação dessas lesões. Foram utilizados dados do conjunto ISIC 2019 e, dentre os classificadores testados, o melhor desempenho foi do ResNet50 que alcançou acurácia de 83%. Para analisar os resultados, aplicaram-se SHAP, LIME e Grad-CAM, que evidenciaram as regiões das imagens com maior influência na classificação. Conclui-se, assim, que a aplicação de XAI se mostra um recurso promissor para o processo de classificação de doenças dermatológicas e podendo ampliar a confiança nos resultados obtidos.

Palavras-chave: Lesões de pele; Redes neurais; Inteligência artificial explicável.

Instituto Federal de Educação, Ciências e Tecnologia de Pernambuco. Campus Paulista. Curso de Análise e Desenvolvimento de Sistemas, 29 de maio de 2025.

ABSTRACT

The skin is the largest organ of the human body and plays a crucial role in protection against external agents, temperature regulation, and sensory perception. Skin lesions compromise this organ and, in severe cases, may indicate dermatological diseases. The use of Artificial Intelligence (AI) has become increasingly common in assisting the diagnosis and prevention of malignant skin lesions. However, its results have sparked intense ethical and moral discussions, particularly when algorithms reproduce biases that directly influence outcomes. In this context, the adoption of Explainable Artificial Intelligence (XAI) stands out by offering greater transparency and understanding of AI decision-making processes. This article aims to demonstrate the classification of skin lesions using artificial neural networks through the classifiers EfficientNetB0, VGG16, and ResNet50, combined with XAI techniques to identify the most relevant features in lesion classification. The ISIC 2019 dataset was used, and among the tested classifiers, ResNet50 achieved the best performance with an accuracy of 83%. To analyze the results, SHAP, LIME, and Grad-CAM were applied, highlighting the regions of the images that had the greatest influence on classification. It is concluded that the application of XAI proves to be a promising resource for the disease classification process, potentially increasing trust in the results obtained.

Keywords: Skin lesions; Neural networks; Explainable artificial intelligence.

1. INTRODUÇÃO

As doenças de pele englobam uma série de condições que afetam a saúde dérmica, com muitas delas representando sérios desafios para a saúde pública. Entre as doenças cutâneas mais prevalentes, destaca-se o câncer de pele, sendo o melanoma um dos tipos mais graves. De acordo com o Ministério da Saúde (2025), o câncer de pele corresponde a cerca de 30% de todos os tumores malignos registrados no Brasil. Dessa forma, a detecção precoce torna-se ainda mais crucial, pois estudos indicam que as chances de cura podem chegar a 90% quando o câncer é identificado nos estágios iniciais (Sociedade Brasileira de Dermatologia, 2023).

Entretanto, a escassez de profissionais especializados, como dermatologistas, impacta diretamente esse processo, evidenciando a necessidade de fortalecer a atenção primária para ampliar o acesso ao diagnóstico precoce, especialmente em regiões com menor acesso à atenção especializada (World Health Organization, 2025).

Nesse âmbito, lesões de pele associadas a sintomas como coceira, ardência, descamação e feridas com cicatrização lenta, sem causa aparente, devem ser investigadas por um dermatologista, pois podem ser indícios de câncer de pele (INCA, 2023). Além disso, fatores como exposição prolongada ao sol, pele clara, olhos claros e histórico familiar aumentam o risco de desenvolver a doença, sendo a conscientização sobre cuidados preventivos fundamental (Brasil, 2023).

Diante dos fatores de risco e da necessidade de prevenção, é indispensável aprimorar a precisão e a agilidade no diagnóstico. A Inteligência Artificial (IA) tem evoluído rapidamente, fornecendo sistemas que complementam e expandem as

habilidades médicas por meio de modelos preditivos (Haug; Drazen, 2023). No entanto, a efetividade dessas tecnologias, em especial em contextos de saúde pública, depende diretamente de sua integração com a atenção primária, responsável pela triagem inicial, pelo rastreamento e pelo encaminhamento oportuno dos casos. Sem essa articulação, o uso da IA tende a permanecer restrito a contextos especializados, o que limita seu alcance populacional e, consequentemente, sua contribuição para o diagnóstico precoce em larga escala.

A versatilidade dos sistemas baseados em Inteligência Artificial permite sua aplicação em diferentes tipos de câncer, destacando-se, por exemplo, os avanços no diagnóstico do câncer de mama. Nesse caso, algoritmos de aprendizado de máquina, como redes neurais convolucionais (do inglês, *Convolutional neural network* - CNNs), têm apresentado alto desempenho na análise de imagens histopatológicas. Em um estudo que propôs uma abordagem de classificação automatizada de tecidos baseada em CNN, obtiveram-se resultados promissores na detecção de câncer de mama, mostrando-se semelhantes ou superiores aos métodos de última geração (Araújo *et al.*, 2017). Essa automação pode superar métodos tradicionais em termos de confiabilidade e rapidez, além de auxiliar especialistas ao reduzir inconsistências no processo de avaliações visuais.

Apesar dos avanços tecnológicos, ainda existem desafios significativos para a aplicação efetiva da IA na detecção precoce do câncer de pele. Embora redes neurais e outras técnicas de IA venham demonstrando desempenho relevante em diagnósticos, a falta de transparência sobre como esses modelos tomam decisões pode limitar a aceitação e a confiança tanto de médicos quanto de pacientes. Para contornar essa limitação, surge a Inteligência Artificial Explicável (do inglês, *Explainable Artificial Intelligence* – XAI), que busca tornar esses modelos mais compreensíveis, de modo que os usuários entendam os critérios por trás de cada decisão – algo essencial para a adoção e eficácia dessas tecnologias na medicina (Linardatos *et al.*, 2021).

Dessa forma, no contexto da Inteligência Artificial, uma das abordagens mais promissoras é o uso de redes neurais artificiais, que tem se destacado na análise de imagens dermatológicas para a detecção antecipada de lesões malignas na pele. As CNNs são especialmente eficazes na identificação de padrões em imagens e, ao serem combinadas com técnicas de XAI, podem fornecer maior clareza sobre as características que influenciam suas previsões (Maron *et al.*, 2021).

Considerando esses aspectos, este trabalho visa contribuir para essa discussão ao comparar, de forma sistemática, três modelos para a classificação de imagens — EfficientNetB0, VGG16 e ResNet50 — e avaliar três métodos distintos de Inteligência Artificial Explicável: SHAP, LIME e Grad-CAM. Esses métodos foram aplicados a Redes Neurais Convolucionais treinadas com dados da *The International Skin Imaging Collaboration* (ISIC). O objetivo principal é identificar qual método se mostra mais eficaz para interpretar, de maneira confiável, os resultados dos diagnósticos de lesões cutâneas malignas. Na análise realizada, o modelo ResNet50 em conjunto com o Grad-CAM destacou-se por apresentar melhor desempenho na classificação e fornecer explicações mais claras, facilitando a interpretação do modelo por especialistas clínicos. Ao priorizar a transparência nas redes neurais, espera-se facilitar a adoção desses modelos na prática clínica, contribuindo para um diagnóstico mais precoce e preciso das lesões de pele malignas.

Por fim, visando a clareza e a organização, o artigo encontra-se estruturado

da seguinte forma: na Seção 2, apresenta-se o referencial teórico, abrangendo conceitos de Aprendizagem de Máquina, Redes Neurais e XAI, bem como os métodos SHAP, LIME e Grad-CAM. Em seguida, a Seção 3 descreve a metodologia adotada, incluindo pré-processamento de dados, análise exploratória e modelagem dos classificadores. Na Seção 4, são discutidos os resultados dos classificadores e das técnicas de XAI. Finalmente, a Seção 5 traz as considerações finais do estudo.

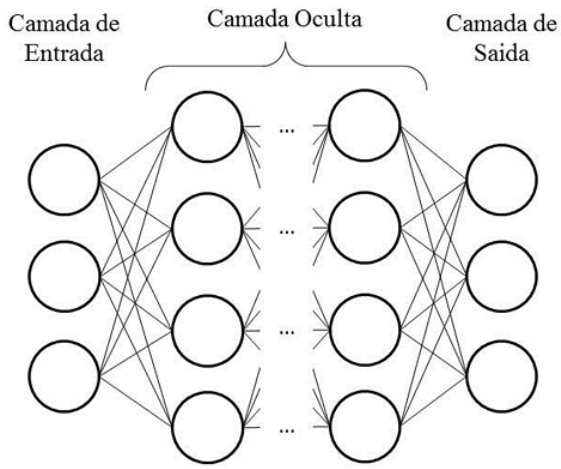
2. FUNDAMENTAÇÃO TEÓRICA

O referencial teórico presente no artigo foi estruturado em tópicos que sustentam o projeto, com o propósito de explorar conceitos fundamentais relacionados ao diagnóstico de lesões de pele com o uso de Inteligência Artificial e de técnicas de Inteligência Artificial Explicável.

2.1 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados no funcionamento do cérebro humano, compostos por unidades de processamento interconectadas, conhecidas como neurônios artificiais que se unem por meio de conexões sinápticas. De uma forma simplificada, uma rede neural artificial pode ser vista como um grafo onde os nós são os neurônios e as ligações fazem a função das sinapses (Ferneda, 2006). Como ilustrado na Figura 1, cada camada de uma rede neural recebe e processa informações de maneira sequencial, transmitindo-as para as camadas seguintes. Esse fluxo contínuo, aliado ao ajuste de pesos e limiares, permite que a rede neural refine suas decisões ao longo do treinamento. Sendo assim, essas redes são capazes de aprender e generalizar a partir de dados, ajustando seus parâmetros internos para resolver problemas complexos.

Figura 1. Representação gráfica de uma rede neural.



Fonte: Pereira e Centeno (2017).

2.1.1 Redes neurais convolucionais

Dentre os diversos tipos de RNAs, destacam-se as Redes Neurais Convolucionais, amplamente utilizadas em tarefas de visão computacional, como classificação e detecção de padrões. Essas redes são compostas por camadas especializadas na extração progressiva de características relevantes das imagens, possibilitando a identificação de representações visuais complexas de forma análoga ao funcionamento dos neurônios biológicos. Por sua arquitetura adaptada ao processamento de dados visuais, as CNNs são especialmente eficazes no processamento de imagens dermatológicas, sendo capazes de apresentar desempenho similar ao de dermatologistas no diagnóstico por meio de fotografias de lesões cutâneas (Silva et al., 2021).

No contexto do diagnóstico médico por imagem, as Redes Neurais Convolucionais têm se destacado como ferramentas altamente eficazes em tarefas de reconhecimento e análise visual. Essas redes são amplamente empregadas em sistemas para detectar padrões complexos em exames clínicos, como radiografias, tomografias e imagens dermatológicas. No caso específico da dermatologia, as CNNs têm se mostrado promissoras na identificação precoce de anomalias cutâneas. Essa capacidade torna as CNNs fundamentais em estudos e aplicações que visam auxiliar no diagnóstico de doenças como o câncer de pele, por meio de uma triagem.

No cenário de classificação de imagens médicas, o modelo EfficientNetB0 tem ganhado destaque por sua eficácia e leveza computacional. De acordo com Iqbal et al. (2024) essa arquitetura se mostrou altamente promissora na tarefa de classificação multicategorias de tumores cerebrais, aliando precisão diagnóstica à eficiência de processamento em ambientes de Internet das Coisas Médicas (IoMT). O EfficientNetB0 alcança esse desempenho por meio de uma técnica chamada *compound scaling*, que equilibra de forma coordenada profundidade, largura e resolução da rede. Essa abordagem resulta em modelos compactos e com alta acurácia, o que os torna especialmente úteis em aplicações clínicas que exigem rapidez e baixo custo computacional.

A ResNet50 é uma arquitetura de rede neural convolucional profunda que utiliza conexões residuais para evitar o problema do desaparecimento do gradiente, sendo especialmente eficaz em tarefas complexas de classificação de imagens (He et al., 2015). No estudo de Murdiyanto et al. (2024), a ResNet50 foi comparada a uma CNN tradicional na classificação de câncer de pele. Os resultados mostraram que a ResNet50 apresentou melhor desempenho na detecção de carcinoma basocelular, destacando-se como uma alternativa promissora em diagnósticos automatizados, apesar de apresentar desafios na classificação de classes menos representadas.

Já o VGG16, com sua arquitetura de rede neural convolucional amplamente utilizada em tarefas de classificação de imagens, especialmente na área médica. Composta por 13 camadas convolucionais e 3 camadas totalmente conectadas, totalizando 16 camadas com pesos ajustáveis, a arquitetura se destaca por sua simplicidade e profundidade. Utiliza filtros pequenos de 3×3 e operações de max pooling para reduzir a dimensionalidade dos dados, o que a torna eficiente na extração de características visuais. Devido a essas qualidades, o VGG16 tem sido amplamente empregado em aprendizado por transferência (do inglês, *Transfer*

learning), contribuindo de forma eficaz em diagnósticos baseados em imagens, como o câncer de pele (Bashir *et al.*, 2024).

A análise comparativa desses modelos visa compreender seus desempenhos na tarefa de reconhecimento de lesões cutâneas. Assim como a implementação de explicabilidade de inteligência artificial ao final do processo, buscando avaliar não apenas os resultados, mas também a interpretabilidade das decisões tomadas pelos modelos, promovendo maior confiança e adoção dos resultados.

2.2 *Explainable Artificial Intelligence* (XAI)

A Inteligência Artificial Explicável (do inglês, *Explainable Artificial Intelligence* - XAI) é uma subárea da IA que visa o desenvolvimento de métodos capazes de fornecer explicações compreensíveis por humanos sobre as decisões tomadas por sistemas de IA complexos, especialmente os modelos do tipo "caixa-preta" (Salih *et al.*, 2024). Essa explicabilidade é essencial em contextos sensíveis, como a saúde, onde entender por que um modelo chegou a determinada decisão é tão importante quanto a própria precisão da predição.

De acordo com Xu *et al.* (2019), a explicabilidade é uma necessidade nas aplicações da IA em processos decisórios críticos, como na medicina, onde as recomendações dos algoritmos precisam ser compreensíveis para os humanos que os utilizam. A falta de transparência nos modelos, pode gerar desconfiança e dificultar a aceitação desses sistemas.

No âmbito dermatológico, onde a precisão e confiabilidade dos diagnósticos são essenciais, a aplicação de XAI permite que os médicos compreendam melhor como as decisões sobre o diagnóstico de câncer de pele são feitas. Isso é particularmente importante, pois a IA pode fornecer *insights* e justificar suas decisões, permitindo uma análise mais consistente é fundamental.

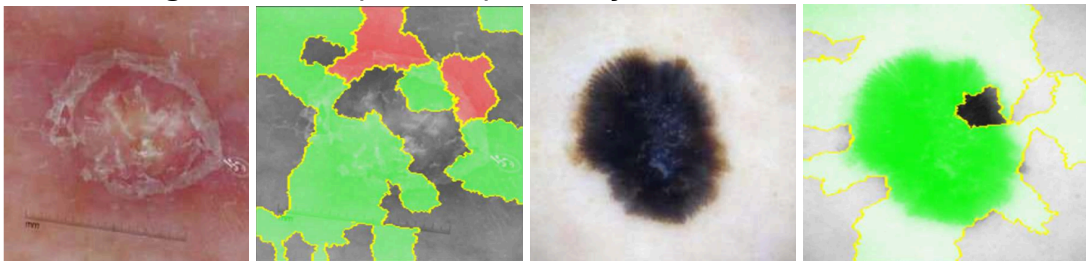
Existem diversas técnicas de explicabilidade na IA, que buscam proporcionar maior clareza sobre como os modelos chegam às suas conclusões. Este trabalho se concentra nas técnicas LIME, Grad-CAM e SHAPE, que são amplamente utilizadas para interpretar e visualizar as decisões tomadas por modelos de redes neurais.

2.2.1 LIME

O LIME (do inglês, *Local Interpretable Model-Agnostic Explanations*) é uma técnica de inteligência artificial explicável que permite gerar explicações locais para as decisões de modelos complexos, como redes neurais convolucionais. Segundo Bader *et al.* (2023), essa técnica aproxima o comportamento do modelo por meio de explicações simplificadas e interpretáveis, destacando as características mais relevantes para uma determinada previsão. Por ser agnóstico ao modelo, o LIME pode ser aplicado a qualquer tipo de classificador, o que o torna útil, por exemplo, para interpretar predições de modelos que distinguem entre imagens de cães e gatos, indicando quais regiões da imagem mais influenciaram a decisão do modelo. A Figura 2 A imagem ilustra a aplicação da técnica LIME para explicar a predição de um modelo de classificação de lesões cutâneas. A imagem original da lesão é processada para gerar uma segmentação, identificando regiões distintas da pele. O LIME, então, destaca em verde as áreas que mais influenciaram a decisão do

modelo, proporcionando uma explicação visual que aumenta a transparência do diagnóstico automatizado.

Figura 2. Exemplo de representação visual usando Lime.

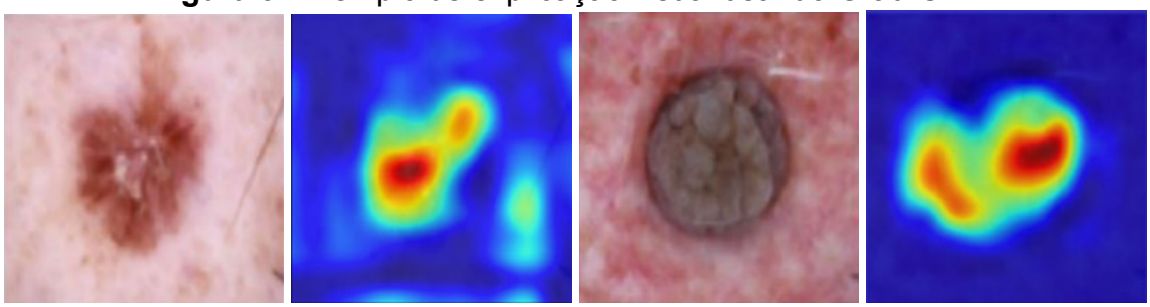


Fonte: Nigar et al. (2022).

2.2.2 Grand-CAM

O Grad-CAM (do inglês, *Gradient-weighted Class Activation Mapping*) é uma técnica de explicação visual que utiliza os gradientes das camadas convolucionais finais de uma rede neural para gerar mapas de ativação, destacando as regiões da imagem que mais influenciaram a predição. Segundo Linardatos *et al.* (2021), essa abordagem permite visualizar de forma intuitiva o foco da rede durante a classificação, tornando o processo mais interpretável. A Figura 3 ilustra a aplicação da técnica Grad-CAM na interpretação das decisões de modelos de classificação de lesões cutâneas. Nesse processo, as imagens originais são sobrepostas por mapas de calor que destacam em tons de vermelho, as regiões nas quais a rede neural concentrou maior atenção durante a predição. Essas áreas contribuíram significativamente para a predição do modelo, oferecendo uma explicação visual mais clara e compreensível de seu raciocínio, o que pode ajudar a reforçar a confiabilidade clínica dos sistemas de inteligência artificial.

Figura 3. Exemplo de explicação visual usando Grad-CAM.



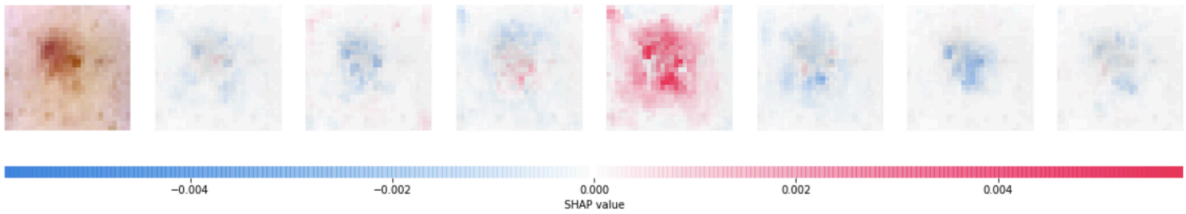
Fonte: Mridha et al. (2023).

2.2.2 SHAP

O SHAP (do inglês, *SHapley Additive exPlanations*) é uma técnica baseada na teoria dos jogos que atribui valores a cada característica de entrada de um modelo, indicando sua contribuição para uma predição específica. Conforme descrito por Bader et al. (2023), o SHAP oferece explicações consistentes e quantitativas, permitindo visualizar claramente o impacto de cada variável nas

decisões do modelo. Essa abordagem é especialmente útil na classificação de imagens, pois identifica quais regiões tiveram maior peso na predição, aumentando a interpretabilidade e a confiança nos resultados. A Figura 4 mostra a aplicação do SHAP em uma imagem de lesão cutânea. Ao lado da imagem original, diferentes áreas são destacadas por uma escala de cores que varia do azul (impacto negativo) ao vermelho (impacto positivo), evidenciando a contribuição de cada região para a decisão do modelo. Essa representação possibilita uma interpretação local detalhada, favorecendo uma análise crítica e transparente do comportamento da rede neural. Dessa forma, o SHAP destaca-se por fornecer explicações localmente fieis e adaptáveis, sendo especialmente valioso em aplicações sensíveis que exigem alta interpretabilidade, como diagnósticos médicos assistidos por IA (Salih et al., 2024).

Figura 4. Exemplo de explicação visual usando SHAP.

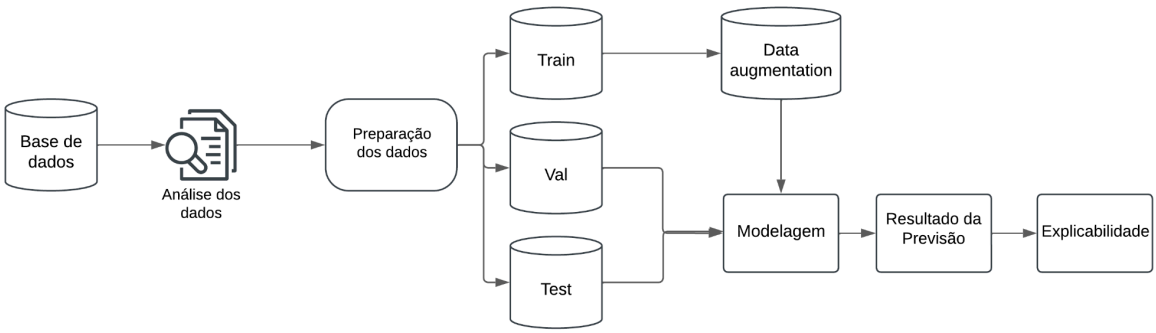


Fonte: Homa (2023).

3. METODOLOGIA

A metodologia deste trabalho consiste na utilização de redes neurais artificiais para classificar imagens de lesões de pele como malignas e benignas, seguida pela aplicação de técnicas da Inteligência Artificial Explicável para interpretar os resultados. O processo abrange a aquisição da base de dados, preparação e divisão dos dados, aplicação de *data augmentation* para ampliar o conjunto de treino, modelagem com redes neurais e, por fim, o uso de métodos explicáveis. Na Figura 5, é apresentado o fluxograma que representa as etapas da metodologia adotada.

Figura 5. Fluxograma da Metodologia empregada neste trabalho.



Fonte: Elaboração própria (2025).

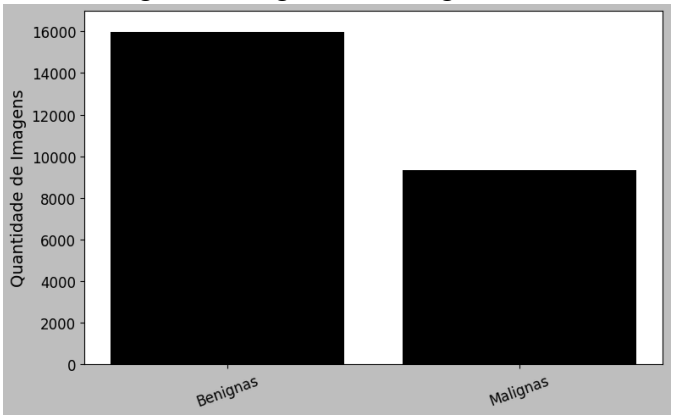
3.1 Entendimento dos dados

Neste projeto foram utilizados dados da *International Skin Imaging Collaboration*, um repositório público de lesões de pele criado em parceria com a academia e a indústria com o intuito de reduzir a mortalidade por câncer de pele (International Skin Imaging Collaboration, 2024). O ISIC, desde 2016 realiza competições nas quais disponibiliza bases de dados para que os participantes possam cumprir um desafio estabelecido.

3.1 Análise Exploratória

A base de dados utilizada é a ISIC 2019, que possui 25.331 imagens obtidas a partir de dermatoscópios. As imagens são distribuídas em 8 classes (*Basal cell carcinoma*, *Dermatofibroma*, *Vascular lesion*, *Melanoma*, *Melanocytic nevus*, *Actinic keratosis*, *Benign keratosis* e *Squamous cell carcinoma*), dessas 15.991 são Benignas e 9.340 simbolizam as Malignas, representados graficamente na Figura 6.

Figura 6. Proporção de imagens malignas e benignas na base de dados ISIC 2019.



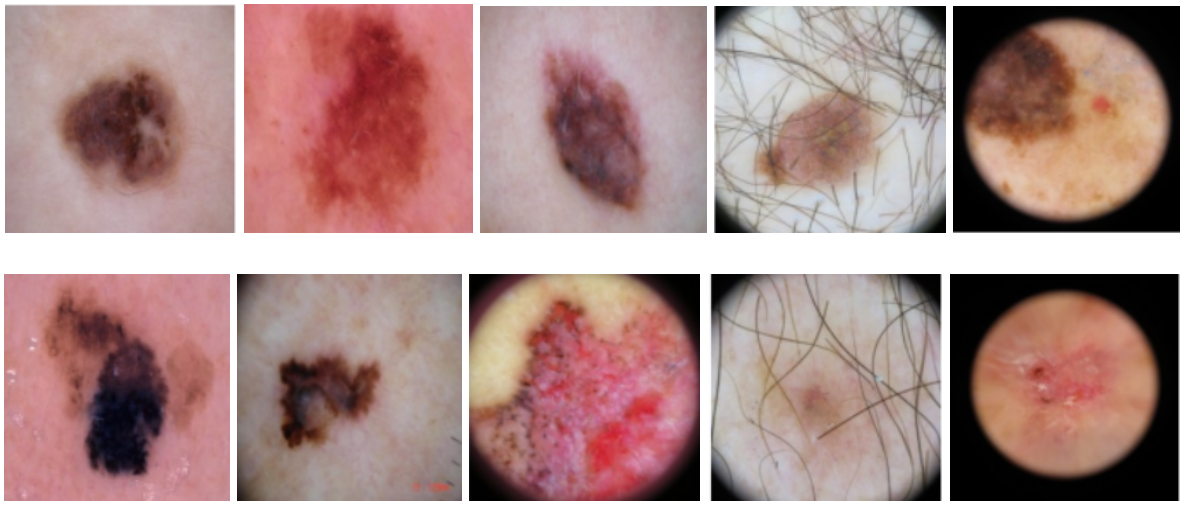
Fonte: Elaboração própria (2024).

Durante a análise, verificou-se que todos os arquivos de imagem são válidos, sem arquivos corrompidos, e estão no formato JPG, o que simplifica o processamento por não exigir tratamento de diferentes extensões. O conjunto é composto por imagens em modo RGB, compatíveis com a maioria dos modelos usados em classificação.

A Figura 7 apresenta amostras representativas das imagens benignas e malignas do conjunto de dados utilizado. Ao analisar essas imagens, nota-se uma grande heterogeneidade em suas dimensões, tanto em altura quanto em largura, o que pode comprometer o desempenho das redes neurais convolucionais durante o treinamento. Em contextos clínicos, onde a precisão é essencial, padronizar tamanho e resolução é fundamental para evitar que variações de escala prejudiquem a extração de características e introduzam vieses nas previsões. Estudos indicam que imagens com resolução igual ou superior a 128×128 pixels melhoram o desempenho na classificação de lesões cutâneas (Mahbod et al., 2021). Portanto, optou-se por redimensionar todas as imagens, garantindo proporções e

resolução uniformes, o que otimiza o processamento e contribui para resultados mais consistentes e confiáveis em diagnósticos assistidos por inteligência artificial.

Figura 7. Imagens de amostras de lesões da base de dados ISIC 2019.



Fonte: International Skin Imaging Collaboration (2019).

3.2. Preparação dos dados

3.2.1 Redimensionamento e Normalização

As imagens da base de dados foram redimensionadas e normalizadas para garantir que estejam no formato adequado para treinamento de modelos de aprendizado de máquina. O tamanho das imagens foi definido para 224 x 224 *pixels* utilizando o algoritmo LANCZOS, que preserva a qualidade ao reduzir as dimensões. Além disso, o código converte as imagens para o formato RGB e as normaliza para o intervalo [0, 1], o que facilita o treinamento de redes neurais.

Após isso, as imagens são salvas em um diretório de saída com o nome original, utilizando o formato JPEG e uma qualidade de 90%. A normalização é revertida antes do salvamento, assegurando que as imagens sejam adequadas para o modelo.

3.2.2. Balanceamento dos dados

Com base nos resultados obtidos na etapa de análise exploratória, foi possível inferir que o conjunto de dados é desbalanceado, uma vez que a categoria de imagem benignas tem o número mais significativo de imagens em comparação com as outras. O desbalanceamento pode causar enviesamento no modelo aplicado para classificação das imagens, por isso foi necessário fazer o balanceamento das classes.

O balanceamento das classes foi feito utilizando a técnica de *data augmentation*, que consiste em criar amostras sintéticas a partir de um conjunto limitado, sem coletar novos dados (Mumuni Alhassan, 2022). Isso é feito por meio de

transformações como rotação, alteração de cor e variação na luminosidade das imagens originais, criando dados sintéticos que mantêm as características essenciais das amostras. Essa estratégia aumenta a diversidade e variabilidade do conjunto de treinamento, melhorando a generalização do modelo. Além disso, essa técnica atende a duas necessidades: aumentar a quantidade de dados e melhorar a capacidade de generalização do modelo, o que contribui para reduzir o risco de *overfitting* (Maharana Kiran, 2022).

Existem diversas formas de aplicar a técnica de data augmentation. Neste trabalho, utilizou-se a classe *ImageDataGenerator* da biblioteca Keras. A partir do total inicial de imagens, definiu-se um limite de 12.000 imagens para cada classe no conjunto de treinamento. Para cada categoria, um número específico de novas imagens sintéticas foi gerado, garantindo um balanceamento adequado entre as classes. A Tabela 1 apresenta os hiperparâmetros utilizados para a geração dos dados.

Tabela 1. Hiperparâmetros do *data augmentation*.

Parâmetro	Valor
rescale	1,0/255,0
rotation_range	180
width_shift_range	0,1
height_shift_range	0,1
zoom_range	0,1
horizontal_flip	TRUE
vertical_flip	TRUE
brightness_range	[0,45 , 0,55]
fill_mode	reflect

Fonte: Elaboração própria.

3.3. Modelagem

Os modelos selecionados para a classificação foram o EfficientNetB0, VGG16 e ResNet50. O EfficientNetB0 se destaca por seu equilíbrio entre precisão e eficiência computacional, sendo projetado para maximizar o desempenho com menor consumo de recursos. Essa característica o torna especialmente eficaz para tarefas complexas, como a análise de imagens médicas. Estudos recentes demonstraram seu potencial em cenários clínicos, como na detecção e classificação de tumores cerebrais, alcançando altas taxas de acurácia, precisão e sensibilidade

ao utilizar abordagens de *fine-tuning* com *transfer learning* e visualizações com Grad-CAM para interpretação dos resultados (Baiju Babu Vimala *et al.*, 2023).

Por sua vez, o VGG16, com uma arquitetura composta por 16 camadas, é projetado para tarefas de classificação de imagens, utilizando filtros pequenos de 3x3 e *max pooling*. Embora seja computacionalmente intensivo, o VGG16 é amplamente utilizado em *transfer learning*. De acordo com Mascarenhas e Agarwal (2021), a profundidade adicional da rede, aliada à sua estrutura simples, pode melhorar significativamente o desempenho em tarefas de visão computacional, mostrando que redes mais profundas, como a VGG16, podem alcançar resultados excepcionais em classificação de imagens.

Em comparação, o ResNet-50, uma variação da arquitetura ResNet com 50 camadas profundas, é conhecido por sua introdução de blocos residuais, que permitem o treinamento de redes muito profundas sem problemas de degradação de desempenho. Essa característica o torna uma escolha popular para tarefas que exigem alta precisão. No estudo de Agughasi *et al.* (2021), o ResNet-50 se destacou ao alcançar 97,3% de acurácia e 99,2% de recall na classificação de imagens de raios-X, demonstrando sua eficácia e desempenho superior em tarefas de diagnóstico por imagem. Esses fatores contribuíram para a escolha desses modelos, visando otimizar o desempenho e a eficiência no processo de classificação.

A implementação foi realizada em Python 3.10, utilizando bibliotecas como TensorFlow, Keras, NumPy, Pandas e Matplotlib, executadas na plataforma Google Colab com acelerador de hardware (GPU), proporcionando o suporte necessário para o treinamento dos modelos.

As imagens foram organizadas em três conjuntos distintos: treinamento, validação e teste, com base na classificação binária entre lesões benignas e malignas. As imagens foram identificadas e separadas de acordo com as classes presentes no arquivo CSV de rótulos, considerando como malignas as classes MEL, BCC, AK e SCC, e como benignas as classes NV, BKL, DF e VASC. A proporção final utilizada foi de 75% para treinamento, 15% para validação e 10% para teste, garantindo a consistência na avaliação do desempenho dos modelos e evitando sobreposição entre os conjuntos.

Durante o treinamento, foi aplicada a técnica de transferência de aprendizado utilizando as arquiteturas previamente mencionadas, carregadas com os pesos do obtidos no conjunto de dados ImageNet com suas camadas convolucionais inicialmente congeladas para preservar o conhecimento adquirido anteriormente.

As imagens foram processadas com a classe *ImageDataGenerator*, redimensionadas para 224×224 pixels e organizadas em lotes de 32, otimizando o uso da memória e facilitando o treinamento em etapas. Foram adicionadas camadas densas customizadas no topo de cada arquitetura, com o objetivo de adaptar os modelos à tarefa de classificação proposta, compostas por uma camada *GlobalAveragePooling2D*, que reduz a dimensionalidade do mapa de características — ou seja, transforma a saída tridimensional da base convolucional em uma estrutura vetorial mais compacta, preservando a informação espacial mais relevante. Em seguida, foram adicionadas duas camadas *Dropout* com taxa de 0,5, atuando como técnica de regularização para reduzir o risco de *overfitting*, intercaladas por uma camada densa com 256 neurônios e ativação ReLU, o que possibilita a aprendizagem de padrões mais complexos e não lineares por parte do modelo. A

camada final, com ativação *softmax*, foi responsável por realizar a classificação entre as classes presentes no conjunto de dados.

A compilação dos modelos foi realizada com o otimizador Adam, selecionado por sua eficiência e estabilidade durante o treinamento, utilizando uma taxa de aprendizagem de 0,0001. As métricas de avaliação incluíram *accuracy*, *mean squared error*, *precision* e *recall*, fornecendo uma análise abrangente do desempenho dos modelos, especialmente relevante em cenários com possíveis desequilíbrios entre classes. A Tabela 2 apresenta os principais hiperparâmetros utilizados no experimento, incluindo o tamanho do batch, o otimizador, as camadas customizadas aplicadas ao topo da arquitetura base, a taxa de *dropout*, os callbacks utilizados e os parâmetros.

Tabela 2. Conjunto de Hiperparâmetros utilizados nos modelos.

Hiperparâmetros	**Valores**
Batch size	32
Algoritmo de aprendizado	Adam Optimizer
Taxa de aprendizado	1e-4
Tipo de transferência	*Transfer Learning* com *Fine-Tuning* Parcial
Camadas personalizadas	*GlobalAveragePooling2D, Dropout (rate = 0.5), Dense (256 units, activation = ReLU), Dropout (rate = 0.5), Dense (number of classes, activation = Softmax)*
Função de ativação	ReLU e Softmax
Função de perda	*binary-cross-entropy*
Callbacks	*ModelCheckpoint e EarlyStopping*
Early stopping	*Patience* = 30 *epochs*
Épocas	150

Fonte: Elaboração própria.

O treinamento foi conduzido por até 150 épocas, com aplicação da técnica de *EarlyStopping*, configurada para interromper o processo automaticamente caso não houvesse melhora consistente na validação após 30 épocas consecutivas. Nesse caso, os melhores pesos obtidos durante o treinamento foram restaurados. Além disso, foi utilizado o *callback ModelCheckpoint*, responsável por salvar o modelo com melhor desempenho ao longo das épocas. Visando assegurar uma comparação

justa entre as arquiteturas, todos os modelos foram treinados sob as mesmas condições experimentais e com os mesmos hiperparâmetros, permitindo uma avaliação consistente dos resultados.

4 RESULTADOS

Os experimentos foram conduzidos com três arquiteturas de CNNs — EfficientNetB0, VGG16 e ResNet50 — utilizando o conjunto de dados ISIC 2019. Após o treinamento e avaliação dos modelos, os melhores resultados foram obtidos com pelo ResNet50, que alcançou uma acurácia de 83%, demonstrando sua eficácia na tarefa de classificação entre lesões malignas e benignas. Os valores de *Accuracy*, *F1-Score*, Precisão e *Recall* obtidos para cada modelo são apresentados na Tabela 3, reforçando o desempenho superior do ResNet50 em relação aos demais.

Tabela 3. Comparativo dos algoritmos analisados na base de dados ISIC 2019.

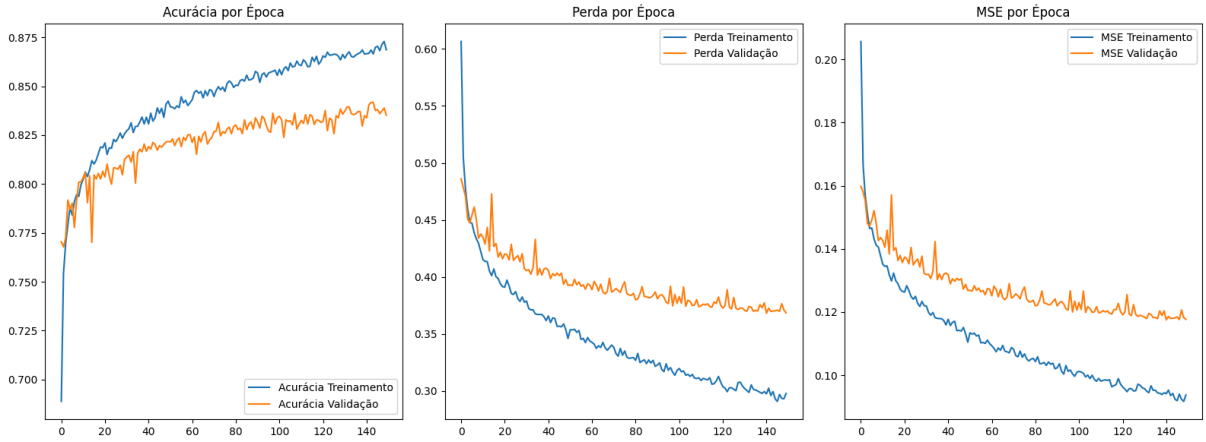
Modelo	Accuracy	F1-Score	Precision	Recall
ResNet50	0,83	0,83	0,83	0,83
EfficientNetB0	0,78	0,77	0,79	0,77
VGG16	0,78	0,78	0,78	0,79

Fonte: Elaboração própria (2025).

Para aprofundar a análise dos resultados, as Figuras 8, 9 e 10 apresentam, respectivamente, as curvas de acurácia, perda *(loss)* e erro quadrático médio (MSE) dos modelos EfficientNetB0, ResNet50 e VGG16 ao longo do treinamento. O ResNet50 também obteve acurácia em torno de 0,83. As curvas de validação e treinamento mantiveram-se próximas, evidenciando estabilidade e consistência. O EfficientNetB0 destacou-se com uma acurácia de validação em torno de 0,85, apresentando curvas de perda e MSE consistentemente decrescentes, o que indica boa generalização e ausência de *overfitting*. O VGG16, embora tenha alcançado desempenho final comparável, exigiu mais de 50 épocas para estabilização, refletindo uma curva de aprendizado mais lenta e maior demanda computacional. Ainda assim, as métricas obtidas foram satisfatórias, com bom desempenho na classificação.

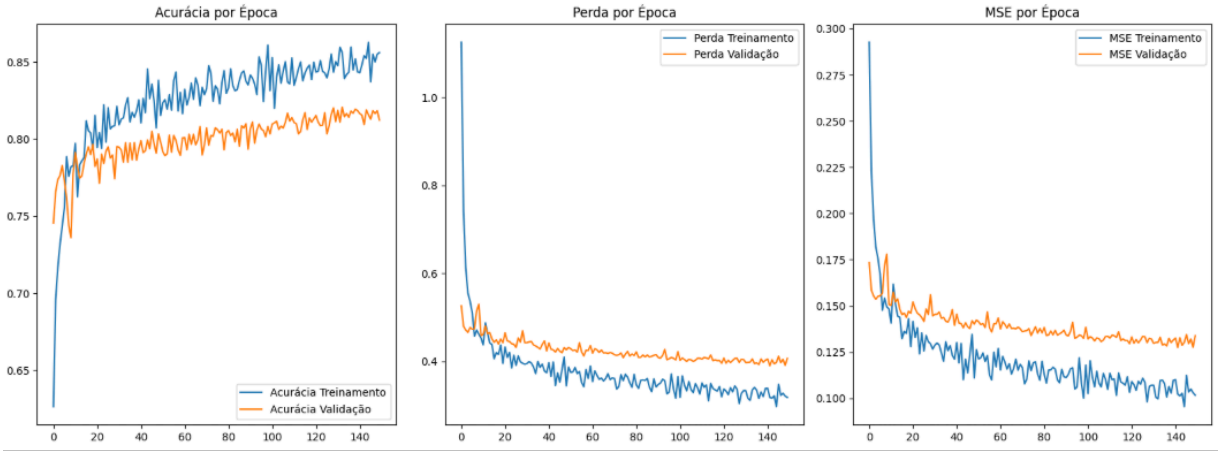
De forma geral, os três modelos apresentaram resultados consistentes em termos de acurácia. No entanto, ResNet50 e o EfficientNetB0 se sobressaem por combinarem desempenho e eficiência computacional, sendo, portanto, as alternativas mais adequadas para a tarefa proposta.

Figura 8. Curva de acurácia, perda e MSE do modelo EfficientNetB0 na ISIC 2019.



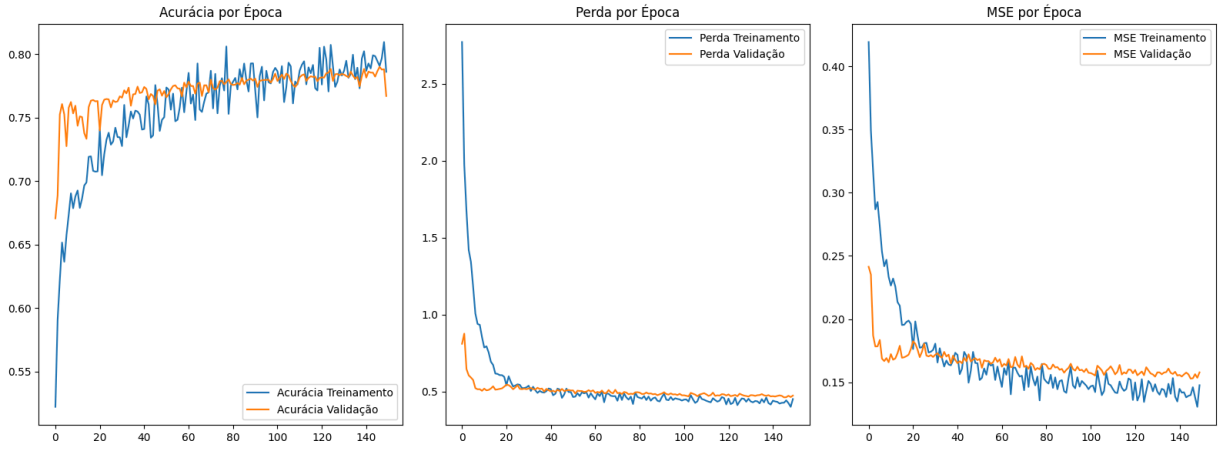
Fonte: Elaboração própria (2025).

Figura 9. Curva de acurácia, perda e MSE do modelo ResNet50 na ISIC 2019.



Fonte: Elaboração própria (2025).

Figura 10. Curva de acurácia, perda e MSE do modelo VGG16 na ISIC 2019.

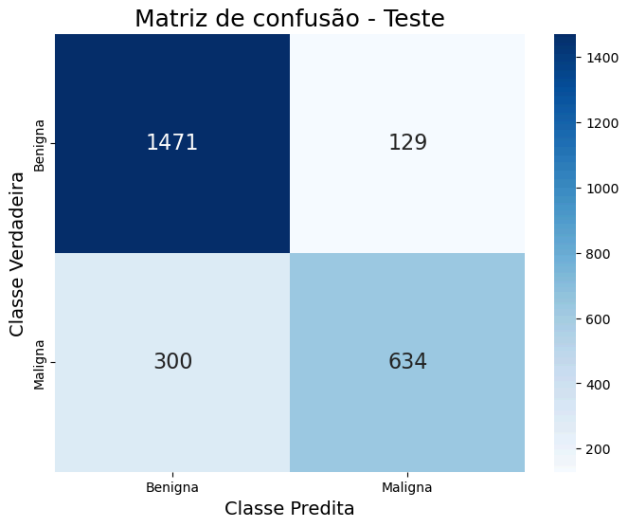


Fonte: Elaboração própria (2025).

A matriz de confusão permite uma avaliação detalhada da capacidade de cada arquitetura em discriminar entre lesões de pele benignas e malignas Na Figura

11 é apresentada a matriz de confusão do modelo ResNet50, que demonstrou um desempenho robusto, com resultados expressivos tanto na sensibilidade quanto na especificidade. No conjunto de teste, o modelo classificou corretamente 1.471 lesões benignas (verdadeiros negativos) e 634 lesões malignas (verdadeiros positivos), apresentando 300 falsos negativos e 129 falsos positivos. Esse equilíbrio entre os tipos de acertos e erros reflete uma boa capacidade do modelo em identificar corretamente casos malignos e minimizar alarmes falsos, especialmente relevantes no contexto clínico. Embora o número de falsos negativos não seja o menor entre os modelos, ele representa um equilíbrio, pois possui um baixo número de falsos positivos que contribui para maior precisão. Além disso, o desempenho consistente entre os conjuntos de treinamento e teste reforça a capacidade de generalização do ResNet50.

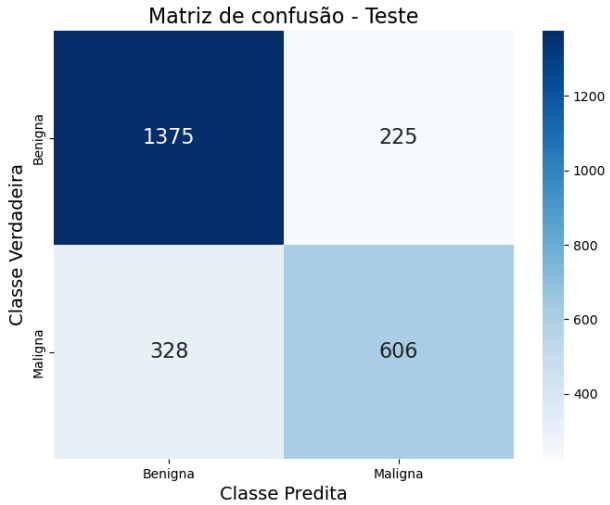
Figura 11. Matriz de confusão ResNet50 na base de dados ISIC 2019.



Fonte: Elaboração própria (2025).

O EfficientNetB0, embora tenha alcançado bons resultados em termos absolutos, com 1.375 verdadeiros negativos e 606 verdadeiros positivos, registrou 328 falsos negativos e 225 falsos positivos, como exibido na Figura 12, indicando uma sensibilidade reduzida em comparação ao ResNet50, o que pode comprometer a detecção precoce de lesões malignas. Ainda assim, o modelo mostrou estabilidade entre os conjuntos de treinamento e teste, demonstrando previsibilidade no comportamento geral.

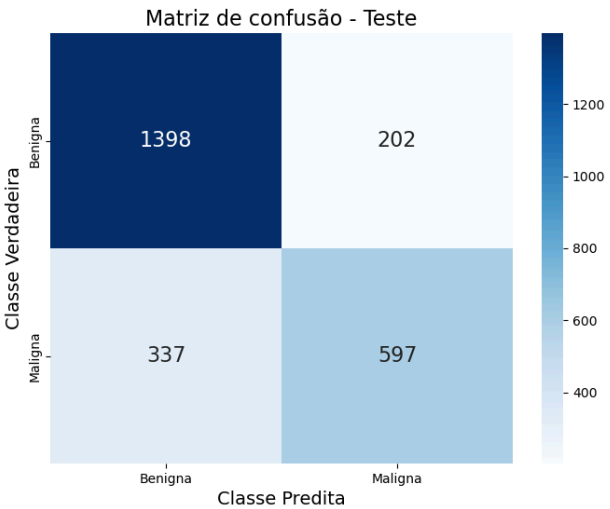
Figura 12. Matriz de confusão EfficienteNetB0 na base de dados ISIC 2019.



Fonte: Elaboração própria (2025).

Por sua vez, o modelo VGG16 obteve desempenho intermediário, com algumas particularidades. Conforme mostrado na Figura 13, onde é apresentada a matriz de confusão do modelo, foram registrados 202 falsos negativos, o menor valor entre os três modelos, o que indica maior sensibilidade, ou seja, maior capacidade de identificar corretamente lesões malignas. No entanto, esse resultado veio acompanhado de 337 falsos positivos, o maior número de falsos positivos entre os modelos analisados. Essa alta taxa de falsos positivos reduz a especificidade e pode levar a um aumento de diagnósticos incorretos de malignidade, acarretando possíveis consequências clínicas indesejadas, como exames adicionais desnecessários.

Figura 13. Matriz de confusão VGG16 na base de dados ISIC 2019.



Fonte: Elaboração própria (2025).

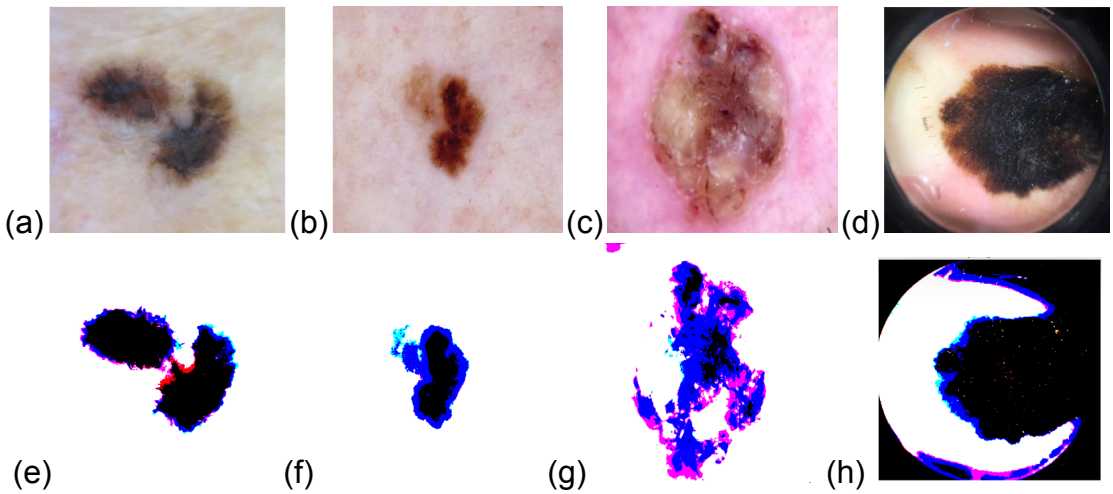
Dessa forma, com base na análise das matrizes de confusão, o modelo ResNet50 apresentou o melhor equilíbrio entre sensibilidade e especificidade, demonstrando maior confiabilidade na tarefa de classificação de lesões de pele. Estes resultados reforçam seu potencial para aplicação em ambientes clínicos, nos quais a precisão diagnóstica e a minimização de erros críticos são fundamentais.

4.1 Explicabilidade

Com o objetivo de tornar as decisões dos modelos mais interpretáveis, foram aplicadas três técnicas de XAI: LIME, Grad-CAM e SHAP. As imagens das Figuras 14, 15 e 16 ilustram, respectivamente, a aplicação dessas abordagens sobre imagens do conjunto de dados, destacando as regiões consideradas mais relevantes para a classificação. Para essa etapa, o modelo escolhido foi o ResNet50, uma vez que apresentou o melhor desempenho em termos de acurácia, *F1-Score*, *recall* e *precison*, sendo assim, tornou-se particularmente relevante analisar seu comportamento por meio das técnicas de XAI, a fim de compreender quais características visuais ele considera mais importante.

Cada técnica de XAI foi avaliada quanto à clareza visual das regiões de interesse. Essa análise contribui para uma melhor compreensão do comportamento dos modelos, especialmente em contextos clínicos sensíveis. A Figura 14 apresenta os resultados obtidos com o uso da técnica LIME sobre o conjunto de dados.

Figura 14. Explicabilidade com Lime para o modelo ResNet50.

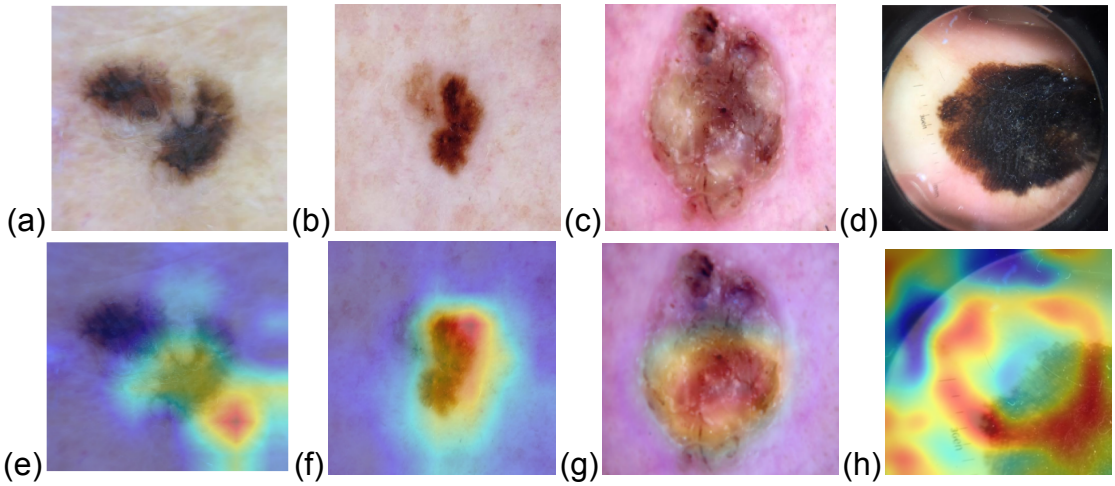


Fonte: Elaboração própria (2025).

Com o uso do LIME, foram geradas explicações locais com base na segmentação da imagem em *superpixels*. A técnica avalia a variação na predição ao ocultar ou manter determinadas regiões, permitindo identificar quais partes da imagem influenciaram positivamente ou negativamente a decisão. Nas visualizações, as regiões destacadas em azul representam contribuição positiva

significativa para a classe prevista, enquanto as em vermelho ou neutras indicaram menor relevância, ou possível influência negativa. Dando continuidade à análise explicativa, a Figura 15 apresenta a aplicação da técnica Grad-CAM.

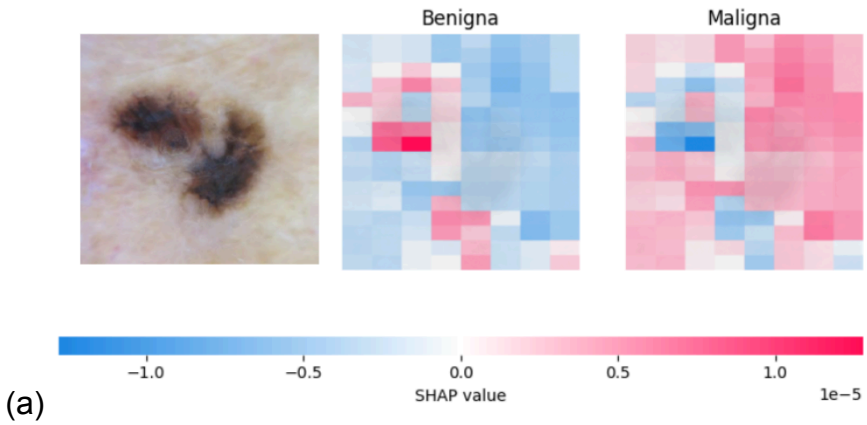
Figura 15. Explicabilidade com Grad-CAM para o modelo ResNet50.

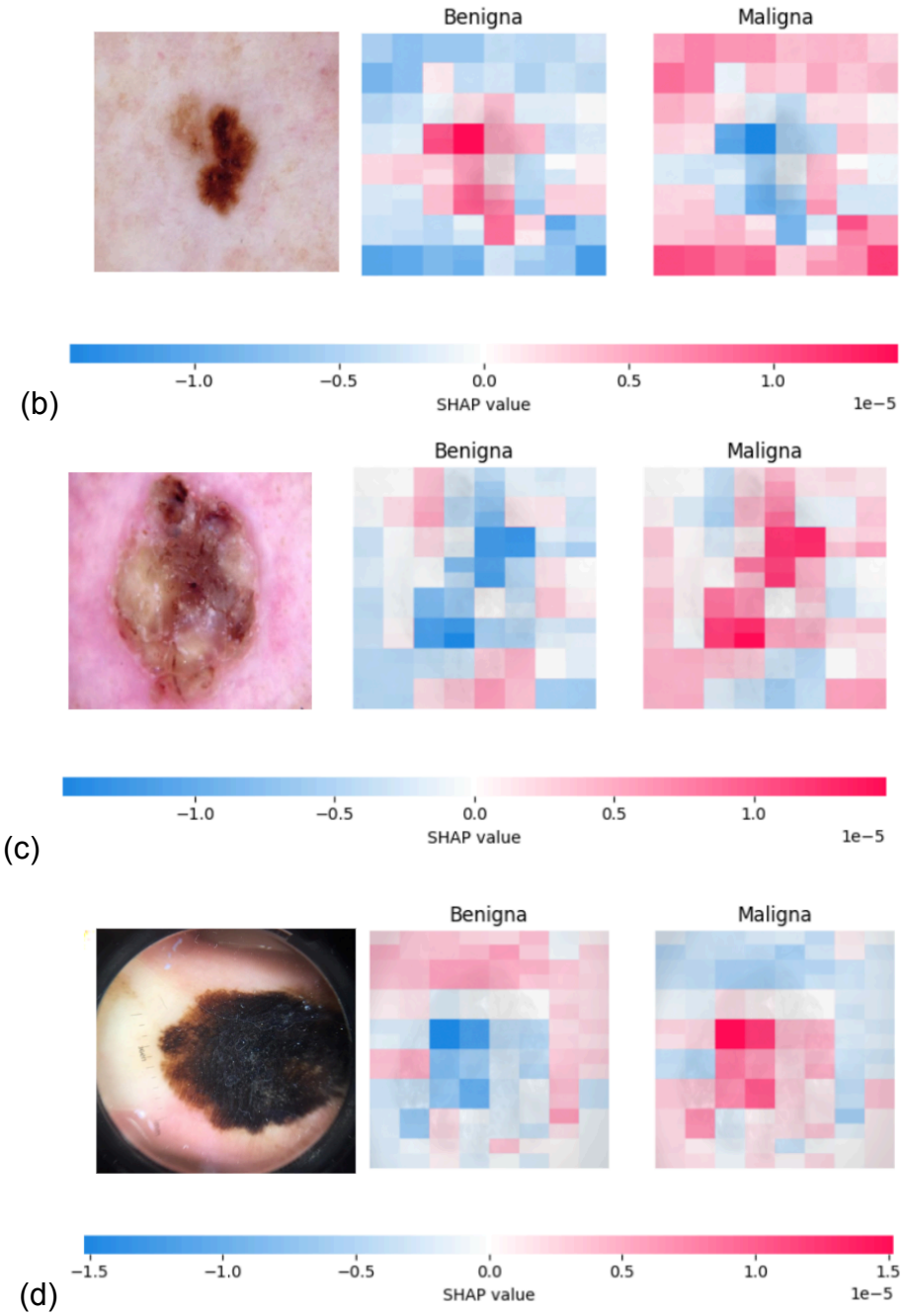


Fonte: Elaboração própria (2025).

Por meio do Grad-CAM, foi possível gerar mapas de calor que evidenciam as áreas com maior influência sobre a decisão do modelo. Regiões com tons mais quentes, como vermelho e amarelo, indicam forte ativação dos neurônios da rede, enquanto áreas em azul e roxo indicam menor relevância ou até mesmo ausência de contribuição. Em sequência, a Figura 16 ilustra os resultados do método SHAP, que oferece uma perspectiva mais detalhada ao quantificar a contribuição individual de cada região da imagem para a previsão realizada.

Figura 16. Explicabilidade com SHAP para o modelo ResNet50.





Fonte: Elaboração própria (2025).

O SHAP é fundamentado na teoria dos jogos e calcula o valor de Shapley para cada pixel ou região, representando a contribuição marginal daquela parte da imagem em diferentes combinações possíveis de entrada. Áreas em vermelho ou rosa indicam impacto positivo na predição, enquanto as em azul representam influência negativa.

As regiões com coloração mais intensa, textura irregular ou presença de lesões bem definidas foram consistentemente destacadas pelas três abordagens,

indicando sua relevância para a tomada de decisão. Essas áreas frequentemente coincidem com padrões visuais associados a características malignas. Por outro lado, imagens contendo lesões benignas ou sinais menos evidentes tendem a apresentar menor ativação nas visualizações, com destaque mais disperso ou até ausente, refletindo uma contribuição reduzida para a predição final do modelo.

A partir de uma análise não clínica das saídas das técnicas de explicabilidade, observou-se que, em geral, observou-se uma correspondência visual clara entre as regiões destacadas e áreas da pele com alterações visuais, sugerindo que as explicações fornecem uma base confiável para a interpretação do modelo. Assim, as três técnicas demonstraram coerência na identificação das regiões visualmente relevantes, indicando que o modelo ResNet50 aprendeu padrões morfológicos significativos para a tarefa.

Entretanto, também foram observadas situações em que o modelo direcionou atenção para regiões secundárias, como áreas periféricas ou o fundo da imagem, conforme ilustrado na Figura 15 (e) e na Figura 16 (a). No caso do LIME, verificou-se que, em imagens em que o modelo errou a classificação, as regiões destacadas nem sempre correspondiam a áreas visualmente relevantes, evidenciando limitações inerentes à explicabilidade local que podem comprometer a confiança nas predições. Já o Grad-CAM apresentou, mesmo em classificações incorretas, ativações em regiões indicativas de alterações visuais, o que pode indicar que o modelo foca em padrões que não necessariamente correspondem a sinais clinicamente relevantes, apontando para possíveis vieses ou necessidade de refinamento dos dados e da arquitetura. O SHAP forneceu uma visão detalhada da contribuição individual das regiões, mas em alguns casos exibiu distribuições de importância dispersas ou incoerentes.Esses achados evidenciam possíveis vieses no processo de treinamento e ressaltam a necessidade de investigações adicionais para compreender suas causas, além de ajustes nos dados ou na arquitetura do modelo para aprimorar a robustez das predições.

Por fim, é importante destacar que, embora as técnicas de explicabilidade aumentem significativamente a transparência dos modelos de inteligência artificial, elas não substituem o julgamento especializado. A interpretação dos resultados deve sempre ser realizada em conjunto com o conhecimento clínico apropriado. Nesse contexto, a Inteligência Artificial Explicável se configura não apenas como uma ferramenta técnica, mas também como um recurso valioso para o desenvolvimento de sistemas de suporte à decisão médica que sejam mais seguros, auditáveis e centrados no benefício do paciente.

5. CONCLUSÃO

A aplicação de redes neurais convolucionais para a classificação de lesões de pele mostrou-se eficaz, especialmente quando combinada a técnicas de Inteligência Artificial Explicável, conforme proposto neste estudo. Entre os modelos avaliados, o ResNet50 apresentou o melhor desempenho, destacando-se em métricas como acurácia, precisão e *F1-score*, revelando-se uma alternativa promissora para apoio clínico no diagnóstico de doenças dermatológicas.

Dentre as técnicas de XAI utilizadas — LIME, Grad-CAM e SHAP —, Grad-CAM e LIME foram os que evidenciaram com mais clareza as regiões das

imagens que influenciaram a decisão do modelo. Essas técnicas permitiram identificar os padrões visuais mais relevantes para a classificação, o que pode contribuir diretamente para o aumento da confiança dos profissionais da saúde no uso da IA. No entanto, algumas inconsistências foram observadas nas explicações, como o destaque de regiões irrelevantes ou fora da lesão, sobretudo em casos de erro do modelo. Essas falhas foram identificadas principalmente na presença de artefatos nas imagens (pelos, sombras, reflexos), ou ainda à presença de padrões visuais no fundo, que confundem o modelo e geram correlações equivocadas.

Esses fatores apontam oportunidades de melhorias e direções futuras, como incorporar técnicas e métricas que permitem uma avaliação mais robusta das explicações fornecidas, pelas abordagens de XAI. Ademais, propõe-se a integração de dados clínicos (como idade, histórico familiar, tipo de pele, grau de exposição solar), às imagens, de modo a enriquecer o processo de classificação e tornar os modelos mais completos e contextualizados. Outras possibilidades de expansão do trabalho incluem explorar estratégias para melhorar a acurácia do modelo e nos hiperparâmetros utilizados.

Além dos aspectos técnicos, este estudo reforça a importância de desenvolver sistemas de IA socialmente responsáveis. Isso envolve promover diversidade nas equipes de desenvolvimento, avaliar os impactos de previsões incorretas — especialmente em áreas sensíveis como a saúde — e contar com o apoio de especialistas humanos para uma camada adicional de segurança. A construção de soluções éticas, confiáveis e centradas no paciente exige um esforço interdisciplinar que une conhecimento técnico, sensibilidade ética e perspectiva humana.

Por fim, o objetivo central deste trabalho — identificar a abordagem mais eficaz para interpretar os resultados dos modelos de diagnóstico — foi alcançado com a comparação sistemática entre arquiteturas de CNN e métodos de XAI, destacando a combinação ResNet50 e Grad-CAM como uma solução promissora no apoio ao diagnóstico médico.

REFERÊNCIAS

AGUGHASI, V.; MURALI, S.; DEEPU, R.; SHIVAMURTHY, R. C. ResNet-50 vs VGG-19 vs training from scratch: a comparative analysis of the segmentation and classification of pneumonia from chest X-ray images. **Global Transitions Proceedings**, [S.l.], v. 2, n. 2, p. 375–381, 2021. Disponível em: https://doi.org/10.1016/j.gltp.2021.08.027. Acesso em: 16 maio 2025.

ARAUJO, T.; ARESTA, G.; CASTRO, E.; ROUCO, J.; AGUIAR, P.; ELOY, C.; POLÓNIA, A.; CAMPILHO, A. Classification of breast cancer histology images using Convolutional Neural Networks. **Plos One**, v. 12, n. 6, e0177544, 2017. Disponível em: https://doi.org/10.1371/journal.pone.0177544. Acesso em: 02 mar. 2025.

BADER, F.; KOPRINKA, I.; SHATZ, S. et al. Explainable Artificial Intelligence for Skin Lesion Classification Using LIME and SHAP: a comparative study. **Diagnostics**, v. 13, n. 11, 1932, 2023. Disponível em: https://www.mdpi.com/2075-4418/13/11/1932. Acesso em: 16 mai. 2025.

BASHIR, A.; ALI, H. M.; AHMED, F.; et al. Skin cancer diagnosis using VGG16 and transfer learning: analyzing the effects of data quality over quantity on model efficiency. **Applied Sciences**, v. 14, n. 17, 7447, 2024. Disponível em: https://www.mdpi.com/2076-3417/14/17/7447. Acesso em: 17 abr. 2025.

BRASIL. Ministério da Saúde. Exposição solar: saiba como se proteger e prevenir o câncer de pele, 05 jan. 2023. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-brasil/prevencao-ao-cancer/dezembro-laranja-prevencao-e-deteccao-precoce-do-cancer-de-pele. Acesso em: 22 jan. 2025.

BRASIL. Ministério da Saúde. Câncer de pele representa 30% de todos os tumores malignos registrados no país. 2025. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-com-ciencia/noticias/2025/fevereiro/cancer-de-pele-representa-30-de-todos-os-tumores-malignos-registrados-no-pais. Acesso em: 30 maio 2025.

FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 25-30, jan./abr, 2006. Disponível em: https://www.scielo.br/j/ci/a/SQ9myjZWLxnyXfstXMgCdcH/?format=pdf&lang=pt. Acesso em: 4 mar. 2025.

HAUG, C. J.; DRAZEN, J. M. Artificial intelligence and machine learning in clinical medicine. **New England Journal of Medicine**, v. 388, n. 13, p. 1201-1208, 2023. Disponível em: https://www.nejm.org/doi/full/10.1056/NEJMra2302038. Acesso em: 03 jan. 2025.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. **IEEE Conference on Computer Vision and Pattern Recognition**, 2015. Disponível em: https://arxiv.org/abs/1512.03385. Acesso em: 16 mai. 2025.

HOMA, Celina. Evaluating XAI algorithms in skin cancer classification: a path towards trustworthy AI systems. Trabalho de Conclusão de Curso (Bacharelado) – School of Engineering, Zurich University of Applied Sciences (ZHAW), Winterthur, 2023. Disponível em: https://digitalcollection.zhaw.ch/handle/11475/29936. Acesso em: 02 jun. 2025.

INCA. INSTITUTO NACIONAL DO CÂNCER. Estimativa de 2023: incidência de câncer no Brasil, 2023. Disponível em: https://www.inca.gov.br. Acesso em: 05 fev. 2025.

ISIC. International Skin Imaging Collaboration Project, 2018. Disponível em: https://www.isic-archive.com/. Acesso em: 15 jan. 2025.

IQBAL, A.; JAFFAR, M. A.; JAHANGIR, R. Enhancing brain tumour multi-classification using Efficient-Net B0-based intelligent diagnosis for Internet of Medical Things (IoMT) applications. **Information**, v. 15, n. 8, p. 489, 2024. Disponível em: https://www.mdpi.com/2078-2489/15/8/489. Acesso em: 19 abr. 2025.

LINARDATOS, P.; PAPASTEFANOPOULOS, V.; KOTSIANTIS, S. Explainable AI: a review of machine learning interpretability methods. **Entropy**, v. 23, n. 1, p. 18, 2021. Disponível em: https://www.mdpi.com/1099-4300/23/1/18. Acesso em: 24 jan. 2025.

MAHBOD, A.; SCHAEFER, G.; WANG, C.; ECKER, R.; DORFFNER, G.; ELLINGER, I.Investigating and Exploiting Image Resolution for Transfer Learning-based Skin Lesion Classification, 2020 25th **International Conference on Pattern Recognition (ICPR)**, Milan, Italy, pp. 4047-4053, 2021. Disponível em: https://doi.org/10.1109/ICPR48806.2021.9412307. Acesso em: 4 jun. 2025.

MAHARANA, K.; MONDAL, S.; NEMADE, B. A review: data pre-processing and data augmentation techniques. **Global Transitions Proceedings**, v. 3, n. 1, p. 91–99, 2022. Disponível em: https://www.sciencedirect.com/science/article/pii/S2666285X22000565. Acesso em: 05 jan. 2025.

MASCARENHAS, S.; AGARWAL, M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. **International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications**, 2021. Disponível em: https://www.researchgate.net/publication/370467299. Acesso em: 16 mai. 2025.

MARON, R. C. et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. **European Journal of Cancer**, v. 156, p. 202–216, 2021. ISSN 0959-8049. Disponível em: https://www.sciencedirect.com/science/article/pii/S0959804921004445. Acesso em: 24 jan. 2025.

MRIDHA, K.; UDDIN, M. M.; SHIN, J.; KHADKA, S.; MRIDHA, M. F. An interpretable skin cancer classification using optimized convolutional neural network for a smart healthcare system. **IEEE Access**, v. 11, p. 41003–41018, 2023. Disponível em: https://doi.org/10.1109/ACCESS.2023.3269694. Acesso em: 4 jun. 2025.

MUMUNI, A.; MUMUNI, F. Data augmentation: a comprehensive survey of modern techniques. **Array**, v. 16, p. 100258, 2022. Disponível em: https://doi.org/10.1016/j.array.2022.100258. Acesso em: 7 jan. 2025.

MURDIYANTO, R.; ASYHARI, A. T.; PRAMONO, R. E.; REZAL, M. Performance comparison of CNN and ResNet50 for skin cancer classification using U-Net segmented images. **Indonesian Journal of Data and Science**, 2024. Disponível em: https://www.researchgate.net/publication/387582441. Acesso em: 16 mai. 2025.

NIGAR, N.; UMAR, M.; SHAHZAD, M. K.; ISLAM, S.; ABALO, D. A deep learning approach based on explainable artificial intelligence for skin lesion classification. **IEEE Access**, v. 10, p. 113715–113725, 2022. Disponível em: https://doi.org/10.1109/ACCESS.2022.3217217. Acesso em: 4 jun. 2025.

PEREIRA, K. et al. Fatores que influenciam a aceitação de Tecnologias de Inteligência Artificial na Saúde. **Revista Gestão & Saúde**, v. 13, p. 02–20, 2022.

Disponível em: https://www.researchgate.net/publication/378383096. Acesso em: 29 jan. 2025.

SALIH, Ahmed M. et al. A perspective on explainable artificial intelligence methods: SHAP and LIME. **Artificial Intelligence Systems**, 2024. Disponível em: https://doi.org/10.1002/aisy.202400304. Acesso em: 11 abr. 2025.

SILVA, T. A. M. et al. Aprendizado de máquina aplicado ao diagnóstico de câncer de pele. **Anais do Congresso Brasileiro de Inteligência Computacional (CBIC)**. 2021. Disponível em: https://sbic.org.br/wp-content/uploads/2021/09/pdf/CBIC_2021_paper_55.pdf. Acesso em: 26 jan. 2025.

SOCIEDADE BRASILEIRA DE DERMATOLOGIA. Melanoma. 2023. Disponível em: https://www.sbd.org.br. Acesso em: 03 jan. 2025.

VIMALA, B. B.; SRINIVASAN, S.; MATHIVANAN, S. K. et al. Detection and classification of brain tumor using hybrid deep learning models. **Scientific Reports**, v. 13, p. 23029, 2023. Disponível em: https://doi.org/10.1038/s41598-023-50505-6. Acesso em: 16 mai. 2025.

WORLD HEALTH ORGANIZATION. Skin diseases as a global public health priority: Executive Board Decision EB156(24). 2025. Disponível em: https://apps.who.int/gb/ebwha/pdf_files/EB156/B156_(24)-en.pdf. Acesso em: 30 maio 2025.

XU, F. et al. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. **International Conference on Artificial Intelligence for Industries (AI4I)**, 2019. Disponível em: https://www.researchgate.net/publication/336131051. Acesso em: 29 jan. 2025.