

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**

**HENRY GABRIEL REIS**

**LUIS FELIPE TAKAZONO HATSUMURA**

**LUIZ HENRIQUE LIPPMANN SALES**

**DESENVOLVIMENTO DE SISTEMA INTELIGENTE PARA DIAGNÓSTICO DE  
LESÕES DERMATOLÓGICAS**

**CURITIBA**

**2026**

**HENRY GABRIEL REIS  
LUIS FELIPE TAKAZONO HATSUMURA  
LUIZ HENRIQUE LIPPMANN SALES**

**DESENVOLVIMENTO DE SISTEMA INTELIGENTE PARA DIAGNÓSTICO DE  
LESÕES DERMATOLÓGICAS**

**DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR THE DIAGNOSIS OF  
DERMATOLOGICAL LESIONS**

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Elétrica do curso de Engenharia Elétrica e de Bacharel em Engenharia de Controle e Automação do curso de Engenharia de Controle e Automação da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Narco Afonso Ravazzoli Maciejewski

**CURITIBA**

**2026**



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

**HENRY GABRIEL REIS  
LUIS FELIPE TAKAZONO HATSUMURA  
LUIZ HENRIQUE LIPPMANN SALES**

**DESENVOLVIMENTO DE SISTEMA INTELIGENTE PARA DIAGNÓSTICO DE  
LESÕES DERMATOLÓGICAS**

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Elétrica do curso de Engenharia Elétrica e de Bacharel em Engenharia de Controle e Automação do curso de Engenharia de Controle e Automação da Universidade Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 25 / junho / 2026

---

Narco Afonso Ravazzoli Maciejewski  
Doutorado  
Universidade Tecnológica Federal do Paraná

---

Roberto Zanetti Freire  
Doutorado  
Universidade Tecnológica Federal do Paraná

---

Mariana Antonia Aguiar Furucho  
Doutorado  
Universidade Tecnológica Federal do Paraná

**CURITIBA  
2026**

*“Um ser humano deve transformar informação em inteligência ou conhecimento. Tendemos a esquecer que nenhum computador fará a pergunta ‘Por quê?’”*

(Grace Hopper, 1987)

## RESUMO

O câncer de pele representa um desafio significativo para a saúde pública mundial e brasileira, destacando-se o melanoma pela elevada agressividade e letalidade quando não identificado precocemente. Diante da subjetividade inerente à avaliação visual de lesões cutâneas e das barreiras de acesso a especialistas, este trabalho apresenta o desenvolvimento de uma abordagem computacional de apoio à análise de imagens dermatoscópicas baseada em técnicas de aprendizado profundo. O objetivo consiste em desenvolver e avaliar um classificador multi-classe capaz de diferenciar três categorias diagnósticas: melanoma, carcinoma basocelular e carcinoma espinocelular. Portanto, o sistema não realiza uma classificação binária entre pele saudável e não saudável, nem contempla uma classe de lesões benignas. A metodologia foi organizada em etapas de pré-processamento, segmentação semântica, extração da região de interesse e classificação. Para a segmentação, foram utilizadas imagens e máscaras manuais da base ISIC 2018 Task 1, empregando-se uma arquitetura U-Net associada ao codificador ResNet-50. Após o treinamento, o modelo segmentador foi aplicado às imagens selecionadas da base ISIC 2019, utilizadas na etapa de classificação. A classificação foi realizada com a arquitetura EfficientNet-B4, por meio de transferência de aprendizado, considerando três configurações experimentais com diferentes estratégias de aumento e balanceamento dos dados. O desempenho da segmentação foi avaliado pelo Coeficiente Dice e pelo Índice de Jaccard, alcançando valores de validação próximos de 0,75 e 0,64, respectivamente. Na classificação, a configuração inicial apresentou a maior área sob a curva média entre as classes, com AUC macro de 0,9421, e F1-Scores de 88,92% para melanoma e 83,59% para carcinoma basocelular. Entretanto, verificou-se menor desempenho para o carcinoma espinocelular, classe menos representada na base original. Em uma configuração experimental complementar com balanceamento sintético, a sensibilidade e o F1-Score dessa classe atingiram 85,10% e 78,13%, respectivamente. Os resultados evidenciam que o desbalanceamento e a composição dos conjuntos de dados influenciam o comportamento do modelo. Conclui-se que a abordagem desenvolvida apresenta potencial como base para ferramentas de apoio à diferenciação entre as três neoplasias estudadas, embora não deva ser interpretada como um sistema geral de detecção de lesões saudáveis, benignas ou malignas.

Palavras-chave: câncer de pele; classificação multiclasse; dermatoscopia; aprendizado profundo; segmentação semântica.

## ABSTRACT

Skin cancer represents a significant challenge to global and Brazilian public health, with melanoma standing out because of its high aggressiveness and lethality when not detected at an early stage. Given the inherent subjectivity of the visual assessment of skin lesions and the barriers to accessing specialists, this work presents the development of a computational approach based on deep learning techniques to support the analysis of dermoscopic images. The objective is to develop and evaluate a multiclass classifier capable of differentiating among three diagnostic categories: melanoma, basal cell carcinoma, and squamous cell carcinoma. Therefore, the proposed system does not perform binary classification between healthy and unhealthy skin and does not include a class representing benign lesions. The methodology was organized into preprocessing, semantic segmentation, region-of-interest extraction, and classification stages. For the segmentation stage, images and manually annotated masks from the ISIC 2018 Task 1 dataset were used to train a U-Net architecture with a ResNet-50 encoder. After training, the segmentation model was applied to selected images from the ISIC 2019 dataset, which were subsequently used in the classification stage. Classification was performed using the EfficientNet-B4 architecture through transfer learning, considering three experimental configurations with different data augmentation and class-balancing strategies. Segmentation performance was evaluated using the Dice Coefficient and the Jaccard Index, achieving validation values of approximately 0.75 and 0.64, respectively. In the classification stage, the initial configuration achieved the highest mean area under the curve across the classes, with a macro-AUC of 0.9421 and F1-scores of 88.92% for melanoma and 83.59% for basal cell carcinoma. However, lower performance was observed for squamous cell carcinoma, which was the least represented class in the original dataset. In a complementary experimental configuration employing synthetic balancing, the sensitivity and F1-score for this class reached 85.10% and 78.13%, respectively. The results indicate that class imbalance and dataset composition directly influence model behavior. It is concluded that the developed approach has potential as a foundation for tools designed to support the differentiation among the three investigated neoplasms, although it should not be interpreted as a general system for detecting healthy, benign, or malignant skin lesions.

Keywords: deep learning; dermoscopy; multiclass classification; semantic segmentation; skin cancer.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura anatômica da pele humana e suas camadas .....	17
Figura 2 – Camadas da epiderme e seus tipos celulares especializados .....	18
Figura 3 – Variantes clínicas do carcinoma basocelular .....	19
Figura 4 – Aspecto clínico do carcinoma espinocelular .....	20
Figura 5 – Aspecto clínico do melanoma maligno .....	21
Figura 6 – Estágios de crescimento do melanoma .....	22
Figura 7 – Subtipos clínicos de melanoma .....	24
Figura 8 – Critérios da regra ABCDE para detecção de melanoma .....	26
Figura 9 – Comparação entre a inspeção visual e a dermatoscopia .....	28
Figura 10 – Análise de uma lesão de melanoma e seu histograma .....	31
Figura 11 – Comparativo das funções de distribuição de brilho: melanoma vs. nevo benigno .....	31
Figura 12 – Processo de pré-processamento e segmentação em lesão com artefatos (pelos) .....	33
Figura 13 – Etapas típicas de um sistema de diagnóstico assistido por computador ..	34
Figura 14 – Comparativo de desempenho do modelo conforme sua complexidade ...	39
Figura 15 – Modelo matemático neurônio artificial .....	42
Figura 16 – Representação gráfica das principais funções de ativação .....	43
Figura 17 – Estrutura de uma rede neural de múltiplas camadas .....	44
Figura 18 – Representação da operação de convolução: entrada, kernel e mapa de características .....	45
Figura 19 – Fluxo metodológico proposto para segmentação e classificação das imagens dermatoscópicas .....	66
Figura 20 – Etapas completas de pré-processamento aplicadas a uma amostra de Carcinoma Espinocelular .....	72
Figura 21 – Exemplos de transformações aplicadas durante o aumento de dados para a amostra ISIC_0053457 .....	74
Figura 22 – Progresso das etapas de processamento e segmentação de uma lesão dermatoscópica da base ISIC .....	76
Figura 23 – Curvas de treinamento e validação da U-Net .....	83
Figura 24 – Exemplo de máscara com delimitação relativamente adequada .....	85
Figura 25 – Exemplo de máscara com menor precisão na delimitação da lesão .....	85
Figura 26 – Curvas de treinamento e validação do classificador na configuração inicial .....	87
Figura 27 – Matriz de confusão do classificador na configuração inicial .....	88
Figura 28 – Curvas de treinamento e validação do classificador sem aumento de dados .....	89
Figura 29 – Matriz de confusão do classificador sem aumento de dados .....	90
Figura 30 – Curvas de treinamento e validação do classificador com balanceamento sintético .....	91
Figura 31 – Matriz de confusão do classificador com balanceamento sintético .....	92

## LISTA DE TABELAS

<b>Tabela 1 – Comparação das métricas de desempenho entre as bases de Ferreira (2018) .....</b>	<b>59</b>
<b>Tabela 2 – Distribuição das imagens selecionadas para a etapa de classificação. ...</b>	<b>68</b>
<b>Tabela 3 – Organização dos pares imagem-máscara utilizados na etapa de segmentação.....</b>	<b>70</b>
<b>Tabela 4 – Distribuição inicial das imagens da etapa de classificação nos conjuntos de treinamento, validação e teste.....</b>	<b>70</b>
<b>Tabela 5 – Principais hiperparâmetros utilizados no treinamento do classificador. ..</b>	<b>78</b>
<b>Tabela 6 – Principais diferenças entre as configurações experimentais do classificador.....</b>	<b>80</b>
<b>Tabela 7 – Distribuição das imagens nas três configurações experimentais do classificador.....</b>	<b>81</b>
<b>Tabela 8 – Resultados comparativos dos três classificadores no conjunto de teste .</b>	<b>93</b>

## LISTA DE ABREVIATURAS E SIGLAS

ABCDE	Regra clínica: Assimetria, Borda, Cor, Diâmetro, Evolução
AC	Acurácia, do inglês <i>Accuracy</i>
AJCC	<i>American Joint Committee on Cancer</i>
ALM	Melanoma Lentiginoso Acral, do inglês <i>Acral Lentiginous Melanoma</i>
API	<i>Application Programming Interface</i>
AUC	Área sob a Curva, do inglês <i>Area Under the Curve</i>
AUC-ROC	Área sob a Curva de Característica de Operação do Receptor, do inglês <i>Area Under the Receiver Operating Characteristic Curve</i>
BCC	<i>Basal Cell Carcinoma</i> (Carcinoma Basocelular)
BCE	Entropia Cruzada Binária, do inglês <i>Binary Cross-Entropy</i>
BKL	Ceratose Benigna, do inglês <i>Benign Keratosis</i>
CAD	Diagnóstico Auxiliado por Computador, do inglês <i>Computer-Aided Diagnosis</i>
CBC	Carcinoma Basocelular
CCE	Entropia Cruzada Categórica, do inglês <i>Categorical Cross-Entropy</i>
CEC	Carcinoma Espinocelular
CSS	<i>Cascading Style Sheets</i>
CSV	Valores separados por vírgula, do inglês <i>Comma-Separated Values</i>
DA	Aumento de Dados, do inglês <i>Data Augmentation</i>
DL	Aprendizado Profundo, do inglês <i>Deep Learning</i>
DNA	Ácido Desoxirribonucleico
FC6 / FC7	<i>Fully Connected</i> (Camadas totalmente conectadas)
FCM	<i>Fuzzy C-Means</i>
FL	Perda Focal, do inglês <i>Focal Loss</i>
FN	Falso Negativo, do inglês <i>False Negative</i>
FP	Falso Positivo, do inglês <i>False Positive</i>
GPU	Unidade de Processamento Gráfico, do inglês <i>Graphics Processing Unit</i>
GT	<i>Ground Truth</i>
HPV	Papilomavírus Humano
HTML	<i>Hypertext Markup Language</i>
IA	Inteligência Artificial
IARC	Agência Internacional de Pesquisa em Câncer
IHP	<i>Independent Histogram Pursuit</i>
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge</i>
IoU	Índice de Jaccard ou Intersecção sobre União, do inglês <i>Intersection over Union</i>
ISIC	<i>International Skin Imaging Collaboration</i>
K-NN	<i>K-Nearest Neighbor</i>

LMM	Melanoma Lentigo Maligno, do inglês <i>Lentigo Maligna Melanoma</i>
ML	Aprendizado de Máquina, do inglês <i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
NM	Melanoma Nodular, do inglês <i>Nodular Melanoma</i>
NV	Nevo Melanocítico, do inglês <i>Melanocytic Nevus</i>
ODS	Objetivos de Desenvolvimento Sustentável
OVR	<i>One-Vs-Rest</i>
P	Precisão, do inglês <i>Precision</i>
PDI	Processamento Digital de Imagens
PH <sup>2</sup>	Base de dados do Hospital Pedro Hispano
RGB	<i>Red, Green, Blue</i>
RGP	Fase do Crescimento Radial, do inglês <i>Radial Growth Phase</i>
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
SBD	Sociedade Brasileira de Dermatologia
SCC	<i>Squamous Cell Carcinoma</i> (Carcinoma Espinocelular)
SE	Sensibilidade, do inglês <i>Sensitivity</i>
SP	Especificidade, do inglês <i>Specificity</i>
SSM	Melanoma de Disseminação Superficial, do inglês <i>Superficial Spreading Melanoma</i>
SVM	Máquina de Vetor de Suporte, do inglês <i>Support Vector Machine</i>
TDS	<i>Total Dermatoscopy Score</i>
TL	Transferência de Aprendizado, do inglês <i>Transfer Learning</i>
TN	Verdadeiro Negativo, do inglês <i>True Negative</i>
TP	Verdadeiro Positivo, do inglês <i>True Positive</i>
UV	Ultravioleta
VC	Visão Computacional
VGG	<i>Visual Geometry Group</i>
VGP	Fase de Crescimento Vertical, do inglês <i>Vertical Growth Phase</i>
W-EFO-E-	(Arquitetura híbrida de rede neural)
CNN	
YIQ	<i>Luma, In-phase, Quadrature</i>

## **Acrônimos**

Adam	<i>Adaptive Moment Estimation</i>
BCN20000	<i>Barcelona 20000 Dataset</i>
CUDA	<i>Compute Unified Device Architecture</i>
cuDNN	Biblioteca de Redes Neurais Profundas CUDA, do inglês <i>CUDA Deep Neural Network library</i>
FLOPS	Operações de Ponto Flutuante por Segundo, do inglês <i>Floating Point Operations Per Second</i>
GLOBOCAN	<i>Global Cancer Observatory</i>
Grad-CAM	<i>Gradient-weighted Class Activation Mapping</i>
HAM10000	<i>Human Against Machine with 10,000 Training Images</i>
INCA	Instituto Nacional de Câncer José Alencar Gomes da Silva

MEL	Melanoma
NaN	Não é um número, do inglês <i>Not a Number</i>
ONU	Organização das Nações Unidas
ReLU	<i>Rectified Linear Unit</i>
ROC	Característica de Operação do Receptor, do inglês <i>Receiver Operating Characteristic</i>
ROI	Região de Interesse, do inglês <i>Region of Interest</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
SUS	Sistema Único de Saúde
UFES	Universidade Federal do Espírito Santo
UFMA	Universidade Federal do Maranhão
ViT	<i>Vision Transformers</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>13</b>
<b>1.1</b>	<b>Problema</b> .....	<b>13</b>
<b>1.2</b>	<b>Justificativa</b> .....	<b>14</b>
<b>1.3</b>	<b>Objetivos</b> .....	<b>14</b>
1.3.1	Objetivo Geral .....	14
1.3.2	Objetivos Específicos .....	14
<b>1.4</b>	<b>Estrutura do Trabalho</b> .....	<b>15</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>16</b>
<b>2.1</b>	<b>Câncer de Pele</b> .....	<b>16</b>
2.1.1	Pele Humana .....	16
2.1.2	Tipos de Câncer de Pele .....	18
2.1.3	Melanócitos e Formação do Melanoma .....	21
2.1.4	Classificação dos Melanomas .....	23
2.1.5	Diagnóstico .....	25
2.1.6	Dermatoscopia .....	28
<b>2.2</b>	<b>Visão Computacional</b> .....	<b>29</b>
2.2.1	Imagens Digitais .....	30
2.2.2	Reconhecimento de Padrões .....	33
<b>2.3</b>	<b>Inteligência Artificial e Aprendizado de Máquina</b> .....	<b>37</b>
2.3.1	Paradigma do Aprendizado de Máquina .....	37
2.3.2	Aprendizado Supervisionado .....	38
2.3.3	Generalização, <i>Overfitting</i> e Divisão de Dados .....	39
2.3.4	Desbalanceamento de Classes e a Técnica SMOTE .....	40
<b>2.4</b>	<b>Redes Neurais e Aprendizado Profundo</b> .....	<b>41</b>
2.4.1	Estrutura e Treinamento .....	42
2.4.2	Redes Neurais Convolucionais .....	44
2.4.2.1	Camada Convolucional .....	45
2.4.2.2	Camada de <i>Pooling</i> .....	45
2.4.2.3	Camada Totalmente Conectada .....	45
2.4.2.4	<i>Dropout</i> e <i>Batch Normalization</i> (Camadas auxiliares) .....	46
<b>2.5</b>	<b>Arquiteturas de Redes Neurais Convolucionais</b> .....	<b>46</b>
2.5.1	Arquitetura Escolhida .....	47
<b>2.6</b>	<b>Métricas de Desempenho</b> .....	<b>48</b>
2.6.1	Métrica Dice .....	48
2.6.2	Métricas Complementares .....	49
2.6.3	Função de Perda .....	51
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b> .....	<b>53</b>
<b>3.1</b>	<b>Dermatoscopia Digital e Evolução Tecnológica</b> .....	<b>53</b>
<b>3.2</b>	<b>Bases de Dados Dermatoscópicas Públicas</b> .....	<b>54</b>
<b>3.3</b>	<b>Critério Clínico: Regra ABCDE</b> .....	<b>56</b>
<b>3.4</b>	<b>Revisão de Arquiteturas</b> .....	<b>58</b>
3.4.1	Segmentação Semântica .....	58
3.4.2	Classificação e <i>Transfer Learning</i> .....	61
3.4.3	Precursos e Desafios no Diagnóstico Móvel .....	63
<b>4</b>	<b>METODOLOGIA</b> .....	<b>65</b>
<b>4.1</b>	<b>Ambiente de desenvolvimento e ferramentas computacionais</b> .....	<b>66</b>

4.2	Bases de dados utilizadas .....	68
4.3	Divisão dos conjuntos de treinamento, validação e teste .....	70
4.4	Pré-processamento das imagens .....	71
4.5	Estratégias de balanceamento e aumento de dados .....	73
4.6	Segmentação com U-Net e geração das regiões de interesse.....	75
4.7	Configurações experimentais do classificador .....	76
5	RESULTADOS E DISCUSSÕES .....	81
5.1	Organização dos Experimentos e Distribuição de Dados .....	81
5.2	Resultados da Segmentação.....	82
5.3	Resultados do Classificador .....	86
5.4	Recursos Computacionais e Tempos de Processamento .....	93
5.5	Discussão Geral, Limitações e Implicações .....	94
6	CONCLUSÃO .....	97
	REFERÊNCIAS.....	100

## 1 INTRODUÇÃO

O câncer é caracterizado pelo crescimento desenfreado de células capazes de invadir tecidos e órgãos, podendo inclusive se espalhar por todo o corpo em seu estágio final (Salvajoli; Souhami; Faria, 2021; INCA, 2022). Embora categórica, essa definição não traz consigo a experiência emocional de quem recebe um diagnóstico. Para essas pessoas, o câncer não se limita a termos médicos, mas a uma vivência marcada por angústias, medos e incertezas. Em seu relato à *Melanoma Research Foundation*, a influenciadora digital Leah Alexis Adams (Orcione, 2025) compartilha: “Lembro do meu coração e estômago afundarem ao mesmo tempo, todos os meus hábitos ruins com sol e cuidados com a pele passando diante dos meus olhos, e me sentindo completamente perdida.”

Esse sentimento de desespero não é isolado. Ele reflete uma realidade alarmante: a cada ano, mais de 1,5 milhões de pessoas são diagnosticadas com câncer de pele em todo o mundo, conforme o relatório global mais recente do GLOBOCAN 2022, disponibilizados pelo *Global Cancer Observatory* (IARC, 2024). No Brasil, o cenário é ainda mais grave: o câncer de pele é o tipo mais incidente, representando quase um terço de todos os diagnósticos oncológicos (INCA, 2022; Amaral, 2023). Destes, cerca de 1% a 3% são classificados como melanoma, forma mais agressiva e letal do câncer de pele (Coelho, 2016; INCA, 2022).

Responsável por cerca de 75% das mortes por câncer de pele, o melanoma é uma neoplasia maligna cujo diagnóstico precoce é fundamental para um prognóstico positivo (INCA, 2022; SBD, 2020). O impacto do diagnóstico pode ser observado nas taxas de sobrevivência: quando identificado em estágio inicial, a probabilidade de sobrevivência em 5 anos é de 99%. Por outro lado, quando a detecção ocorre em estágio tardio, essa taxa cai para apenas 25% (ACS, 2025; ONCOGUIA, 2025).

### 1.1 Problema

Diante desse cenário, diversos estudos e pesquisas na área de dermatoscopia têm buscado desenvolver técnicas que auxiliem os profissionais da saúde a aumentar a agilidade e a precisão dos diagnósticos. Contudo, a identificação do melanoma por meio de imagens dermatoscópicas é um grande desafio, exigindo tempo, experiência e esforço computacional. As precisões típicas variam entre 75% a 85%, dependendo da habilidade do profissional (Coelho, 2016). Outro fator crítico é que, mesmo com o auxílio dessas imagens, o diagnóstico definitivo permanece dependente da confirmação via biópsia (exame histopatológico), onerando o processo em custo e tempo (BRASIL, 2022). Além disso, muitas avaliações em locais mais afastados, sobretudo àqueles atendidos pelo SUS (Sistema Único de Saúde), acabam sendo realizadas por pessoas não especializadas, aumentando os riscos de atraso no tratamento e erros no diagnóstico (Alencar, 2015; Coelho, 2016).

## 1.2 Justificativa

Nesse sentido, a crescente evolução das tecnologias relacionadas a processamento de imagem e Inteligência Artificial (IA) tem se mostrado um fator decisivo na busca por soluções inovadoras aos desafios da dermatoscopia (Alencar, 2015). Em especial, técnicas de Aprendizado Profundo (*Deep Learning* ou DL) e Redes Neurais Convolucionais (RNCs) têm se apresentado bastante eficazes na classificação de imagens dermatoscópicas, possibilitando a extração automática de padrões e características relevantes (Santos *et al.*, 2017; Júnior; Bezerra; Andrade, 2020).

Com base nos desafios identificados e nas possibilidades emergentes, torna-se evidente a necessidade do desenvolvimento de métodos capazes de aprimorar a dermatoscopia e, por consequência, beneficiar os pacientes. Nesse contexto, a construção de um classificador de imagens baseado em redes neurais surge como uma alternativa factível para ampliar o alcance da dermatoscopia de qualidade, permitindo identificar as lesões e classificá-las como melanoma, carcinoma basocelular (CBC) e carcinoma espinocelular (CEC). Essa abordagem contribui para a otimização do processo de triagem e para a melhoria dos resultados clínicos, sobretudo em regiões com acesso limitado a especialistas (Alencar, 2015).

Dessa forma, o trabalho também se soma aos esforços do Objetivo de Desenvolvimento Sustentável 3 (ODS 3) da Organização das Nações Unidas (ONU, 2015). Ao viabilizar o diagnóstico precoce e democratizar o acesso à saúde, o projeto deixa de ser apenas uma ferramenta técnica para se tornar um aliado real na redução da mortalidade e na promoção do bem-estar da população.

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

Este trabalho objetiva desenvolver um classificador inteligente de imagens dermatoscópicas, utilizando técnicas de aprendizado profundo, capaz de identificar melanoma, CBC e CEC.

### 1.3.2 Objetivos Específicos

- Levantar informações sobre técnicas de DL, com ênfase em arquiteturas convolucionais voltadas à segmentação e classificação de imagens dermatoscópicas.
- Estruturar um subconjunto do banco de dados da *International Skin Imaging Collaboration* (ISIC) 2019, contemplando lesões de melanoma, CBC e CEC, para o treinamento e validação dos modelos.

- Aplicar estratégias de pré-processamento e segmentação a fim de melhorar as imagens, delimitar a região das lesões e remover ruídos ou artefatos indesejados.
- Desenvolver um modelo de inteligência artificial baseado em redes neurais para classificação automática de lesões de pele, com otimização de hiperparâmetros.
- Validar o desempenho do sistema e discutir sua aplicabilidade como ferramenta de apoio ao diagnóstico precoce de melanoma.

#### 1.4 Estrutura do Trabalho

A fim de facilitar a compreensão lógica e guiar a leitura, este documento foi estruturado de forma sequencial em seis capítulos principais. Na etapa inicial, o Capítulo 1 apresenta a contextualização do tema, definindo o problema de pesquisa, a justificativa para a realização do estudo e os objetivos gerais e específicos que norteiam o projeto.

Em seguida, a fundamentação teórica é estabelecida no Capítulo 2, o qual consolida os conceitos médicos fundamentais sobre neoplasias cutâneas e aprofunda as bases matemáticas e computacionais do Aprendizado Profundo (*Deep Learning*), das Redes Neurais Convolucionais e das métricas de avaliação de desempenho.

Para situar a pesquisa no cenário científico atual, o Capítulo 3 apresenta uma revisão da literatura e dos trabalhos correlatos, traçando um panorama dos métodos mais avançados de segmentação e classificação automatizadas de lesões de pele.

A essência técnica e prática do projeto concentra-se no Capítulo 4, dedicado ao Desenvolvimento e à Metodologia. Nesta etapa, é detalhado todo o fluxo computacional empregado, passando pelas técnicas de pré-processamento e limpeza das imagens com o algoritmo *Dull Razor*, pela segmentação semântica via U-Net, até chegar à arquitetura de classificação utilizando a EfficientNet-B4.

Já a eficácia dessas escolhas metodológicas é validada no Capítulo 5, espaço reservado para a apresentação dos Resultados e Discussão. Nesse ponto, o desempenho dos modelos treinados é estudado com análise crítica sob a ótica das métricas estatísticas levantadas, avaliando o resultado do algoritmo na identificação correta das classes (melanoma, CBC e CEC) e sua capacidade de generalização.

Por fim, o Capítulo 6 encerra o documento com as Considerações Finais, retomando as hipóteses e objetivos iniciais, pontuando as limitações técnicas encontradas durante a execução e propondo direcionamentos para trabalhos e integrações futuras.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo organiza a base teórica do trabalho em três frentes principais. No início, abordam-se os aspectos biológicos e clínicos do câncer de pele, incluindo a anatomia cutânea e os métodos tradicionais de diagnóstico. Na sequência, exploram-se a visão computacional e a representação das imagens digitais. Esses conceitos permitem entender como o sistema processa informações visuais para identificar padrões de forma automatizada. Por fim, detalham-se o aprendizado profundo e as arquiteturas EfficientNet-B4 e U-Net, além das métricas que validam a eficácia técnica e clínica do classificador. Essa base é essencial para compreender a convergência entre o processamento digital de imagens e o suporte ao diagnóstico dermatológico.

### 2.1 Câncer de Pele

O câncer de pele compreende um amplo espectro de neoplasias, benignas e malignas, que se desenvolvem nas diferentes camadas da pele e apresentam elevada incidência em populações de diversas regiões do mundo. Trata-se do tipo de câncer mais prevalente em escala global, superando, em número de casos, as demais neoplasias (Sathe; Zito, 2025).

Segundo a Agência Internacional de Pesquisa em Câncer (IARC, 2024), o câncer de pele é o mais incidente no mundo. Em 2022, foram registrados mais de 1,5 milhão de novos casos globais. Destes, cerca de 330 mil foram melanomas, resultando em aproximadamente 60 mil óbitos (IARC, 2024). No Brasil, a doença também lidera as estatísticas, afetando sobretudo pessoas de pele clara acima de 40 anos. Contudo, nota-se um aumento da incidência em faixas etárias mais jovens devido à exposição solar inadequada (INCA, 2022). As estimativas do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) para o triênio 2023-2025 preveem 220.490 novos casos anuais de não melanoma e 8.980 de melanoma, evidenciando a magnitude do problema de saúde pública (INCA, 2022).

O desenvolvimento do câncer de pele está relacionado à proliferação anormal e descontrolada das células cutâneas, que, em decorrência de mutações genéticas, passam a se dividir de maneira desordenada, podendo invadir tecidos adjacentes e, nos casos mais graves, disseminar-se para outros órgãos. Para entender melhor essa condição, exige-se um estudo minucioso sobre a anatomia e as características da pele (Coelho, 2016; Horiuchi, 2022).

#### 2.1.1 Pele Humana

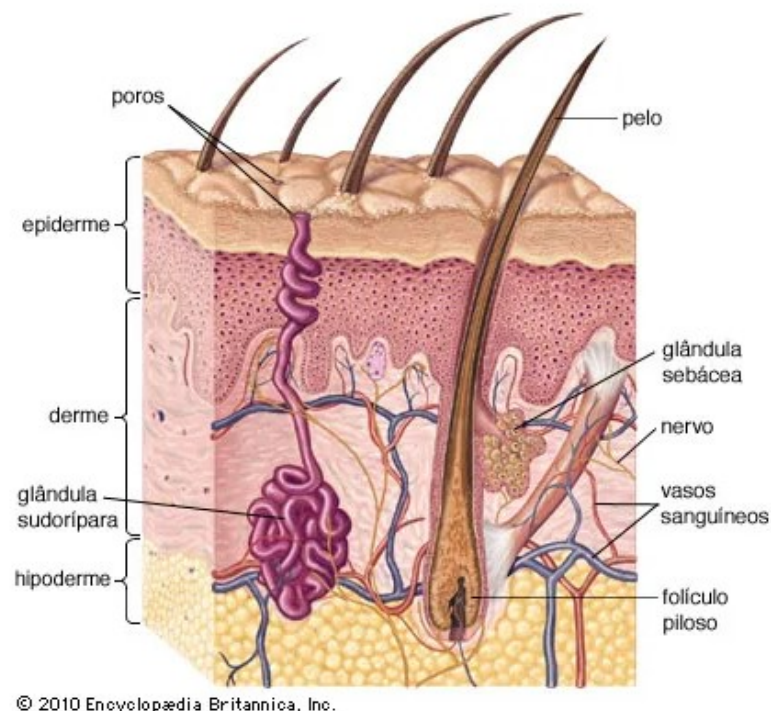
A pele, também denominada *cútis*, é o revestimento externo do organismo e constitui um órgão complexo, essencial para a manutenção da vida. Representa mais de 15% do peso corporal e funciona como a principal interface entre o corpo e o meio ambiente. Por estar, de forma contínua, exposta a fatores externos, entre eles a radiação ultravioleta (UV), agentes físicos,

químicos e biológicos, a pele desempenha papel fundamental como barreira protetora (Rivitti, 2014).

Graças à sua arquitetura físico-química especializada, a pele exerce múltiplas funções necessárias à homeostase orgânica, definida como a capacidade de manter a estabilidade do meio interno frente às alterações externas (Hall, 2017). Sua principal função é a de barreira, impedindo a entrada de microrganismos, substâncias nocivas e traumas físicos, além de reduzir a perda transepidérmica de água e eletrólitos. Atua também como componente do sistema imunológico inato, apresentando elevada atividade imunológica. A melanina sintetizada na epiderme desempenha papel central na fotoproteção, absorvendo e difundindo a radiação UV. Ademais, a pele participa da termorregulação, por meio da sudorese e do controle da rede vascular cutânea; da percepção sensorial (tato, dor, pressão e temperatura); e da síntese cutânea de vitamina D, iniciada a partir de precursores epidérmicos fotoativados (Bohjanen, 2017).

Quanto à estrutura, a pele é constituída por três camadas principais: a epiderme (superior), a derme ou cório (intermediária) e a hipoderme ou tecido celular subcutâneo (profunda), como é possível observar na Figura 1.

**Figura 1 – Estrutura anatômica da pele humana e suas camadas**



**Fonte: Adaptado de Encyclopedia Britannica (2010).**

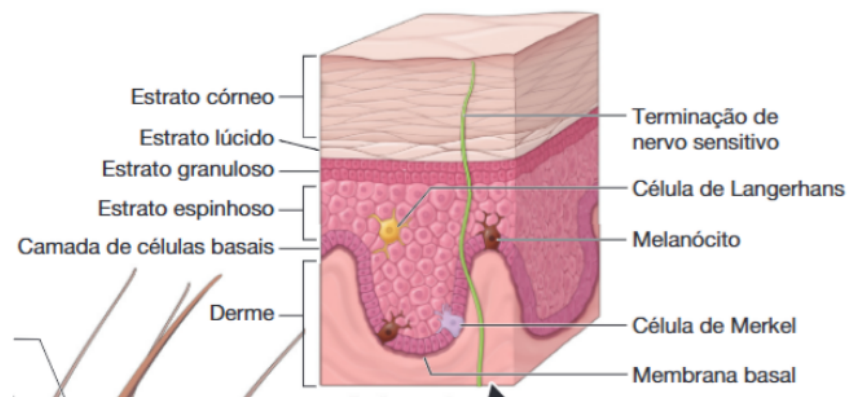
A derme, camada intermediária, consiste em tecido conjuntivo denso, rico em fibras colágenas e, em menor proporção, fibras elásticas, responsáveis pela resistência, elasticidade e flexibilidade cutâneas. Nessa camada encontram-se vasos sanguíneos, terminações nervosas, glândulas sudoríparas, glândulas sebáceas e folículos pilosos (Benedetti, 2024).

A hipoderme, camada mais profunda, é formada, em sua maior parte, por tecido adiposo. Atua como reserva energética, isolamento térmico e amortecimento mecânico. Contém

lóbulos adiposos, vasos sanguíneos, neurônios sensoriais e escassos anexos cutâneos (Lotfolahi, 2024; Yousef *et al.*, 2025).

A epiderme, camada mais superficial e externa, é um epitélio estratificado queratinizado, fino e resistente, separado da derme pela membrana basal. Essa camada organiza-se em estratos celulares sobrepostos, do mais profundo ao mais superficial: estrato basal, estrato espinhoso, estrato granuloso, estrato lúcido e estrato córneo, conforme demonstrado na Figura 2.

**Figura 2 – Camadas da epiderme e seus tipos celulares especializados**



Fonte: Bohjanen (2017).

No estrato basal localizam-se as células-tronco com atividade mitótica ativas, responsáveis pela produção contínua de queratinócitos. Estes representam cerca de 90% das células epidérmicas e, à medida que migram para a superfície, formam a barreira física da pele (Yousef *et al.*, 2025; Benedetti, 2024).

Embora os queratinócitos sejam fundamentais e origem dos carcinomas, como o carcinoma basocelular, o foco central para o diagnóstico do câncer de pele mais letal recai sobre outra célula residente na camada basal: o melanócito (ver Figura 2). Derivados da crista neural, os melanócitos têm a função de sintetizar a melanina e transferi-la aos queratinócitos vizinhos, criando um escudo protetor contra a radiação UV (Rivitti, 2014; Coelho, 2016).

Sob a perspectiva oncológica, a identificação e análise dessas estruturas são cruciais, pois os melanócitos são as células de origem do melanoma. Este é o tipo mais agressivo de câncer de pele, responsável por cerca de 75% dos óbitos associados à doença, apesar de representar apenas 3% dos diagnósticos (Júnior; Bezerra; Andrade, 2020; Horiuchi, 2022). Portanto, a compreensão da localização dessas células na junção dermoepidérmica é fundamental para o desenvolvimento de estratégias de diagnóstico precoce automatizado.

### 2.1.2 Tipos de Câncer de Pele

O câncer de pele é a neoplasia mais prevalente no Brasil e no mundo. O principal fator de risco para o seu desenvolvimento, em todos os tipos, são os raios UV, que causam lesões no DNA (Ácido Desoxirribonucleico) com dano cumulativo ao longo da vida. Do ponto de vista

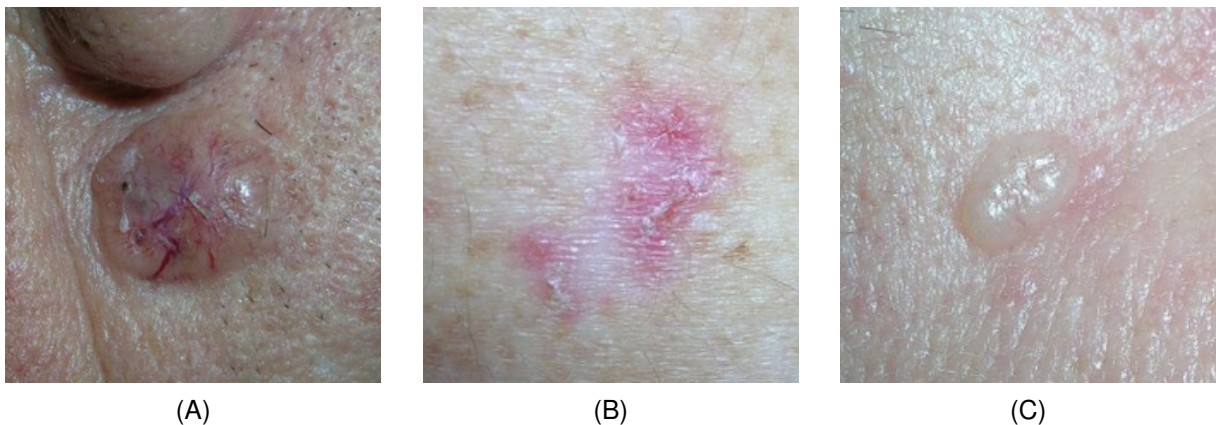
oncológico, os tumores cutâneos primários são, em diversos estudos, classificados em dois grandes grupos: melanocíticos (melanoma) e não melanocíticos (INCA, 2022).

Os tumores de pele não melanoma são os mais frequentes e, embora apresentem menor mortalidade, podem causar mutilações significativas quando não tratados de forma precoce. Dentro deste grupo, destacam-se o carcinoma basocelular (CBC) e o carcinoma espinocelular (CEC), responsáveis pela maioria dos casos (Amaral, 2023).

O CBC é o tipo mais comum de câncer de pele, correspondendo a 75%–80% dos tumores não melanoma em pessoas de pele clara. Originado da epiderme, é composto por células que se assemelham aos queratinócitos basais da camada germinativa. Sob o ponto de vista histológico, apresenta células basofílicas com pouco citoplasma, com frequência organizadas em paliçadas ou ninhos celulares. A radiação UV é o principal fator etiológico, estando associada a mutações nos genes *PTCH1* (*patched 1*) e *TP53*, que codifica a proteína p53. O CBC possui, em geral, excelente prognóstico, com altas taxas de cura e baixa recorrência após tratamento cirúrgico (Sathe; Zito, 2025).

Sob o ponto de vista clínico, o CBC ocorre de preferência nos dois terços superiores da face, sendo a forma nódulo-ulcerativa a apresentação mais típica. Essa condição inicia-se como uma pápula rósea e perlada, evoluindo para um nódulo com bordas translúcidas (perláceas), ulceração central e telangiectasias finas, como ilustrado na Figura 3-A. Entre suas variantes, destaca-se o tipo superficial (Figura 3-B), caracterizado por placas eritemato-descamativas no tronco, e o tipo esclerosante (Figura 3-C), que se manifesta como uma placa endurecida branco-amarelada semelhante à esclerodermia (Rivitti, 2014).

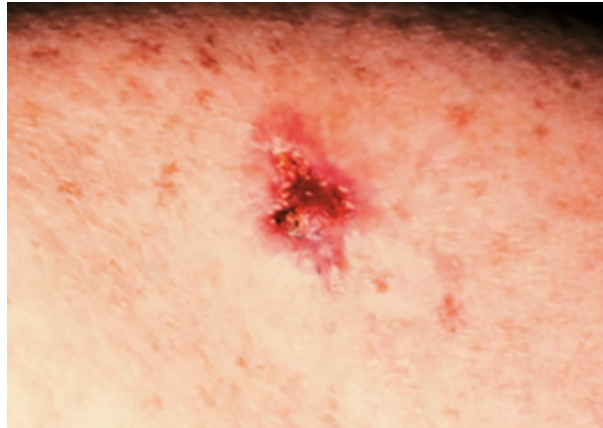
**Figura 3 – Variantes clínicas do Carcinoma Basocelular**



**Fonte: Adaptado de The Skin Cancer Foundation (2024a).**

O CEC surge da proliferação atípica dos queratinócitos da camada espinhosa da epiderme, apresentando-se com frequência como pápulas ou nódulos hiperqueratóticos (com descamação ou crostas espessas), conforme ilustrado na Figura 4.

**Figura 4 – Aspecto clínico do Carcinoma Espinocelular**



**Fonte: Adaptado de The Skin Cancer Foundation (2024b).**

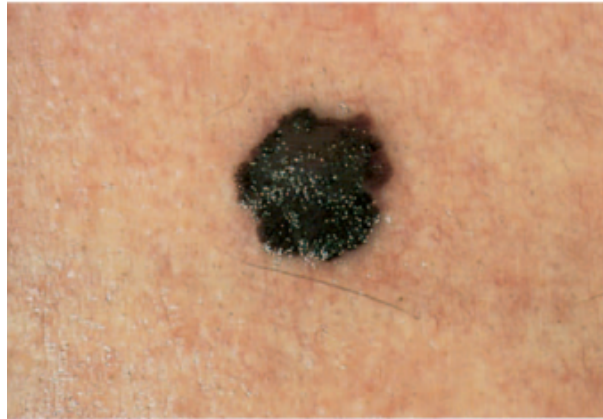
O carcinoma espinocelular é um tumor invasivo, com maior potencial metastático que o CBC. Entre as alterações moleculares associadas à doença, destacam-se mutações no gene supressor tumoral TP53, responsável pela codificação da proteína p53, que participa do controle do ciclo celular, do reparo do DNA e da indução da apoptose. Alterações nesse gene estão presentes em quase todos os casos de CEC de alto risco. Também são observadas mutações nos genes NOTCH1 e NOTCH2, responsáveis pela codificação de receptores da via de sinalização Notch, envolvida na regulação da diferenciação, da proliferação e do destino das células. No CEC cutâneo, a perda da função desses genes pode comprometer a diferenciação normal dos queratinócitos e favorecer o desenvolvimento tumoral (Jasmine *et al.*, 2024).

O CEC é o segundo tipo mais prevalente entre caucasianos, asiáticos e hispânicos, mas constitui o câncer de pele mais frequente em pessoas de pele negra. Excluindo o melanoma, o CEC responde por cerca de 75% das mortes por câncer de pele. Em pessoas de pele escura, surge sobretudo em áreas pouco expostas ao sol, como pernas, com frequência associado a cicatrização crônica, condição que confere risco metastático elevado (20%–40%). Outros fatores de risco incluem imunossupressão prolongada e infecção pelo papilomavírus humano (HPV) (Bradford, 2009).

As localizações mais comuns do CEC incluem lábio inferior, orelhas, face e dorso das mãos. A lesão inicial costuma apresentar-se como área queratósica endurecida ou como nódulo que, ao evoluir, pode ulcerar com borda infiltrada ou assumir aspecto vegetante ou córneo. Uma variante menos agressiva é o carcinoma verrucoso, que pode manifestar-se como a papilomatose oral florida (Rivitti, 2014).

O melanoma é o tipo de câncer de pele mais agressivo e responsável pela maioria das mortes associadas a tumores cutâneos. Origina-se dos melanócitos, células produtoras de melanina, e embora seja menos frequente que os tumores não melanoma, apresenta elevada taxa de letalidade devido ao seu alto potencial metastático. Sem tratamento adequado e precoce, o melanoma pode ser fatal, apresentando-se em geral como uma lesão pigmentada de cores variadas e bordas irregulares, conforme exemplificado na Figura 5 (Siqueira; Fernandes, 2015).

**Figura 5 – Aspecto clínico do melanoma maligno**



**Fonte: Adaptado de The Skin Cancer Foundation (2024a).**

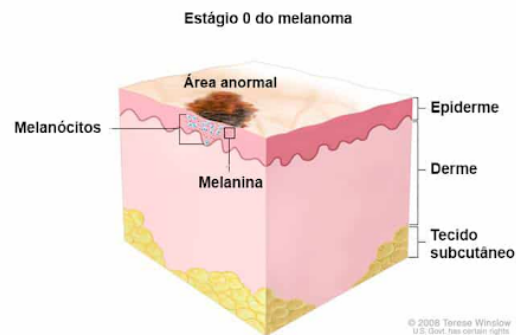
### 2.1.3 Melanócitos e Formação do Melanoma

O melanoma tem origem nos melanócitos, células responsáveis pela produção de melanina, o pigmento que confere coloração à pele. Sob o ponto de vista embriológico, esses elementos celulares derivam do neuroectoderma ou crista neural, migrando em seguida para se alojarem, em sua maior parte, na camada basal da epiderme (Vieira, 2016). A função primária do melanócito consiste na síntese de melanina, um pigmento amorfo de coloração escura produzido em organelas especializadas denominadas melanossomas. O processo de síntese, conhecido como melanogênese, inicia-se a partir do aminoácido essencial tirosina, o qual é oxidado pela enzima tirosinase, resultando na formação dos diferentes tipos de melanina (Figueiredo *et al.*, 2003; Rivitti, 2014).

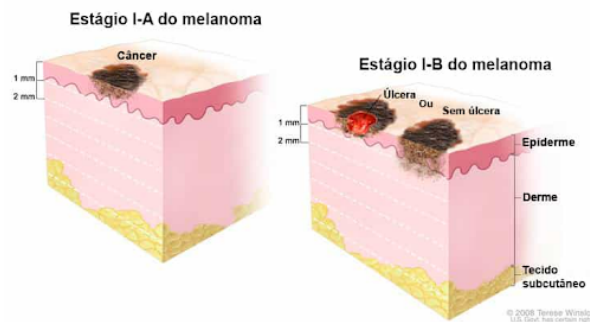
Do ponto de vista funcional, os melanócitos, em conjunto com os queratinócitos, constituem as unidades epidermomelânicas da pele. A melanina exerce papel fundamental na fotoproteção celular, sendo transferida para os queratinócitos, onde se posiciona como uma barreira sobre o núcleo, protegendo-o dos efeitos dos raios UV. Os melanócitos sintetizam duas formas de melanina: a eumelanina (marrom a preta) e a feomelanina (amarelada a avermelhada). A variação na pigmentação cutânea entre indivíduos decorre da regulação do processo de melanogênese, e não da quantidade de melanócitos presentes na pele (Rivitti, 2014; Júnior; Bezerra; Andrade, 2020).

O melanoma caracteriza-se pelo crescimento anormal e descontrolado de melanócitos, podendo surgir tanto em pele previamente íntegra quanto a partir de lesões melanocíticas pré-existentes, como nevos. A progressão tumoral é impulsionada por mutações genéticas, com frequência associadas à radiação UV, agente que provoca danos ao DNA e proliferação celular desregulada (Vazquez, 2019). O avanço clínico dessa neoplasia é, de forma tradicional, dividido em duas etapas de crescimento distintas, conforme ilustrado em forma de esquema na Figura 6.

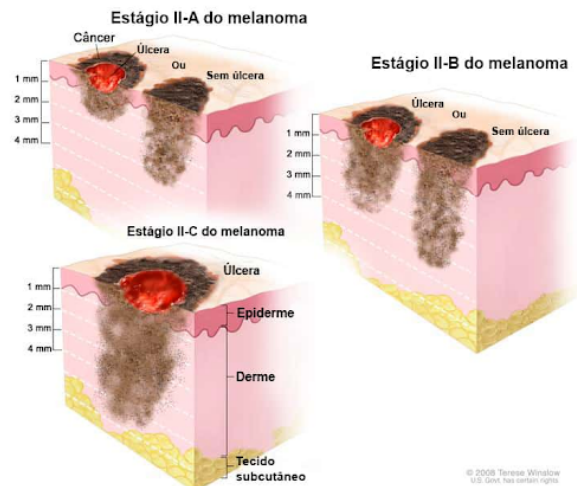
**Figura 6 – Estágios de crescimento do melanoma**



(A) Fase de Crescimento Radial (Estágio 0/*in situ*)



(B) Início da Fase de Crescimento Vertical (Invasão Dérmica)



(C) Fase Vertical Avançada (Maior espessura de Breslow)

**Fonte: Adaptado de National Cancer Institute (2019).**

A primeira etapa, denominada Fase de Crescimento Radial (*Radial Growth Phase*, RGP), conforme representado na Figura 6(A), é caracterizada pela expansão horizontal das células neoplásicas, permanecendo restrita às camadas mais superficiais da epiderme. Essa etapa corresponde ao melanoma *in situ* (representado como Estágio 0 na Figura 6(A)), no qual

o risco de metástase é mínimo e o tratamento cirúrgico apresenta altas taxas de cura (Coelho, 2016; Elder *et al.*, 2020).

A segunda fase, denominada Fase de Crescimento Vertical (*Vertical Growth Phase*, VGP), conforme ilustrado na Figura 6(B), envolve a progressão do tumor para a derme e, em alguns casos, para a hipoderme (representada pelos estágios invasivos na Figura 6(C)). Durante essa fase, a profundidade tumoral, em particular a espessura de Breslow, torna-se um dos mais importantes indicadores prognósticos, uma vez que as camadas profundas da pele contêm vasos sanguíneos e linfáticos que facilitam a disseminação metastática (Coelho, 2016; Elder *et al.*, 2020).

A invasão dérmica define o caráter agressivo do melanoma e está, de modo direto, associada ao pior prognóstico clínico. O VGP incipiente é identificado quando um agrupamento celular na derme supera, em tamanho, o maior agrupamento presente na epiderme, ou quando há atividade mitótica evidente na derme, indicando a transição do foco proliferativo para camadas mais profundas da pele (Elder *et al.*, 2020; Júnior; Bezerra; Andrade, 2020).

#### 2.1.4 Classificação dos Melanomas

Por tradição, o melanoma cutâneo primário é classificado em subtipos clínico-patológicos que apresentam diferenças marcantes quanto ao padrão de crescimento, distribuição anatômica e velocidade de progressão. Apesar dessas particularidades, o principal determinante prognóstico é a profundidade de invasão tumoral no momento do diagnóstico, mensurada pelo Índice de Breslow. Os quatro subtipos histológicos predominantes são: Melanoma de Disseminação Superficial, Melanoma Nodular, Melanoma Lentigo Maligno e Melanoma Lentiginoso Acral (Vazquez, 2019; Horiuchi, 2022).

Todos os melanomas são neoplasias malignas, independentemente do subtipo clínico-patológico. Dessa forma, a presença do termo “maligno” na denominação melanoma lentigo maligno não significa que os demais subtipos sejam benignos, mas corresponde à nomenclatura específica dessa variante.

Constituindo o subtipo mais prevalente, o melanoma de disseminação superficial (*superficial spreading melanoma*, SSM) corresponde a cerca de 60% a 70% dos casos. Caracteriza-se por apresentar RGP prolongada, que pode evoluir durante meses ou anos, em sua maior parte confinada à epiderme. Com frequência, o SSM surge a partir da transformação maligna de nevos antigos. Na prática clínica, manifesta alterações compatíveis com o mnemônico ABCDE (Regra clínica: Assimetria, Borda, Cor, Diâmetro e Evolução), sobretudo assimetria, irregularidade das bordas e variação de cores, conforme observado na Figura 7 (Rivitti, 2014; Vazquez, 2019).

**Figura 7 – Subtipos clínicos de melanoma**



(A) Melanoma de disseminação superficial



(B) Melanoma nodular



(C) Melanoma lentigo maligno



(D) Melanoma lentiginoso acral

**Fonte: Adaptado de The Skin Cancer Foundation (2024c).**

Reconhecido como um dos subtipos mais agressivos, o melanoma nodular (*nodular melanoma*, NM) representa entre 15% e 30% dos casos e destaca-se pelo predomínio da VGP, em geral de evolução rápida. Comumente surge de novo, sem lesões precursoras evidentes, e é definido, sob o ponto de vista histológico, pela ausência da fase de crescimento radial. Em virtude dessa progressão acelerada, o NM tende a ser diagnosticado com maior espessura de Breslow, o que resulta em pior prognóstico. Clinicamente, manifesta-se como nódulo exofítico, polipóide ou em placa, de coloração escura (ver Figura 7-B); entretanto, pode ser amelanótico em cerca de 5% dos casos, apresentando tonalidades rósea, bege ou esbranquiçada, o que, com frequência, atrasa o diagnóstico (Rivitti, 2014; Coelho, 2016; Vazquez, 2019).

Já o melanoma lentigo maligno (*lentigo maligna melanoma*, LMM) corresponde a cerca de 5% (Vieira, 2016) a 10% (Coelho, 2016) dos casos e acomete em sua maior parte idosos, sobretudo em áreas, de forma crônica, fotoexpostas, como face e pescoço. O LMM representa a forma invasiva do lentigo maligno, considerado melanoma *in situ* de evolução lenta. Sob o ponto de vista clínico, manifesta-se como mácula pigmentada de bordas irregulares e crescimento centrífugo lento (Figura 7-C). A transição para o estágio invasivo é sugerida pelo surgimento de nódulos pigmentados ou áreas de elevação irregular na superfície da lesão (Rivitti, 2014; Vieira, 2016; Horiuchi, 2022).

Por sua vez, o melanoma lentiginoso acral (*acral lentiginous melanoma*, ALM) é menos prevalente em indivíduos de pele clara (2% a 8%), mas representa a forma mais comum en-

tre pessoas de pele negra e asiática (35% a 60%) (Rivitti, 2014). Localiza-se de preferência em regiões palmoplantares e subungueais (Figura 7-D). O diagnóstico com frequência é tardio devido à semelhança com lesões benignas e ao menor índice de suspeição clínica nessas áreas (Vazquez, 2019).

No que diz respeito ao estadiamento, parâmetros histopatológicos específicos são essenciais. O índice de Breslow, que mede em milímetros a profundidade de invasão a partir da camada granular da epiderme até o ponto mais profundo atingido pelo tumor, constitui o principal fator prognóstico e orienta a conduta terapêutica. A classificação varia desde melanoma *in situ* até lesões com mais de 4 mm de espessura (Rivitti, 2014; Horiuchi, 2022).

Além disso, o nível de Clark descreve a invasão em relação às camadas anatômicas da pele, variando do nível I (*in situ*) ao nível V (invasão da hipoderme). Embora tenha menor impacto prognóstico que o índice de Breslow, esse parâmetro pode contribuir para o estadiamento em casos selecionados. A presença de ulceração e o índice mitótico (mitoses por mm<sup>2</sup>) também são fatores prognósticos independentemente significativos. Embora o subtipo histológico seja item obrigatório no laudo anatomopatológico, ele não integra os critérios prognósticos da 8ª edição do *American Joint Committee on Cancer* (AJCC) (Figueiredo *et al.*, 2003; Rivitti, 2014; Horiuchi, 2022).

Considerando a expressiva diversidade morfológica e prognóstica dos melanomas, torna-se fundamental a utilização de métodos diagnósticos precisos e padronizados. Assim, a seção seguinte abordará os principais procedimentos empregados no diagnóstico clínico e histopatológico, que estabelecem a base para a identificação precoce da doença e para o suporte ao desenvolvimento de modelos computacionais de apoio ao diagnóstico.

### 2.1.5 Diagnóstico

A avaliação clínica inicial constitui etapa fundamental no diagnóstico de lesões cutâneas suspeitas de malignidade. O primeiro objetivo do exame é distinguir lesões melanocíticas de não melanocíticas e, depois, classificá-las como benignas, suspeitas ou malignas. A técnica semiológica empregada no exame dermatológico segue princípios semelhantes aos da prática clínica geral, porém se caracteriza pela centralidade do exame objetivo, que precede e direciona a anamnese.

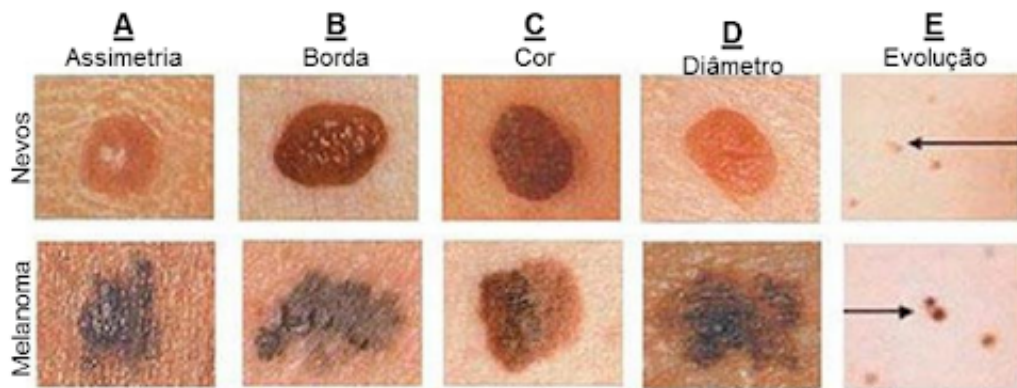
O exame clínico inicia-se pela inspeção, realizada de maneira sistemática em todo o tegumento, incluindo couro cabeludo, unhas e mucosas. Para uma visualização adequada, recomenda-se que o procedimento seja executado em ambiente bem iluminado, de preferência com luz natural ou fluorescente posicionada atrás do examinador. A inspeção deve ser, na etapa inicial, panorâmica, com observação a uma distância de 1 a 2 metros, seguida de análise aproximada a 20 ou 30 centímetros, utilizando-se lupa quando necessário para avaliar detalhes morfológicos (Rivitti, 2014).

Conforme descrito em estudos como os de Horiuchi (2022), Siqueira e Fernandes (2015) e Alencar (2015), a regra mnemônica ABCDE permanece como uma das ferramentas mais efi-

cazes para a triagem de lesões melanocíticas suspeitas, sobretudo no rastreamento do melanoma. Os critérios avaliados, detalhados sob o aspecto visual na Figura 8, incluem:

- Assimetria (A): falta de correspondência geométrica entre as metades da lesão;
- Bordas (B): contornos irregulares, dentados ou mal definidos;
- Cor (C): presença de múltiplas tonalidades (preto, castanho, vermelho, azul ou branco) em uma mesma lesão;
- Diâmetro (D): lesões com diâmetro geralmente superior a 6 mm;
- Evolução (E): alterações recentes em tamanho, forma, cor ou sintomas.

**Figura 8 – Critérios da regra ABCDE para detecção de melanoma**



**Fonte: Adaptado de Tsao et al. (2015).**

A inspeção deve identificar sinais clínicos de alarme que justifiquem encaminhamento especializado. Lesões que não cicatrizam, em cerca de quatro semanas ou em 15 dias, conforme o tipo de tumor, configuram importante sinal de alerta. O sangramento a mínimos atritos, frequente em tumores cutâneos não melanoma avançados, merece atenção (INCA, 2022). A ulceração é apresentação comum em CBC e CEC, enquanto, no caso do melanoma, ulceração e presença de nódulo têm implicações prognósticas relevantes (Vieira, 2016).

Alterações no tamanho, forma ou coloração também demandam investigação. O melanoma com frequência manifesta crescimento progressivo, bordas irregulares e aumento em profundidade, fator associado ao risco de metástase (Coelho, 2016). Sinais como elevação da lesão, prurido, descamação e sangramento são também relevantes. No CBC, por exemplo, a lesão costuma ser elevada, brilhante, translúcida ou rósea, podendo apresentar crosta central e sangramento fácil (Vazquez, 2019; Amaral, 2023).

Embora a inspeção dermatológica seja o ponto inicial, a anamnese permanece componente essencial do processo diagnóstico. A história clínica orienta hipóteses diagnósticas e complementa a avaliação morfológica (Vieira, 2016). A exposição excessiva aos raios UV constitui o principal fator etiológico da maioria dos cânceres de pele. O dano induzido ao DNA pela

radiação é cumulativo; episódios de queimaduras solares intermitentes, sobretudo na infância e adolescência, representam importante fator de risco para o melanoma (Vazquez, 2019). Por outro lado, a exposição crônica e contínua tende a estar mais associada aos tumores não melanoma (Vieira, 2016). Tanto o espectro UV-B (280–320 nm) quanto o UV-A (320–400 nm) desempenham papel relevante na carcinogênese cutânea (Sathe; Zito, 2025). Além da radiação, a exposição ocupacional a agrotóxicos e solventes, como o tricloroetileno, também está associada ao risco aumentado de neoplasias cutâneas (INCA, 2022).

A investigação clínica deve contemplar antecedentes pessoais e familiares, uma vez que o melanoma pode surgir tanto sobre nevos preexistentes quanto em pele íntegra. A presença de múltiplos nevos melanocíticos ou nevos atípicos justifica acompanhamento periódico, assim como o histórico familiar de melanoma, reconhecido como fator de risco relevante (Coelho, 2016). Aspectos fenotípicos, como pele clara, olhos claros, cabelos claros e presença de efélides, constituem fatores predisponentes. O uso de câmaras de bronzeamento artificial constitui fonte adicional de radiação UV e está associado ao aumento do risco de melanoma (INCA, 2022).

A predisposição genética também deve ser considerada no processo diagnóstico. O histórico familiar de melanoma, sobretudo em parentes de primeiro grau, e o antecedente pessoal da doença elevam o risco. Em termos genéticos, mutações germinativas em genes supressores tumorais, como o CDKN2A, do inglês *Cyclin-Dependent Kinase Inhibitor 2A*, podem conferir susceptibilidade aumentada (Figueiredo *et al.*, 2003; Coelho, 2016). Além disso, condições hereditárias, como o xeroderma pigmentoso, caracterizado por falhas no reparo de DNA após exposição à radiação UV, aumentam de forma expressiva o risco de neoplasias cutâneas. Quadros de imunossupressão, tanto por doenças quanto por uso de imunossupressores, em particular em transplantados, também elevam a susceptibilidade ao desenvolvimento de tumores cutâneos (Rivitti, 2014; Vieira, 2016).

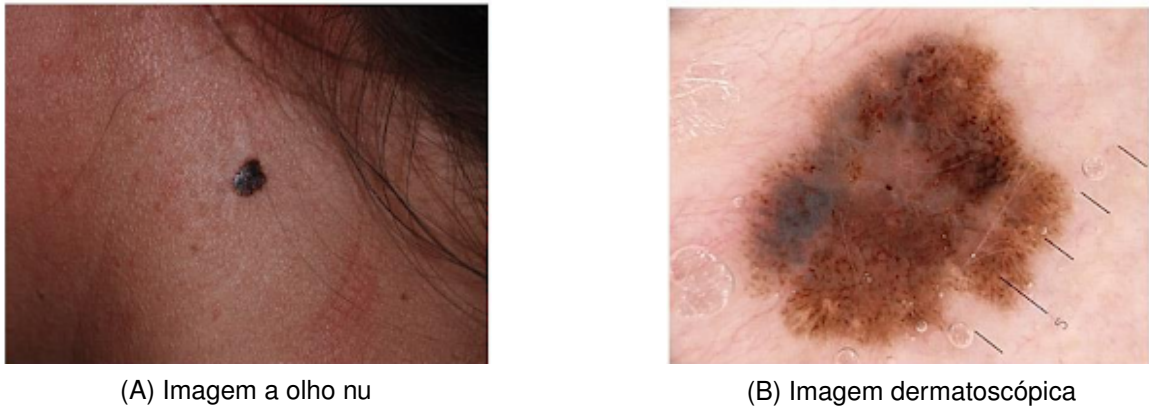
Apesar de não haver evidências conclusivas de que uma técnica específica de autoexame reduza a morbimortalidade, grande parte dos melanomas é identificada no início pelo próprio paciente ou por familiares, o que ressalta a importância de conhecer a própria pele e observar alterações recentes (INCA, 2022). A acurácia diagnóstica de médicos não especialistas, quando utilizam apenas imagens clínicas, mantém-se em torno de 60% (Júnior; Bezerra; Andrade, 2020). Por esse motivo, o diagnóstico precoce é determinante: quando o melanoma é identificado em estágio localizado, as taxas de cura podem atingir de 90% a 95% (Siqueira; Fernandes, 2015; Amaral, 2023).

O conhecimento detalhado dessas características, sobretudo morfológicas e de pigmentação, é a base para o desenvolvimento de sistemas de classificação inteligentes, que objetivam extrair características de maneira automatizada para auxiliar no diagnóstico médico.

### 2.1.6 Dermatoscopia

A dermatoscopia, também denominada microscopia de epiluminescência, é um método diagnóstico não invasivo em diversos estudos empregado na avaliação de lesões cutâneas pigmentadas (Sarrafzadeh; Baygi; Ghassemi, 2010). A técnica estabelece uma interface entre o exame clínico a olho nu e a análise histopatológica, permitindo a visualização de estruturas subsuperficiais que não são acessíveis pela inspeção direta, conforme demonstrado na Figura 9. Essa característica eleva de em grande medida a precisão diagnóstica, aumentando a acurácia de cerca de 60%, obtida no exame clínico convencional, para valores entre 75% e 85% (Júnior; Bezerra; Andrade, 2020).

**Figura 9 – Comparação entre a inspeção visual e a dermatoscopia**



**Fonte: Ávila (2024).**

A utilidade da dermatoscopia decorre de sua capacidade de identificar padrões e estruturas morfológicas presentes na epiderme, na junção dermoepidérmica e na derme papilar. O exame é realizado com o dermatoscópio, instrumento óptico que utiliza iluminação direta ou polarizada, combinado a lentes de ampliação. Com frequência utiliza-se um líquido de imersão, como óleo, gel ou glicerina, com o objetivo de reduzir a reflexão da luz na camada córnea, aumentando a translucidez e permitindo melhor visualização das camadas profundas (Carmo; Silva, 2008).

Do ponto de vista clínico, a dermatoscopia constitui ferramenta essencial para diferenciar lesões pigmentadas benignas e malignas, sendo em particular relevante para o diagnóstico precoce do melanoma. Além de reduzir erros diagnósticos, a técnica orienta condutas terapêuticas, auxilia na decisão sobre necessidade de biópsia e melhora a capacidade de acompanhamento evolutivo de lesões suspeitas.

Quanto à instrumentação, a técnica pode ser classificada em manual e digital, caracterizando diferentes abordagens no processo de aquisição e análise das imagens. A dermatoscopia manual, ou simples, baseia-se no uso de um dermatoscópio portátil, cuja interpretação depende apenas da experiência e do treinamento do examinador. Assim, sua eficácia está, de modo direto, relacionada ao domínio técnico e à familiaridade do profissional com os padrões dermatoscópicos (Coelho, 2016).

Por outro lado, a dermatoscopia digital integra o dermatoscópio a sistemas de captura e armazenamento de imagens, como câmeras digitais ou computadores equipados com videodermatoscopia. Essa modalidade permite registrar imagens em alta resolução, possibilitando seu arquivamento, comparação evolutiva e processamento por softwares especializados. A digitalização das imagens favorece a incorporação de técnicas de Processamento Digital de Imagens (PDI), fundamentais para o desenvolvimento de sistemas de Diagnóstico Auxiliado por Computador (CAD). Esses sistemas reduzem a subjetividade inerente à avaliação humana, contribuindo para a análise padronizada e a classificação automatizada das lesões cutâneas (Coelho, 2016; Júnior; Bezerra; Andrade, 2020).

Os avanços tecnológicos recentes também possibilitaram integrar dermatoscópios a dispositivos móveis, como smartphones, mediante adaptadores ópticos específicos. Essa portabilidade amplia o acesso ao pré-diagnóstico, potencializa a tele dermatologia e permite que especialistas realizem avaliação remota. Além disso, possibilita que algoritmos embarcados realizem triagem inicial de lesões suspeitas, apoiando a identificação precoce de casos que demandam intervenção (Alencar, 2015).

A compreensão das bases técnicas e diagnósticas da dermatoscopia é fundamental para o desenvolvimento de sistemas computacionais capazes de interpretar imagens dermatoscópicas de forma padronizada. A identificação dos padrões estruturais característicos das lesões constitui o alicerce para a extração automatizada de atributos e para o treinamento de modelos de DL. Nesse sentido, a próxima seção abordará os principais fundamentos da Visão Computacional (VC), área que fornece as técnicas essenciais para o processamento, análise e classificação automática das imagens utilizadas neste estudo.

## 2.2 Visão Computacional

A visão computacional constitui uma área da ciência da computação dedicada ao desenvolvimento de métodos capazes de extrair, interpretar e transformar informações relevantes a partir de imagens ou sequências de imagens. Seu propósito central é automatizar tarefas por natureza associadas à visão humana, tais como reconhecimento, classificação, análise estrutural e compreensão de padrões visuais presentes no ambiente (Alencar, 2015; Coelho, 2016). Para isso, um sistema de VC em geral é estruturado em etapas sequenciais, que incluem aquisição da imagem, pré-processamento, segmentação, extração ou seleção de características e, por fim, reconhecimento ou interpretação dos padrões identificados. Ao contrário do PDI, cujo foco principal é o aprimoramento visual visando à percepção humana, a VC busca compreender o conteúdo semântico presente na imagem, permitindo a análise automatizada (Júnior; Bezerra; Andrade, 2020).

No ecossistema da inteligência artificial, a VC integra-se de forma estratégica ao campo mais amplo da IA, sobretudo nas subáreas de Aprendizado de Máquina (*Machine Learning*, ML) e DL. O ML engloba métodos que permitem aos sistemas computacionais aprenderem padrões a partir de dados, enquanto o DL utiliza arquiteturas compostas por múltiplas camadas de redes

neurais, capazes de extrair representações hierarquizadas e complexas. Nesse contexto, as RNCs tornaram-se a abordagem predominante em VC devido à sua capacidade de capturar por processo automático padrões espaciais e texturais das imagens, dispensando a necessidade de extração manual de características (Júnior; Bezerra; Andrade, 2020; Horiuchi, 2022).

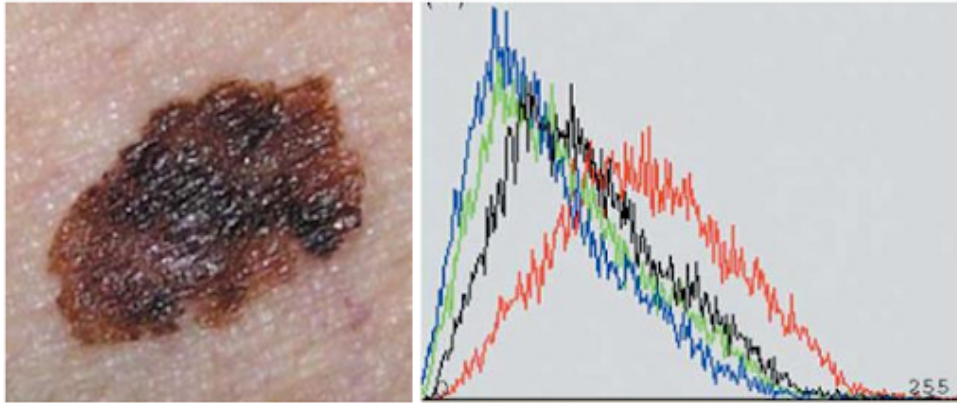
No domínio das aplicações médicas, a VC e o PDI desempenham papel fundamental no desenvolvimento de sistemas de CAD. Tais sistemas visam apoiar a prática clínica reduzindo a subjetividade inerente à interpretação humana, padronizando análises e aumentando a acurácia diagnóstica (Júnior; Bezerra; Andrade, 2020). Na dermatologia, em particular, a aplicação da VC na análise de imagens dermatoscópicas tem apresentado resultados expressivos, possibilitando a identificação precoce de lesões malignas e a diferenciação precisa entre múltiplos tipos de neoplasias, como o melanoma e os carcinomas (Alencar, 2015). Considerando que o melanoma é o tipo mais letal entre os cânceres de pele, a incorporação de técnicas avançadas de análise automatizada torna-se relevante, contribuindo para decisões clínicas mais rápidas, seguras e fundamentadas.

### 2.2.1 Imagens Digitais

O desenvolvimento de classificadores inteligentes baseados em DL exige a compreensão precisa de como a informação visual é representada em formato digital, uma vez que as RNCs operam de modo direto sobre estruturas matriciais de dados. A representação matemática das imagens constitui, portanto, a base conceitual para as etapas de pré-processamento, segmentação e extração da Região de Interesse (ROI) (Santos *et al.*, 2017).

Uma imagem digital pode ser, em termos formais, descrita como uma função bidimensional de intensidade luminosa  $f(x, y)$ , na qual  $x$  e  $y$  correspondem às coordenadas espaciais, e  $f(x, y)$  ao valor de brilho naquele ponto. Para possibilitar sua interpretação por sistemas computacionais, essa função contínua é submetida aos processos de amostragem, discretização espacial que converte o plano contínuo em uma matriz, e quantização, que transforma os níveis contínuos de intensidade em valores discretos (Coelho, 2016). A Figura 10 exemplifica essa representação, demonstrando como a informação visual de um melanoma é decomposta em distribuições de intensidade nos canais de cor RGB (Red, Green, Blue) e em tons de cinza.

**Figura 10 – Análise de uma lesão de melanoma e seu histograma**

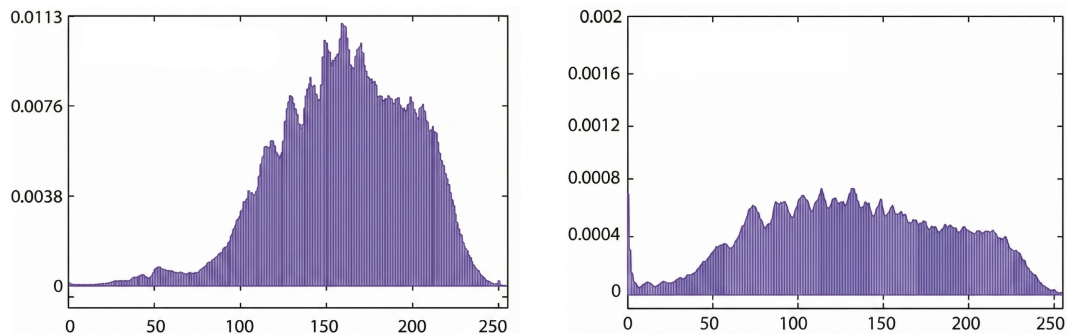


**(A) Imagem dermatoscópica da lesão. (B) Histograma RGB e tons de cinza.**

**Fonte: Adaptado de Manousaki et al. (2006).**

Em contrapartida, as lesões benignas manifestam um perfil estatístico mais estável e uniforme. Com o intuito de evidenciar essa distinção, a Figura 11 apresenta um comparativo entre as funções de distribuição de brilho de duas lesões cutâneas distintas.

**Figura 11 – Comparativo das funções de distribuição de brilho: melanoma vs. nevo benigno**



**(A) Histograma de melanoma maligno (elevada variância). (B) Histograma de nevo pigmentado (unimodal e concentrado).**

**Fonte: Adaptado de Orel et al. (2019).**

O histograma à esquerda refere-se a um melanoma maligno, evidenciando uma elevada variância e heterogeneidade interna, características da natureza caótica associada ao crescimento tumoral. Por outro lado, o gráfico à direita representa um nevo benigno, caracterizado por uma distribuição unimodal e concentrada, o que denota maior homogeneidade cromática. Essa estabilidade estatística é um fator determinante para a convergência dos algoritmos de segmentação e para o desempenho discriminatório dos classificadores baseados em DL (Gopal *et al.*, 2023).

Sob a ótica da representação computacional, uma imagem digital é representada como uma matriz discreta de dimensão  $N \times M$ , ou  $H \times W$  (altura  $\times$  largura), cujos elementos são denominados *pixels* (*picture elements*). Cada pixel armazena um valor de intensidade luminosa, que pode corresponder tanto a uma escala de cinza quanto a múltiplos canais de cor (Alencar, 2015; Júnior; Bezerra; Andrade, 2020).

Nas imagens coloridas, o modelo mais difundido é o RGB, em que a imagem é constituída por três matrizes  $H \times W$  empilhadas, cada uma representando um canal de cor primária. A combinação dos valores desses canais define a cor final de cada pixel. De forma geral, utiliza-se profundidade de 8 bits por canal, possibilitando 256 níveis de intensidade para cada cor e mais de 16 milhões de combinações possíveis (Coelho, 2016; Júnior; Bezerra; Andrade, 2020).

Em contraste, imagens em escala de cinza apresentam apenas um canal ( $C = 1$ ), sendo com frequência obtidas pela conversão a partir do modelo RGB. Entre os métodos de conversão, destaca-se o da luminância, baseado no modelo YIQ, composto pela luminância (Y) e pelas componentes de crominância (I) (*In-phase*) e (Q) (*Quadrature*). Desenvolvido originalmente para a transmissão televisiva, este modelo separa a informação de brilho (Y) da informação de cor (I e Q). No processamento de imagens, a componente Y é obtida através de uma média ponderada dos canais RGB ( $Y = 0,299R + 0,587G + 0,114B$ ), garantindo que a conversão para tons de cinza preserve a percepção humana de luminosidade sem perda de informação estrutural (Coelho, 2016).

A resolução de uma imagem, definida por suas dimensões  $H \times W$ , exerce influência direta no custo computacional do treinamento de modelos de DL. Bases de dados dermatoscópicas costumam apresentar resoluções variadas, o que exige padronização prévia. O redimensionamento para tamanhos menores, como  $256 \times 256$  pixels, reduz o custo computacional e acelera o processo de treinamento sem comprometer de forma significativa a capacidade de generalização das RNCs (Ferreira, 2018).

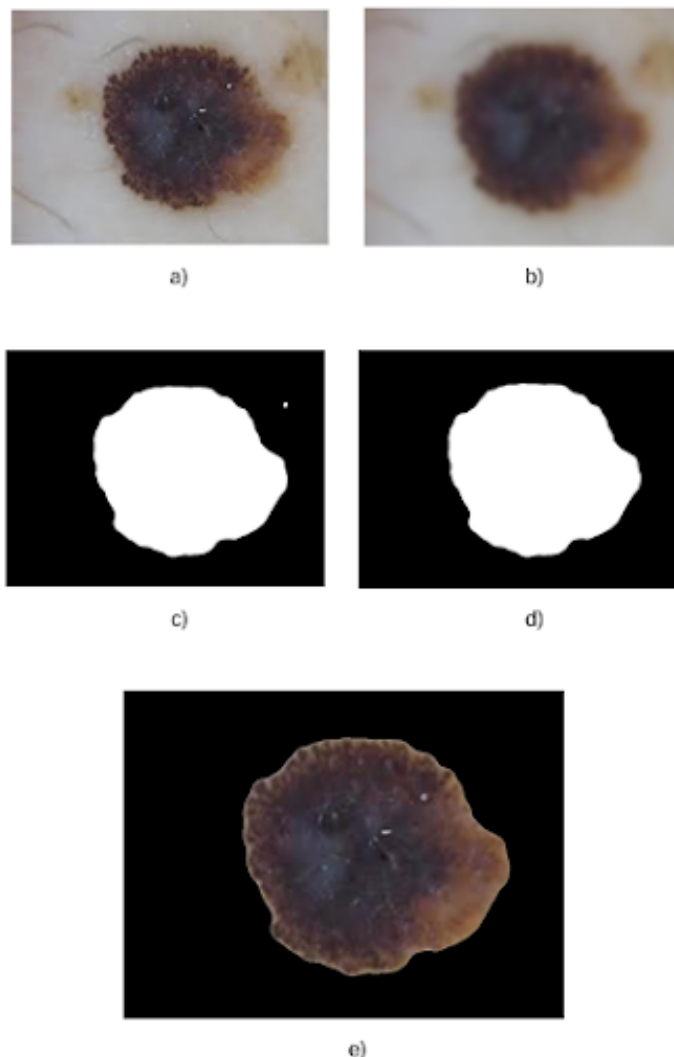
A qualidade das imagens utilizadas em sistemas de Diagnóstico Auxiliado por Computador (*Computer-Aided Diagnosis – CAD*) constitui um fator crítico para o desempenho dos classificadores. Formatos que empregam compressão com perda podem introduzir artefatos indesejados, degradando informações relevantes, em particular em estruturas finas da lesão, como bordas e padrões pigmentares (Horiuchi, 2022).

A análise automatizada de imagens dermatoscópicas, embora promissora, apresenta desafios intrínsecos decorrentes tanto de artefatos do processo de aquisição quanto da própria morfologia das lesões. Um dos principais obstáculos é o baixo contraste entre a lesão e a pele circundante, que compromete a segmentação e dificulta a delimitação precisa do contorno (Sarrafzadeh; Baygi; Ghassemi, 2010). Além disso, melanomas e lesões benignas podem apresentar elevada similaridade visual, com bordas irregulares ou difusas, dificultando ainda mais a separação entre áreas lesionadas e pele saudável (Araújo; Junior, 2017).

Artefatos físicos também impactam a interpretação visual automatizada. Pelos sobre a área avaliada podem invadir a lesão e gerar falsas bordas; fluidos de imersão utilizados durante o exame (óleo, gel, glicerina ou água) podem provocar bolhas e ruídos ópticos; e variações de iluminação ou falhas no uso de luz polarizada produzem sombras e reflexos, prejudicando o contraste e a uniformidade cromática (Sarrafzadeh; Baygi; Ghassemi, 2010; Araújo; Junior, 2017; Popescu; El-khatib; Ichim, 2022). Elementos externos, como as bordas escuras do dermatoscópio (*black frames*) e marcadores de escala, também interferem no processo de segmentação e devem ser removidos (Sarrafzadeh; Baygi; Ghassemi, 2010).

Nesse contexto, o pré-processamento torna-se etapa essencial. Seu objetivo é mitigar degradações e eliminar artefatos, garantindo que as características relevantes da lesão sejam preservadas. Técnicas como filtros de mediana, métodos de remoção de pelos e algoritmos de correção de iluminação são aplicados antes da segmentação, a qual isola a ROI da pele saudável circundante e prepara a imagem para etapas posteriores de análise (Mahmoud; Al-jumaily; Takruri, 2011; Sarrafzadeh; Baygi; Ghassemi, 2010). A Figura 12 exemplifica esse processo, demonstrando como a aplicação de filtros suavizadores remove a interferência dos pelos, permitindo uma segmentação mais precisa da lesão.

**Figura 12 – Processo de pré-processamento e segmentação em lesão com artefatos (pelos)**



**(A) Imagem original com presença de pelos. (B) Resultado da aplicação do filtro Gaussiano para remoção de ruídos. (C) Segmentação inicial. (D) Aplicação de crescimento de região. (E) Segmentação final da lesão.**

**Fonte: Adaptado de Coelho (2016).**

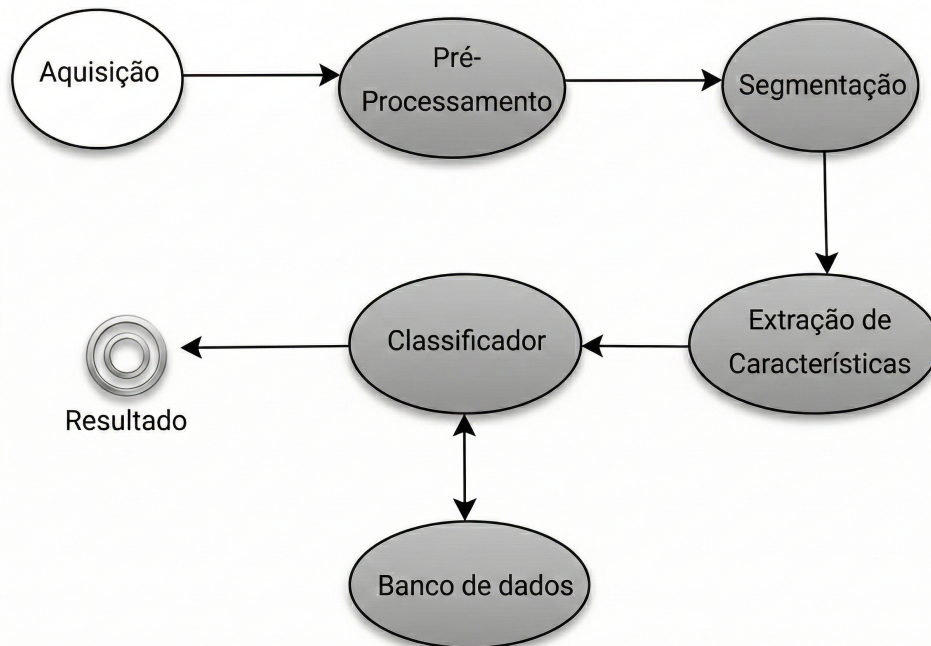
### 2.2.2 Reconhecimento de Padrões

O reconhecimento de padrões consiste no processo de identificação e extração de características significativas em dados brutos, objetivando sua classificação com base em um

modelo preestabelecido. No contexto da visão computacional aplicada à dermatologia, isso se traduz na capacidade de sistemas de IA analisarem matrizes de *pixels* para distinguir classes de interesse clínico, em particular na diferenciação entre patologias cutâneas como o melanoma e os carcinomas (Zhang *et al.*, 2023).

Ao longo da história, os sistemas de CAD para lesões cutâneas seguem um fluxo de trabalho sequencial (*pipeline*), na descrição clássica, dividido em pré-processamento, segmentação, extração de características e classificação, conforme ilustrado na Figura 13.

**Figura 13 – Etapas típicas de um sistema de diagnóstico assistido por computador**



**Fonte: Adaptado de Coelho (2016).**

A etapa inicial, o pré-processamento, é crucial para a mitigação de ruídos inerentes à aquisição dermatoscópica, tais como reflexos, bolhas de ar e pelos. Técnicas como a filtragem Gaussiana (suavização) e filtros de mediana são comumente aplicados para a remoção de artefatos, enquanto a equalização de histograma ou métodos como o *Independent Histogram Pursuit* (IHP) são empregados para realce de contraste em lesões pouco delimitadas (Sarrafzadeh; Baygi; Ghassemi, 2010; Coelho, 2016). Visando a redução do custo computacional em abordagens tradicionais, a conversão do espaço RGB para escala de cinza ou a seleção de canais específicos, como o canal Azul, que apresenta maior discriminação da lesão pigmentada, são práticas documentadas na literatura (Rosado, 2009; Siqueira; Fernandes, 2015).

Subsequente ao pré-processamento, a segmentação visa isolar a ROI da pele saudável. Abordagens clássicas baseiam-se na detecção de descontinuidades (bordas) ou similaridades (regiões). Métodos de limiarização (*thresholding*), como o algoritmo de Otsu, buscam minimizar a variância intra-classe para separar objeto e fundo. Como alternativa, técnicas de *clustering*, como o *Fuzzy C-Means* (FCM), e filtros detectores de bordas (Sobel, Laplaciano) foram em larga escala utilizados para delimitar a lesão e permitir o cálculo de atributos geométricos e clínicos, essenciais para a precisão diagnóstica (Araújo; Junior, 2017; Ferreira, 2018).

No paradigma clássico de ML, a etapa de extração de características é realizada por intervenção manual, transformando a imagem segmentada em um vetor de atributos quantificáveis. Na dermatoscopia, esses atributos são com frequência inspirados na regra ABCDE. Algoritmos como a fórmula TDS (*Total Dermatoscopy Score*) operacionalizam esses conceitos atribuindo pesos a fatores como assimetria e variabilidade de cores para estimar a probabilidade de malignidade (Coelho, 2016; Santos *et al.*, 2017). Pontuações abaixo de 4,8 indicam lesão benigna; entre 4,8 e 5,45, lesão suspeita; e acima de 5,45, alta probabilidade de melanoma. A fórmula TDS é ponderada conforme a Equação 1:

$$TDS = (A \cdot 1,3) + (B \cdot 0,1) + (C \cdot 0,5) + (D \cdot 0,5), \quad (1)$$

em que  $A$  representa o escore de assimetria,  $B$  indica a irregularidade das bordas,  $C$  quantifica a variação de cores e  $D$  corresponde à contagem de estruturas dermatoscópicas diferenciais.

Para a etapa final de classificação nesse modelo tradicional, o vetor de características alimenta algoritmos supervisionados. Dentre os mais prevalentes na literatura, destacam-se o *K-Nearest Neighbor* (K-NN), Máquinas de Vetor de Suporte (SVM) e Redes Neurais Artificiais rasas, como o *Multilayer Perceptron* (MLP). Estudos comparativos indicam que o SVM e o MLP, de forma regular, superam classificadores mais interpretáveis, como K-NN e Árvores de Decisão, na tarefa de discriminar lesões baseadas em características extraídas manualmente (Alencar, 2015; Zhang *et al.*, 2023).

Entretanto, a área observou uma mudança de paradigma com a consolidação do DL. Diferente do fluxo clássico, onde a extração de características é um processo manual e propenso a viés, arquiteturas modernas como as RNCs unificam essa etapa e a classificação em um processo de ponta a ponta (*end-to-end*).

Essas redes realizam a aprendizagem de representações de forma hierárquica: camadas iniciais detectam padrões simples (bordas, texturas), enquanto camadas profundas abstraem conceitos semânticos complexos da lesão, mimetizando o processamento visual biológico (Ferreira, 2018). Em estudos recentes, arquiteturas baseadas em *Vision Transformers* (ViTs) introduziram mecanismos de autoatenção (*self-attention*), permitindo que o modelo pondere a relevância global de *pixels* distantes entre si, superando limitações de campo receptivo dos modelos convolucionais convencionais e aprimorando a capacidade de generalização em diagnósticos complexos (Zhang *et al.*, 2023).

A aplicação dessas arquiteturas em problemas de classificação de imagens, como o diagnóstico de melanoma e dos carcinomas, é em diversos estudos validada por sua capacidade de extrair características (*features*) hierárquicas de alto nível. Contudo, essa abordagem exige grandes volumes de dados para otimizar de fato seus milhões de parâmetros e evitar o *overfitting*. Embora sejam muito eficazes, seu treinamento do zero demanda bases de dados extensas, o que contrasta com a realidade de bases de dado médicos, com frequência pequenos e desbalanceados. Treinar modelos como *Visual Geometry Group* (VGG-19), *Residual Network* (ResNet-50) ou *Densely Connected Convolutional Network* (DenseNet-121) a partir do zero é,

portanto, pouco viável (Júnior; Bezerra; Andrade, 2020). Para contornar essa limitação, aplica-se a Transferência de Aprendizado (*Transfer Learning* – TL). Nessa abordagem, utiliza-se um modelo pré-treinado em grandes bases como o ImageNet, cujas camadas convolucionais já capturam padrões visuais essenciais, como bordas, texturas e formas geométricas (Cavalcante, 2019; Júnior; Bezerra; Andrade, 2020). Em seguida, realiza-se o ajuste fino, substituindo e ajustando as camadas finais para a tarefa de classificação multiclasse proposta (melanoma, CBC e CEC), enquanto as camadas iniciais são preservadas (Gouda *et al.*, 2022). Por conseguinte, essa abordagem não apenas acelera a convergência do treinamento, mas também confere ao classificador uma robusta capacidade de generalização, elevando a performance mesmo diante de bases de dado limitados. Modelos como ResNet-50, InceptionV3 e Inception-ResNet figuram entre os mais utilizados na literatura (Gouda *et al.*, 2022; Zhang *et al.*, 2023).

Outro desafio recorrente é o desbalanceamento dos dados, com destaque no conjunto ISIC (*International Skin Imaging Collaboration*), no qual o melanoma, o CBC e o CEC em geral representam a minoria dos casos em comparação com nevos e outras lesões benignas. Essa disparidade tende a induzir um viés severo no treinamento, levando o modelo a priorizar a classe majoritária para minimizar o erro global, o que resulta em um desempenho enganoso e incapacidade de generalização (Júnior; Bezerra; Andrade, 2020). Para mitigar esse problema e expandir a representatividade do espaço de características, técnicas de Aumento de Dados (*Data Augmentation* - DA) são essenciais. Transformações como rotações, espelhamentos, cortes e ajustes cromáticos ampliam a diversidade das amostras. Em domínios específicos, abordagens como a simulação sintética de pelos (*hair augmentation*) são cruciais para replicar a variabilidade de aquisição dos diferentes dermatoscópios e ampliar a resistência do modelo a oclusões (Santos *et al.*, 2017; Mateen *et al.*, 2024). Dessa forma, a rede aprende características mais robustas, forçando a extração de características de alto nível essenciais para a discriminação semântica da lesão, e reduz o risco de *overfitting*.

No âmbito da avaliação de desempenho, métricas como acurácia tornam-se insuficientes, já que o desbalanceamento pode encobrir falhas na detecção das lesões malignas (melanoma, CBC e CEC), que são as classes de maior relevância clínica. Por isso, a sensibilidade é priorizada, pois mede a proporção de Verdadeiros Positivos corretamente identificados. Minimizar falsos negativos é crítico em um diagnóstico oncológico (Júnior; Bezerra; Andrade, 2020; Mateen *et al.*, 2024). Como complemento, utiliza-se F1-Score e a AUC-ROC (Área sob a Curva Característica de Operação do Receptor), que avaliam a capacidade discriminativa do modelo em diferentes limiares (Gouda *et al.*, 2022; Raza *et al.*, 2025).

Por fim, ao integrar DA, TL e métricas adequadas, busca-se desenvolver um classificador robusto, capaz de superar limitações dos métodos manuais e fornecer suporte confiável ao diagnóstico clínico. Assim, torna-se fundamental compreender as bases teóricas que sustentam esses algoritmos, motivando a discussão apresentada na seção seguinte sobre os fundamentos da IA e do ML.

## 2.3 Inteligência Artificial e Aprendizado de Máquina

A IA constitui um campo multidisciplinar da ciência da computação dedicado ao desenvolvimento de sistemas capazes de emular funções cognitivas humanas, tais como o aprendizado, o raciocínio, a percepção e a resolução de problemas (Coelho, 2016). De maneira abrangente, a IA representa o esforço científico para automatizar tarefas intelectuais que, tradicionalmente, demandam intervenção humana. Enquanto campo de estudo, a IA é vasta e hierárquica, englobando subáreas como o ML e o DL, conforme ilustrado na literatura (Chollet, 2018).

Sob perspectiva histórica, o desenvolvimento da IA bifurcou-se em dois paradigmas fundamentais. A IA Simbólica (ou Clássica), predominante nas primeiras décadas da área, opera sob a premissa de manipulação de símbolos explícitos. Nessa abordagem, programadores codificam regras lógicas e determinísticas ("se-então") para processar informações. Embora eficaz para domínios fechados e bem definidos, como jogos de xadrez, a abordagem simbólica demonstrou-se insuficiente para resolver problemas complexos e não estruturados, como o reconhecimento de fala ou a classificação de imagens, onde as regras são difusas e difíceis de formalizar manualmente (Chollet, 2018).

Em contrapartida, a IA conexionista ou estatística fundamenta o paradigma moderno do ML. Inspirada nos mecanismos biológicos de processamento de informação, notadamente nas Redes Neurais Artificiais (RNA), essa abordagem busca desenvolver algoritmos capazes de induzir modelos de conhecimento a partir de dados empíricos (amostras), permitindo a identificação de padrões sem a necessidade de programação explícita de regras lógicas (Júnior; Bezerra; Andrade, 2020).

No contexto médico, a IA consolidou-se como a base dos sistemas CAD. Tais sistemas atuam como ferramentas de suporte à decisão clínica, visando mitigar a subjetividade e a variabilidade interobservadora inerentes à avaliação humana. A integração da dermatoscopia digital com técnicas de PDI e IA possibilita a análise quantitativa das lesões cutâneas (Gouda *et al.*, 2022). Ao processar grandes volumes de dados e identificar nuances morfológicas imperceptíveis ao olho nu, como variações sutis nos parâmetros da regra ABCDE, esses algoritmos oferecem uma segunda opinião robusta, contribuindo para a redução de biópsias desnecessárias e para o aumento da precisão no diagnóstico precoce do melanoma e dos carcinomas (Coelho, 2016).

### 2.3.1 Paradigma do Aprendizado de Máquina

O ML é a subárea da IA que confere aos sistemas computacionais a capacidade de aprender e melhorar seu desempenho a partir da experiência (dados), sem serem, de forma explícita, programados para cada situação específica. Em essência, o ML é uma técnica utilizada para analisar e prever padrões a partir de uma amostra de dados, que serão em seguida aplicados em novos dados para realizar inferências (Júnior; Bezerra; Andrade, 2020).

O ML representa uma inversão fundamental no paradigma de programação tradicional. No modelo clássico, o fluxo lógico exige que regras e dados sejam fornecidos para gerar respostas. No aprendizado de máquina, o sistema recebe os Dados (imagens) e as Respostas esperadas (os rótulos ou diagnósticos corretos), e o algoritmo é responsável por inferir as Regras matemáticas que correlacionam as entradas às saídas (Júnior; Bezerra; Andrade, 2020).

Para o problema de classificação de neoplasias cutâneas, este paradigma é essencial. As características visuais que distinguem uma lesão maligna, como assimetria complexa, bordas irregulares e variabilidade de cores, são difíceis de descrever através de regras rígidas de programação. No entanto, algoritmos de ML podem aprender a ponderar esses atributos de forma estatística.

O objetivo final não é apenas memorizar os exemplos de treino, mas alcançar a generalização: a capacidade de classificar de forma adequada novos casos clínicos nunca antes vistos pelo sistema. Algoritmos clássicos como K-NN, SVM e RNAs exemplificam técnicas que, historicamente, exploraram esse paradigma para automatizar a análise de padrões dermatoscópicos (Coelho, 2016; Júnior; Bezerra; Andrade, 2020).

### 2.3.2 Aprendizado Supervisionado

O aprendizado supervisionado constitui o paradigma metodológico central para a tarefa de classificação de imagens dermatoscópicas proposta neste trabalho. Esta abordagem fundamenta-se no treinamento de algoritmos mediante o fornecimento de um conjunto de dados rotulados, permitindo que o modelo aprenda, por indução, a função de mapeamento entre as variáveis de entrada e as saídas esperadas (Ferreira, 2018). No contexto deste estudo, a entrada ( $x$ ) corresponde à imagem digital da lesão cutânea, enquanto a saída ( $y$ ) representa o rótulo da classe, categorizando a lesão em múltiplos diagnósticos possíveis: melanoma, CBC ou CEC.

A premissa fundamental deste paradigma é a existência de um *Ground Truth* (GT) confiável. Em aplicações de IA Médica, a veracidade dos rótulos é crítica e deve ser estabelecida por padrões-ouro, como o exame histopatológico (biópsia) ou o consenso de especialistas. Bases de dados de referência, como a *PH2* e o arquivo do ISIC, garantem essa confiabilidade, fornecendo diagnósticos confirmados na prática clínica. O processo de treinamento de RNAs, como o MLP ou as RNCs, ocorre através da comparação iterativa entre a predição do modelo e este rótulo verdadeiro (Ferreira, 2018; Gouda *et al.*, 2022).

Em termos formais, o objetivo do aprendizado supervisionado é aproximar uma função que, parametrizada por um conjunto de pesos e vieses, minimize o erro de predição em dados não vistos. Esse processo é formulado como um problema de otimização, onde busca-se encontrar os parâmetros que minimizam uma Função de Custo ou Perda (*Loss Function*). Esta função quantifica a divergência entre a previsão da rede e o rótulo real. Durante o treinamento, algoritmos de otimização baseados em gradiente, como o *backpropagation* (retropropagação do erro), ajustam os pesos da rede na direção oposta ao gradiente da função de perda, refinando

de forma gradual a capacidade discriminativa do modelo para a detecção precoce das lesões abordadas (Ferreira, 2018; Cavalcante, 2019).

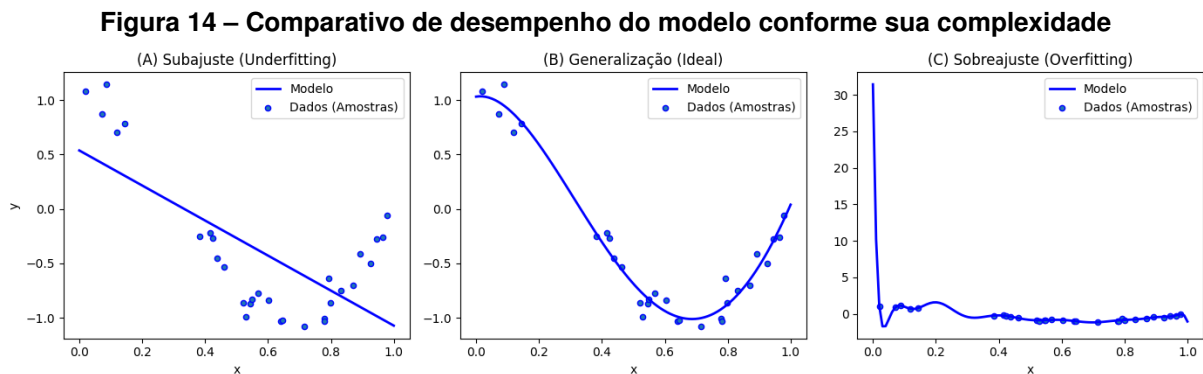
### 2.3.3 Generalização, *Overfitting* e Divisão de Dados

A capacidade de Generalização é o objetivo final de qualquer modelo de ML aplicado ao diagnóstico, definindo-se como a habilidade do sistema em manter alta acurácia quando confrontado com dados inéditos (*unseen data*) que não fizeram parte do processo de treinamento. Um modelo generalista é aquele que capturou os padrões latentes da doença, e não apenas memorizou os exemplos específicos apresentados (Ferreira, 2018).

Entretanto, o processo de aprendizagem está sujeito a dois modos principais de falha que comprometem esse objetivo. O primeiro, denominado subajuste (*Underfitting*), ocorre quando o modelo possui complexidade insuficiente para capturar a estrutura dos dados, resultando em baixo desempenho tanto no treinamento quanto no teste. Em redes neurais, isso em geral denota uma arquitetura, em excesso, rasa ou treinamento insuficiente (Cavalcante, 2019).

Em contrapartida, o sobreajuste (*Overfitting*) representa o fenômeno crítico onde o modelo se torna excessivamente especializado nos dados de treinamento, capturando ruídos estocásticos e particularidades irrelevantes da amostra. Nesse cenário, observa-se uma alta acurácia no treino, mas uma degradação severa no teste, conforme ilustrado na representação esquemática da Figura 14, simulação matemática para fins ilustrativos, fundamentada em dados sintéticos para fins demonstrativos. O *overfitting* é induzido por treinamento excessivo (muitas épocas) ou falta de regularização. Para mitigá-lo, técnicas como DA e *Dropout* (inativação aleatória de neurônios) são indispensáveis para forçar a rede a aprender características robustas e distribuídas (Gouda *et al.*, 2022; Amaral, 2023; Raza *et al.*, 2025).

Na Figura 14, o eixo horizontal (x) representa os valores da variável de entrada sintética, enquanto o eixo vertical (y) representa a variável de resposta associada a cada entrada. Os pontos correspondem às amostras utilizadas, e a linha contínua representa o ajuste produzido pelo modelo.



**(A) Modelo com baixa complexidade incapaz de capturar a curva (Subajuste). (B) Modelo que captura o padrão real dos dados (Generalização). (C) Modelo excessivamente complexo que incorpora o ruído dos dados (Sobreajuste).**

Fonte: Autoria própria (2025).

Para assegurar a generalização e evitar vieses de avaliação, é imperativo o particionamento rigoroso do *dataset* em três subconjuntos distintos: treinamento, validação e teste. O conjunto de treinamento é utilizado para o aprendizado do modelo, onde os pesos e *bias* da rede são ajustados por iterações. A validação é utilizada para o monitoramento do desempenho a cada época e para o ajuste de hiperparâmetros (ex: taxa de aprendizado). O erro de validação é a métrica guia para interromper o treinamento (*early stopping*) antes que o *overfitting* ocorra (Mateen *et al.*, 2024).

Teste é uma partição de dados isolada, utilizada apenas ao final do desenvolvimento para fornecer uma estimativa imparcial da performance do modelo no mundo real. Métodos como a Validação Cruzada (*k-fold Cross-Validation*) são continuamente empregados para garantir, em termos estatísticos, que os resultados não dependem de uma partição específica dos dados (Santos *et al.*, 2017; Raza *et al.*, 2025).

Historicamente, antes da consolidação do DL, os sistemas CAD dependiam da extração manual de características, como descritores de textura e histogramas, definidos por engenharia humana. Essa abordagem limitava a escalabilidade e a adaptação dos modelos. Em contraste, o DL supera essa barreira ao permitir o aprendizado de representações (*feature learning*), onde algoritmos extraem, por processo automático, características hierárquicas e discriminativas dos dados brutos (Zhang *et al.*, 2023). Essa evolução tecnológica fundamenta o uso de RNCs neste trabalho, cujos princípios arquiteturais e funcionais serão detalhados na próxima seção.

#### 2.3.4 Desbalanceamento de Classes e a Técnica SMOTE

Um dos desafios mais críticos no desenvolvimento de modelos de Aprendizado de Máquina para aplicações médicas é o desbalanceamento inerente aos conjuntos de dados. Em repositórios de imagens dermatoscópicas, as diferentes categorias diagnósticas em poucos casos apresentam uma distribuição uniforme, de modo que patologias específicas, como o CEC, representam uma minoria acentuada (Gouda *et al.*, 2022). Essa disparidade quantitativa tende a induzir um viés severo durante o treinamento, levando o algoritmo a priorizar o acerto na classe majoritária com o intuito de minimizar com rapidez a função de custo global. Como consequência, o classificador pode atingir uma alta acurácia ilusória, falhando por processo sistemático na detecção da classe minoritária e comprometendo a capacidade de generalização do sistema (Buda; Maki; Mazurowski, 2018).

O impacto dessa assimetria no contexto das RNCs foi investigado em profundidade por Buda, Maki e Mazurowski (Buda; Maki; Mazurowski, 2018). Em seu estudo sistemático sobre o problema do desbalanceamento de classes, os autores comprovaram que a assimetria afeta de forma adversa e significativa o desempenho de arquiteturas profundas voltadas à classificação de imagens. Ademais, a pesquisa evidenciou que a aplicação de métodos em nível de dados, em particular a superamostragem das classes minoritárias, desponta como uma das abordagens empíricas mais consistentes e eficazes para mitigar esse viés em RNCs, supe-

rando alternativas consolidadas, como o ajuste de limiares ou a modificação manual de pesos nas funções de perda (Buda; Maki; Mazurowski, 2018).

Contudo, a simples replicação exata de amostras originais da classe minoritária pode exacerbar o risco de *overfitting*, uma vez que a rede neural acaba memorizando instâncias idênticas repetidas vezes, sem que haja a extração de nova variabilidade morfológica (Chawla *et al.*, 2002). Para transpor essa limitação matemática e estatística, Chawla *et al.* propuseram o algoritmo SMOTE (*Synthetic Minority Over-sampling Technique*), que substitui a replicação por geração inteligente de dados.

Em contraste das abordagens clássicas de aumento de dados que operam de modo direto no espaço da imagem bidimensional (por meio de transformações geométricas ou fotométricas controladas), o SMOTE fundamenta-se na criação de amostras apenas sintéticas por meio de interpolação algébrica. A técnica atua selecionando instâncias da classe minoritária e identificando seus vizinhos mais próximos (*k-Nearest Neighbors*) no espaço de características (*feature space*). Em seguida, instâncias artificiais são geradas matematicamente ao longo dos segmentos de reta que unem a amostra analisada aos seus vizinhos, introduzindo novas observações plausíveis na região de decisão da classe em desvantagem (Chawla *et al.*, 2002).

No contexto de sistemas CAD voltados para a oncologia cutânea, a aplicação do balanceamento sintético induzido pelo SMOTE confere à rede neural a oportunidade de extrair padrões discriminativos de maneira equitativa. Ao expandir o espaço de representação de categorias com menor suporte clínico (como o CEC), a técnica força o modelo convolucional a estabelecer uma fronteira de decisão mais ampla e robusta, reduzindo a prevalência de falsas predições em favor da classe majoritária e, por consequência, elevando a utilidade clínica do diagnóstico automatizado (Chawla *et al.*, 2002; Buda; Maki; Mazurowski, 2018).

## 2.4 Redes Neurais e Aprendizado Profundo

O avanço tecnológico e o surgimento de computadores cada vez mais potentes trouxeram consigo uma forte onda de pesquisa sobre o funcionamento do cérebro humano, na tentativa de desvendar como este consegue fazer tomadas de decisões de forma tão rápida e precisa, enquanto no mundo digital isso ainda representa um grande desafio (Alencar, 2015). Para tentar reproduzir o funcionamento do órgão humano, foram desenvolvidos modelos matemáticos denominados neurônios artificiais, que conseguem executar diversas funções lógicas. Ao combinar o funcionamento de diversos desses neurônios, formam-se sistemas com grande capacidade de processamento, conhecidos como RNAs (Júnior; Bezerra; Andrade, 2020). O desenvolvimento dessas redes acompanhou a evolução dos recursos computacionais, permitindo a criação de algoritmos mais sofisticados para aumentar ainda mais o poder computacional dessas redes (Alencar, 2015).

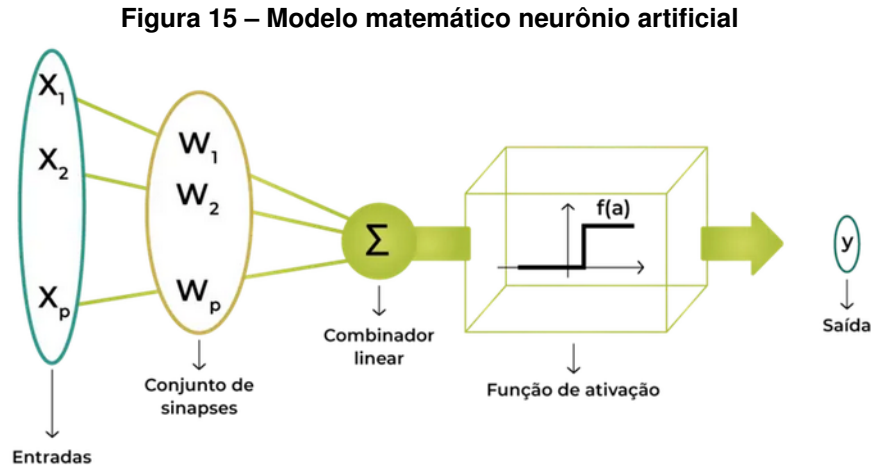
O *Deep Learning* diz respeito a um subconjunto das RNAs que utiliza múltiplas camadas ocultas para extrair características complexas dos dados de entrada. Essa abordagem é sobretudo eficaz para identificar padrões de alto nível; por esse motivo, é utilizada em tarefas

como reconhecimento de imagens, voz e processamento de linguagem natural (Júnior; Bezerra; Andrade, 2020).

Em aplicações mais simples, as RNAs compostas de apenas algumas camadas ocultas costumam apresentar um alto grau de confiabilidade. No entanto, em situações mais complexas, como a de análise de imagens, torna-se necessário o uso de uma rede neural profunda, que possui uma complexidade mais alta, com dezenas ou até mesmo centenas de camadas em que as camadas sucessivas processam e refinam as representações obtidas nas etapas anteriores (Júnior; Bezerra; Andrade, 2020).

#### 2.4.1 Estrutura e Treinamento

A estrutura de uma RNA é composta por componentes chamados de neurônios artificiais, inspirados no funcionamento de um neurônio presente no cérebro humano. Cada um deles realiza operações matemáticas simples, e que quando são interconectados em grandes quantidades, conseguem realizar tarefas complexas para reconhecimento de padrão e tomada de decisões (Alencar, 2015). A estrutura básica de um neurônio artificial é ilustrada na Figura 15, que apresenta o fluxo de processamento e as variáveis envolvidas nas operações de cada neurônio.



Fonte: Leone (2022).

De maneira análoga à rede neural do cérebro humano, cada neurônio recebe um conjunto de  $p$  entradas ( $X$ ), provenientes de fontes externas ou da saída de neurônios anteriores. Cada valor desse vetor de entrada possui um peso sináptico correspondente, utilizado para ponderar a importância de cada informação. Além disso, pode-se incluir um termo de viés (*bias*) para ajustar o limiar de decisão do neurônio, de forma semelhante ao termo independente de uma função afim (Alencar, 2015). O somatório ponderado gera uma função linear ( $a$ ), que é em seguida aplicada a uma função de ativação ( $f(a)$ ), resultando na saída do neurônio ( $y$ ).

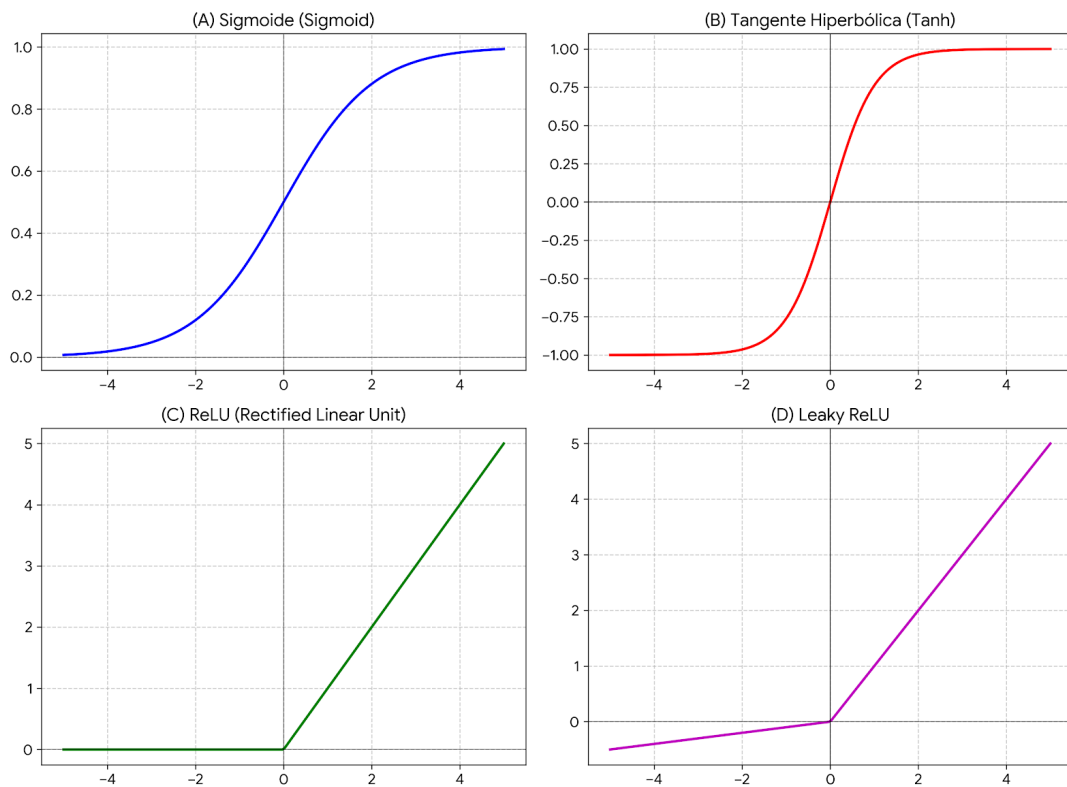
Quanto às funções de ativação, existem diversas opções utilizadas para introduzir não-linearidade à rede. Dentre elas destacam-se a linear, o limite rígido, a sigmoide e a tangente

hiperbólica, cujos comportamentos gráficos são apresentados na Figura 16 (Júnior; Bezerra; Andrade, 2020).

Na Figura 16, o eixo horizontal representa o valor de entrada ( $a$ ) da função de ativação, obtido pelo somatório ponderado das entradas do neurônio acrescido do termo de viés. O eixo vertical representa a saída ( $f(a)$ ) produzida pela respectiva função de ativação.

**Figura 16 – Representação gráfica das principais funções de ativação**

Principais Funções de Ativação



**(A) Sigmoide:** comprime a saída entre 0 e 1. **(B) Tanh:** centrada em zero, saída entre -1 e 1. **(C) ReLU (Rectified Linear Unit):** linear para positivos e nula para negativos. **(D) Leaky ReLU:** variante da ReLU para evitar neurônios mortos.

**Fonte: Autoria própria (2025).**

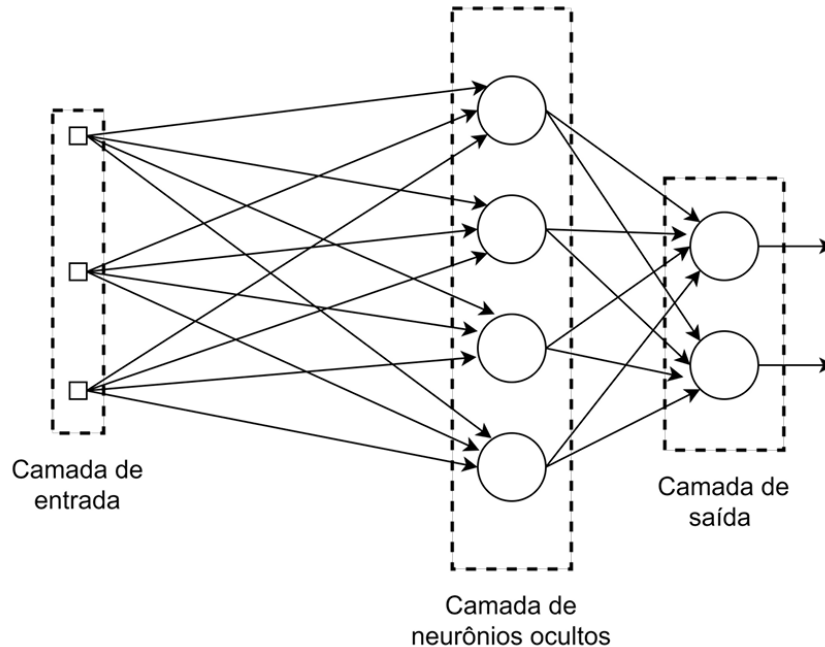
A saída representa como o neurônio avaliou as entradas e realizou sua decisão; esta informação será passada adiante como uma entrada para um próximo ramo da RNA, onde este cálculo explicado será executado outra vez para gerar uma próxima saída até que encerre o processo e gere uma decisão final (Júnior; Bezerra; Andrade, 2020).

De forma resumida, essas redes neurais podem ser divididas em 3 tipos de camadas, conforme estruturado na Figura 17:

- Camada de entrada: responsável por receber os dados;
- Camadas ocultas: realizam o processamento interno e o aprendizado das relações entre os neurônios, sendo chamadas assim por não possuírem conexão direta com o ambiente externo;

- Camada de saída: gera o resultado do sistema, onde obtemos a "resposta" do problema proposto.

**Figura 17 – Estrutura de uma rede neural de múltiplas camadas**



**Entrada: vetor de atributos  $x = [x_1, x_2, \dots, x_n]$  fornecido à rede neural. Saída: vetor de respostas ou previsões  $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m]$  produzido após o processamento pelas camadas ocultas.**

**Fonte: Adaptado de Prado Júnior et al. (2020).**

#### 2.4.2 Redes Neurais Convolucionais

As RNCs representam uma evolução das RNAs, sendo projetadas para processar dados com estrutura espacial, como imagens e vídeos. Enquanto os modelos tradicionais tratam cada entrada de forma independente, essas redes exploram a disposição bidimensional dos pixels, extraindo automaticamente padrões visuais, como bordas, formas e texturas, em diferentes níveis de complexidade (Júnior; Bezerra; Andrade, 2020).

Essa capacidade de detectar e combinar características locais faz dessa arquitetura uma das técnicas mais eficazes em reconhecimento de imagens, detecção de objetos e classificação de padrões visuais. Em vez de exigir que um especialista defina os atributos mais relevantes, elas aprendem essas representações dos dados de entrada, reduzindo a necessidade de pré-processamento (Santos *et al.*, 2017).

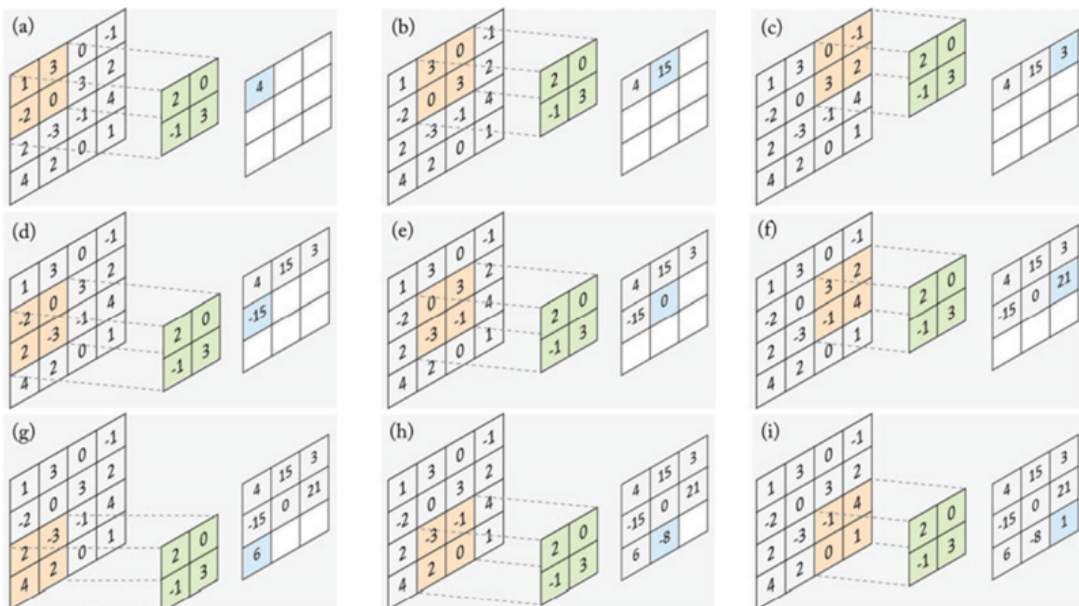
De modo geral, esse tipo de rede é composto por uma sequência de camadas, cada uma responsável por uma função específica no fluxo de processamento.

### 2.4.2.1 Camada Convolutiva

É a camada fundamental de uma RNC. Sua função é aplicar filtros (*kernels*) sobre a imagem de entrada, realizando uma operação chamada convolução, conforme demonstrado na Figura 18. Esses filtros percorrem toda a imagem e identificam padrões locais, como linhas, cantos e texturas (Alencar, 2015).

Nas camadas mais profundas, os filtros passam a detectar características mais abstratas, por exemplo, partes de objetos ou formas complexas, a partir da combinação das informações extraídas pelas camadas anteriores (Alencar, 2015).

**Figura 18 – Representação da operação de convolução: entrada, kernel e mapa de características**



As etapas de (A) a (I) representam o deslocamento da janela de *Max Pooling* sobre o mapa de entrada. Em cada posição, seleciona-se o maior valor da região analisada, que é então armazenado no mapa de características de saída.

Fonte: Adaptado de Prado Júnior et al. (2020).

### 2.4.2.2 Camada de Pooling

A camada de *Pooling* tem como objetivo reduzir a dimensão dos mapas de características gerados pela convolução, preservando as informações mais importantes. Essa operação diminui a quantidade de parâmetros e o custo computacional, além de aumentar a robustez da rede contra variações de posição ou ruído na imagem (Alencar, 2015).

O tipo mais comum é o *Max Pooling*, que seleciona o valor máximo dentro de uma região da imagem, representando a característica dominante daquele local (Santos *et al.*, 2017).

### 2.4.2.3 Camada Totalmente Conectada

Após as etapas de convolução e *pooling*, os mapas de características são transformados em um vetor unidimensional por meio de uma etapa chamada *flattening*, que os conecta a uma ou mais camadas totalmente conectadas (*fully connected layers*) (Lenc; Vedaldi, 2015).

Essas camadas funcionam de forma semelhante às das RNAs tradicionais, combinando os pesos e gerando a predição final.

#### 2.4.2.4 *Dropout e Batch Normalization (Camadas auxiliares)*

Além das camadas principais, as RNCs modernas incorporam camadas auxiliares que melhoram o desempenho e a estabilidade do treinamento.

- *Dropout*: desativa de forma aleatória uma porcentagem de neurônios durante o treinamento, reduzindo o risco de *overfitting*, ou seja, quando a rede se ajusta demais aos dados de treinamento e perde capacidade de generalização (Júnior; Bezerra; Andrade, 2020).
- *Batch Normalization*: normaliza as ativações intermediárias da rede, garantindo que os valores fiquem dentro de uma faixa controlada. Isso acelera o treinamento, permite taxas de aprendizado mais altas e torna o processo mais estável (Santos *et al.*, 2017).

Essas camadas são utilizadas em arquiteturas modernas, como VGG, ResNet e DenseNet, contribuindo para o sucesso das RNCs em tarefas de reconhecimento de imagens.

## 2.5 Arquiteturas de Redes Neurais Convolucionais

Com o avanço das RNCs, diversas arquiteturas foram desenvolvidas com o objetivo de otimizar o aprendizado, reduzir a complexidade computacional e melhorar o desempenho em tarefas de reconhecimento de imagens. Essas arquiteturas diferem em número de camadas, forma de conexão entre elas e estratégias de processamento dos dados, representando diferentes etapas da evolução das RNCs ao longo do tempo.

Um dos primeiros marcos foi a LeNet-5, proposta por Lecun em 1998, que introduziu de forma estruturada o uso de camadas convolucionais e operações de *pooling* para reconhecimento de dígitos manuscritos. Embora simples quando comparada aos modelos atuais, estabeleceu a base conceitual das RNCs modernas (Júnior; Bezerra; Andrade, 2020).

Em 2014, a arquitetura VGG consolidou o uso de convoluções 3×3 empilhadas, priorizando profundidade como forma de aumentar a capacidade de representação da rede. Com versões de 16 e 19 camadas, a VGG tornou-se referência pela extração robusta de características, apesar do elevado número de parâmetros e custo computacional associado (Júnior; Bezerra; Andrade, 2020).

A evolução seguinte veio com a ResNet, vencedora do ImageNet em 2015, que introduziu as conexões residuais (*skip connections*). Essa inovação solucionou o problema do desaparecimento do gradiente em redes muito profundas, permitindo treinamentos estáveis mesmo com centenas de camadas. Como consequência, a ResNet tornou-se uma das arquiteturas mais utilizadas em TL para aplicações médicas (Júnior; Bezerra; Andrade, 2020).

Depois, a DenseNet aperfeiçoou esse conceito ao conectar cada camada a todas as camadas subsequentes dentro de um bloco denso, promovendo forte reutilização de características e reduzindo a redundância de parâmetros. Essa abordagem aumentou a eficiência da rede e favoreceu o aprendizado com menos dados, tornando-se muito comum em aplicações biomédicas (Júnior; Bezerra; Andrade, 2020).

No contexto de segmentação de imagens médicas, a arquitetura U-Net destacou-se por seu formato *encoder-decoder* com conexões simétricas de *skip*. Essa estrutura permite preservar informações espaciais finas durante o processo de reconstrução da imagem segmentada, tornando a U-Net adotada em tarefas como segmentação de lesões dermatoscópicas (Júnior; Bezerra; Andrade, 2020).

Em estudos recentes, arquiteturas baseadas em autoatenção, como os ViTs, introduziram mecanismos capazes de modelar dependências globais entre regiões da imagem, superando algumas limitações do campo receptivo das RNCs tradicionais. Embora ainda demandem grandes conjuntos de dados, os ViTs têm se mostrado competitivos em aplicações médicas complexas (Júnior; Bezerra; Andrade, 2020).

Assim, a evolução histórica dessas arquiteturas reflete a busca contínua por modelos mais profundos, eficientes e capazes de generalizar em cenários de alta variabilidade visual, características essenciais para a análise automática de imagens dermatoscópicas. A arquitetura selecionada para este trabalho é apresentada na seção seguinte.

### 2.5.1 Arquitetura Escolhida

Para este trabalho, optou-se pelo uso de duas arquiteturas consolidadas na literatura atual de visão computacional: EfficientNet-B4 para a etapa de classificação de imagens dermatoscópicas e U-Net com codificador (*backbone*) ResNet-50 para a etapa de segmentação. A escolha dessas arquiteturas se baseia no equilíbrio entre desempenho, eficiência computacional e ampla utilização em aplicações médicas (Ronneberger; Fischer; Brox, 2015; Tan; Le, 2020).

A EfficientNet-B4 é uma arquitetura moderna de classificação que utiliza uma estratégia sistemática de escalonamento de profundidade, largura e resolução, permitindo alcançar alta precisão com um número reduzido de parâmetros em comparação com redes convencionais. Essa eficiência a torna adequada para bases de imagens médicas, nas quais detalhes sutis precisam ser identificados sem exigir recursos computacionais excessivos (Tan; Le, 2020).

Já para a segmentação, adotou-se a U-Net, uma arquitetura empregada em aplicações biomédicas por sua capacidade de preservar detalhes espaciais finos por meio de conexões de atalho entre a fase de contração e a fase de expansão da rede (Ronneberger; Fischer; Brox, 2015). Para melhorar ainda mais sua capacidade de extração de características, é utilizada uma ResNet-50 como codificador, permitindo que a rede combine a estrutura da U-Net com camadas residuais mais profundas e robustas.

Com essa combinação de EfficientNet-B4 para classificação e U-Net com ResNet-50 para segmentação, busca-se construir um fluxo de processamento sequencial e integrado, capaz de lidar primeiro com a delimitação precisa da lesão (isolando-a da pele saudável) para, em seguida, realizar o reconhecimento e a classificação exata entre as três patologias abordadas (melanoma, CBC e CEC), oferecendo resultados confiáveis dentro das exigências do contexto biomédico.

## 2.6 Métricas de Desempenho

A avaliação do desempenho de modelos baseados em redes neurais é essencial para verificar sua capacidade de generalização e assegurar que os resultados sejam, sob o ponto de vista clínico, confiáveis. Em aplicações médicas, a escolha das métricas deve levar em conta condições particulares dos conjuntos de dados, como desbalanceamento entre classes e grande variabilidade morfológica das lesões (Júnior; Bezerra; Andrade, 2020).

Para classificação, a qualidade das previsões é analisada por métricas como acurácia, precisão, sensibilidade e F1-score, que quantificam o comportamento do modelo ao distinguir de forma adequada as classes de interesse (Júnior; Bezerra; Andrade, 2020). Já em tarefas de segmentação, onde o objetivo é delimitar a área exata da lesão em relação ao tecido saudável, são necessárias métricas que avaliem a similaridade espacial entre as máscaras preditas e as máscaras reais. Nesse contexto, o coeficiente Dice é utilizado por medir o grau de sobreposição entre as regiões segmentadas (Júnior; Bezerra; Andrade, 2020).

Além das métricas, o processo de aprendizagem do modelo depende da otimização conduzida pela função de perda, responsável por orientar os ajustes nos pesos durante o treinamento. Embora não seja uma métrica de avaliação, a função de perda exerce influência direta sobre a convergência e sobre a capacidade do modelo de identificar padrões relevantes (Júnior; Bezerra; Andrade, 2020).

Nos tópicos seguintes, são descritos o coeficiente Dice e a função de perda adotada neste trabalho, destacando seus papéis na avaliação e no treinamento dos modelos de classificação e segmentação de imagens dermatoscópicas.

### 2.6.1 Métrica Dice

A métrica Dice é utilizada para avaliar o desempenho de modelos de segmentação de imagens, sobretudo em aplicações médicas, onde a precisão na delimitação da região lesionada é essencial (Júnior et al., 2020). Seu objetivo é quantificar o grau de sobreposição entre a segmentação produzida pelo modelo e a segmentação de referência (GT ou verdade fundamental) fornecida por especialistas (Ronneberger; Fischer; Brox, 2015; Araújo; Junior, 2017; Amaral, 2023).

O índice Dice é definido em notação de conjuntos, onde  $A$  e  $B$  são os conjuntos (amostras) a serem comparados, conforme expresso na Equação 2 (Araújo; Junior, 2017):

$$Dice(A,B) = \frac{2|A \cap B|}{|A| + |B|}. \quad (2)$$

O valor resultante varia entre 0 (totalmente diferente) e 1 (totalmente igual), sendo que valores próximos de 1 indicam forte concordância entre o modelo e o padrão-ouro (Araújo; Junior, 2017).

A relevância do Dice no contexto biomédico está associada à necessidade de medir a fidelidade estrutural da área segmentada (Júnior; Bezerra; Andrade, 2020). Uma segmentação útil deve manter o formato, o contorno e a extensão da lesão, propriedades capturadas por essa métrica, sendo o Dice considerado um dos indicadores mais adequados para avaliar algoritmos de segmentação (Júnior; Bezerra; Andrade, 2020; Amaral, 2023).

Em cenários com classes desbalanceadas, como as lesões cutâneas, em que a área da lesão (*foreground*) é muito menor que o restante da imagem, o Dice é preferido sobre métricas baseadas apenas em contagem de acertos, como a acurácia (Júnior; Bezerra; Andrade, 2020), que representa a proporção total de classificações corretas (Silva; Peres; Boscaroli, 2016). Em *datasets* onde as manchas são menores em proporção ao fundo, o modelo é mais assertivo na classificação do fundo (verdadeiros negativos), o que eleva a AC. O Dice, ao se concentrar na sobreposição, que mede a similaridade entre duas amostras, oferece uma medida mais coerente da qualidade da segmentação (Júnior; Bezerra; Andrade, 2020).

### 2.6.2 Métricas Complementares

Além do coeficiente Dice e da função de perda, a avaliação quantitativa de modelos de diagnóstico auxiliado por computador baseia-se na análise da matriz de confusão. Esta matriz relaciona as predições do modelo com GT, classificando os resultados em: Verdadeiro Positivo (TP), *pixels* da lesão corretamente identificados; Verdadeiro Negativo (TN), *pixels* do fundo corretamente identificados; Falso Positivo (FP), *pixels* do fundo incorretamente classificados como lesão; e Falso Negativo (FN), *pixels* da lesão incorretamente classificados como fundo (Júnior; Bezerra; Andrade, 2020). A partir desses valores, derivam-se métricas fundamentais para a validação clínica:

- Acurácia (AC): representa a proporção total de classificações corretas realizadas pelo modelo em relação ao total de amostras. Embora forneça uma visão geral, deve ser analisada com cautela em bases desbalanceadas, onde a classe majoritária (fundo) pode inflar o resultado (Júnior; Bezerra; Andrade, 2020). É calculada por meio da Equação 3:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

- Precisão (P): indica a proporção de pixels classificados como positivos (lesão) que realmente pertencem à classe de interesse. É uma métrica que avalia a confiabilidade das detecções positivas do modelo, como demonstrado pela Equação 4 (Júnior; Bezerra;

Andrade, 2020):

$$P = \frac{TP}{TP + FP}. \quad (4)$$

- Sensibilidade (SE): também conhecida como revocação, descreve a capacidade do modelo de prever corretamente as amostras que realmente contêm a doença, definida a partir da Equação 5. No contexto das neoplasias cutâneas, uma alta sensibilidade é crítica para evitar falsos negativos, garantindo que lesões malignas severas (como o melanoma e os carcinomas) não sejam ignoradas (Júnior; Bezerra; Andrade, 2020):

$$SE = \frac{TP}{TP + FN}. \quad (5)$$

- Especificidade (SP): descreve a capacidade de prever corretamente as amostras negativas para a classe em análise (como a pele saudável delimitada no fundo da imagem durante a segmentação, ou a correta exclusão de uma patologia durante a classificação). Em conjunto com a SE, permite avaliar o equilíbrio do modelo entre detectar a doença e não gerar alarmes falsos, definida pela Equação 6 (Júnior; Bezerra; Andrade, 2020):

$$SP = \frac{TN}{TN + FP}. \quad (6)$$

- F1-score: representa a média harmônica da P e da SE, é uma métrica fundamental para medir o desempenho da classificação correta, sendo crucial porque a eficácia dos sistemas não pode ser avaliada apenas com base na P ou na SE, em particular em cenários desbalanceados (Gouda *et al.*, 2022). O F1-score é calculado pela Equação 7:

$$F1 - score = 2 \times \left( \frac{P \times SE}{P + SE} \right). \quad (7)$$

- Índice de Jaccard (IoU): semelhante ao coeficiente Dice, o índice Jaccard, ou Intersecção sobre União (IoU), mensura a similaridade entre a máscara predita (A) e a máscara real (B). É uma métrica rigorosa para segmentação, penalizando erros de não-sobreposição mais severamente que o Dice, definido pela Equação 8 (Júnior; Bezerra; Andrade, 2020):

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}. \quad (8)$$

- Área Sob a Curva ROC (AUC): a curva ROC (*Receiver Operating Characteristic*) relaciona a taxa de verdadeiros positivos (Sensibilidade) com a taxa de falsos positivos (1 - Especificidade) em diferentes limiares de decisão. A AUC resume essa curva em um único valor escalar, indicando a capacidade global do modelo em distinguir entre as classes (tanto na segmentação da lesão quanto na separação das patologias melanoma, CBC e CEC), independentemente do limiar de classificação escolhido.

As métricas de desempenho fundamentais para a validação do classificador, AC, P, SE, SP e IoU, operam dentro de um intervalo contínuo fechado variando de 0 a 1 (ou, percentualmente, de 0% a 100%). Neste espectro, o valor 1 representa o desempenho ideal, indicando a concordância absoluta entre a predição do modelo e o GT, enquanto o valor 0 denota a total incapacidade do sistema em realizar a classificação correta.

No entanto, a interpretação clínica desses valores exige contextualização, embora uma AC próxima de 1 sugira excelência global, em bases desbalanceadas esse número pode ser enganoso se o modelo apenas predizer a classe majoritária. Portanto, a análise crítica deve focar no equilíbrio entre SE e SP. Um valor de SE próximo a 1 é o requisito prioritário em oncologia, pois significa que o sistema é capaz de detectar a quase totalidade das neoplasias malignas, minimizando o risco fatal dos falsos negativos. Em contrapartida, valores elevados de P e SP (próximos a 1) indicam que o modelo é robusto contra falsos alarmes, evitando que pacientes saudáveis sejam submetidos a procedimentos invasivos desnecessários, estabelecendo assim um *trade-off* que deve ser ajustado conforme o rigor clínico desejado.

### 2.6.3 Função de Perda

A função de perda é o componente que orienta o processo de aprendizado de uma rede neural. É ela que indica o quanto as previsões do modelo se afastam dos valores corretos e, a partir disso, direciona os ajustes necessários nos pesos durante o treinamento. Escolher a função de perda adequada é essencial, pois diferentes tarefas exigem formas diferentes de medir esse erro (Júnior; Bezerra; Andrade, 2020).

No caso da classificação multiclasse, como distinguir entre melanoma, CBC e CEC, a *Categorical Cross-Entropy* (CCE) constitui uma função de perda comumente empregada. Ela compara a probabilidade prevista pelo modelo com o rótulo real e penaliza erros de forma proporcional à confiança da rede (Júnior; Bezerra; Andrade, 2020). A CCE costuma apresentar bom desempenho em problemas médicos, sobretudo quando combinada com estratégias de transferência de aprendizado e técnicas para lidar com desbalanceamento entre classes.

Já para a segmentação, a escolha da função de perda precisa levar em conta o fato de que a área da lesão, em geral, ocupa uma porção pequena da imagem. Em situações assim, funções tradicionais como a *Cross-Entropy* podem favorecer o fundo da imagem, prejudicando a detecção da ROI. Para contornar esse problema, perdas como a *Dice Loss* ou a *Focal Loss* são bastante usadas na área da saúde.

A *Dice Loss* se baseia na Métrica Dice e busca maximizar a sobreposição entre a máscara prevista e a máscara real, funcionando muito bem quando a lesão é pequena ou tem contornos pouco definidos. A *Focal Loss*, por sua vez, dá mais peso aos *pixels* mais difíceis de classificar, ajudando a rede a lidar com casos mais desafiadores (Amaral, 2023).

Em muitos estudos, combinações dessas perdas são aplicadas, como CCE + *Dice Loss*, para equilibrar o aprendizado de detalhes locais (*pixel a pixel*) com o aprendizado da estrutura

global da lesão. Esse tipo de abordagem resulta em segmentações mais precisas e confiáveis em contextos clínicos onde a preservação dos contornos e da forma da lesão é fundamental.

### 3 REVISÃO BIBLIOGRÁFICA

A literatura sobre o CAD de lesões cutâneas evoluiu drasticamente das abordagens clássicas de PDI, que dependiam da extração manual de características baseadas em regras clínicas como o ABCDE, para o aprendizado profundo. No cenário atual, o uso de RNCs se consolidou como o padrão ouro, superando a subjetividade inerente à interpretação humana.

#### 3.1 Dermatoscopia Digital e Evolução Tecnológica

A transição da dermatoscopia convencional (microscopia de epiluminescência) para a dermatoscopia digital representa um marco tecnológico decisivo para o diagnóstico precoce do melanoma, CBC e CEC. Esse avanço converteu um procedimento em sua maior parte manual em um processo padronizado, quantificável e compatível com análises automatizadas baseadas em PDI e IA. Os primeiros métodos de digitalização, que consistiam na captura das lesões com câmeras específicas (dermafoto) e posterior digitalização das fotografias reveladas, mostraram-se limitados devido à perda de resolução e de detalhes clínicos essenciais (Coelho, 2016).

A infraestrutura tecnológica da dermatoscopia digital é heterogênea e tem evoluído para garantir maior portabilidade, precisão e acessibilidade. Entre os dispositivos utilizados, destacam-se o fotodermatoscópio, composto por uma câmera digital acoplada a um dermatoscópio, e os sistemas de videodermatoscopia digital, nos quais o equipamento é conectado a um computador que possibilita a digitalização, o armazenamento e a manipulação das imagens por softwares especializados. Esses sistemas facilitam o acompanhamento longitudinal das lesões, permitindo comparar morfologia, coloração e evolução temporal (Coelho, 2016).

A teledermatologia constitui uma modalidade de atendimento na qual imagens dermatológicas e informações clínicas são transmitidas para avaliação à distância. A integração dessa prática com dispositivos móveis e adaptadores ópticos permite registrar e encaminhar imagens de lesões cutâneas para análise por especialistas, ampliando o alcance da triagem e do acompanhamento clínico em regiões com acesso limitado a dermatologistas. Quando associadas a sistemas de apoio à decisão, essas soluções também podem empregar algoritmos de IA para auxiliar na priorização dos casos que exigem avaliação especializada.

Soluções baseadas em adaptadores ópticos acoplados a smartphones, como o dermatoscópio para iPhone citado por (Coelho, 2016), democratizam o acesso ao pré-diagnóstico, especialmente em regiões remotas. Tais dispositivos permitem que algoritmos embarcados de IA realizem triagens iniciais, possibilitando encaminhamento rápido para avaliação especializada. Exemplos dessa integração entre *hardware* móvel e IA incluem o *software* SKINVIEWER, desenvolvido por (Coelho, 2016), e aplicativos mais recentes, como o Serpens, que incorpora arquiteturas de DL baseadas em ViT (Amaral, 2023). Também se destacam sistemas multiplataforma que utilizam redes neurais como a InceptionV3 integrada via *Application Programming Interface* (API) em aplicativos desenvolvidos em Flutter (Hermes; Moreira; Lima, 2024).

A principal contribuição desses sistemas digitais reside na elevação da acurácia diagnóstica. Estudos indicam que a dermatoscopia digital associada ao PDI pode aumentar a sensibilidade diagnóstica em 10 a 30% quando comparada apenas ao exame clínico visual, como retratado em Blum, 2004 citado por Coelho, 2016). Sistemas de apoio ao diagnóstico baseados em IA têm demonstrado resultados superiores aos obtidos por dermatologistas menos experientes. Por exemplo, Ruíz et al. (2011), analisado por Alencar (2015), reportaram especificidade de 97,87% e sensibilidade de 78,43% utilizando classificadores como K-NN, Bayesiano e MLP combinados. Barata et al. (2013) alcançaram sensibilidade de 96% e especificidade de 80% com o uso de AdaBoost. Métodos baseados em extração automática de parâmetros geométricos, como a razão perímetro/área, também se mostraram eficazes: Fikrle (2007), citado por Coelho (2016), obteve 84,7% de sucesso na detecção de melanomas utilizando essa abordagem morfológica.

Com o advento do DL, os sistemas de segmentação e classificação avançaram significativamente. A segmentação semântica, etapa crucial no fluxo CAD, obteve aumentos expressivos de desempenho com arquiteturas como a U-Net, empregada na segmentação de lesões dermatoscópicas devido à sua capacidade de preservar informações espaciais finas. Em contraste, sistemas tradicionais baseados na regra ABCDE, como o SKINVIEWER (versão 1), alcançaram apenas 63% de acurácia (Coelho, 2016), inferior aos valores entre 70% e 90% reportados em estudos contemporâneos.

Apesar dos avanços tecnológicos, persistem desafios relevantes. A ausência de padronização na aquisição das imagens, incluindo variações de iluminação, angulação, distância e foco, compromete a consistência dos modelos de IA e pode induzir erros classificatórios. Coelho (2016) demonstrou que o uso de um dermatoscópio de baixo custo acoplado a um iPhone 5s (modelo NO. 9882-IP5) resultou em imagens distorcidas e de baixa qualidade, tornando o dispositivo inadequado para fins clínicos. Essa limitação evidencia a necessidade de *hardware* de captura de alta precisão, mesmo em abordagens móveis. Além disso, os altos custos de *softwares* comerciais de análise dermatoscópica dificultam sua adoção por clínicas e hospitais, sobretudo em contextos de menor investimento. Soma-se a isso o fato de que muitos sistemas proprietários apresentam ecossistemas fechados, o que implica custos elevados de manutenção, atualizações e suporte técnico.

### **3.2 Bases de Dados Dermatoscópicas Públicas**

O desenvolvimento e a validação de algoritmos de IA e DL aplicados à dermatoscopia digital dependem da disponibilidade de grandes coleções de imagens de alta qualidade. A eficácia preditiva de um modelo está relacionada à robustez e à diversidade da base de dados empregada em seu treinamento, validação e teste. Nesse contexto, o uso de bases públicas tornou-se essencial para o avanço das pesquisas em sistemas de apoio ao diagnóstico nas tarefas de segmentação e classificação de lesões cutâneas.

Diversos repositórios dermatoscópicos ganharam destaque na literatura por oferecerem imagens padronizadas, acompanhadas de anotações por especialistas (GT), possibilitando avaliações comparáveis e replicáveis entre diferentes estudos. Entre esses repositórios, destacam-se iniciativas internacionais que integram centros hospitalares, universidades e laboratórios de pesquisa.

A principal iniciativa global nesse domínio é o ISIC, responsável pelo *ISIC Archive*, considerado o maior repositório público de imagens dermatoscópicas disponível. O projeto representa uma colaboração entre instituições acadêmicas e industriais com o objetivo de reduzir a mortalidade por melanoma ao fomentar o desenvolvimento de sistemas computacionais de detecção precoce (Ferreira, 2018). Desde 2016, o ISIC promove anualmente o *ISIC Challenge*, competição internacional que estimula o desenvolvimento de modelos para segmentação de lesões, detecção de atributos dermatoscópicos e classificação de doenças cutâneas.

O ISIC 2017, uma das bases de dados mais utilizadas para segmentação, serviu como base para diversos estudos comparativos de arquiteturas de DL. Júnior, Bezerra e Andrade (2020) avaliaram arquiteturas U-Net com diferentes codificadores (DenseNet-121, ResNet-50 e VGG-19), sendo o modelo U-Net + ResNet-50 o mais eficaz, com IoU médio de 81,94%. Além disso, estudos como o de Jahanifar et al. (2018) citado por Júnior et al. (2020), demonstraram que abordagens baseadas em *ensemble* de múltiplas redes alcançam desempenho superior, obtendo até 80,6% de IoU no mesmo desafio.

O ISIC 2018 ampliou o escopo e o volume das imagens, tornando-se outra base de dados de referência. Cavalcante (2019), utilizando 2.594 imagens para treinamento e validação, observou que o modelo U-Net com codificador DenseNet-121 obteve IoU de 0,772 no conjunto de teste. Entre os melhores desempenhos da competição, observou-se a predominância de estratégias de *ensemble* aliadas ao TL, combinando redes como DeepLab, PSPNet, U-Net e VGG.

Outro repositório utilizado é o PH<sup>2</sup>, desenvolvido pelo Hospital Pedro Hispano (Portugal). Ele contém 200 imagens de lesões melanocíticas, 80 nevos comuns, 80 nevos atípicos e 40 melanomas, acompanhadas de segmentações, diagnóstico clínico e histopatológico e critérios dermatoscópicos estruturados (rede pigmentar, áreas de regressão, véu azul-esbranquiçado, entre outros). Devido à sua alta qualidade e padronização, o PH<sup>2</sup> é empregado como conjunto de validação externa (*out-of-sample testing*).

Outra base de dados de grande relevância é o HAM10000 (*Human Against Machine with 10,000 Training Images*), atualmente composto por mais de 11.000 imagens dermatoscópicas. Ele abrange sete categorias diagnósticas de grande importância clínica: queratoses actínicas e carcinoma intraepitelial, carcinoma basocelular, lesões benignas semelhantes à queratose, dermatofibroma, melanoma, nevos melanocíticos e lesões vasculares. Essa base de dados é considerado essencial para treinar classificadores multiclasse mais robustos. Estudos como o de Raza et al. (2025) demonstraram que arquiteturas otimizadas de redes neurais convolucionais podem superar até mesmo modelos profundos como a ResNet152V2, chegando a acurácia de 98,12% em validação.

Embora menos abrangentes, outras bases também desempenham papel relevante na literatura. O trabalho de Horiuchi (2022), por exemplo, apresenta uma base com imagens clínicas (fotografias convencionais) de lesões hipercrômicas, totalizando 51 melanomas e 800 controles não melanoma, todos com confirmação anatomopatológica. Apesar de não se tratar de imagens dermatoscópicas, a base de dados é valiosa por incluir informações prognósticas e histológicas detalhadas, como índice de Breslow e presença de ulceração, características raras nas bases públicas, que não dispõem de metadados clínicos completos.

A introdução da base de dados ISIC 2019 (associado ao repositório BCN20000) representou um marco essencial para o treinamento de classificadores multiclasse. Esse repositório ampliou o volume de imagens complexas, introduzindo e balanceando categorias críticas, como o Carcinoma Espinocelular (CEC) e o Carcinoma Basocelular (CBC), além do melanoma. Devido a essa complexidade morfológica e ao severo desbalanceamento intrínseco aos dados do mundo real, essa base tornou-se o padrão ouro atual para avaliar a robustez de modelos baseados em IA.

### 3.3 Critério Clínico: Regra ABCDE

A evolução da dermatoscopia para sistemas CAD teve como ponto de partida a sistematização dos critérios clínicos observados pelos dermatologistas. Regras como o ABCDE, o Método de Menzies e a Lista dos Sete Pontos estabeleceram a base conceitual que os algoritmos clássicos de VC e PDI buscaram quantificar. Como a avaliação dermatoscópica depende da experiência do especialista, introduzindo subjetividade e variabilidade interobservador, o uso de técnicas computacionais surgiu como alternativa para padronizar a análise e reduzir vieses diagnósticos. Nesse contexto, a extração de características consiste em traduzir os critérios clínicos em atributos numéricos, utilizados por modelos clássicos de ML.

A assimetria refere-se à ausência de correspondência entre as metades da lesão quando dividida ao longo de um ou mais eixos. Computacionalmente, ela é estimada a partir de medições geométricas que quantificam quanto a forma da lesão se desvia da simetria ideal. Coelho (2016) descreve que a assimetria pode ser calculada pela diferença percentual entre a quantidade de *pixels* nas direções horizontal (A1) e vertical (A2). Quanto maior a diferença, maior a pontuação atribuída, sendo melanomas em geral mais assimétricos que nevos benignos. O trabalho de Fikrle e Pizinger (2007), citado por Coelho (2016), analisou 260 imagens utilizando parâmetros morfológicos, incluindo a razão perímetro/área (*taxa*) e o termo *Assym1*, que representa a assimetria no eixo inercial *A1*. A probabilidade de malignidade (*P*) proposta pelo modelo é dada pela Equação 9:

$$P = \frac{1}{1 + e^{-4,738 + 7,474 \cdot taxa - 0,190 \cdot Assym1}} \quad (9)$$

Siqueira e Fernandes (2015) propuseram um método específico para mensurar a assimetria, baseado no cálculo do centroide *pc* da lesão segmentada e na comparação da corres-

pondência de *pixels* após sua divisão. A assimetria é expressa como a porcentagem de *pixels* não coincidentes entre as metades. Avaliando bases clínicas como *Atlas of Clinical Dermatology* e DermNet NZ, os autores observaram que essa característica isolada identificou corretamente 69,23% dos melanomas avaliados.

Bordas irregulares e transições abruptas são indicadores importantes de malignidade. A análise da borda depende da etapa de segmentação, considerada a etapa crítica dos sistemas CAD, pois erros nessa fase propagam-se para a extração de características. O algoritmo proposto por Siqueira e Fernandes (2015) calcula o conjunto de distâncias  $d_i$  entre o centroide  $p_c(l_c, c_c)$  da lesão e cada ponto  $p_i(l_i, c_i)$  que forma o contorno do sinal. Essa distância, Equação 10, é obtida com base no teorema de Pitágoras:

$$d_i = \sqrt{(l_c - l_i)^2 + (c_c - c_i)^2}. \quad (10)$$

A métrica utilizada para quantificar a irregularidade da borda é o desvio padrão ( $\sigma_d$ ) das distâncias. Lesões benignas apresentam contornos mais uniformes, resultando em baixo desvio padrão, enquanto melanomas tendem a apresentar bordas assimétricas e abruptas, elevando o valor da métrica. No estudo de Siqueira e Fernandes (2015), o desvio padrão da borda alcançou 89,23% de acerto na identificação de melanomas, utilizando um limiar de 7,6832, sendo a característica de maior poder discriminativo do estudo.

A presença de múltiplas cores é outro critério visual relevante na dermatoscopia. Sob o ponto de vista clínico, as cores mais associadas ao melanoma incluem preto, branco, azul-acinzentado, vermelho e diferentes tonalidades de marrom. Quanto maior a heterogeneidade cromática, maior a suspeição. O terceiro algoritmo proposto por Siqueira e Fernandes (2015) quantifica a variância das intensidades do histograma da lesão segmentada. Para reduzir o custo computacional, utiliza-se a imagem em escala de cinza, preservando a distribuição tonal e mantendo a sensibilidade à heterogeneidade. Lesões benignas em geral apresentam baixa variância, enquanto melanomas exibem distribuição tonal mais dispersa. Esse método apresentou um índice de identificação de melanoma de 64,62% (limiar de 0,009204), representando o critério com menor poder discriminativo entre os três analisados, mas ainda relevante quando combinado com outras características.

O critério do diâmetro estabelece que lesões maiores que 6 mm devem ser avaliadas com maior cautela. No entanto, sua mensuração em sistemas CAD baseados apenas em imagens 2D enfrenta limitações, devido à ausência de referência física que permita converter *pixels* em milímetros. Por esse motivo, Siqueira e Fernandes (2015) desconsideraram essa característica, assim como a elevação, em seus experimentos. Em sistemas clínicos completos, o diâmetro pode ser estimado quando há calibração do equipamento de captura ou presença de marcadores físicos na imagem, possibilitando analisar outras características morfogeométricas relevantes.

Na prática, a dimensão da lesão, da qual o diâmetro seria um estimador direto, é avaliada por meio de atributos computacionais como área, perímetro e circularidade, que não dependem

de unidades absolutas para serem comparados entre diferentes sinais. A área corresponde ao total de *pixels* pertencentes à região segmentada da lesão, enquanto o perímetro representa a soma dos *pixels* que compõem seu contorno. Esses descritores básicos permitem análises geométricas mais robustas, como a circularidade, que expressa o grau de semelhança entre a forma da lesão e um círculo de mesma área, uma métrica que captura deformações características de lesões malignas.

O software SKINVIEWER (Versão 3), desenvolvido por Coelho (2016), exemplifica a aplicação estruturada desses procedimentos em um ambiente computacional. A ferramenta foi projetada para extrair, de maneira automática, atributos como diâmetro estimado, área, perímetro e circularidade, que alimentam módulos de classificação. Esses atributos podem ser utilizados tanto em comparação direta com um banco de dados de lesões rotuladas quanto em modelos que atribuem escores clínicos ponderados a cada característica. Assim, medidas geométricas tornam-se elementos fundamentais para a construção de sistemas CAD capazes de reproduzir, de forma objetiva, critérios que na dermatologia tradicional dependiam da inspeção visual especializada.

A evolução é um dos critérios mais críticos da regra ABCDE, pois monitora mudanças no tamanho, forma, cor ou sintomatologia da lesão ao longo do tempo. No entanto, sua implementação em sistemas CAD é complexa, pois exige o acompanhamento longitudinal de imagens capturadas em diferentes consultas. Amaral (2023) propôs uma ferramenta denominada Serpens, desenvolvida para auxiliar dermatoscopistas na análise do critério E. O sistema utiliza um modelo ViT integrado à arquitetura UNETR (do inglês *UNET Transformers*) para realizar a segmentação semântica das imagens. O Serpens organiza o histórico de lesões do paciente e permite comparar máscaras segmentadas ao longo das consultas, com intervalos de três meses. Em um caso clínico real, o sistema evidenciou crescimento significativo de um nevo melanocítico dentro desse intervalo, facilitando a tomada de decisão do médico e demonstrando o potencial da evolução temporal como característica computacional.

### 3.4 Revisão de Arquiteturas

O DL representa o avanço tecnológico que permitiu a transição dos sistemas CAD da extração manual de características geométricas (baseadas na engenharia de características) para o aprendizado automático de representações hierárquicas. As RNCs e os modelos baseados em *transformers* são as tecnologias que dominam o estado da arte na análise de imagens dermatoscópicas, sendo aplicadas primariamente em duas tarefas interdependentes: Segmentação Semântica e Classificação.

#### 3.4.1 Segmentação Semântica

A segmentação semântica constitui a etapa mais crítica do fluxo de um sistema CAD, pois determina quais *pixels* pertencem à ROI da lesão. Uma segmentação imprecisa compro-

mete todo o processamento subsequente em um cenário marcado por forte variabilidade de textura, cor, iluminação e pela presença de artefatos, como pelos, reflexos e sombras. Nesse contexto, a arquitetura U-Net consolidou-se como a principal referência para segmentação de imagens biomédicas, devido à sua estrutura *encoder-decoder* com conexões de atalho que preservam informações espaciais essenciais.

O trabalho desenvolvido por Ferreira (2018) na Universidade Federal do Maranhão (UFMA) teve como foco a segmentação automática de lesões cutâneas utilizando Redes Neurais Convolucionais, com ênfase na arquitetura U-Net. O objetivo central foi propor uma metodologia de segmentação semântica capaz de isolar, *pixel a pixel*, a região da lesão, servindo como etapa prévia fundamental para sistemas de diagnóstico assistido por computador.

A arquitetura adotada baseou-se na U-Net clássica proposta por Ronneberger, Fischer e Brox (2015), diferenciando-se apenas pelo ajuste na quantidade de filtros em cada camada, mantendo, portanto, a estrutura de contração e expansão original. O treinamento foi conduzido utilizando duas bases públicas validadas: PH<sup>2</sup> e um conjunto de imagens derivado do *ISIC-Archive*, garantindo maior robustez e confiabilidade dos resultados.

Como estratégia de otimização, o método empregou o otimizador Adam, cujo nome deriva da expressão *Adaptive Moment Estimation*, com taxa de aprendizado inicial de 0,001. A Entropia Cruzada foi escolhida como função de perda, adequada ao problema de segmentação binária e utilizada em cenários biomédicos. Para reduzir o custo computacional e acelerar o processo de treinamento, as imagens foram redimensionadas, decisão que, segundo o autor, não comprometeu a capacidade de generalização do modelo.

O desempenho foi avaliado por métricas clássicas de sobreposição, consideradas essenciais em tarefas de segmentação, tais como Sensibilidade (SE), Precisão (P), Coeficiente de Dice e Índice de Interseção sobre União (IoU). Essas métricas permitem mensurar o grau de concordância entre a segmentação predita pelo modelo e a segmentação de referência (*ground truth*), sendo adotadas na literatura biomédica.

A Tabela 1 sintetiza os resultados obtidos nas bases de dados avaliadas por Ferreira (2018), permitindo uma análise comparativa do desempenho do modelo em diferentes cenários experimentais.

**Tabela 1 – Comparação das métricas de desempenho entre as bases de Ferreira (2018)**

Métrica	Base PH <sup>2</sup>	Base ISIC
SE	98,02%	97,54%
P	98,22%	97,83%
Dice	97,66%	97,58%
IoU	95,07%	95,13%

**Fonte: Adaptada de Ferreira (2018).**

A partir dos resultados apresentados na Tabela 1, observa-se que o Coeficiente de Dice e o IoU apresentaram valores elevados e consistentes entre as bases analisadas, indicando boa concordância entre as máscaras preditas e as máscaras de referência. Entretanto, a análise isolada dessas métricas não é suficiente para inferir impacto clínico direto.

No contexto médico, a sensibilidade (SE) assume papel prioritário, pois está relacionada à capacidade do modelo em identificar corretamente as regiões lesionadas. Uma sensibilidade reduzida implica maior ocorrência de falsos negativos, o que pode comprometer etapas subsequentes do diagnóstico assistido por computador. Assim, embora Dice e IoU sejam métricas robustas para avaliação técnica da segmentação, a sensibilidade tende a ser a métrica de maior relevância quando se considera a aplicação clínica do sistema.

Ferreira concluiu que a segmentação constitui uma etapa crítica para o desempenho das fases subsequentes do CAD, como a extração de características e a classificação. Como recomendações para trabalhos futuros, o autor destacou a importância de incorporar técnicas adicionais de pré-processamento, como remoção de pelos e melhor exploração de canais de cor, além da adoção de validação cruzada e de métodos sistemáticos para otimização de hiperparâmetros, visando aprimorar ainda mais a capacidade de generalização dos modelos.

O trabalho de monografia desenvolvido por Cavalcante (2019) na Universidade Federal do Espírito Santo (UFES) investigou soluções para o problema de segmentação semântica de lesões cutâneas utilizando RNCs. O autor adotou a arquitetura U-Net como base, dada sua capacidade comprovada de produzir segmentações robustas mesmo quando o conjunto de treinamento é limitado, característica particularmente relevante em aplicações biomédicas. A segmentação visa rotular cada *pixel* da imagem como pertencente à classe lesão ou fundo, preservando informações espaciais e estruturais da região analisada.

O estudo avaliou uma U-Net modificada, integrando diferentes redes classificadoras (*backbones*) no *encoder*. As arquiteturas investigadas foram: VGG-16, ResNet-18 e DenseNet-121. Todas foram inicializadas com pesos pré-treinados no *ImageNet Large Scale Visual Recognition Challenge 2012* (ILSVRC 2012), aplicando princípios de TL para acelerar o aprendizado e melhorar a generalização do modelo. O experimento utilizou o conjunto de dados do ISIC 2018, amplamente empregado em pesquisas de segmentação dermatológica.

Para reduzir *overfitting* e aumentar a variabilidade das amostras, o autor aplicou técnicas de DA, incluindo rotações, deslocamentos e zoom. O desempenho das arquiteturas foi medido por meio do IoU Limiarizado, uma métrica baseada na sobreposição entre a máscara predita e a máscara real, com a aplicação de um limiar para avaliar apenas segmentações acima de determinado nível de confiança.

Os resultados obtidos evidenciaram diferenças claras entre os codificadores. A U-Net com DenseNet-121 apresentou o melhor desempenho, alcançando IoU limiarizado de 0,772. Em seguida, a U-Net com ResNet-18 registrou valor de 0,741, enquanto a U-Net com VGG-16 apresentou o pior resultado, com 0,694. Segundo a análise do autor, o desempenho inferior da VGG-16 pode estar associado à sua menor profundidade e menor capacidade de extração de características discriminativas quando comparada às arquiteturas mais modernas, como ResNet e DenseNet. O estudo reforça a importância da escolha do codificador no desempenho final da U-Net, destacando que arquiteturas mais profundas e com conexões avançadas, como o encadeamento denso da DenseNet, favorecem a extração de representações mais ricas e adequadas à segmentação de lesões cutâneas.

O estudo de Júnior, Bezerra e Andrade (2020) avaliou o impacto do uso de diferentes codificadores pré-treinados no *encoder* da U-Net para a segmentação de lesões nas bases ISIC 2017 e PH<sup>2</sup>, explorando um protocolo experimental rigoroso. Na ISIC 2017, o conjunto foi dividido em 70% para treinamento (1400 imagens), 25% para teste (500 imagens) e 5% para validação (100 imagens). A base PH<sup>2</sup> foi utilizada como conjunto externo, permitindo avaliar a capacidade de generalização do modelo. Os codificadores comparados, DenseNet-121, ResNet-50 e VGG-19, foram inicializados com pesos pré-treinados no ImageNet (ILSVRC 2012), seguindo a prática de TL para acelerar a convergência e melhorar o desempenho.

No estudo de Júnior, Bezerra e Andrade (2020), a U-Net foi implementada com função de ativação ReLU e operações de *Max Pooling* no processo de redução espacial. Os autores aplicaram técnicas de DA a todas as amostras de treino, incluindo rotações de 0° a 360°, variações aleatórias de zoom, adição de ruído Gaussiano, alterações de brilho e contraste, ajustes de nitidez, filtros de desfoque, transformações de perspectiva e mudanças de saturação. Esses procedimentos foram essenciais para reduzir *overfitting*, dada a limitada quantidade de imagens dermatológicas rotuladas. Os hiperparâmetros (número de épocas, *batch size* e taxa de aprendizado) foram ajustados de forma empírica, seguindo critérios de estabilidade e desempenho no conjunto de validação.

A função de perda ( $L$ ) adotada foi uma combinação híbrida, definida pela Equação 11:

$$L = (1 - Dice) + FL. \quad (11)$$

Esta formulação associa a perda de Dice ( $1 - Dice$ ), adequada para avaliar a sobreposição em segmentação, com a perda focal ( $FL$ , do inglês *Focal Loss*), que ajusta o peso das amostras conforme o grau de dificuldade de classificação. Essa combinação mostrou-se particularmente eficiente para lidar com o desbalanceamento de classes, uma característica típica em imagens de pele onde a área da lesão ocupa proporções reduzidas em relação ao fundo.

Entre as configurações avaliadas, a U-Net com ResNet-50 como codificador apresentou o melhor desempenho médio na ISIC 2017 após o aumento de dados, atingindo métricas expressivas: IoU de 81,94%, Dice de 90,07%, AC de 96,42%, P de 95,33%, SP de 99,01% e SE de 85,37%. Os autores destacam ainda que ResNet-50 e DenseNet-121 superaram VGG-19, que obteve o pior desempenho, devido às capacidades superiores de extração de características profundas e à regularização intrínseca das arquiteturas mais modernas.

### 3.4.2 Classificação e *Transfer Learning*

A tarefa de classificação de lesões cutâneas requer a extração automática de características discriminativas capazes de diferenciar lesões benignas de melanomas. Diante da escassez de grandes bases de dados médicas, o TL consolidou-se como abordagem dominante para aprimorar o desempenho de modelos profundos, reduzindo custos de anotação e mitigando problemas de sobreajuste.

Santos *et al.* (2017) apresentaram um dos primeiros trabalhos a empregar TL na classificação de imagens dermatoscópicas. O estudo utilizou as bases PH2 e ISIC, ambas compostas por imagens diagnosticadas por dermatologistas, totalizando 426 amostras (213 melanomas e 213 não-melanomas). Os dados foram divididos em 80% para treino (340 imagens) e 20% para teste (86 imagens).

As arquiteturas AlexNet e VGG-F foram utilizadas como extratoras de características, explorando as ativações das camadas FC6 e FC7 (FC, do inglês *Fully Connected*). Para aumentar a diversidade do conjunto de treino, aplicaram-se técnicas robustas de DA, incluindo rotações, translações, espelhamentos, alterações de cor, brilho, contraste, inclusão de ruído e recorte da ROI, resultando em 5.780 imagens balanceadas. As características extraídas foram classificadas por métodos tradicionais, sendo o K-NN o mais eficaz, atingindo 91,5% de acurácia, com sensibilidade e precisão superiores a 0,9. A melhor configuração foi obtida com a rede VGG-F, utilizando a camada FC7 normalizada. O estudo demonstrou que redes pré-treinadas podem fornecer representações discriminativas mesmo com bases pequenas.

Tan e Le (2020) introduziram um método inovador de escalonamento de arquiteturas profundas, que impactou em aplicações médicas, incluindo dermatoscopia. O trabalho propõe uma abordagem sistemática para otimizar acurácia e custo computacional, utilizando uma busca neural multiobjetivo de acordo com a Equação 12:

$$AC(m) \times [FLOPS(m)/T]^w, \quad (12)$$

onde  $T$  define o orçamento-alvo de operações de ponto flutuante por segundo (FLOPS, do inglês *Floating-Point Operations per Second*), e  $w$  controla o peso entre precisão e eficiência. Esse processo resultou na arquitetura EfficientNet-B0, base da família EfficientNet.

O estudo demonstra que escalar apenas profundidade, largura ou resolução é ineficiente; o escalonamento composto otimiza os três. A família EfficientNet alcançou desempenho superior com menor número de parâmetros. Por exemplo, a EfficientNet-B4 obteve 82,9% Top-1 com 19M parâmetros, enquanto a B7 atingiu 84,3% com 66M, e o GPipe necessitou de 556M para desempenho equivalente.

Gouda *et al.* (2022) desenvolveram um *framework* de DL para detecção de câncer de pele utilizando a base ISIC 2018, avaliando a aplicação de TL e procedimentos de otimização em cenários reais. O estudo empregou modelos pré-treinados, como ResNet-50 e InceptionV3, terceira versão da família de arquiteturas Inception, submetidos a um processo de ajuste fino com o uso das bibliotecas *TensorFlow/Keras* em ambiente *Linux*.

O trabalho lidou explicitamente com o desbalanceamento de classes presente no ISIC (predominância de lesões benignas) aplicando DA e *Dropout*, essenciais para aumentar a robustez do modelo e reduzir bias. O desempenho foi medido por métricas clínicas importantes, incluindo F1-Score e AUC-ROC, reforçando a necessidade de avaliações além da acurácia bruta em bases desbalanceadas.

Zhang *et al.* (2023) realizaram uma revisão abrangente de avanços em ML e DL aplicados ao diagnóstico dermatológico entre 2018 e 2023, analisando estudos com bases como HAM10000 e ISIC 2017. A revisão confirma a predominância de modelos DL modernos em detrimento da extração manual de características. Entre as arquiteturas mais recorrentes destacam-se ResNet-50, InceptionV3, variantes aprimoradas da U-Net para segmentação e modelos híbridos compostos por DenseNet + ResNet. Tecnologias emergentes incluem modelos de ViT equipados com *Soft Label Encoder*.

Resultados notáveis incluem: método baseado em AlexNet pré-treinada, classificação de 7 classes no ISIC 2018, com 98,70% acurácia, 95,60% sensibilidade, 99,27% especificidade e 95,06% precisão; em segmentação, métodos como W-EFO-E-CNN, do inglês *Whale-Electric Fish Optimization-based Ensemble Convolutional Neural Network*, combinados a uma U-Net aprimorada atingiram 98% de acurácia.

A revisão reforça problemas recorrentes das bases médicas, como o desbalanceamento de classes (crítico no HAM10000) e a alta variabilidade intra/interclasse, que podem induzir bias e reduzir a capacidade de generalização. Tais limitações exigem estratégias robustas de TL, regularização e DA.

Mais recentemente, Houssein *et al.* (2024) aplicaram um modelo convolucional profundo focado na extração rigorosa de atributos (*features*), obtendo 97,11% de Acurácia e 97,12% de Sensibilidade. A literatura demonstra que, em contextos oncológicos, a Sensibilidade e o F1-Score são os verdadeiros indicadores de sucesso na distinção crítica entre as lesões. Tais referenciais justificam a necessidade de explorar arquiteturas de alta eficiência, como a Efficient-Net abordada neste trabalho, combinadas a métodos rigorosos de aumento de dados, visando elevar a eficácia diagnóstica sem onerar o custo computacional.

### 3.4.3 Precursores e Desafios no Diagnóstico Móvel

A literatura demonstra a viabilidade técnica de integrar algoritmos de diagnóstico a dispositivos móveis, embora apresente desafios críticos que balizam o desenvolvimento de futuras soluções. Alencar (2015) e Coelho (2016) comprovaram, por meio de redes neurais e algoritmos de extração geométrica (como a fórmula TDS), a possibilidade de classificar lesões em tempo real. Contudo, ambos os estudos identificaram limitações severas na aquisição de imagens via *smartphones* de baixo custo, que introduzem distorções, sombras e ruídos incompatíveis com o rigor exigido para triagens clínicas.

Mais recentemente, Hermes, Moreira e Lima (2024) e Amaral (2023) exploraram a integração de modelos de aprendizado profundo (utilizando arquiteturas como InceptionV3 e UNETR) em ecossistemas móveis via APIs e estruturas de desenvolvimento (*frameworks*) modernos. Embora tais abordagens proporcionem escalabilidade e facilidade de acesso, os resultados indicam que o desempenho dos modelos é altamente dependente da padronização das imagens e da robustez da segmentação semântica. Falhas de generalização manifestadas por

dificuldades em lidar com artefatos de captura e variações na qualidade do *hardware* permanecem como obstáculos significativos para a adoção prática dessas ferramentas.

As experiências supracitadas fundamentam que, para a implementação futura de ferramentas de triagem móvel, é imperativo que o modelo computacional seja validado com alta robustez em ambiente controlado, utilizando bases de dados padronizados e arquiteturas de alta capacidade extrativa, conforme o método desenvolvido neste trabalho. Tais precedentes reforçam a necessidade de priorizar a precisão diagnóstica em cenários clínicos antes de transpor a solução para dispositivos portáteis.

## 4 METODOLOGIA

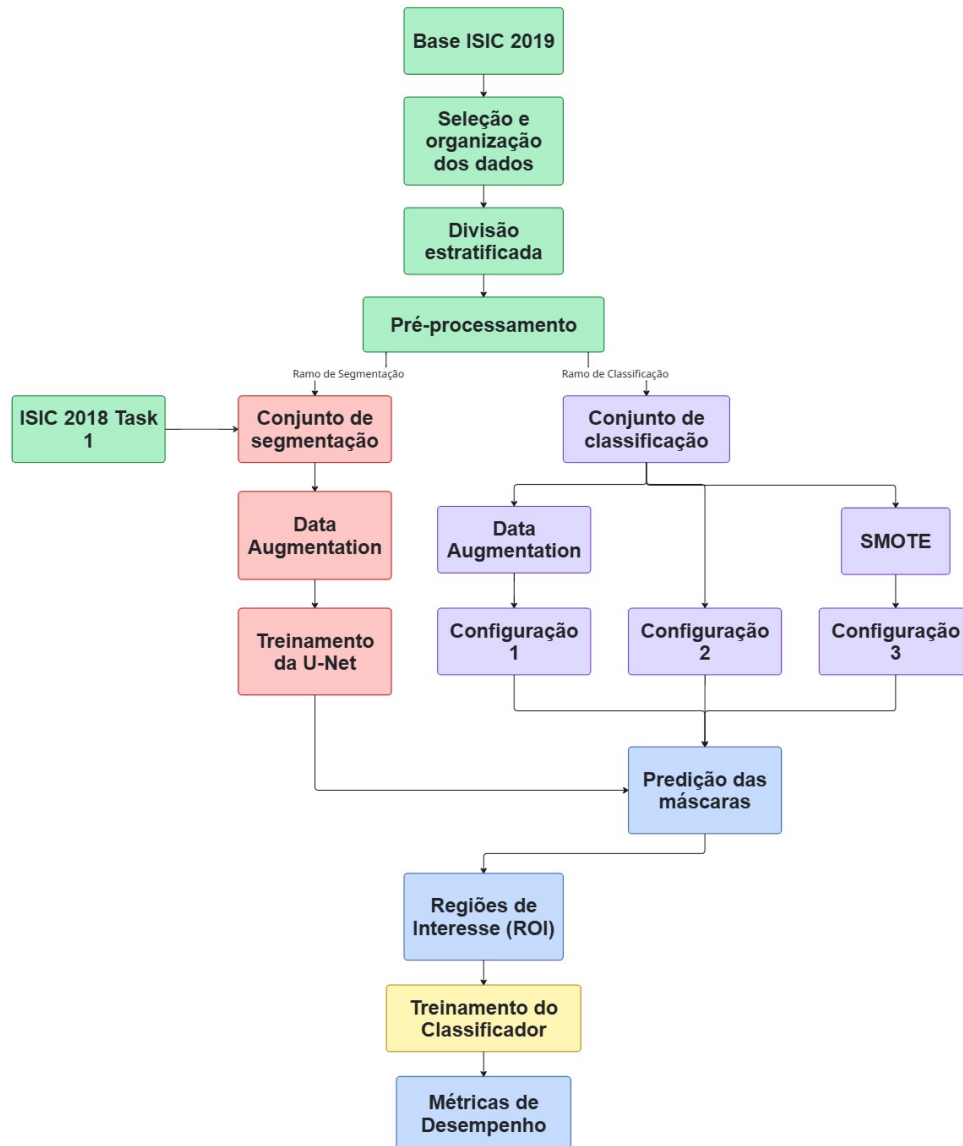
Este capítulo apresenta os procedimentos adotados para o desenvolvimento do sistema de apoio à análise de lesões dermatológicas em imagens dermatoscópicas. A metodologia foi estruturada de forma sequencial, abrangendo a seleção e organização das bases de dados, o pré-processamento das imagens, o treinamento do modelo de segmentação, a geração das regiões de interesse e, por fim, o treinamento e a avaliação do modelo de classificação.

O método proposto foi organizado em dois ramos complementares: segmentação e classificação. O primeiro teve como objetivo delimitar a região da lesão na imagem, separando-a do fundo e de elementos externos que poderiam interferir na análise. O segundo utilizou as regiões de interesse geradas a partir das máscaras preditas para classificar as imagens entre melanoma, CBC e CEC.

Na etapa de classificação, foram avaliadas três configurações experimentais de treinamento. Todas utilizaram a mesma arquitetura principal, a mesma U-Net para geração das regiões de interesse e as mesmas classes de saída. As diferenças entre as configurações concentraram-se na composição dos conjuntos de treinamento, validação e teste, bem como nas estratégias adotadas para lidar com o desbalanceamento entre as classes. Essa comparação permitiu avaliar o impacto da preparação dos dados no desempenho final do classificador.

A Figura 19 apresenta a visão geral do fluxo metodológico adotado. A base ISIC 2019 foi utilizada para a seleção das imagens e dos rótulos diagnósticos empregados na classificação. Em seguida, os dados foram organizados e divididos de forma estratificada em conjuntos de treinamento, validação e teste. Paralelamente, a base ISIC 2018 Task 1 foi utilizada para a etapa de segmentação, por disponibilizar máscaras manuais da região da lesão.

**Figura 19 – Fluxo metodológico proposto para segmentação e classificação das imagens dermatoscópicas.**



**Fonte: Autoria própria (2026).**

A adoção dessa estrutura em duas etapas foi motivada pela necessidade de reduzir a interferência de informações visuais que não pertencem à lesão. Em imagens dermatoscópicas, é comum a presença de pelos, bordas pretas, vinhetas, reflexos, variações de iluminação e regiões de pele saudável ao redor da área analisada. Embora esses elementos façam parte da imagem original, eles não representam, necessariamente, características clínicas da lesão. Assim, a segmentação foi empregada como uma etapa intermediária para destacar a região de maior interesse antes da classificação.

#### 4.1 Ambiente de desenvolvimento e ferramentas computacionais

O sistema foi desenvolvido integralmente na linguagem de programação *Python* (versão 3.10.11), escolhida por sua ampla consolidação no ecossistema de visão computacional e

aprendizado profundo. Para assegurar a viabilidade computacional no treinamento dos modelos, o ambiente foi configurado com suporte a processamento paralelo por Unidade de Processamento Gráfico (GPU, do inglês *Graphics Processing Unit*), utilizando a biblioteca *TensorFlow* (versão 2.10.1) otimizada para compatibilidade com a arquitetura CUDA (do inglês *Compute Unified Device Architecture*) e com a biblioteca *cuDNN* (do inglês *CUDA Deep Neural Network library*). Além disso, a fim de garantir modularidade e reprodutibilidade, todo o fluxo experimental (desde o pré-processamento até a avaliação de métricas) foi estruturado em *scripts* independentes, estratégia fundamental para viabilizar o desenvolvimento paralelo das diferentes etapas do projeto pelos integrantes da equipe.

A infraestrutura de software do projeto fundamentou-se em um ecossistema de bibliotecas consolidadas. A construção, a otimização e o treinamento das arquiteturas de aprendizado profundo (U-Net com codificador ResNet-50 e EfficientNet-B4) foram gerenciados pelo *TensorFlow/Keras* (Abadi *et al.*, 2015; Chollet *et al.*, 2015), tirando proveito de sua eficiência na manipulação de grafos computacionais. As manipulações matriciais de baixo nível e o processamento digital das imagens apoiaram-se no *OpenCV* (Bradski, 2000) e no *NumPy* (Harris *et al.*, 2020), garantindo o manuseio rápido e preciso de tensores multidimensionais. A estruturação, a filtragem e a organização tabular dos metadados clínicos foram realizadas via *Pandas* (McKinney, 2010), viabilizando o pareamento exato entre as imagens e seus respectivos rótulos diagnósticos. Para a etapa de aumento de dados, adotou-se a *Albumentations* (Buslaev *et al.*, 2020), cuja escolha foi essencial para garantir a sincronização espacial e matemática das transformações geométricas entre as imagens originais e suas respectivas máscaras de segmentação. Adicionalmente, funções auxiliares para a visualização gráfica de métricas de desempenho e o monitoramento iterativo do processamento computacional foram supridas pelas bibliotecas *Matplotlib* (Hunter, 2007) e *tqdm*.

Durante as etapas críticas de pré-processamento, a prioridade em preservar a integridade original das imagens médicas exigiu o desenvolvimento de algoritmos sob medida, evitando a dependência de funções genéricas de terceiros. A remoção de bordas escuras, por exemplo, foi estruturada por uma sequência de limiarização, análise matemática e recorte dinâmico operada puramente por matrizes do *OpenCV*. Com esse mesmo rigor, a atenuação de pelos reproduziu a lógica do método *Dull Razor* ao encadear operações morfológicas específicas para o mapeamento desses artefatos, filtros de suavização gaussiana e algoritmos de reconstrução de textura. Essa premissa de controle computacional guiou também a geração da ROI. Para assegurar um reenquadramento exato da lesão antes do redimensionamento final para  $380 \times 380$  pixels, foi codificada uma rotina própria que vinculou o mapa de predição da U-Net à busca direta pelo maior componente conexo da imagem.

Esse mesmo nível de controle pautou o tratamento e a estruturação da base de dados. Na configuração experimental com balanceamento sintético, a aplicação do método SMOTE dispensou o uso de pacotes externos dedicados à superamostragem. O algoritmo foi estruturado de maneira dedicada: o pacote *scikit-learn* (Pedregosa *et al.*, 2011) atuou de forma restrita na busca espacial dos vizinhos mais próximos, enquanto a interpolação algébrica para a síntese

das novas amostras ocorreu inteiramente sobre matrizes do *NumPy*. Além disso, as rotinas de particionamento estratificado dos conjuntos (treino, validação e teste) e a extração de métricas de avaliação também se apoiaram em implementações próprias. Com essa abordagem, as ferramentas de terceiros serviram primordialmente como infraestrutura de cálculo numérico, garantindo que o núcleo metodológico e o fluxo de dados permanecessem sob o domínio do próprio projeto.

## 4.2 Bases de dados utilizadas

Foram utilizadas duas bases públicas de imagens dermatoscópicas, cada uma com uma função específica dentro da metodologia. Para a etapa de classificação, foi utilizada a base ISIC 2019, que contém imagens associadas a rótulos diagnósticos disponíveis em arquivo de metadados. Essa base foi empregada por apresentar um número expressivo de imagens e por permitir a seleção das classes de interesse deste trabalho.

A base ISIC 2019 é composta por uma ampla diversidade de amostras, englobando lesões benignas (como nevos melanocíticos, ceratoses benignas e dermatofibromas) e malignas. Contudo, na classificação, o escopo deste estudo foi delimitado à distinção entre as três categorias de maior risco oncológico: melanoma, CBC e CEC. Essa delimitação foi motivada não apenas pela urgência clínica dessas patologias, mas também por um fator arquitetural crítico: as lesões benignas apresentam uma altíssima heterogeneidade morfológica entre si. Agrupar patologias tão distintas em uma única macrocategoria "benigna" introduziria uma severa variância intra-classe nos dados, dificultando a convergência matemática da rede. Sendo assim, as classes de lesões não cancerígenas foram excluídas do classificador final, garantindo que o modelo concentrasse sua capacidade extrativa na diferenciação das neoplasias abordadas.

Após a filtragem da base ISIC 2019, foram selecionadas 8.473 imagens, distribuídas em 4.522 imagens de melanoma, 3.323 imagens de CBC e 628 imagens de CEC. A Tabela 2 apresenta a distribuição das imagens utilizadas na classificação.

**Tabela 2 – Distribuição das imagens selecionadas para a etapa de classificação.**

Classe	Descrição	Quantidade de imagens
MEL	Melanoma	4.522
CBC	Carcinoma basocelular	3.323
CEC	Carcinoma espinocelular	628
<b>Total</b>	–	<b>8.473</b>

Observa-se que a base apresenta desbalanceamento entre as classes, principalmente pela menor quantidade de imagens de CEC em comparação com melanoma e CBC. Esse fator foi considerado durante a organização dos dados, pois uma base desbalanceada pode levar o modelo a favorecer as classes mais numerosas. Por esse motivo, foram avaliadas diferentes estratégias de preparação e balanceamento dos dados, descritas na seção de configurações experimentais do classificador. Dessa forma, a distribuição original da base serviu como ponto

de partida, enquanto cada configuração experimental adotou uma composição específica para os conjuntos de treinamento, validação e teste.

Para a etapa de segmentação, foi utilizada a base ISIC 2018 *Task 1*, pois ela disponibiliza máscaras manuais da região das lesões. Essas máscaras foram utilizadas como referência para o treinamento da U-Net, permitindo que o modelo aprendesse a separar a lesão do fundo da imagem. Diferentemente da classificação, a segmentação não teve o objetivo de distinguir a patologia. Esse processo foi estruturado para realizar a separação espacial entre a lesão e a pele saudável. Portanto, o modelo segmentador avalia a imagem em nível de pixel, determinando se cada pixel pertence ou não à ROI.

Como a base ISIC 2019 utilizada para classificação não disponibilizava máscaras manuais oficiais para as imagens, e considerando a natureza cumulativa dos repositórios do ISIC (onde bases mais recentes englobam imagens de edições anteriores), a construção do conjunto de segmentação exigiu uma etapa prévia de verificação de correspondência entre imagem, máscara e rótulo diagnóstico. Para isso, foram considerados apenas os casos em que havia uma imagem dermatoscópica, uma máscara manual associada (proveniente da ISIC 2018) e um rótulo válido no arquivo de metadados.

A verificação foi realizada a partir dos identificadores das imagens. O sistema percorreu o conjunto de máscaras disponíveis e, para cada uma delas, buscou a imagem correspondente na base local. Quando necessário, também foram consideradas versões reduzidas das imagens, uma vez que alguns arquivos estavam disponíveis apenas nessa forma. Após a localização da imagem, o identificador era comparado aos registros do arquivo de metadados, garantindo que apenas amostras com rótulo diagnóstico válido fossem incluídas.

Dessa forma, o processo de inclusão no conjunto de segmentação seguiu uma sequência rigorosa de validação: disponibilidade da máscara manual, localização da imagem correspondente, verificação de possíveis versões reduzidas e confirmação do rótulo tabular. Essa arquitetura de cruzamento não recorreu a bibliotecas genéricas de gerenciamento de *datasets*. Em vez disso, a lógica de emparelhamento foi codificada com o módulo nativo *pathlib* para a varredura dos diretórios, acoplada à leitura estruturada do arquivo *CSV* (*Comma-Separated Values*) via *Pandas* (*pandas.read\_csv*). Essa verificação explícita mitigou a inserção de imagens sem máscara e assegurou que o treinamento da U-Net recebesse somente pares perfeitamente correspondentes.

Na estrutura local dos dados utilizada neste trabalho, o conjunto de segmentação foi organizado em 2.593 pares imagem-máscara. Esses pares estavam distribuídos entre as categorias de Ceratose Benigna (BKL, do inglês *Benign Keratosis*), Melanoma (MEL) e Nevo Melanocítico (NV, do inglês *Melanocytic Nevus*), conforme apresentado na Tabela 3. Essa distribuição, entretanto, representa apenas a organização dos pares disponíveis e não indica que o segmentador tenha sido treinado para classificar essas categorias. A saída da U-Net manteve-se focada na delimitação espacial, distinguindo apenas os pixels pertencentes à lesão daqueles pertencentes ao fundo da imagem.

**Tabela 3 – Organização dos pares imagem-máscara utilizados na etapa de segmentação.**

<b>Categoria</b>	<b>Finalidade no estudo</b>	<b>Quantidade de pares</b>
BKL	Segmentação binária da lesão	208
MEL	Segmentação binária da lesão	519
NV	Segmentação binária da lesão	1.866
<b>Total</b>	–	<b>2.593</b>

Dessa forma, as duas bases foram empregadas de maneira complementar. A ISIC 2018 *Task 1* forneceu os pares imagem-máscara necessários para o treinamento do modelo de segmentação. A ISIC 2019, por sua vez, forneceu as imagens e os rótulos diagnósticos utilizados no treinamento do classificador. Após o treinamento da U-Net, o modelo segmentador foi aplicado às imagens do conjunto de classificação, gerando máscaras preditas para a extração das regiões de interesse utilizadas pela EfficientNet-B4.

### 4.3 Divisão dos conjuntos de treinamento, validação e teste

A partição estocástica do conjunto de dados foi projetada para separar as imagens utilizadas na retropropagação do erro (treinamento) daquelas empregadas no monitoramento (validação) e na avaliação isolada do modelo (teste). No fluxo principal, a divisão estratificada foi implementada combinando a função *train\_test\_split* (da biblioteca *scikit-learn*) a estruturas de agrupamento nativas (como *defaultdict* e *Counter*), utilizando sempre uma semente de geração pseudoaleatória fixa (*seed = 42*) por meio do *NumPy*. Esse rigor metodológico assegurou a reprodutibilidade dos experimentos e suprimiu o risco de contaminação cruzada entre os subconjuntos, prevenindo o viés de vazamento de dados. Ao executar o particionamento antes de qualquer técnica de superamostragem ou aumento de dados, evitou-se que versões sintéticas de uma mesma imagem original habitassem simultaneamente os conjuntos de treinamento e teste, cenário que mascararia a real capacidade de generalização da rede.

Inicialmente, a base filtrada da ISIC 2019 foi organizada de forma estratificada, preservando a proporção original das classes. Essa organização resultou em uma divisão aproximada de 70% para treinamento, 15% para validação e 15% para teste, conforme apresentado na Tabela 4. Essa distribuição representa a separação inicial dos dados e serviu como referência para a construção das configurações experimentais.

**Tabela 4 – Distribuição inicial das imagens da etapa de classificação nos conjuntos de treinamento, validação e teste.**

<b>Classe</b>	<b>Treinamento</b>	<b>Validação</b>	<b>Teste</b>	<b>Total</b>
MEL	3.166	678	678	4.522
CBC	2.327	498	498	3.323
CEC	440	94	94	628
<b>Total</b>	<b>5.933</b>	<b>1.270</b>	<b>1.270</b>	<b>8.473</b>

O conjunto de treinamento foi utilizado para o ajuste dos pesos da rede neural. O conjunto de validação foi empregado para acompanhar o comportamento do modelo durante o

treinamento e auxiliar na identificação de possíveis sinais de sobreajuste. Já o conjunto de teste foi reservado para a avaliação final de cada configuração experimental.

Como foram avaliadas três configurações distintas do classificador, a composição final dos conjuntos variou entre as execuções. A primeira configuração manteve validação e teste com a distribuição inicial desbalanceada. A segunda utilizou conjuntos balanceados menores, formados apenas por imagens originais. A terceira adotou uma estratégia de balanceamento mais ampla, alterando a composição do treinamento e do teste. Essas diferenças são detalhadas na seção de configurações experimentais do classificador.

No ramo de segmentação, os pares imagem-máscara também foram separados antes da aplicação de transformações de aumento de dados. Nesse caso, quando utilizadas, as transformações geométricas aplicadas às imagens de treinamento foram replicadas de forma sincronizada nas máscaras correspondentes, garantindo que a marcação da lesão permanecesse alinhada à imagem após cada transformação.

#### 4.4 Pré-processamento das imagens

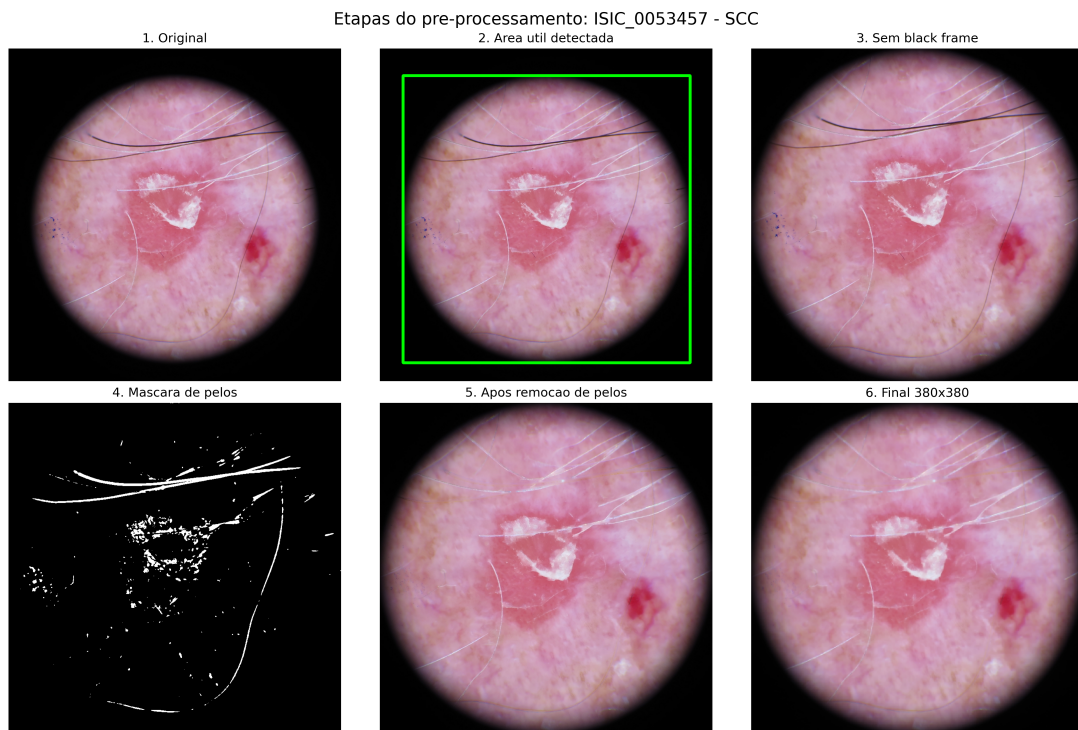
O pré-processamento teve como objetivo padronizar as imagens e reduzir interferências visuais que poderiam prejudicar tanto a segmentação quanto a classificação. As imagens dermatoscópicas podem apresentar diferenças de resolução, bordas escuras, vinhetas periféricas, pelos, reflexos e variações de iluminação. Embora essas características sejam comuns no processo de aquisição, elas podem dificultar a identificação da região da lesão e introduzir padrões irrelevantes ao diagnóstico. Para otimizar o tempo computacional dessa fase e viabilizar a execução em larga escala, o processamento de todo o conjunto de dados foi paralelizado utilizando a biblioteca nativa *concurrent.futures*, por meio da classe *ThreadPoolExecutor*.

A primeira etapa corretiva consistiu na remoção de bordas pretas e vinhetas periféricas, visando eliminar regiões escuras causadas pelo próprio dispositivo de aquisição, as quais poderiam levar a rede a interpretar esses artefatos externos como características relevantes. Essa rotina foi implementada analiticamente via *OpenCV*. Inicialmente, as imagens foram convertidas para escala de cinza e submetidas a uma operação de limiarização, separando a área útil da lesão do fundo escuro indesejado. A máscara binária resultante foi refinada por operações morfológicas de abertura e fechamento para mitigar ruídos internos. Na sequência, a detecção do perímetro da área válida foi realizada (via *cv2.findContours*), permitindo o cálculo automatizado de uma caixa delimitadora que orientou o recorte final da imagem original.

Em seguida, aplicou-se a supressão de artefatos capilares, reproduzindo a lógica do algoritmo *Dull Razor*. Este método clássico da literatura dermatoscópica consiste em três fases fundamentais: a identificação estrutural dos fios, a geração de uma máscara binária e a reconstrução digital do tecido oculto. Para viabilizar esse comportamento sob o domínio da própria implementação, a rotina foi integralmente codificada combinando primitivas matriciais: na primeira fase, a imagem recebeu a aplicação de uma operação morfológica de *black-hat*, técnica que calcula a diferença entre o fechamento morfológico e a imagem original, utilizando um ele-

mento estruturante retangular para evidenciar matematicamente os artefatos lineares escuros em relação ao fundo mais claro da pele. Na segunda fase, o mapa de ruído gerado foi submetido a uma suavização espacial (filtro *GaussianBlur*) e a uma nova limiarização, extraindo a máscara exata de localização dos pelos. Por fim, a reconstrução das áreas ocluídas foi realizada pela técnica de restauração de imagem (*inpainting* via *cv2.inpaint*), baseada no método de Telea. Esse algoritmo preenche as regiões mascaradas calculando a média ponderada dos *pixels* vizinhos conhecidos, permitindo propagar gradualmente as características e texturas da pele saudável adjacente das bordas para o centro da falha. O impacto visual dessa sequência metodológica é apresentado na Figura 20, que ilustra o ganho de visibilidade da lesão.

**Figura 20 – Etapas completas de pré-processamento aplicadas a uma amostra de Carcinoma Espinocelular**



(1) Imagem original com presença de artefatos; (2) Detecção da área útil da lesão; (3) Remoção das bordas escuras; (4) Máscara binária para detecção de pelos; (5) Imagem resultante após a remoção de pelos via *Dull Razor*; e (6) Imagem final redimensionada e pronta para as etapas seguintes.

Fonte: Autoria própria (2026).

No ramo de segmentação, as imagens limpas foram redimensionadas para  $256 \times 256$  *pixels* (utilizando a função *cv2.resize*), dimensão adotada para compatibilizar os tensores de entrada com a arquitetura da U-Net e controlar o custo computacional. Para garantir o alinhamento espacial, as máscaras correspondentes foram ajustadas para a mesma resolução empregando interpolação por vizinho mais próximo, estratégia que preservou o caráter discreto dos *pixels* de rotulação (fundo *versus* lesão). Adicionalmente, as imagens *RGB* deste ramo foram escalonadas para o intervalo contínuo de 0 a 1, favorecendo a estabilidade numérica da descida do gradiente durante o treinamento.

No ramo de classificação, as imagens foram redimensionadas para  $380 \times 380$  *pixels*, resolução nativa compatível com a entrada da arquitetura EfficientNet-B4. Diferentemente da abordagem adotada na segmentação, evitou-se a aplicação de uma normalização manual adicional nesta etapa, uma vez que o carregador da rede EfficientNet no *Keras* já incorpora as transformações algébricas esperadas pelo modelo. Esse cuidado estrutural inibiu a dupla normalização dos dados, o que poderia alterar de forma indesejada a distribuição estatística dos valores de entrada.

Por fim, as variações intrínsecas de brilho e contraste foram tratadas exclusivamente dentro das estratégias de aumento de dados aplicadas às configurações experimentais correspondentes, conforme detalhado na seção seguinte.

#### 4.5 Estratégias de balanceamento e aumento de dados

Devido ao desbalanceamento observado entre as classes melanoma, CBC e CEC, foram avaliadas diferentes estratégias de preparação dos dados para o treinamento do classificador. O objetivo dessas estratégias foi reduzir a tendência do modelo a favorecer as classes com maior número de amostras e verificar se o uso de aumento de dados ou de amostras sintéticas alterava de forma expressiva o desempenho final.

Na configuração inicial, foi aplicado aumento de dados ao conjunto de treinamento. Essa técnica consiste na geração de versões modificadas das imagens originais por meio de transformações controladas, preservando o rótulo diagnóstico da amostra. Foram utilizadas transformações geométricas e fotométricas, como rotações, deslocamentos, aproximações, inversões e ajustes de brilho e contraste. Essas transformações simulam variações que podem ocorrer no processo real de aquisição das imagens, contribuindo para que o classificador dependa menos de padrões específicos de posição, escala ou iluminação.

A segunda configuração avaliou o modelo sob um regime de subamostragem. Nessa execução, evitou-se a introdução de variações artificiais, e o treinamento ocorreu estritamente com imagens originais. O balanceamento foi obtido pela limitação aleatória do número de amostras das classes majoritárias (melanoma e CBC) ao tamanho exato da classe minoritária (CEC). Essa redução foi codificada utilizando a biblioteca *NumPy* e estruturas de agrupamento nativas do *Python*, operando sempre sob uma semente fixa para assegurar a reprodutibilidade da seleção. Esse cenário forneceu uma linha de base rigorosa sobre o comportamento do classificador perante um conjunto perfeitamente equilibrado, porém estatisticamente reduzido.

Para a terceira configuração, adotou-se o balanceamento sintético inspirado no método SMOTE. Em vez de recorrer a pacotes externos genéricos dedicados a essa tarefa, a técnica foi implementada nativamente e aplicada sobre as ROIs já pré-processadas. Para viabilizar o cálculo analítico, as imagens foram compactadas em representações vetoriais normalizadas utilizando matrizes *NumPy*. O pacote *scikit-learn* foi acionado exclusivamente na etapa de mapeamento geométrico para localizar as imagens mais semelhantes por meio da busca de vizinhos mais próximos. A partir dessas âncoras, as novas amostras (e suas respectivas máscaras,

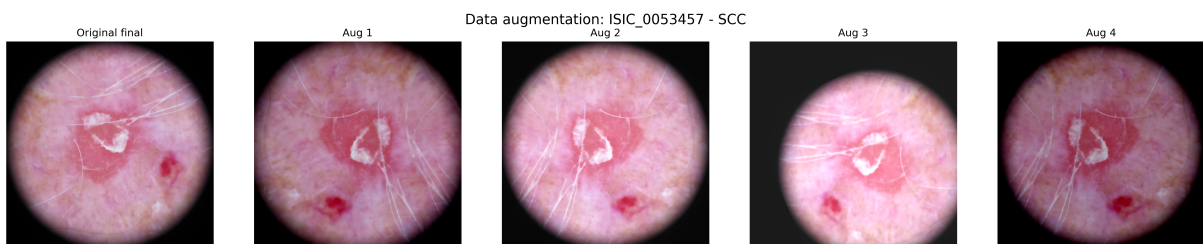
no caso da segmentação) foram geradas por meio de interpolação algébrica, seguida de uma nova binarização para as máscaras sintéticas. Essa formulação teve como objetivo ampliar a representatividade das classes menos numerosas, especialmente o CEC, reduzindo a diferença de suporte no espaço vetorial e permitindo verificar se essa abordagem modificava de forma relevante o comportamento do classificador.

A validação foi mantida sem amostras geradas artificialmente, preservando seu papel de acompanhamento do treinamento com imagens reais. No caso do conjunto de teste, as configurações apresentaram composições diferentes, conforme detalhado posteriormente na Tabela 6. Essa escolha permitiu comparar o comportamento do classificador sob diferentes condições de balanceamento.

Embora o uso de amostras artificiais no conjunto de teste não represente o protocolo mais indicado para uma validação clínica definitiva, essa configuração foi mantida como análise experimental complementar. O objetivo foi avaliar a influência das técnicas de balanceamento sobre os resultados e verificar se o desempenho do classificador permanecia coerente quando as classes eram apresentadas com suportes mais próximos.

Na etapa de segmentação, quando utilizado, o aumento de dados foi aplicado de forma sincronizada à imagem e à respectiva máscara. Para garantir esse alinhamento espacial de forma robusta e automatizada, utilizou-se a biblioteca *Albumentations*, adotada na literatura de visão computacional para processamento médico. Esse cuidado é indispensável, pois qualquer transformação geométrica (como rotações ou deslocamentos) realizada na imagem deve ser replicada matematicamente na máscara correspondente. Caso contrário, a marcação da lesão deixaria de coincidir com a região correta da imagem, prejudicando o treinamento supervisionado da U-Net, como ilustrado na Figura 21.

**Figura 21 – Exemplos de transformações aplicadas durante o aumento de dados para a amostra ISIC\_0053457**



**À esquerda, a imagem original processada, seguida por quatro variações geradas por rotações, deslocamentos espaciais e aproximações. No ramo de segmentação, essas transformações foram aplicadas de forma sincronizada às respectivas máscaras.**

**Fonte: A autoria própria (2026).**

O uso dessas transformações contribui para que o modelo aprenda características menos dependentes de escala, rotação e variações de iluminação, aumentando sua robustez frente a diferenças naturais de aquisição das imagens dermatoscópicas.

#### 4.6 Segmentação com U-Net e geração das regiões de interesse

A etapa de segmentação foi realizada com o objetivo de delimitar automaticamente a região da lesão nas imagens dermatoscópicas. Para isso, foi utilizada a arquitetura U-Net, implementada com um codificador ResNet-50 inicializado com pesos pré-treinados na base ImageNet, e treinada com os pares imagem-máscara válidos provenientes da ISIC 2018 *Task 1*. As máscaras manuais foram empregadas como referência durante o treinamento, permitindo que o modelo aprendesse a identificar a área correspondente à lesão. O ajuste matemático dessa rede foi guiado por uma função de perda híbrida, codificada no próprio projeto, que combinou o Coeficiente Dice à Perda Focal (*Focal Loss*, com parâmetros  $\gamma = 2.0$  e  $\alpha = 0.25$ ).

O carregamento dos dados dessa etapa foi orquestrado por uma estrutura baseada na classe *Sequence* (da biblioteca *Keras*). Por meio dela, as imagens utilizadas no treinamento da U-Net foram lidas, convertidas para *RGB* e redimensionadas para  $256 \times 256$  *pixels*, conforme descrito na etapa de pré-processamento. As máscaras correspondentes foram ajustadas para a mesma dimensão e convertidas para representação binária. Durante o treinamento, quando aplicadas transformações de aumento de dados, estas foram realizadas de forma sincronizada entre imagem e máscara por meio da biblioteca *Albumentations*, preservando o alinhamento espacial exato entre a lesão e sua marcação.

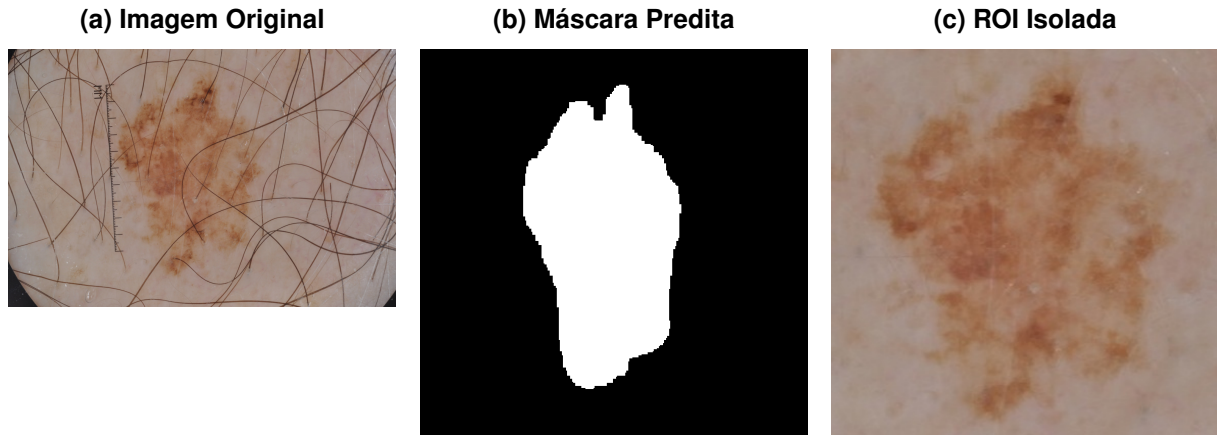
Após o treinamento, a U-Net foi aplicada às imagens utilizadas no ramo de classificação. Como essas imagens não possuíam máscaras manuais disponíveis, o modelo segmentador foi empregado para gerar máscaras preditas da região da lesão. O mapa de probabilidades retornado pela rede foi limiarizado com um valor padrão de 0,5 e submetido a operações morfológicas de abertura e fechamento para redução de ruídos. Em seguida, utilizou-se a rotina de componentes conexos (*connectedComponentsWithStats* do *OpenCV*) para isolar estritamente a maior área contínua. Essas máscaras validadas permitiram identificar a área de maior interesse clínico dentro da imagem, reduzindo a influência de elementos externos, como pele ao redor, bordas escuras e artefatos de aquisição.

A partir de cada máscara predita, o sistema extraiu as coordenadas espaciais da lesão e calculou uma caixa delimitadora acrescida de uma margem de segurança espacial para gerar a ROI. Como mecanismo de contingência, caso a máscara gerada se mostrasse excessivamente pequena ou estruturalmente inválida, o sistema previu a utilização da imagem original inteira, evitando o descarte da amostra.

Diferentemente de abordagens que aplicam a máscara de forma direta para ocultar o fundo, atribuindo cor preta ou valores nulos (do inglês *Not a Number* - NaN) aos *pixels* externos, este trabalho utilizou a segmentação semântica estritamente para o reenquadramento geométrico da lesão. Dessa forma, a pele adjacente contida dentro da caixa delimitadora foi preservada. Essa estratégia metodológica foi adotada para evitar a perda de informações morfológicas cruciais nas bordas da patologia, mitigando o impacto de possíveis falhas de subsegmentação da U-Net e mantendo o contexto cutâneo local para a análise da EfficientNet-B4. Por fim, as ROIs reenquadradas foram redimensionadas para  $380 \times 380$  *pixels*, dimensão nativa da

entrada do classificador. O progresso visual dessas etapas, desde a predição da máscara até o isolamento da ROI, é ilustrado na Figura 22.

**Figura 22 – Progresso das etapas de processamento e segmentação de uma lesão dermatoscópica da base ISIC**



**(a) Imagem clínica original; (b) Máscara binária predita pela arquitetura U-Net; (c) Região de Interesse (ROI) resultante após o recorte espacial e redimensionamento, evidenciando a preservação do tecido adjacente.**

**Fonte: Autoria própria (2026).**

Esse procedimento estabeleceu a ligação entre os dois ramos da metodologia, cuja avaliação geométrica baseou-se nas métricas de Coeficiente Dice e IoU, implementadas via *TensorFlow*. A segmentação forneceu uma estimativa da localização da lesão, enquanto a classificação utilizou essa informação para concentrar o processo de extração de características nos padrões morfológicos e texturais mais relevantes da imagem. Assim, as imagens originalmente obtidas da base ISIC 2019 foram transformadas em regiões de interesse padronizadas, com dimensão de  $380 \times 380$  *pixels*, utilizadas como entrada do classificador.

Com as regiões de interesse definidas, a etapa seguinte consistiu no treinamento da rede responsável pela distinção entre melanoma, CBC e CEC. Para isso, foram avaliadas diferentes configurações experimentais, descritas na próxima seção, com o objetivo de analisar o impacto das estratégias de balanceamento e aumento de dados sobre o desempenho do classificador.

#### **4.7 Configurações experimentais do classificador**

Após a geração das regiões de interesse, foi realizado o treinamento do classificador responsável pela distinção entre melanoma, CBC e CEC. A arquitetura utilizada nas execuções foi mantida estruturalmente equivalente, instanciada a partir do módulo *keras.applications.EfficientNetB4*, com entrada de  $380 \times 380$  *pixels*. A camada de saída foi configurada com três neurônios, correspondentes às patologias avaliadas. Para a etapa preditiva, adotou-se a ativação *softmax*, uma função matemática que normaliza os valores brutos da rede

em uma distribuição de probabilidades, assegurando que a soma das saídas resulte exatamente em 1.

A escolha da EfficientNet-B4 foi motivada por sua capacidade de extrair características visuais relevantes com boa relação entre desempenho e custo computacional. Como as imagens dermatoscópicas apresentam padrões sutis de textura, borda, assimetria e variação cromática, a utilização de uma arquitetura convolucional previamente treinada em uma base ampla de imagens (empregando os pesos da base ImageNet e excluindo a camada de classificação original via parâmetro *include\_top=False*) contribui para aproveitar representações visuais já aprendidas, caracterizando o uso de transferência de aprendizado. Em seguida, essas representações foram ajustadas ao problema específico deste trabalho por meio do treinamento supervisionado com as imagens segmentadas.

Para orquestrar o fluxo de alimentação da rede, codificou-se um gerador de dados baseado na classe *Sequence* do *Keras*. Essa estrutura foi responsável por ler as imagens via *OpenCV* (realizando a conversão do espaço de cores *BGR* para *RGB*), converter os rótulos diagnósticos para o formato *one-hot* (*tf.keras.utils.to\_categorical*), uma representação vetorial binária na qual a classe verdadeira da lesão recebe o valor 1 e as demais recebem 0, e calcular pesos por classe de forma inversamente proporcional à quantidade de amostras, auxiliando o modelo a não negligenciar as categorias minoritárias.

O treinamento foi organizado em duas fases. Na primeira, a rede-base EfficientNet-B4 foi mantida congelada, sendo treinadas apenas as camadas finais responsáveis pela classificação. Essa estratégia permite adaptar a saída do modelo às três classes de interesse, reduzindo o risco de alterações bruscas nos pesos previamente aprendidos. Na segunda fase, foi realizado o ajuste fino (*fine-tuning*), técnica de transferência de aprendizado que consiste em descongelar e atualizar sutilmente os pesos de uma rede pré-treinada para adaptá-la a uma nova tarefa, liberando aproximadamente os 30% finais das camadas da rede-base. Esse procedimento permitiu que o modelo refinasse suas representações visuais especificamente para o domínio das imagens dermatoscópicas. Durante essa etapa, as camadas de normalização em lote (*BatchNormalization*) da rede-base foram mantidas congeladas, a fim de preservar a estabilidade estatística aprendida durante o pré-treinamento.

O módulo final de classificação, acoplado ao modelo base, foi composto por uma camada de média global (*GlobalAveragePooling2D*), seguida de normalização em lote. Para reduzir o sobreajuste e tornar a predição menos dependente de uma única configuração interna da rede, adotou-se a estratégia de *Multi-Sample Dropout*. Essa técnica consiste na criação de ramificações paralelas que aplicam diferentes máscaras de descarte sobre o mesmo mapeamento de características, simulando múltiplos cenários de regularização em uma única passagem. No código desenvolvido, essa arquitetura paralela utilizou quatro camadas de *dropout* com taxas progressivas (0,2; 0,3; 0,4 e 0,5), que geraram predições independentes por meio de camadas densas. Ao final, as saídas geradas por essas diferentes ramificações foram combinadas por média, resultando na predição definitiva do classificador.

Como função de perda, foi utilizada a entropia cruzada categórica, adequada para problemas de classificação multiclasse. Também foi empregada a suavização dos rótulos (*label smoothing* de 0,1), com o objetivo de evitar que o modelo se tornasse excessivamente confiante em suas predições durante o treinamento. O otimizador utilizado foi o Adam, escolhido por sua adaptação automática da taxa de aprendizado ao longo das atualizações dos pesos. O controle do treinamento foi orquestrado por retornos de chamada do *Keras*: *EarlyStopping* (interrupção antecipada), *ReduceLROnPlateau* (redução automática da taxa de aprendizado), *ModelCheckpoint* (salvamento dos melhores pesos), *CSVLogger* (registro do histórico), além de um *callback* customizado com a biblioteca *tqdm* para o acompanhamento visual do progresso.

A Tabela 5 apresenta os principais hiperparâmetros adotados no treinamento do classificador.

**Tabela 5 – Principais hiperparâmetros utilizados no treinamento do classificador.**

Parâmetro	Valor
Arquitetura	EfficientNet-B4
Tamanho da imagem	380 × 380
<i>batch size</i>	4
Épocas da fase inicial	100
Épocas de ajuste fino	15
Taxa de aprendizado da fase inicial	0,001
Taxa de aprendizado do ajuste fino	0,000001
Suavização dos rótulos	0,1
<i>Semente aleatória</i>	42

A adoção de um *batch size* reduzido, parâmetro que define a quantidade de amostras processadas simultaneamente antes de cada atualização dos pesos da rede, foi motivada pela dimensão das imagens de entrada e pelas limitações de memória durante o treinamento. A taxa de aprendizado mais elevada na fase inicial permitiu ajustar as camadas finais do classificador de forma mais rápida, enquanto a taxa reduzida no ajuste fino foi utilizada para evitar alterações bruscas nos pesos aprendidos pela EfficientNet-B4.

Com o objetivo de avaliar o impacto da organização dos dados e das estratégias de balanceamento sobre o desempenho do classificador, foram realizadas três configurações experimentais. Todas utilizaram a mesma arquitetura principal e as regiões de interesse geradas a partir da mesma U-Net, mas diferiram quanto à composição dos conjuntos de dados e às estratégias aplicadas para lidar com o desbalanceamento entre as classes.

A primeira configuração correspondeu ao treinamento adotado no trabalho. Nessa execução, o conjunto de treinamento foi balanceado por aumento de dados, utilizando transformações geométricas e fotométricas, como rotações, inversões, deslocamentos, aproximações e ajustes de brilho e contraste. Essa estratégia ampliou a quantidade de amostras disponíveis durante o treinamento e buscou reduzir a tendência do modelo a favorecer as classes originalmente mais numerosas. Entretanto, os conjuntos de validação e teste permaneceram com distribuição desbalanceada entre as classes.

A segunda configuração foi realizada sem aumento de dados, utilizando apenas imagens originais. Nesse caso, o balanceamento foi feito por redução das classes majoritárias, limitando a quantidade de imagens de melanoma e CBC ao número disponível para CEC. Essa configuração permitiu avaliar o comportamento do classificador em um conjunto balanceado, porém menor, sem a introdução de imagens modificadas sinteticamente. Por utilizar menos amostras no treinamento, essa execução apresenta menor diversidade visual, mas serve como comparação para verificar o efeito do balanceamento por redução.

A terceira configuração utilizou o SMOTE como estratégia de balanceamento sintético. Embora a utilização de amostras aumentadas ou sintéticas em conjuntos de avaliação não seja a prática mais indicada para estimar a generalização clínica definitiva de um modelo, essa configuração foi mantida como análise experimental complementar. O objetivo foi verificar se o aumento e o balanceamento artificial dos dados alteravam de forma expressiva o comportamento do classificador. Dessa forma, a comparação entre as três configurações permitiu avaliar se as estratégias de aumento de dados estavam introduzindo distorções relevantes ou se contribuíam para ampliar a variabilidade das amostras de forma coerente entre as classes.

Essa técnica, SMOTE, foi aplicada sobre os dados já pré-processados e organizados como regiões de interesse, com o objetivo de ampliar a quantidade de amostras disponíveis para as classes menos representadas. A arquitetura do classificador e a U-Net utilizada para a geração das regiões de interesse foram mantidas em relação às demais configurações, de modo que a principal diferença dessa execução esteve na composição dos conjuntos de dados. A validação permaneceu sem amostras sintéticas, permitindo acompanhar o treinamento com imagens reais.

Todas as configurações experimentais utilizaram um classificador baseado na EfficientNet-B4, com entrada de  $380 \times 380$  *pixels* e saída correspondente às três classes de interesse: melanoma, CBC e CEC. Também foi mantida a mesma U-Net para a geração das regiões de interesse utilizadas como entrada do classificador. Dessa forma, as execuções não representam arquiteturas distintas, mas diferentes formas de preparação, balanceamento e divisão dos dados.

A Tabela 6 apresenta apenas os aspectos que diferenciaram as três configurações avaliadas.

**Tabela 6 – Principais diferenças entre as configurações experimentais do classificador.**

Aspecto	Configuração inicial	Sem aumento de dados	Com balanceamento sintético
Estratégia principal	Aumento de dados no conjunto de treinamento	Uso apenas de imagens originais	Balanceamento sintético dos dados
Composição do treinamento	10.000 imagens por classe	440 imagens por classe	2.824 imagens por classe
Composição da validação	MEL: 678; CBC: 498; CEC: 94	94 imagens por classe	MEL: 678; CBC: 498; CEC: 94
Composição do teste	MEL: 678; CBC: 498; CEC: 94	94 imagens por classe	678 imagens por classe
Forma de balanceamento	Geração de variações por transformações geométricas e fotométricas	Redução das classes majoritárias ao tamanho da classe CEC	Geração de amostras sintéticas para equilibrar as classes
Principal limitação	Teste desbalanceado, com menor suporte para CEC	Conjunto de treinamento reduzido	Teste balanceado com amostras sintéticas, usado como avaliação experimental complementar

A comparação entre essas configurações não deve ser interpretada como a avaliação de arquiteturas diferentes, mas como a análise do efeito da preparação dos dados sobre um classificador estruturalmente equivalente. Assim, as diferenças observadas nos resultados estão associadas à forma de balanceamento, ao tamanho dos conjuntos e à composição das amostras utilizadas em treinamento, validação e teste.

Também foi adotado um rigor metodológico para evitar o vazamento de dados entre os conjuntos. Para garantir isso, a divisão dos dados foi realizada antes da aplicação de qualquer técnica de aumento de dados ou balanceamento sintético. Por meio computacional, essa divisão estratificada foi implementada utilizando a função *train\_test\_split* da biblioteca *scikit-learn*, combinada a operações da biblioteca *NumPy* com semente de geração pseudoaleatória fixa (*seed = 42*) para garantir a reprodutibilidade. Essa abordagem assegurou que uma imagem original e suas futuras variações sintéticas permanecessem ancoradas a um único subconjunto, impedindo que o modelo fosse avaliado em imagens derivadas de amostras já vistas na fase de treinamento, o que elevaria sinteticamente as métricas de desempenho.

A configuração final considerada mais adequada foi escolhida não apenas pelo maior valor global de AUC, mas pela consistência do desempenho entre as três classes. Essa escolha é importante porque, em bases desbalanceadas, métricas globais podem mascarar dificuldades do modelo em classes menos representadas. Assim, a análise priorizou métricas por classe, extraídas por meio da biblioteca *scikit-learn* (utilizando funções como *classification\_report*, *confusion\_matrix* e *roc\_auc\_score*), em particular sensibilidade, precisão e F1-Score. Adicionalmente, a especificidade foi computada via código próprio a partir da matriz de confusão sob uma abordagem um-contra-todos, de modo a verificar matematicamente se o classificador apresentava desempenho equilibrado para melanoma, CBC e CEC.

## 5 RESULTADOS E DISCUSSÕES

Este capítulo apresenta e discute os resultados obtidos a partir do fluxo metodológico desenvolvido, contemplando a etapa de segmentação das lesões e a etapa de classificação diagnóstica. A análise foi baseada nos artefatos gerados ao final dos treinamentos, incluindo curvas de treinamento e validação da U-Net, curvas de treinamento do classificador, matrizes de confusão, métricas de desempenho por classe e registros de divisão dos dados.

A análise dos resultados considera as três configurações experimentais descritas na metodologia, permitindo comparar o efeito do aumento de dados e do balanceamento sintético sobre o desempenho do classificador.

### 5.1 Organização dos Experimentos e Distribuição de Dados

Conforme apresentado na Seção 4.6, o classificador foi avaliado em três configurações experimentais, mantendo-se a mesma arquitetura principal, baseada na EfficientNet-B4, e as mesmas regiões de interesse geradas pela U-Net. Dessa forma, as diferenças observadas entre os resultados estão associadas principalmente à forma de preparação, aumento e balanceamento dos dados, e não à alteração da arquitetura do modelo.

A Tabela 7 apresenta a distribuição das imagens utilizadas em cada configuração. Essa organização é importante porque os conjuntos de treino, validação e teste não tiveram a mesma composição nos três experimentos, o que influencia diretamente a interpretação das métricas obtidas.

**Tabela 7 – Distribuição das imagens nas três configurações experimentais do classificador**

Configuração	Subconjunto	Melanoma	CBC	CEC
Configuração inicial	Treinamento	10.000	10.000	10.000
	Validação	678	498	94
	Teste	678	498	94
Sem aumento de dados	Treinamento	440	440	440
	Validação	94	94	94
	Teste	94	94	94
Com balanceamento sintético	Treinamento	2.824	2.824	2.824
	Validação	678	498	94
	Teste	678	678	678

**Fonte: Autoria própria (2026).**

A distribuição apresentada na Tabela 7 mostra que cada configuração impõe uma condição distinta de avaliação. Na configuração inicial, o treinamento foi amplamente balanceado por aumento de dados, mas a validação e o teste permaneceram desbalanceados, com menor suporte para a classe CEC. Assim, métricas globais obtidas nessa configuração tendem a refletir com maior peso o desempenho nas classes mais numerosas. Na configuração sem aumento de dados, os conjuntos ficaram balanceados, porém com número reduzido de imagens, o que torna a comparação entre as classes mais direta, mas limita a diversidade visual disponível para

o treinamento. Já na configuração com balanceamento sintético, o conjunto de teste foi equilibrado entre as três classes, permitindo observar melhor o comportamento do modelo em CEC, embora esse resultado deva ser interpretado como uma avaliação experimental complementar, por envolver amostras sintéticas.

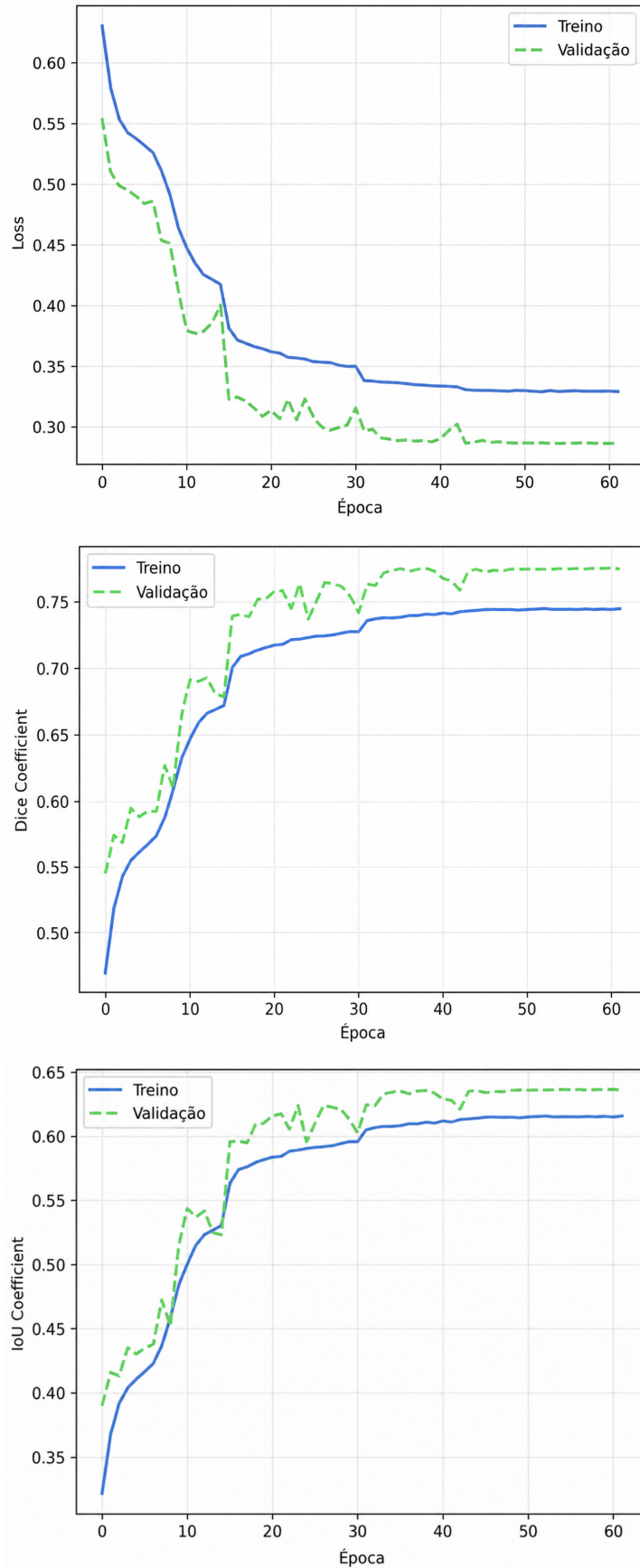
Em especial, a comparação entre as configurações permite verificar se as estratégias de aumento e balanceamento de dados alteraram de forma expressiva o comportamento do classificador ou se contribuíram para ampliar a variabilidade das amostras de maneira coerente.

Portanto, a análise dos resultados não se limita à acurácia global. Como a base original apresenta desbalanceamento entre melanoma, CBC e CEC, também são consideradas métricas por classe, como precisão, sensibilidade, especificidade e F1-Score. Essa abordagem permite avaliar se o classificador apresentou desempenho equilibrado entre as classes ou se continuou favorecendo categorias com maior representatividade.

## **5.2 Resultados da Segmentação**

A Figura 23 apresenta as curvas de treinamento e validação obtidas para a U-Net. A análise da segmentação foi realizada a partir da evolução da função de perda, do Coeficiente Dice e do IoU, pois essas métricas permitem avaliar a aproximação entre as máscaras preditas pelo modelo e as máscaras de referência.

**Figura 23 – Curvas de treinamento e validação da U-Net**



Curvas de treinamento e validação ao longo das épocas: (A) evolução da função de perda (*Loss*); (B) evolução do Coeficiente Dice; e (C) evolução do Índice de Jaccard (*IoU*).

Fonte: Autoria própria (2026).

As curvas de validação permaneceram próximas das curvas de treinamento e, em alguns momentos, apresentaram desempenho ligeiramente superior. Esse comportamento não indica, necessariamente, erro experimental. Como o conjunto de treinamento recebeu maior diversidade de padrões morfológicos, cromáticos e texturais presentes nas imagens devido ao aumento de dados, ele apresenta maior complexidade de convergência do que o conjunto de validação. Dessa forma, a proximidade entre as curvas sugere que o modelo apresentou convergência adequada e não demonstrou sinais evidentes de sobreajuste durante o treinamento.

Observa-se que a função de perda apresentou queda consistente ao longo das épocas, tanto no conjunto de treinamento quanto no conjunto de validação. A perda de treinamento iniciou em valor próximo de 0,63 e estabilizou-se em torno de 0,33, enquanto a perda de validação caiu de aproximadamente 0,55 para valores próximos de 0,29. Esse comportamento indica que o modelo passou a gerar máscaras progressivamente mais próximas das anotações esperadas, reduzindo os erros de segmentação ao longo do treinamento.

Além da redução da perda, as métricas Dice e IoU apresentaram crescimento progressivo. O Coeficiente Dice de validação aproximou-se de valores superiores a 0,75, enquanto o IoU de validação estabilizou-se próximo de 0,64. Esses valores indicam que a U-Net conseguiu aprender uma delimitação funcional das regiões de interesse, embora ainda distante de resultados mais elevados reportados em alguns trabalhos específicos de segmentação. Por exemplo, Júnior, Bezerra e Andrade (2020), ao compararem arquiteturas baseadas em U-Net com diferentes codificadores, relataram valores de Dice superiores a 89% e Jaccard próximos ou superiores a 81% em determinadas configurações. Ferreira (2018) também reportou médias elevadas de Dice e IoU em bases públicas de imagens dermatoscópicas. Em comparação, os resultados obtidos neste trabalho são mais modestos, mas permanecem coerentes com uma segmentação útil para extração da região de interesse.

Essa diferença em relação à literatura pode ser explicada por alguns fatores. Parte dos trabalhos comparativos concentra-se exclusivamente na tarefa de segmentação, com maior controle sobre hiperparâmetros, bases de dados e avaliação visual das máscaras. Neste trabalho, a segmentação foi utilizada como uma etapa intermediária de um fluxo maior, cujo objetivo final é a classificação das lesões. Assim, a U-Net não foi avaliada apenas como produto final, mas como mecanismo para reduzir a influência do fundo da imagem e fornecer regiões de interesse ao classificador.

A Figura 24 apresenta um exemplo visual em que a segmentação produzida pela U-Net foi relativamente adequada. Observa-se que a máscara cobre a maior parte da região pigmentada da lesão e preserva uma delimitação compatível com a área de interesse. Embora a borda da máscara não acompanhe todos os detalhes finos da lesão, a região principal foi corretamente destacada, o que indica uma segmentação funcional para a etapa posterior de classificação.

**Figura 24 – Exemplo de máscara com delimitação relativamente adequada**

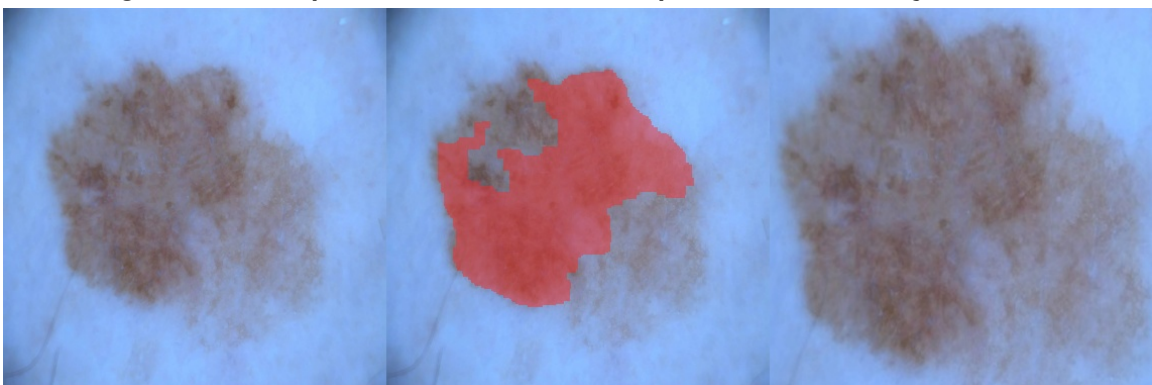


**Fonte: Autoria própria (2026).**

Outro ponto relevante é que a segmentação de lesões dermatoscópicas é sensível à presença de bordas pouco definidas, baixo contraste entre lesão e pele, pelos, artefatos e regiões escuras nas extremidades da imagem. Araújo e Junior (2017), ao compararem métodos tradicionais de segmentação, destacam que imagens com baixo contraste, artefatos e bordas pretas prejudicam a delimitação correta das lesões. Essa observação ajuda a interpretar os resultados obtidos neste trabalho, pois mesmo com pré-processamento e remoção de artefatos, ainda podem existir casos em que a fronteira entre a lesão e a pele saudável não é claramente separável.

A Figura 25 ilustra esse tipo de limitação. Nesse exemplo, a máscara não cobre toda a extensão visual da lesão e permanece concentrada em uma região parcial da imagem. Esse comportamento sugere dificuldade do modelo em reconhecer áreas de transição suave entre a lesão e a pele ao redor, especialmente quando a coloração é menos contrastante. Como consequência, partes relevantes da lesão podem não ser consideradas na região de interesse enviada ao classificador.

**Figura 25 – Exemplo de máscara com menor precisão na delimitação da lesão**



**Fonte: Autoria própria (2026).**

A comparação entre os dois exemplos visuais ajuda a interpretar os valores médios das métricas. O desempenho da segmentação foi suficiente para gerar regiões de interesse úteis ao classificador, mas ainda existem casos em que a delimitação não acompanha toda a extensão visual da lesão. Portanto, os resultados da U-Net devem ser interpretados como satisfatórios para o funcionamento do método, mas não como uma segmentação perfeita.

Do ponto de vista prático, os resultados indicam que a segmentação foi suficiente para cumprir sua função dentro do fluxo proposto. A U-Net conseguiu destacar a região da lesão e gerar máscaras que permitiram a construção das regiões de interesse utilizadas posteriormente pelo classificador. Esse papel é relevante porque a literatura aponta que a segmentação pode melhorar a análise automatizada ao direcionar o modelo para a área clinicamente mais importante da imagem. Mateen *et al.* (2024), por exemplo, observaram que a etapa de segmentação contribuiu positivamente para o desempenho da classificação em imagens dermatoscópicas, reforçando a importância de isolar a região da lesão antes da decisão diagnóstica.

Apesar disso, a segmentação ainda representa uma limitação do sistema. Pequenos erros na delimitação da máscara podem remover partes importantes da lesão ou preservar regiões externas que não contribuem para o diagnóstico. Esse efeito pode se propagar para a etapa de classificação, principalmente em lesões com bordas irregulares, baixo contraste ou padrões morfológicos sutis. Portanto, embora os resultados da U-Net sejam satisfatórios para a geração das regiões de interesse, melhorias futuras na segmentação podem contribuir diretamente para aumentar a estabilidade e a confiabilidade do classificador.

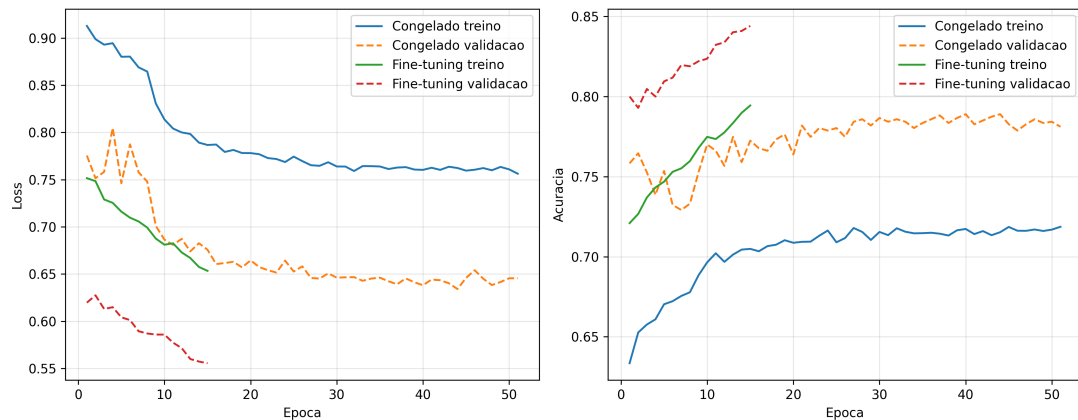
### 5.3 Resultados do Classificador

Os resultados do classificador foram analisados a partir das três configurações experimentais descritas na Seção 4.6. Como a arquitetura principal foi mantida nas três execuções, baseada na EfficientNet-B4, as diferenças observadas estão relacionadas principalmente à forma de preparação, aumento e balanceamento dos dados.

Diferentemente da etapa de segmentação, na qual o Coeficiente Dice foi utilizado para avaliar a sobreposição entre máscaras preditas e máscaras de referência, a etapa de classificação foi avaliada por métricas próprias de problemas multiclasse. Nessa etapa, o modelo não produz uma máscara espacial, mas uma probabilidade associada a cada classe diagnóstica. Por isso, foram utilizadas métricas como precisão, sensibilidade, especificidade, F1-Score, acurácia e AUC macro no esquema um-contra-os-demais (OVR, do inglês *One-vs-Rest*). Dessa forma, o Dice não foi aplicado aos classificadores, pois essa métrica mede sobreposição entre regiões segmentadas e não desempenho de decisão entre classes.

As Figuras 26, 28 e 30 apresentam as curvas de treinamento e validação dos três classificadores. Essas curvas permitem observar o comportamento dos modelos durante o treinamento, enquanto as matrizes de confusão, apresentadas nas Figuras 27, 29 e 31, permitem avaliar como os erros se distribuíram entre melanoma, CBC e CEC.

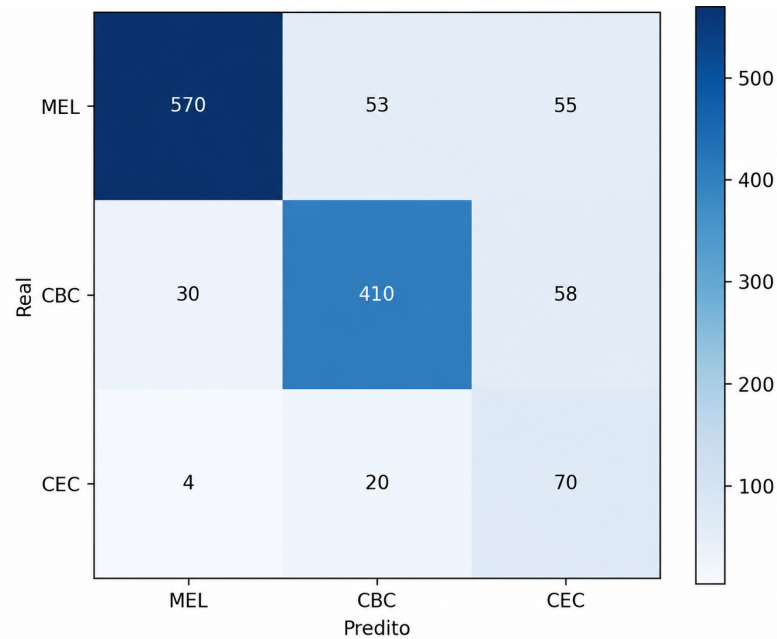
**Figura 26 – Curvas de treinamento e validação do classificador na configuração inicial**



**Curvas de treinamento e validação do classificador na configuração inicial: (A) evolução da função de perda (*Loss*); e (B) evolução da acurácia durante as etapas de treinamento com camadas congeladas e de ajuste fino (*fine-tuning*).**

**Fonte: Autoria própria (2026).**

Na configuração inicial, as curvas indicam redução progressiva da função de perda e aumento da acurácia e da AUC ao longo do treinamento, mostrando que o classificador conseguiu aprender padrões discriminativos a partir das regiões de interesse geradas pela segmentação. Esse comportamento também aparece nas métricas finais, pois essa configuração apresentou a maior AUC macro OVR entre os três experimentos, com valor de 0,9421, além dos melhores F1-Scores para melanoma e CBC, com 88,92% e 83,59%, respectivamente. Entretanto, essa leitura deve ser feita com cautela, pois os conjuntos de validação e teste permaneceram desbalanceados, com menor número de imagens da classe CEC. Como consequência, mesmo com bom desempenho global, o modelo apresentou precisão de apenas 38,25% e F1-Score de 50,54% para CEC. Portanto, a configuração inicial demonstrou boa capacidade geral de separação entre as classes, mas não apresentou desempenho equilibrado entre todas elas.

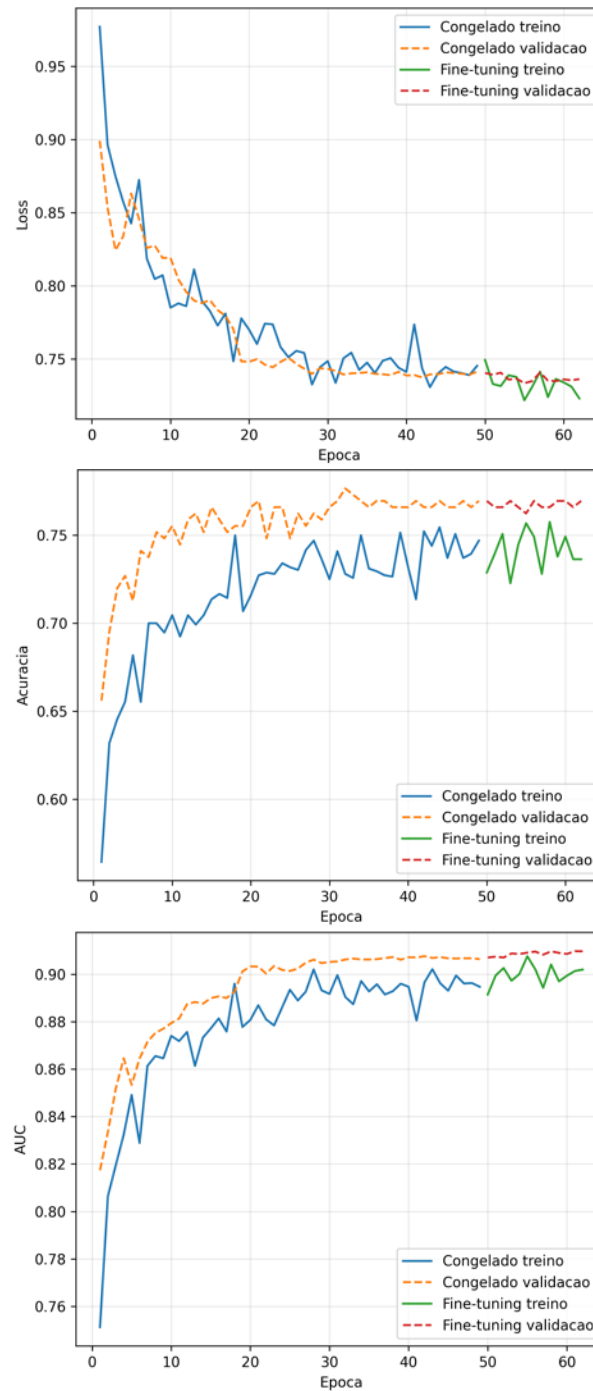
**Figura 27 – Matriz de confusão do classificador na configuração inicial**

**Fonte: Autoria própria (2026).**

Na configuração inicial, a matriz de confusão mostra que o classificador obteve melhor desempenho nas classes melanoma e CBC, com 570 acertos em 678 imagens de melanoma e 410 acertos em 498 imagens de CBC. Para a classe CEC, foram 70 acertos em 94 imagens, indicando que o modelo recuperou parte relevante dos casos dessa classe. No entanto, essa leitura deve ser complementada pela análise dos falsos positivos e falsos negativos.

Considerando a classe CEC como positiva em uma análise um-contra-todos, ocorreram 24 falsos negativos, correspondentes às imagens reais de CEC classificadas como melanoma ou CBC. Por outro lado, ocorreram 113 falsos positivos para CEC, sendo 55 imagens reais de melanoma e 58 de CBC classificadas incorretamente como CEC. Esse comportamento explica a sensibilidade relativamente alta da classe CEC, mas também sua baixa precisão. Em outras palavras, o modelo não ignorou a classe CEC, pois conseguiu reconhecer parte dos seus padrões. Entretanto, também passou a atribuir esse rótulo a muitas imagens de outras classes. Assim, a matriz de confusão confirma que a maior AUC macro dessa configuração não representa, necessariamente, equilíbrio entre as classes.

**Figura 28 – Curvas de treinamento e validação do classificador sem aumento de dados**



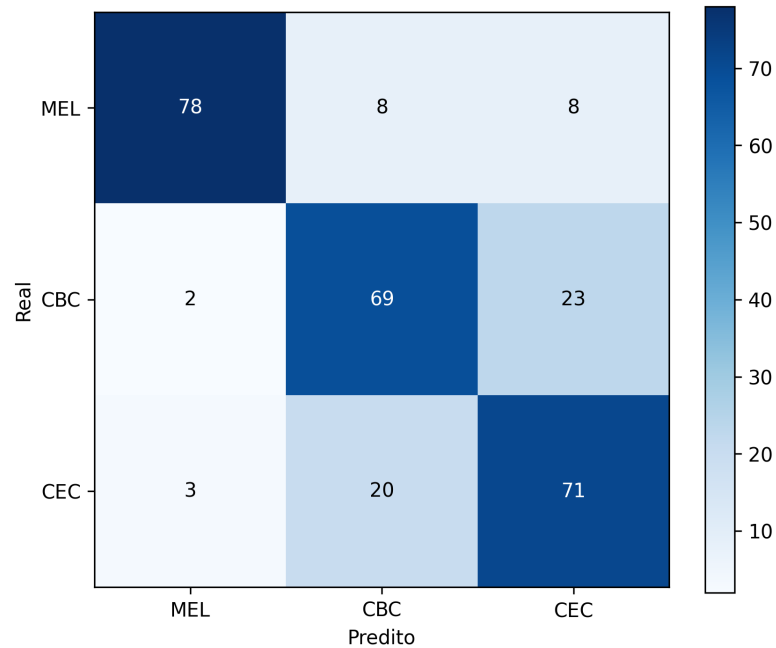
**Curvas de treinamento e validação do classificador na configuração sem aumento de dados: (A) evolução da função de perda (*Loss*); (B) evolução da acurácia; e (C) evolução da área sob a curva (AUC), durante as etapas de treinamento inicial e de ajuste fino (*fine-tuning*).**

**Fonte: Autoria própria (2026).**

Na configuração sem aumento de dados, as curvas mostram uma convergência mais limitada em comparação à configuração inicial. A função de perda diminuiu ao longo das épocas, mas os valores de acurácia e AUC de validação permaneceram inferiores aos observados no primeiro experimento. Esse comportamento é coerente com o menor volume de dados utilizado no treinamento, uma vez que o balanceamento por redução limitou o conjunto a 440 imagens por classe. A AUC macro OVR foi de 0,9090, a menor entre as três configurações. Apesar

disso, o desempenho entre CBC e CEC tornou-se mais equilibrado, com F1-Score de 72,25% para CBC e 72,45% para CEC. Assim, essa configuração mostra que a redução das classes majoritárias diminuiu a diferença de desempenho entre as classes, mas ao custo de menor diversidade visual disponível para treinamento e avaliação.

**Figura 29 – Matriz de confusão do classificador sem aumento de dados**

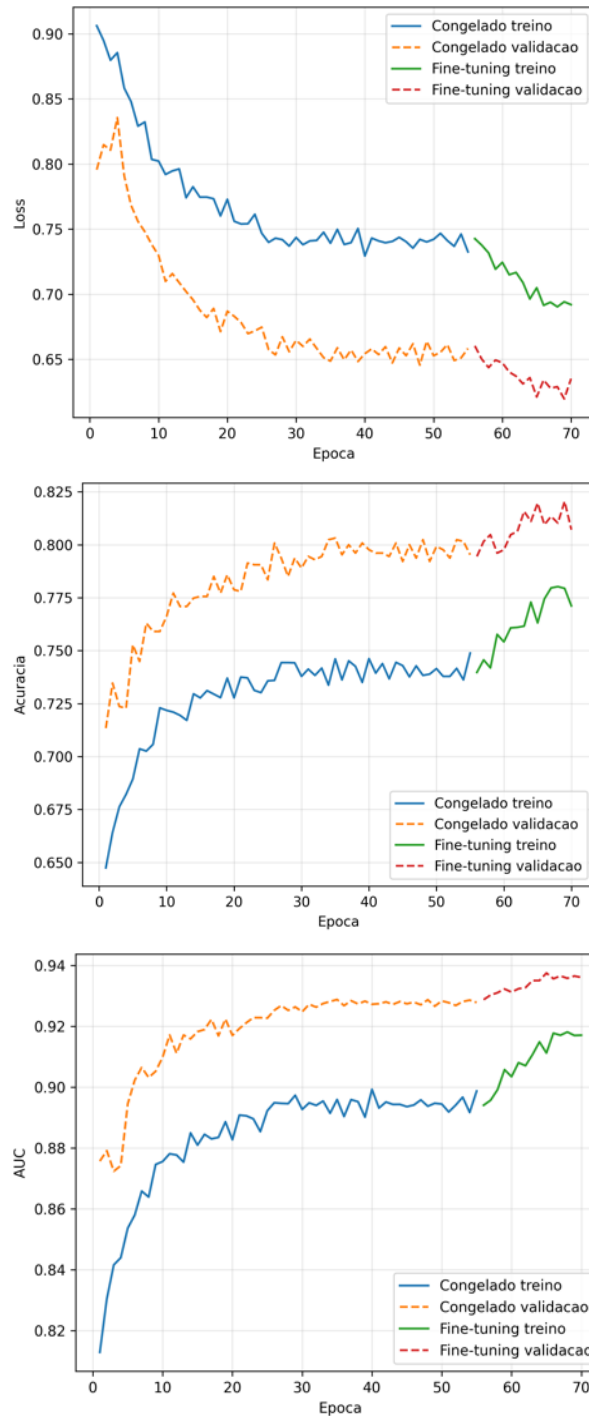


**Fonte: Autoria própria (2026).**

Na configuração sem aumento de dados, a matriz de confusão apresenta uma distribuição de erros mais equilibrada entre as classes. O modelo classificou corretamente 78 das 94 imagens de melanoma, 69 das 94 imagens de CBC e 71 das 94 imagens de CEC. Como o conjunto de teste possui o mesmo número de imagens por classe, a comparação entre acertos e erros torna-se mais direta.

Considerando cada classe em uma análise um-contra-todos, observa-se que melanoma apresentou 16 falsos negativos, correspondentes a 8 imagens classificadas como CBC e 8 como CEC. Além disso, houve 5 falsos positivos para melanoma, sendo 2 imagens reais de CBC e 3 de CEC classificadas incorretamente como melanoma. Para CBC, ocorreram 25 falsos negativos e 28 falsos positivos. Para CEC, ocorreram 23 falsos negativos e 31 falsos positivos. Esses valores mostram que a retirada do aumento de dados e o balanceamento por redução diminuíram a discrepância observada na configuração inicial, mas também tornaram os erros mais distribuídos entre as classes. Assim, o modelo apresentou avaliação mais equilibrada, porém com menor desempenho global e menor diversidade visual disponível para treinamento.

**Figura 30 – Curvas de treinamento e validação do classificador com balanceamento sintético**



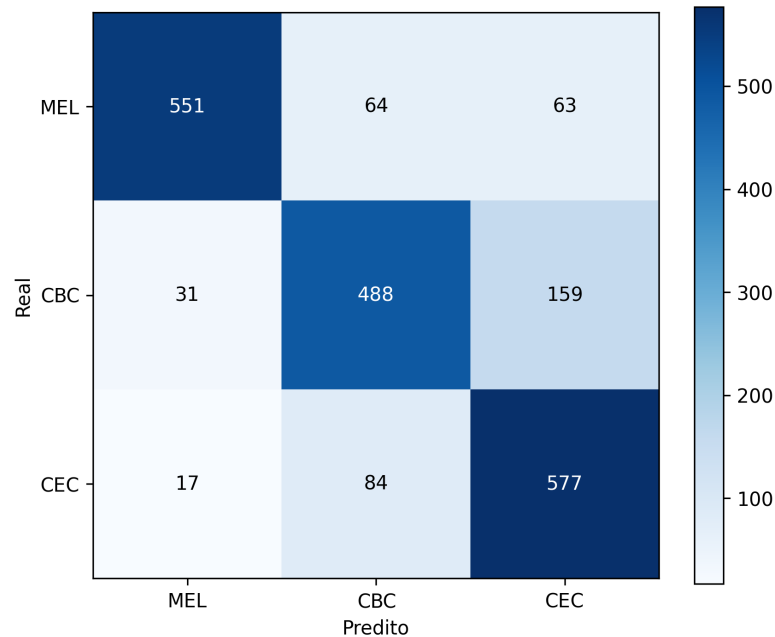
**Curvas de treinamento e validação do classificador na configuração com balanceamento sintético: (A) evolução da função de perda (*Loss*); (B) evolução da acurácia; e (C) evolução da área sob a curva (AUC), durante as etapas de treinamento inicial e de ajuste fino (*fine-tuning*).**

**Fonte: Autoria própria (2026).**

Na configuração com balanceamento sintético, as curvas indicam melhora consistente durante o treinamento, com redução da perda e aumento da acurácia e da AUC. O desempenho de validação ficou em posição intermediária entre as duas configurações anteriores: inferior ao da configuração inicial em termos de AUC, mas superior ao experimento sem aumento de dados. A AUC macro OVR final foi de 0,9265. Embora esse valor não tenha superado o da

configuração inicial, essa execução apresentou o melhor desempenho para a classe CEC, com sensibilidade de 85,10% e F1-Score de 78,13%. Além disso, o conjunto de teste apresentou 678 imagens por classe, permitindo uma avaliação mais equilibrada entre melanoma, CBC e CEC. Dessa forma, o balanceamento sintético contribuiu para reduzir a fragilidade observada na classe CEC sem provocar uma queda acentuada no desempenho global do classificador.

**Figura 31 – Matriz de confusão do classificador com balanceamento sintético**



**Fonte: Autoria própria (2026).**

Na configuração com balanceamento sintético, a matriz de confusão mostra uma melhora importante na identificação da classe CEC. O modelo classificou corretamente 551 das 678 imagens de melanoma, 488 das 678 imagens de CBC e 577 das 678 imagens de CEC. A classe CEC apresentou o maior número absoluto de acertos entre as três classes nessa configuração, o que está de acordo com o aumento de sensibilidade observado nas métricas.

Considerando a classe CEC como positiva em uma análise um-contra-todos, ocorreram 101 falsos negativos, sendo 17 imagens classificadas como melanoma e 84 como CBC. Por outro lado, ocorreram 222 falsos positivos para CEC, sendo 63 imagens reais de melanoma e 159 de CBC classificadas incorretamente como CEC. Esses valores mostram que o balanceamento sintético aumentou a capacidade do modelo de recuperar casos reais de CEC, mas ainda manteve uma quantidade relevante de falsos positivos nessa classe. Em outras palavras, o modelo passou a reconhecer melhor a classe CEC, mas também continuou confundindo parte das imagens de CBC com CEC. Assim, a matriz de confusão indica melhora no equilíbrio entre as classes, embora a separação entre CBC e CEC permaneça como um dos principais desafios do classificador.

De forma geral, as matrizes de confusão mostram que o principal desafio do classificador não foi distinguir melanoma das demais classes, mas separar de forma consistente CBC e CEC. Esse comportamento é relevante do ponto de vista dermatológico, pois ambas pertencem

ao grupo dos cânceres de pele não melanoma e podem apresentar padrões dermatoscópicos parcialmente semelhantes, como textura, coloração e bordas. Além disso, os erros do classificador não possuem o mesmo impacto em um sistema de apoio à triagem: falsos negativos podem ser mais críticos por representarem lesões reais que deixam de ser identificadas pelo modelo, enquanto falsos positivos podem aumentar encaminhamentos, exames complementares ou intervenções desnecessárias. Assim, a análise das matrizes reforça que modelos aplicados à saúde devem ser avaliados não apenas pelo número total de acertos, mas também pelo tipo de erro cometido e por sua possível implicação clínica.

A Tabela 8 resume as métricas obtidas no conjunto de teste para cada configuração.

**Tabela 8 – Resultados comparativos dos três classificadores no conjunto de teste**

Configuração	Classe	Precisão	Sensib.	Espec.	F1	Suporte
Inicial	Melanoma	94,37%	84,07%	94,26%	88,92%	678
	CBC	84,89%	82,33%	90,54%	83,59%	498
	CEC	38,25%	74,47%	90,39%	50,54%	94
Sem aumento de dados	Melanoma	93,98%	82,98%	97,34%	88,14%	94
	CBC	71,13%	73,40%	85,11%	72,25%	94
	CEC	69,61%	75,53%	83,51%	72,45%	94
Com balanceamento sintético	Melanoma	91,99%	81,27%	96,46%	86,30%	678
	CBC	76,73%	71,98%	89,09%	74,28%	678
	CEC	72,22%	85,10%	83,63%	78,13%	678

**Fonte: Autoria própria (2026).**

A comparação entre as três configurações mostra que a maior AUC macro não corresponde, necessariamente, ao desempenho mais equilibrado entre as classes. A configuração inicial apresentou o melhor desempenho global e os maiores F1-Scores para melanoma e CBC, mas teve desempenho limitado em CEC. A configuração sem aumento de dados reduziu essa diferença entre as classes, porém foi avaliada em um conjunto menor. Já a configuração com balanceamento sintético apresentou o melhor desempenho para CEC e permitiu uma comparação mais equilibrada, por utilizar o mesmo suporte para as três classes no conjunto de teste.

A análise conjunta das curvas de treinamento, das matrizes de confusão e das métricas por classe mostra que os gráficos de convergência não são suficientes, isoladamente, para escolher o melhor classificador. A configuração inicial apresentou as melhores curvas globais e a maior AUC macro, mas sua matriz de confusão revelou baixo desempenho para CEC. Por outro lado, a configuração com balanceamento sintético não obteve a maior AUC, mas apresentou melhor sensibilidade e F1-Score para CEC, além de avaliação com suporte equilibrado entre as classes. Isso reforça que, em bases desbalanceadas, a análise por classe é mais informativa do que a observação isolada da acurácia ou da AUC.

#### **5.4 Recursos Computacionais e Tempos de Processamento**

Em relação aos recursos computacionais, o treinamento, a validação e os testes das arquiteturas propostas neste trabalho foram executados em uma estação de trabalho equipada

com processador Intel Core i7-10750H, com 6 núcleos e 12 threads, 15,84 GB de memória RAM e uma Unidade de Processamento Gráfico (GPU) dedicada NVIDIA GeForce RTX 2060, com 6 GB de memória de vídeo. A utilização de GPU foi fundamental para viabilizar os experimentos, permitindo aceleração por hardware por meio do ambiente TensorFlow/Keras com suporte às bibliotecas CUDA e cuDNN.

O tempo de processamento variou substancialmente entre as diferentes etapas do fluxo metodológico. Na etapa de segmentação, o treinamento da arquitetura U-Net com codificador ResNet-50 demandou aproximadamente 19 horas e 54 minutos. Já na etapa de classificação, os tempos de execução da EfficientNet-B4 variaram conforme o volume de dados de cada configuração experimental. A configuração inicial, com aumento de dados, foi a mais custosa, demandando cerca de 10 horas, 33 minutos e 30 segundos de treinamento. A execução com balanceamento sintético exigiu aproximadamente 3 horas, 30 minutos e 54 segundos, enquanto a configuração sem aumento de dados, por utilizar um conjunto substancialmente menor, convergiu de forma mais rápida, em cerca de 31 minutos e 46 segundos.

Os tempos apresentados foram estimados a partir dos registros e metadados dos arquivos gerados durante as execuções, uma vez que os logs de treinamento armazenaram as métricas por época, mas não registraram explicitamente a duração individual de cada época.

Não foram mensurados formalmente o consumo máximo de memória da GPU, o uso médio de CPU/RAM nem o tempo médio de inferência por imagem, sendo esses aspectos indicados como pontos para avaliação em trabalhos futuros.

A diferença entre os tempos de processamento entre as etapas evidencia que o treinamento supervisionado demanda recursos computacionais consideráveis, especialmente nas configurações com maior volume de dados. Ainda assim, após o treinamento, o modelo final pode ser utilizado para inferência de novas imagens de forma mais ágil, reforçando sua viabilidade como ferramenta computacional de apoio à análise de lesões dermatológicas.

## 5.5 Discussão Geral, Limitações e Implicações

A análise conjunta da segmentação e da classificação mostra que o pipeline desenvolvido foi capaz de produzir resultados funcionais para a análise de lesões dermatológicas em imagens dermatoscópicas. A U-Net apresentou convergência adequada e gerou regiões de interesse úteis para a etapa posterior de classificação. Entretanto, os exemplos visuais demonstraram que a segmentação ainda não é perfeita, principalmente em imagens com bordas pouco definidas, baixo contraste entre lesão e pele ou transições cromáticas suaves. Assim, pequenos erros na delimitação espacial da máscara podem influenciar severamente a classificação, seja pela remoção de partes relevantes da lesão (suprimindo, por exemplo, bordas irregulares e transições abruptas que são críticas para o diagnóstico visual de malignidade), seja pela preservação de regiões externas sem valor diagnóstico que adicionam ruído ao modelo.

Na etapa de classificação, os resultados evidenciaram que a escolha da melhor configuração não deve ser baseada apenas em métricas globais. A configuração inicial apresentou a

maior AUC macro OVR e melhor desempenho para melanoma e CBC, mas manteve desempenho inferior para CEC. Esse comportamento reforça que, em bases médicas desbalanceadas, métricas globais encobrem fragilidades importantes em classes menos representadas, induzindo o modelo a atuar, na prática, como um preditor de classes majoritárias para minimizar rapidamente sua função de perda.

A configuração com balanceamento sintético mostrou-se mais adequada para o objetivo deste trabalho, pois melhorou o desempenho da classe CEC e permitiu uma avaliação mais equilibrada entre as três patologias. Embora sua AUC macro tenha sido inferior à configuração inicial, o ganho expressivo em sensibilidade e F1-Score para o CEC indica que o modelo deixou de negligenciar a classe minoritária e passou a reconhecer efetivamente os seus padrões visuais.

Ao contextualizar esses resultados experimentais com a literatura recente, observa-se que o desempenho do classificador se equipara às abordagens mais avançadas da literatura, especialmente no rigor da avaliação em bases desbalanceadas. Trabalhos pioneiros de transferência de aprendizado, como o de Santos *et al.* (2017), reportaram sensibilidade superior a 90% utilizando redes pré-treinadas. Mais recentemente, a revisão de Zhang *et al.* (2023) e o estudo de Houssein *et al.* (2024) demonstraram que modelos convolucionais profundos aplicados a bases oncológicas alcançam sensibilidades na faixa de 95% a 97% para o diagnóstico global. A configuração com balanceamento sintético desenvolvida neste estudo obteve uma AUC macro OVR de 0,9265 e uma sensibilidade de 85,10% especificamente para o CEC, a classe minoritária e de maior dificuldade morfológica. Esse resultado não apenas confirma a competitividade técnica do modelo, mas, alinhando-se às diretrizes de Gouda *et al.* (2022), corrobora que a priorização de métricas por classe é fundamental para evitar a ilusão de uma acurácia global artificialmente elevada.

Apesar desses ganhos estatísticos em ambiente de laboratório, surge um questionamento crítico quanto à viabilidade e à segurança de se aplicar técnicas de aumento de dados ou geração sintética em aplicações de saúde no mundo real. Em sistemas de diagnóstico médico, a margem tolerável para incertezas deve ser mínima, uma vez que as decisões afetam diretamente o prognóstico e a vida dos pacientes. A geração artificial de imagens acarreta o risco inerente de produzir artefatos ou distorções morfológicas que não condizem com o comportamento biológico autêntico das lesões.

Consequentemente, treinar uma rede neural com dados artificiais pode levar o modelo a memorizar padrões clínicos irreais. Portanto, do ponto de vista ético e da segurança do paciente, não se considera adequado fundamentar um sistema de triagem definitiva em amostras sintéticas. A adoção do *data augmentation* e do SMOTE neste estudo deve ser interpretada estritamente como um recurso paliativo e exploratório para contornar limitações computacionais e de desbalanceamento do conjunto atual. O rigor científico exige que a solução definitiva para a escassez de dados não seja a geração sintética, mas sim a ampliação contínua da base com novas imagens reais, devidamente validadas por especialistas.

Essa necessidade crítica por imagens autênticas torna-se ainda mais evidente ao analisarmos as limitações clínicas enfrentadas pelo modelo. O principal desafio do classificador não consistiu em distinguir o melanoma, mas sim em separar de forma consistente o CBC e o CEC. Por pertencerem ao grupo dos carcinomas e compartilharem origens celulares próximas na epiderme, essas patologias apresentam padrões dermatoscópicos visualmente muito semelhantes. O alto índice de confusão entre essas duas categorias comprova, na prática, que a escassez de variabilidade morfológica original do CEC é um fator limitante intrínseco ao problema, que não pode ser solucionado puramente por algoritmos de geração sintética. Além desse obstáculo atrelado à natureza dos dados, a ausência de validação externa em outra base de imagens apresenta-se como uma limitação para atestar a capacidade de generalização do modelo frente a diferentes cenários e equipamentos clínicos.

Em suma, a principal contribuição experimental deste trabalho foi demonstrar que a organização prévia, o rigor no particionamento dos dados e as estratégias de balanceamento influenciam de forma direta a confiabilidade das predições em IA médica. O sistema desenvolvido apresenta inegável potencial como ferramenta de triagem e de segunda opinião diagnóstica (sistemas CAD), fornecendo um diagnóstico transparente sobre as barreiras impostas pelo desbalanceamento, embora ainda necessite de validação clínica com dados reais ampliados para transcender a etapa experimental e ser aplicado com segurança na área da saúde.

## 6 CONCLUSÃO

Este trabalho apresentou o desenvolvimento e a avaliação de uma abordagem computacional voltada ao apoio à análise de lesões cutâneas em imagens dermatoscópicas. A proposta concentrou-se na construção de um fluxo metodológico de ponta a ponta, capaz de organizar os dados, aplicar técnicas de pré-processamento, gerar máscaras de segmentação e estruturar as imagens para o treinamento de modelos de aprendizado profundo. Os cuidados adotados na divisão rigorosa dos conjuntos de treino, validação e teste demonstraram-se fundamentais para evitar o vazamento de dados e garantir estimativas de desempenho confiáveis.

Durante o desenvolvimento, estabeleceu-se uma base experimental sólida para a etapa de segmentação utilizando a arquitetura U-Net. A avaliação comprovou a viabilidade desta rede em destacar a ROI, o que se provou essencial para reduzir a influência de áreas externas à lesão, otimizando o processamento das etapas subsequentes. Contudo, observou-se que a segmentação ainda é sensível a imagens com baixo contraste e bordas pouco definidas, reforçando que imprecisões na delimitação espacial tendem a se propagar para o classificador, limitando a extração de características em casos complexos.

A configuração com balanceamento sintético mostrou-se mais adequada para o objetivo deste trabalho, pois melhorou o desempenho da classe CEC e permitiu uma avaliação mais equilibrada entre as três classes. Embora sua AUC macro tenha sido inferior à configuração inicial, o ganho em sensibilidade e F1-Score para CEC indica que o modelo passou a reconhecer melhor a classe anteriormente mais limitada.

Apesar desses ganhos estatísticos em ambiente de laboratório, surge um questionamento crítico quanto à viabilidade e à segurança de se aplicar técnicas de aumento de dados ou geração sintética em aplicações de saúde no mundo real. Em sistemas de diagnóstico médico, a margem tolerável para incertezas deve ser mínima, uma vez que as decisões afetam diretamente o prognóstico e a vida dos pacientes. A geração artificial de imagens acarreta o risco inerente de produzir artefatos ou distorções morfológicas que não condizem com o comportamento biológico autêntico das lesões.

Consequentemente, treinar uma rede neural com dados artificiais pode levar o modelo a memorizar padrões clínicos irreais. Portanto, do ponto de vista ético e da segurança do paciente, não se considera adequado fundamentar um sistema de triagem definitiva em amostras sintéticas. A adoção do *data augmentation* e do SMOTE neste estudo deve ser interpretada estritamente como um recurso paliativo e exploratório para contornar limitações computacionais e de desbalanceamento do conjunto atual. O rigor científico exige que a solução definitiva para a escassez de dados não seja a geração sintética, mas sim a ampliação contínua da base com novas imagens reais, validadas por especialistas.

Essa necessidade crítica por imagens autênticas torna-se ainda mais evidente ao analisarmos as limitações clínicas enfrentadas pelo modelo. O principal desafio do classificador não consistiu em distinguir o melanoma, mas sim em separar de forma consistente o CBC e o CEC, patologias que apresentam padrões dermatoscópicos visualmente muito semelhantes. O alto

índice de confusão entre essas duas categorias comprova, na prática, que a escassez de variabilidade morfológica original do CEC é um gargalo biológico que não pode ser solucionado puramente por algoritmos de geração sintética. Além desse obstáculo atrelado à natureza dos dados, a restrição de hardware e capacidade computacional apresentou-se como um limitador técnico para a execução de treinamentos e validações ainda mais exaustivas.

Portanto, a principal contribuição experimental deste projeto consistiu em demonstrar quantitativamente que a organização prévia, o particionamento dos dados e as estratégias de balanceamento influenciam de forma direta a confiabilidade das predições. O trabalho cumpriu o objetivo de entregar uma ferramenta de classificação estruturada, fornecendo simultaneamente um diagnóstico transparente sobre as barreiras impostas pelo desbalanceamento de dados no mundo real.

Como desdobramentos lógicos deste estudo, fundamentados nas limitações mapeadas e no potencial de expansão da tecnologia, propõe-se um conjunto de diretrizes para trabalhos futuros:

- **Estudos de Ablação no Pré-processamento:** Conduzir experimentos comparativos para avaliar quantitativamente o impacto de técnicas específicas, como mensurar o desempenho do classificador em conjuntos de dados processados com e sem a remoção de pelos (algoritmo *Dull Razor*), isolando a influência de artefatos na tomada de decisão da rede neural.
- **Expansão e Curadoria da Base de Dados:** Adquirir e integrar novas imagens clínicas reais, com foco absoluto na classe CEC. A ampliação do suporte estatístico minoritário é vital para fornecer à rede neural a diversidade autêntica de padrões patológicos, eliminando a dependência de técnicas de geração artificial de dados e garantindo a confiabilidade e a segurança ética exigidas para futuras aplicações em ambientes clínicos reais.
- **Integração Arquitetural e Novas Métricas:** Explorar a unificação da U-Net e da EfficientNet em uma arquitetura única de treinamento ponta a ponta, aliada à implementação de funções de perda otimizadas para classes minoritárias (como a *Focal Loss*), visando penalizar dinamicamente os erros cometidos na detecção das classes com menor suporte.
- **Avaliação Externa (*Cross-Dataset*):** Submeter o modelo final a bases de dados independentes (como as imagens da série HAM10000) para avaliar a resiliência da classificação frente a diferentes dispositivos de captura, calibrações de iluminação e biotipos de pele.
- **Implantação em Plataforma Móvel (*Edge AI*):** Visando a aplicabilidade clínica, sugere-se a otimização do fluxo computacional (através da quantização de pesos) para embarcá-lo em um aplicativo para *smartphones*. Esta transição viabilizará o uso

da ferramenta em ambientes de triagem, auxiliando profissionais de saúde com uma segunda opinião ágil e descentralizada.

Em suma, espera-se que a continuidade desta pesquisa e a implementação das propostas futuras consolidem o modelo não apenas como um experimento computacional, mas como uma ferramenta de real impacto clínico e social. Dessa forma, o trabalho reafirma o compromisso da engenharia em desenvolver soluções tecnológicas acessíveis, inovadoras e eficazes para os desafios da saúde pública.

## REFERÊNCIAS

- ABADI, M. *et al.* **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**, . 2015. Disponível em: <https://www.tensorflow.org/>.
- ACS. Web Page **Survival Rates for Melanoma Skin Cancer**, . 2025. Disponível em: <https://www.cancer.org/cancer/types/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin-cancer-by-stage.html>. Acesso em: 25 novembro 2025.
- ALENCAR, F. E. S. **Desenvolvimento de um sistema para classificação automática de imagens dermatoscópicas para dispositivos móveis**. 2015. Dissertação (Dissertação (Mestrado em Ciência da Computação)) — Universidade do Estado do Rio Grande do Norte, Biblioteca Central Orlando Teixeira, BR-RN 2015.
- AMARAL, G. V. M. Trabalhos de Conclusão de Curso (TCC) - Bacharelado - CIÊNCIA DA COMPUTAÇÃO- IC **Segmentação semântica de imagens dermatoscópicas de lesões de pele utilizando aprendizado profundo**, . 2023. Disponível em: <http://www.repositorio.ufal.br/jspui/handle/123456789/14168>. Acesso em: 25 novembro 2025.
- ARAÚJO, P. V. M. de; JUNIOR, G. B. Estudo comparativo de métodos de segmentação das bordas de lesões em imagens dermatoscópicas. **VI Encontro Acadêmico de Computação - UFMA**, , 2017. Disponível em: <https://petcompufma.org/eacomp/2017/anais%202017/NEWEstudo-comparativo-de-m%C3%A9todos-de-segmenta%C3%A7%C3%A3o.pdf>.
- BENEDETTI, J. **Structure and Function of the Skin**, . 2024. MSD Manual Consumer Version. Acesso em: 30 nov. 2025. Disponível em: <https://www.msmanuals.com/home/skin-disorders/biology-of-the-skin/structure-and-function-of-the-skin>.
- BOHJANEN, K. Estrutura e funções da pele. **Dermatologia Clínica. Seção I Bases para diagnóstico e tratamento**, , 2017.
- BRADFORD, P. T. Skin cancer in skin of color. **PubMed**, , 2009. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC2757062/>. Acesso em: 25 novembro 2025.
- BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, , 2000.
- BRASIL. **Diretrizes Diagnósticas e Terapêuticas do Melanoma Cutâneo**. Brasília: Ministério da Saúde, 2022. Acesso em: 29 nov. 2025. Disponível em: [https://www.gov.br/conitec/pt-br/midias/protocolos/ddt/20221109\\_ddt\\_melanoma\\_cutaneo.pdf](https://www.gov.br/conitec/pt-br/midias/protocolos/ddt/20221109_ddt_melanoma_cutaneo.pdf).
- BUDA, M.; MAKI, A.; MAZUROWSKI, M. A. A systematic study of the class imbalance problem in convolutional neural networks. **Neural Networks**, v. 106,, p. 249–259, 2018.
- BUSLAEV, A. *et al.* Alumentations: fast and flexible image augmentations. **Information**, Multidisciplinary Digital Publishing Institute v. 11, n. 2, p. 125, 2020.
- CARMO, G. C. do; SILVA, M. R. e. Dermoscopy: basic concepts. **PubMed**, , 2008. Disponível em: <https://doi.org/10.1111/j.1365-4632.2008.03556.x>. Acesso em: 25 novembro 2025.
- CAVALCANTE, M. C. TCC de Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo **SEGMENTAÇÃO DE IMAGENS DE LESÕES DE PELE USANDO A REDE NEURAL CONVOLUCIONAL U-NET**, . 2019.
- CHAWLA, N. V. *et al.* SMOTE: synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, v. 16,, p. 321–357, 2002.

CHOLLET, F. **Deep Learning with Python**. 2. ed. [S.l.]: Manning Shelter Island, 2018. ISBN 9781617294433.

CHOLLET, F. *et al.* **Keras**, . 2015. <https://keras.io>.

COELHO, T. S. **Implementação de algoritmos computacionais para interpretação de imagens dermatoscópicas para diagnóstico de tumores de pele**. 2016. 113 p. Dissertação (Mestrado) — Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo São Paulo 2016.

ELDER, D. E. *et al.* The 2018 world health organization classification of cutaneous, mucosal, and uveal melanoma: Detailed analysis of 9 distinct subtypes defined by their evolutionary pathway. **PubMed**, ,, 2020. Disponível em: <https://doi.org/10.5858/arpa.2019-0561-ra>. Acesso em: 25 novembro 2025.

Encyclopedia Britannica. **Human skin: cutaway view of the human skin**, . 2010. Acesso em: 30 nov. 2025. Disponível em: <https://www.britannica.com/science/human-skin>.

FERREIRA, B. E. S. TCC de Graduação em Ciência da Computação do Campus do Bacanga **Segmentação semântica de lesões de pele utilizando redes neurais convolucionais**, . 2018. Disponível em: <http://hdl.handle.net/123456789/3532>. Acesso em: 25 novembro 2025.

FIGUEIREDO, L. C. *et al.* Câncer de pele: estudo dos principais marcadores moleculares do melanoma cutâneo. **Revista Brasileira de Cancerologia**, v. 49, n. 3, p. 179–183, set. 2003. Disponível em: <https://rbc.inca.gov.br/index.php/revista/article/view/2098>.

GOPAL, A. *et al.* An automatic region based optimal segmentation and detection of features on dermoscopy images using v-shaped waterfall and water ridges. **Traitement du Signal**, v. 40, n. 2, p. 511–522, 2023.

GOUDA, W. *et al.* Detection of skin cancer based on skin lesion images using deep learning. **PubMed**, ,, 2022. Disponível em: <https://doi.org/10.3390/healthcare10071183>. Acesso em: 25 novembro 2025.

HALL, J. E. **Guyton & Hall: tratado de fisiologia médica**. 13. ed. Rio de Janeiro: Elsevier, 2017.

HARRIS, C. R. *et al.* Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357–362, 2020.

HERMES, G.; MOREIRA, F.; LIMA, J. Desenvolvimento de uma aplicação móvel multiplataforma para detecção e avaliação de risco de lesões cutâneas. *In*: ANAIS ESTENDIDOS DO XX SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO. 2024, Porto Alegre, RS, Brasil. **Anais [...]** Porto Alegre, RS, Brasil: SBC, 2024. p. 327–332. ISSN 0000-0000. Disponível em: [https://sol.sbc.org.br/index.php/sbsi\\_estendido/article/view/28635](https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/28635). Acesso em: 30 novembro 2025.

HORIUCHI, R. Monografia **BASE DE DADOS DE FOTOS CLÍNICAS DE LESÕES HIPERCROMICAS DE PELE COM DIAGNÓSTICO ANATOMOPATOLÓGICO**, . 2022. Disponível em: <https://acervodigital.ufpr.br/handle/1884/82137>. Acesso em: 25 novembro 2025.

HOUSSEIN, E. H. *et al.* An effective multiclass skin cancer classification approach based on deep convolutional neural network. **Cluster Computing**, v. 27,, p. 12799–12819, 2024.

HUNTER, J. D. Matplotlib: A 2d graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90–95, 2007.

IARC. **Skin cancer**. Lyon, 2024. Acesso em: 29 nov. 2025. Disponível em: <https://www.iarc.who.int/cancer-type/skin-cancer/>.

INCA. **Estimativa 2023: incidência de câncer no Brasil**. Rio de Janeiro: INCA, 2022. Acesso em: 30 nov. 2025. Disponível em: <https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil>.

JASMINE, F. *et al.* Molecular profiling and the interaction of somatic mutations with transcriptomic profiles in non-melanoma skin cancer (nm-sc) in a population exposed to arsenic. **PubMed**, ,, 2024. Disponível em: <https://doi.org/10.3390/cells13121056>. Acesso em: 25 novembro 2025.

JÚNIOR, A. P. d. S.; BEZERRA, D. G. d. S.; ANDRADE, Y. S. **Comparação de arquiteturas de Deep Learning para segmentação de imagens dermatoscópicas de melanoma**. 2020. Monografia (Trabalho de Conclusão de Curso) — Universidade de Brasília (UnB) Brasília 2020. Acesso em: 30 nov. 2025. Disponível em: <https://bdm.unb.br/handle/10483/27785>.

LENG, K.; VEDALDI, A. **Understanding image representations by measuring their equivariance and equivalence**, . 2015. Disponível em: <https://arxiv.org/abs/1411.5908>. Acesso em: 25 novembro 2025.

LEONE, J. Web Page **Redes Neurais e Zootecnia: Benefícios da Inteligência Artificial na Agropecuária**, . 2022. Disponível em: <https://nutrimosaic.com.br/redes-neurais-artificiais-aplicadas-a-zootecnia/>.

LOTFOLLAHI, Z. The anatomy, physiology and function of all skin layers and the impact of ageing on the skin. **Wound Practice and Research**, v. 32, n. 1, p. 6–10, 2024.

MAHMOUD, M. K. A.; AL-JUMAILY, A.; TAKRURI, M. The automatic identification of melanoma by wavelet and curvelet analysis: Study based on neural network classification. *In*: 2011 11TH INTERNATIONAL CONFERENCE ON HYBRID INTELLIGENT SYSTEMS (HIS). 2011. **Anais [...]** [S.l.: s.n.], 2011. p. 680–685.

MATEEN, M. *et al.* Hybrid deep learning framework for melanoma diagnosis using dermoscopic medical images. **PubMed**, ,, 2024. Disponível em: <https://doi.org/10.3390/diagnostics14192242>. Acesso em: 25 novembro 2025.

MCKINNEY, W. Data structures for statistical computing in python. *In*: WALT, S. van der; MILLMAN, J. (Ed.). PROCEEDINGS OF THE 9TH PYTHON IN SCIENCE CONFERENCE. 2010. **Anais [...]** [S.l.: s.n.], 2010. p. 56 – 61.

National Cancer Institute. **Melanoma Treatment (PDQ®)–Patient Version**, . 2019. Acesso em: 30 nov. 2025. Disponível em: <https://www.cancer.gov/types/skin/patient/melanoma-treatment-pdq>.

ONCOGUIA. Web Page **Estatística para câncer de pele melanoma**, . 2025. Disponível em: <https://www.oncoguia.org.br/conteudo/estatistia-para-cancer-de-pele-melanoma/7066/186/>. Acesso em: 25 novembro 2025.

ONU. **Transformando nosso mundo: a Agenda 2030 para o Desenvolvimento Sustentável**, . 2015. Nações Unidas Brasil. Acesso em: 3 mar. 2026. Disponível em: <https://brasil.un.org/pt-br/sdgs/3>.

ORCIONE, R. Web Page **From Melanoma Patient to Caregiver: Leah adams' story**, . 2025. Disponível em: <https://www.curemelanoma.org/blog/from-melanoma-patient-to-caregiver>. Acesso em: 25 novembro 2025.

PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in Python. **Journal of machine learning research**, v. 12, n. Oct, p. 2825–2830, 2011.

POPESCU, D.; EL-KHATIB, M.; ICHIM, L. Skin lesion classification using collective intelligence of multiple neural networks. **PubMed**, ,, 2022. Disponível em: <https://doi.org/10.3390/s22124399>. Acesso em: 25 novembro 2025.

RAZA, A. *et al.* Optimizing skin cancer screening with convolutional neural networks in smart healthcare systems. **PubMed**, ,, 2025. Disponível em: <https://doi.org/10.1371/journal.pone.0317181>. Acesso em: 25 novembro 2025.

RIVITTI, E. A. **Manual de dermatologia clínica de Sampaio e Rivitti**. SÃO PAULO: Artes Médicas, 2014. E-book.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional Networks for Biomedical Image Segmentation**, . 2015. Disponível em: <https://arxiv.org/abs/1505.04597>. Acesso em: 25 novembro 2025.

SALVAJOLI, J. V.; SOUHAMI, L.; FARIA, S. L. **Radioterapia em Oncologia**. Rio de Janeiro, São Paulo: Atheneu, 2021. Disponível em: [https://cdnv2.moovin.com.br/atheneu/imagens/files/manuais/1670\\_radioterapia-em-oncologia-3-edicao.pdf](https://cdnv2.moovin.com.br/atheneu/imagens/files/manuais/1670_radioterapia-em-oncologia-3-edicao.pdf). Acesso em: 25 novembro 2025.

SANTOS, A. *et al.* Uma abordagem de classificação de imagens dermatoscópicas utilizando aprendizado profundo com redes neurais convolucionais. *In*: ANAIS DO XXXVII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO (WIM - WORKSHOP DE INFORMÁTICA MÉDICA). 2017, São Paulo. **Anais [...]** São Paulo: SBC, 2017. p. 2010–2019. Acesso em: 30 nov. 2025.

SARRAFZADEH, O.; BAYGI, M. M.; GHASSEMI, P. Skin lesion detection in dermoscopy images using wavelet transform and morphology operations. **2010 17th Iranian Conference of Biomedical Engineering, ICBME 2010 - Proceedings**, ,, p. 1–8, 11 2010.

SATHE, P. R.; ZITO, P. M. **Basal Cell Carcinoma**. Treasure Island (FL): StatPearls Publishing, 2025. Acesso em: 30 nov. 2025. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK535374/>.

SBD. **Câncer da pele**, . 2020. Acesso em: 30 nov. 2025. Disponível em: <https://www.sbd.org.br/doencas/cancer-da-pele/>.

SIQUEIRA, H. C.; FERNANDES, B. J. T. Extração de características de sinais na pele para identificação de melanoma. *In*: Bastos Filho, C. J. A.; POZO, A. R.; LOPES, H. S. (Ed.). ANAIS DO 12 CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL. 2015, Curitiba, PR. **Anais [...]** Curitiba, PR: ABRICOM, 2015. p. 1–6.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**, . 2020. Disponível em: <https://arxiv.org/abs/1905.11946>. Acesso em: 25 novembro 2025.

The Skin Cancer Foundation. **Basal Cell Carcinoma Warning Signs**, . 2024. Acesso em: 30 nov. 2025. Disponível em: <https://www.skincancer.org/skin-cancer-information/basal-cell-carcinoma/warning-signs-images/>.

The Skin Cancer Foundation. **Squamous Cell Carcinoma Warning Signs**, . 2024. Acesso em: 30 nov. 2025. Disponível em: <https://www.skincancer.org/skin-cancer-information/squamous-cell-carcinoma/scc-warning-signs-images/>.

The Skin Cancer Foundation. **Types of Melanoma**, . 2024. Acesso em: 30 nov. 2025. Disponível em: <https://www.skincancer.org/skin-cancer-information/melanoma/types-of-melanoma/>.

VAZQUEZ, V. de L. **Atlas de Câncer de Pele**. 1. ed. SÃO PAULO: Hospital de Câncer de Barretos Barretos, 2019.

VIEIRA, S. C. Oncologia básica para profissionais de saúde. *In:* . 1. ed. Piauí: Editora da Universidade Federal do Piauí (EDUFPI), 2016. p. 91–98. E-book.

YOUSEF, H. *et al.* **Anatomy, Skin (Integument), Epidermis**. Treasure Island (FL): StatPearls Publishing, 2025. Acesso em: 30 nov. 2025. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK470464/>.

ZHANG, J. *et al.* Recent advancements and perspectives in the diagnosis of skin diseases using machine learning and deep learning: A review. **PubMed**, ,, 2023. Disponível em: <https://doi.org/10.3390/diagnostics13233506>. Acesso em: 25 novembro 2025.