

**UNIVERSIDADE FEDERAL DO PAMPA**

**EDUARDA MENEZES DA SILVEIRA**

**SUPORTE AO DIAGNÓSTICO DE  
LESÕES CUTÂNEAS UTILIZANDO  
INTELIGÊNCIA ARTIFICIAL**

**Bagé  
2025**

**EDUARDA MENEZES DA SILVEIRA**

**SUORTE AO DIAGNÓSTICO DE  
LESÕES CUTÂNEAS UTILIZANDO  
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia de Computação como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Sandro da Silva Camargo

**Bagé  
2025**

Ficha catalográfica elaborada automaticamente com os dados fornecidos  
pelo(a) autor(a) através do Módulo de Biblioteca do  
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

S587s Silveira, Eduarda Menezes da  
Suporte ao diagnóstico de lesões cutâneas utilizando  
inteligência artificial / Eduarda Menezes da Silveira.  
117 p.

Trabalho de Conclusão de Curso(Graduação)-- Universidade  
Federal do Pampa, ENGENHARIA DE COMPUTAÇÃO, 2025.  
"Orientação: Sandro da Silva Camargo".

1. detecção automatizada. 2. inteligência artificial. 3.  
redes neurais convolucionais. 4. tumores cutâneos. 5. YOLO. I.  
Título.

**EDUARDA MENEZES DA SILVEIRA**

**SUPORTE AO DIAGNÓSTICO DE LESÕES CUTÂNEAS UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso  
apresentado ao curso de Engenharia de  
Computação como requisito parcial  
para a obtenção do grau de Bacharel  
em Engenharia de Computação.

Dissertação defendida e aprovada em: 8 de dezembro de 2025.

Banca examinadora:

---

Prof. Dr. Sandro da Silva Camargo  
Orientador  
(UNIPAMPA)

---

Prof. Dr. Leonardo Bidese de Pinho  
(UNIPAMPA)

---

Prof. Dr. Julio Saraçol Domingues Jr.  
(UNIPAMPA)



Assinado eletronicamente por **SANDRO DA SILVA CAMARGO, PROFESSOR DO MAGISTERIO SUPERIOR**, em 18/12/2025, às 21:12, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **JULIO SARACOL DOMINGUES JUNIOR, PROFESSOR DO MAGISTERIO SUPERIOR**, em 19/12/2025, às 11:34, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.





Assinado eletronicamente por **LEONARDO BIDESE DE PINHO, PROFESSOR DO MAGISTERIO SUPERIOR**, em 19/12/2025, às 17:47, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

---



A autenticidade deste documento pode ser conferida no site [https://sei.unipampa.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1919672** e o código CRC **4843C7DA**.

---

Dedico este trabalho aos meus pais, pelo apoio constante e por sempre acreditarem no meu potencial.

## **AGRADECIMENTO**

A conclusão deste trabalho representa não apenas a realização de uma etapa acadêmica, mas também o resultado de uma jornada repleta de aprendizados, desafios, perdas e conquistas. Por isso, gostaria de expressar minha sincera gratidão a todos que, de alguma forma, contribuíram para que esta trajetória fosse possível ser concluída.

Agradeço ao meu orientador, Prof. Dr. Sandro da Silva Camargo, pela oportunidade deste trabalho, pela confiança, orientação atenta, pelos conselhos valiosos, pela paciência e pelo incentivo constante ao longo de todo o desenvolvimento deste trabalho. Sem o senhor este trabalho não seria possível ser realizado de forma concreta e assertiva.

Aos professores do curso de Engenharia de Computação, em especial ao professor Bruno Neves e ao professor Julio Saraçol, que foram fundamentais não apenas pelo conhecimento compartilhado, mas principalmente pelas palavras de apoio e incentivo nos momentos em que pensei que não conseguiria concluir. Eles acreditaram em mim antes mesmo que eu acreditasse, e isso fez toda a diferença.

Aos meus familiares, especialmente meus pais Ailton e Isabel, pelo incentivo inabalável e pelo apoio durante toda minha graduação.

As minhas amigas Gabriele e Hêmily que estiveram ao meu lado durante essa caminhada, compartilhando conhecimentos, dificuldades, conquistas e entendendo que nem sempre estaria disponível para elas.

Agradeço também a UNIPAMPA, pela oportunidade de concluir um sonho que achei que não seria possível.

Gostaria de também agradecer especialmente à NVIDIA, pelo generoso apoio por meio da concessão de 3000 horas de uso da GPU A100, que possibilitaram o treinamento e teste dos modelos de inteligência artificial aplicados neste trabalho.

Por fim, a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho, deixo aqui meu mais sincero agradecimento.

“A inteligência artificial não substituirá os médicos, mas os médicos que a utilizam substituirão os que não a usam.”

— Eric Topol

## RESUMO

O câncer de pele é a neoplasia mais frequente no Brasil e no mundo, sendo dividido entre o melanoma, mais agressivo, e os cânceres não melanoma, como o carcinoma basocelular (CBC) e o espinocelular (CEC). Diante das limitações dos métodos tradicionais de diagnóstico, como a inspeção clínica e a biópsia, a Inteligência Artificial (IA), especialmente com Redes Neurais Convolucionais (CNN), surge como alternativa promissora para a detecção precoce e precisa dessas lesões. Este trabalho propõe o desenvolvimento de uma solução baseada na arquitetura *You Only Look Once* (YOLO), capaz de identificar automaticamente diferentes tipos de lesões cutâneas malignas e benignas em imagens dermatoscópicas. Utilizando a base de dados *International Skin Imaging Collaboration* (ISIC Archive) 2019, composta por mais de 25 mil imagens dermatoscópicas rotuladas, o modelo foi treinado a partir de anotações realizadas manualmente com a ferramenta LabelImg, com base nos diagnósticos confirmados por biópsia, o que garantiu a qualidade e a padronização necessárias para o processo de treinamento supervisionado. Os resultados finais demonstraram que a versão YOLOv11 alcançou desempenho satisfatório, com alta acurácia e sensibilidade, tornando-se viável para uso clínico. Adicionalmente, o projeto contempla o desenvolvimento de uma interface *mobile* para facilitar o acesso ao sistema por profissionais da saúde. Também são discutidos aspectos éticos e regulatórios, como privacidade de dados, viés algorítmico e conformidade com exigências da ANVISA. Conclui-se que a integração entre IA e diagnóstico dermatológico representa um avanço significativo para a triagem eficiente e acessível do câncer de pele, sobretudo em regiões com escassez de especialistas.

**Palavras-chave:** Detecção automatizada; inteligência artificial; redes neurais convolucionais; tumores cutâneos; YOLO.

## ABSTRACT

Skin cancer is the most common neoplasm in Brazil and worldwide, and it is classified into melanoma, which is more aggressive, and non-melanoma skin cancers, such as basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Given the limitations of traditional diagnostic methods, such as clinical inspection and biopsy, Artificial Intelligence (AI), especially through Convolutional Neural Networks (CNNs), emerges as a promising alternative for the early and accurate detection of these lesions. This work proposes the development of a solution based on the You Only Look Once (YOLO) architecture, capable of automatically identifying different types of malignant and benign skin lesions in dermoscopic images. The model was trained using the International Skin Imaging Collaboration (ISIC Archive) 2019 dataset, which comprises more than 25,000 labeled dermoscopic images. Annotations were manually performed using the LabelImg tool, based on biopsy-confirmed diagnoses, ensuring the quality and standardization required for the supervised training process. The final results demonstrated that the YOLOv11 version achieved satisfactory performance, with high accuracy and sensitivity, making it viable for clinical use. Additionally, the project includes the development of a mobile interface to facilitate access to the system by healthcare professionals. Ethical and regulatory aspects are also discussed, such as data privacy, algorithmic bias, and compliance with ANVISA requirements. It is concluded that the integration of AI and dermatological diagnosis represents a significant advancement toward efficient and accessible skin cancer screening, especially in regions with limited access to specialists.

**Keywords:** automated detection; artificial intelligence; convolutional neural networks; cutaneous tumors; YOLO.

## LISTA DE FIGURAS

|           |                                                                                                                                          |    |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1  | Exemplos de imagens utilizadas no treinamento do modelo e suas respectivas informações clínicas, utilizadas no processo de anotação..... | 42 |
| Figura 2  | Anotação de imagem no LabelImg <sup>1</sup> .....                                                                                        | 45 |
| Figura 3  | Arquivo de anotação no formato .txt .....                                                                                                | 45 |
| Figura 4  | Pipeline de Detecção de Lesões Cutâneas com YOLOv11 <sup>3</sup> .....                                                                   | 47 |
| Figura 5  | Detecção em tempo real utilizando a câmera do aplicativo.....                                                                            | 53 |
| Figura 6  | Detecção em imagem estática utilizando a câmera do aplicativo.....                                                                       | 53 |
| Figura 7  | Execução local do modelo YOLO exibindo as detecções diretamente no ambiente <i>desktop</i> .....                                         | 56 |
| Figura 8  | Matriz de confusão obtida com o modelo YOLOv5 <sup>2</sup> .....                                                                         | 61 |
| Figura 9  | Curva de F1-score em função do limiar de confiança para o modelo YOLOv5 <sup>2</sup> .....                                               | 62 |
| Figura 10 | Curva de Precision em função do limiar de confiança para o modelo YOLOv5 <sup>2</sup> .....                                              | 63 |
| Figura 11 | Curva Precision-Recall por classe para o modelo YOLOv5 <sup>2</sup> .....                                                                | 64 |
| Figura 12 | Curva de Recall em função do limiar de confiança para o modelo YOLOv5 <sup>2</sup> .....                                                 | 65 |
| Figura 13 | Exemplos de detecções realizadas pelo modelo YOLOv5 <sup>2</sup> em imagens dermatoscópicas .....                                        | 66 |
| Figura 14 | Matriz de Confusão Inicial .....                                                                                                         | 68 |
| Figura 15 | Matriz de Confusão Após Redistribuição do Conjunto de Dados.....                                                                         | 70 |
| Figura 16 | Exemplos de detecção de lesões cutâneas realizada pelo modelo YOLOv11s <sup>3</sup> .....                                                | 71 |
| Figura 17 | Matriz de Confusão Final .....                                                                                                           | 74 |
| Figura 18 | Curva de <i>F1-Score</i> por classe .....                                                                                                | 75 |
| Figura 19 | Curva de Precisão por Classe .....                                                                                                       | 76 |
| Figura 20 | Curva de <i>Precision-Recall</i> por Classe .....                                                                                        | 77 |
| Figura 21 | Curva de <i>Recall</i> por Classe .....                                                                                                  | 78 |
| Figura 22 | Exemplos de detecções realizadas pelo modelo YOLOv11s <sup>3</sup> .....                                                                 | 79 |

## LISTA DE TABELAS

|          |                                                                                                   |    |
|----------|---------------------------------------------------------------------------------------------------|----|
| Tabela 1 | Comparação resumida de trabalhos sobre IA no diagnóstico de câncer de pele.....                   | 36 |
| Tabela 2 | Distribuição das imagens por classe de lesão cutânea na base de dados ISIC 2019.....              | 43 |
| Tabela 3 | Comparação entre os modelos YOLOv5 <sup>2</sup> e variantes da família YOLOv11 <sup>3</sup> ..... | 49 |
| Tabela 4 | Distribuição da quantidade de imagens por classe.....                                             | 60 |
| Tabela 5 | Redistribuição da quantidade de imagens por classe.....                                           | 69 |
| Tabela 6 | Distribuição final da quantidade de imagens por classe.....                                       | 73 |
| Tabela 7 | Comparação de desempenho entre YOLOv5 <sup>2</sup> e YOLOv11s <sup>3</sup> .....                  | 81 |



## LISTA DE ABREVIATURAS E SIGLAS

|              |                                                        |
|--------------|--------------------------------------------------------|
| ANVISA       | Agência Nacional de Vigilância Sanitária               |
| API          | <i>Application Programming Interface</i>               |
| AUC          | Área sob a Curva                                       |
| CBC          | Carcinoma Basocelular                                  |
| CEC          | Carcinoma Espinocelular                                |
| CNN          | Redes Neurais Convolucionais                           |
| CNS          | Conselho Nacional de Saúde                             |
| CEP          | Comitê de Ética em Pesquisa                            |
| DSLR         | <i>Digital Single-Lens Reflex</i>                      |
| FDA          | <i>Food and Drug Administration</i>                    |
| GDPR         | Regulamento Geral sobre a Proteção de Dados            |
| GPU          | <i>Graphics Processing Unit</i>                        |
| Grad-CAM     | <i>Gradient-weighted Class Activation Mapping</i>      |
| IA           | Inteligência Artificial                                |
| IDE          | <i>International Development Enterprises</i>           |
| INCA         | Instituto Nacional de Câncer                           |
| ISIC Archive | <i>International Skin Imaging Collaboration</i>        |
| JPG          | <i>Joint Photographic Experts Group</i>                |
| LGPD         | Lei Geral de Proteção de Dados Pessoais                |
| LIME         | <i>Local Interpretable Model-Agnostic Explanations</i> |
| mAP          | <i>mean Average Precision</i>                          |
| NVIDIA       | <i>Nvidia Corporation</i>                              |
| ROC          | Curva Característica de Operação do Receptor           |
| RSL          | Revisão Sistemática da Literatura                      |

|       |                                                    |
|-------|----------------------------------------------------|
| SIEPE | Salão Internacional de Ensino, Pesquisa e Extensão |
| SUS   | Sistema Único de Saúde                             |
| VPP   | Valor Preditivo Positivo                           |
| VPN   | Valor Preditivo Negativo                           |
| YOLO  | <i>You Only Look Once</i>                          |

## SUMÁRIO

|                                                              |           |
|--------------------------------------------------------------|-----------|
| <b>1 INTRODUÇÃO .....</b>                                    | <b>14</b> |
| 1.1 Objetivos .....                                          | 16        |
| 1.1.1 Objetivo Geral .....                                   | 16        |
| 1.1.2 Objetivos Específicos .....                            | 16        |
| 1.2 Organização do Trabalho .....                            | 17        |
| <b>2 REFERENCIAL TEÓRICO .....</b>                           | <b>18</b> |
| 2.1 Melanoma: Aspectos Clínicos e Epidemiológicos .....      | 18        |
| 2.1.1 Definição e Tipos de Câncer de Pele .....              | 18        |
| 2.1.2 Características do Melanoma .....                      | 19        |
| 2.1.3 Incidência, mortalidade e fatores de risco .....       | 19        |
| 2.1.4 Importância do diagnóstico precoce .....               | 20        |
| 2.2 Métodos Convencionais de Diagnóstico .....               | 21        |
| 2.2.1 Avaliação Clínica .....                                | 21        |
| 2.2.2 Biópsia e Análise Histopatológica .....                | 22        |
| 2.2.3 Limitações dos Métodos Tradicionais .....              | 23        |
| 2.3 Imagens de Lesões Cutâneas .....                         | 24        |
| 2.3.1 Tipos de Imagens Utilizadas .....                      | 25        |
| 2.3.2 Bases de Dados Públicas .....                          | 26        |
| 2.3.3 Desafios Associados às Imagens Médicas .....           | 27        |
| 2.4 Inteligência Artificial .....                            | 28        |
| 2.4.1 Histórico do uso de IA em Diagnóstico por Imagem ..... | 29        |
| 2.4.2 Vantagens e limitações da IA na Dermatologia .....     | 29        |
| 2.5 Deep Learning na Detecção de Melanoma .....              | 30        |
| 2.5.1 Fundamentos de Deep Learning .....                     | 30        |
| 2.5.2 Arquiteturas Comuns: CNN .....                         | 31        |
| 2.5.3 Estratégias de Treinamento .....                       | 32        |
| 2.5.4 Avaliação de Desempenho .....                          | 33        |
| 2.6 Trabalhos Correlatos .....                               | 34        |
| 2.7 Aspectos Éticos e Regulatórios .....                     | 37        |
| 2.7.1 Privacidade de Dados de Pacientes .....                | 37        |
| 2.7.2 Viés Algorítmico e Equidade .....                      | 38        |
| 2.7.3 Regulamentações para o Uso Clínico de IA .....         | 38        |
| <b>3 METODOLOGIA .....</b>                                   | <b>39</b> |
| 3.1 Classificação da Pesquisa .....                          | 39        |
| 3.2 Revisão de Literatura .....                              | 39        |
| 3.3 Etapas do Treinamento do Modelo .....                    | 41        |
| 3.4 Base de Dados .....                                      | 43        |
| 3.5 Anotação das Imagens .....                               | 44        |
| 3.6 Processo de Treinamento .....                            | 46        |
| 3.6.1 Arquiteturas Testadas .....                            | 47        |
| 3.6.2 Hiperparâmetros .....                                  | 49        |
| 3.7 Validação dos Resultados .....                           | 50        |
| 3.8 Ambiente de Desenvolvimento .....                        | 50        |
| 3.9 Desenvolvimento do Aplicativo .....                      | 52        |
| 3.10 Desenvolvimento Aplicação Local .....                   | 55        |
| 3.11 Aspectos Éticos .....                                   | 56        |
| <b>4 RESULTADOS E DISCUSSÕES .....</b>                       | <b>58</b> |
| 4.1 Abordagem Proposta .....                                 | 58        |

|              |                                                                         |           |
|--------------|-------------------------------------------------------------------------|-----------|
| <b>4.2</b>   | <b>Resultados Preliminares .....</b>                                    | <b>59</b> |
| <b>4.2.1</b> | <b>Resultados obtidos com o YOLOv5<sup>2</sup> .....</b>                | <b>59</b> |
| <b>4.2.2</b> | <b>Resultados Preliminares Obtidos com o YOLOv11s<sup>3</sup> .....</b> | <b>67</b> |
| <b>4.3</b>   | <b>Resultados Finais Obtidos com o YOLOv11s<sup>3</sup> .....</b>       | <b>72</b> |
| <b>4.3.1</b> | <b>Comparação entre YOLOv5<sup>2</sup> e YOLOv11s<sup>3</sup> .....</b> | <b>80</b> |
| <b>5</b>     | <b>CONSIDERAÇÕES FINAIS .....</b>                                       | <b>82</b> |
|              | <b>REFERÊNCIAS .....</b>                                                | <b>85</b> |
|              | <b>APÊNDICE A – ARTIGO .....</b>                                        | <b>93</b> |

## 1 INTRODUÇÃO

O câncer de pele é atualmente a neoplasia maligna mais frequente no Brasil e no mundo, representando cerca de 30% de todos os tumores malignos registrados no país (DELGADO *et al.*, 2024). Clinicamente, é classificado em dois grandes grupos: o câncer de pele não melanoma, que inclui o carcinoma basocelular (CBC) e o carcinoma espinocelular (CEC), e o melanoma, que, embora menos incidente, é o mais letal devido à sua elevada agressividade e potencial metastático (NETO *et al.*, 2020).

De acordo com o Instituto Nacional de Câncer (INCA), a taxa estimada de incidência de câncer de pele no Brasil gira em torno de 60 casos por 100 mil habitantes. Dentre os tumores não melanoma, o carcinoma basocelular é o mais comum, representando entre 70 a 75% dos casos, seguido pelo carcinoma espinocelular com 15 a 20% dos registros (NETO *et al.*, 2020). Ambos são geralmente localmente invasivos, com baixa mortalidade, mas alta taxa de recorrência. Já o melanoma, embora represente apenas cerca de 3% das neoplasias cutâneas, concentra a maior parte dos óbitos, dada sua alta taxa de invasividade (BORGES *et al.*, 2014).

Além desses, há outras lesões cutâneas de origem não neoplásica que também merecem atenção clínica, especialmente no contexto do diagnóstico diferencial. Os nevus, embora geralmente benignos, podem apresentar transformações malignas, especialmente quando exibem morfologia atípica, mostrada por alterações na dimensão ou coloração (BORGES *et al.*, 2014). As queratoses actínicas, por sua vez, são amplamente reconhecidas como lesões pré-cancerosas com potencial de evolução para CEC (THAMM; WELZEL; SCHUH, 2024), enquanto as queratoses benignas, apesar de inofensivas, muitas vezes mimetizam lesões malignas, exigindo um olhar clínico atento (SALAFRANCA *et al.*, 2022). Já os dermatofibromas e lesões vasculares, como hemangiomas, necessitam de distinção cuidadosa, dada sua semelhança clínica com neoplasias cutâneas, o que reforça a importância de um olhar clínico especializado (PICCOLO *et al.*, 2018).

A detecção precoce de lesões cutâneas potencialmente malignas é crucial para reduzir mortalidade e morbidade associadas ao câncer de pele. A dermatoscopia, técnica não invasiva que permite a visualização de estruturas subepidérmicas imperceptíveis a olho nu, tem se consolidado como elemento chave nessa estratégia (AZULAY; AZULAY; AZULAY, 2013). Desde seu desenvolvimento e aprimoramento no século XX, a dermatoscopia passou a integrar o rol de procedimentos essenciais na avaliação

dermatológica, elevando a acurácia diagnóstica do melanoma para cerca de 90% quando comparada à inspeção clínica convencional, que apresenta sensibilidade na faixa de 75–80% (BOTTON *et al.*, 2020).

Com os avanços tecnológicos, a dermatoscopia evoluiu para formatos digitais e videodermatoscópicos, integrando softwares que permitem armazenar e comparar imagens ao longo do tempo. No entanto, essas tecnologias ainda dependem crucialmente da experiência e treinamento do examinador, o que contribui para a variabilidade interobservador na interpretação das imagens (AZULAY; AZULAY; AZULAY, 2013). Além disso, a alta demanda por exames dermatológicos e a escassez de profissionais em regiões remotas intensificam as disparidades no acesso à detecção precoce, especialmente no âmbito do sistema público de saúde.

Nesse contexto, os métodos baseados em Inteligência Artificial (IA), especialmente aqueles que utilizam Redes Neurais Convolucionais (CNN), têm emergido como ferramentas promissoras para o reconhecimento automatizado de lesões cutâneas (HEINLEIN *et al.*, 2024). Tais algoritmos são capazes de aprender padrões complexos a partir de grandes conjuntos de imagens dermatoscópicas, alcançando níveis de desempenho comparáveis aos de dermatologistas experientes (SILVA *et al.*, 2021).

O uso de CNN, como no modelo *You Only Look Once* (YOLO), tem se destacado por sua capacidade de realizar detecção e classificação em tempo real com alta acurácia. Estudos recentes indicam que esses modelos podem contribuir significativamente para a triagem de lesões, priorizando casos suspeitos para avaliação médica e otimizando recursos de saúde (HEINLEIN *et al.*, 2024). Além disso, a integração entre IA e dermatoscopia pode minimizar erros de interpretação, reduzir a necessidade de biópsias desnecessárias e ampliar o acesso ao diagnóstico especializado, sobretudo em regiões remotas.

Diante disso, o trabalho apresentado visa desenvolver uma solução computacional baseada na arquitetura YOLO, capaz de detectar automaticamente diferentes tipos de câncer de pele a partir de imagens dermatológicas, contribuindo para o diagnóstico precoce da doença. Além do desenvolvimento e treinamento do modelo de detecção, este estudo propõe a criação de uma interface *mobile* intuitiva que permita a interação com o sistema de forma simples, acessível e eficaz, facilitando seu uso tanto por profissionais da saúde do Sistema Único de Saúde (SUS) bem como profissionais não especializados. A proposta busca, portanto, unir avanços em Inteligência Artificial e usabilidade tecnológica para oferecer uma ferramenta de apoio ao diagnóstico clínico, com o potencial de reduzir

erros, acelerar o processo de triagem e ampliar o acesso a exames de avaliação de lesões cutâneas.

## **1.1 Objetivos**

Esta pesquisa tem como objetivo explorar o uso de técnicas de aprendizado profundo aplicadas à detecção automática de lesões cutâneas em imagens dermatológicas. Considerando a elevada incidência do câncer de pele e a importância do diagnóstico precoce para o aumento das chances de tratamento eficaz, o trabalho busca propor uma solução computacional baseada em visão computacional capaz de auxiliar no processo de triagem e análise de lesões cutâneas, servindo como ferramenta de apoio ao diagnóstico clínico.

### **1.1.1 Objetivo Geral**

Desenvolver, testar e avaliar um modelo de detecção de lesões cutâneas baseado na arquitetura YOLO, com o intuito de auxiliar no diagnóstico precoce de câncer de pele a partir de imagens dermatológicas.

### **1.1.2 Objetivos Específicos**

- Coletar, organizar e preparar um banco de dados com imagens dermatológicas rotuladas, abrangendo diferentes tipos de lesões cutâneas, como melanoma, carcinoma basocelular, carcinoma espinocelular, nevos, queratoses e outras alterações relevantes, visando garantir a qualidade dos dados para o treinamento do modelo;
- Desenvolver e treinar um modelo de detecção automática de lesões cutâneas utilizando a arquitetura YOLO, uma abordagem baseada em redes neurais convolucionais que permite a identificação e classificação de múltiplos objetos em uma única imagem de forma rápida e precisa;
- Ajustar hiperparâmetros e realizar o treinamento do modelo YOLO com técnicas de aprendizado supervisionado, buscando maximizar a acurácia da detecção das lesões

e a capacidade de generalização do modelo frente a novos dados;

- Avaliar o desempenho do modelo treinado com métricas estatísticas apropriadas, como acurácia, a fim de validar a eficiência do sistema em um cenário clínico simulado;
- Desenvolver uma interface gráfica (*mobile*), que permita o envio de imagens dermatológicas e a visualização dos resultados de forma clara, promovendo acessibilidade e utilidade prática da ferramenta tanto em contextos educacionais quanto assistenciais.

## 1.2 Organização do Trabalho

O restante deste trabalho está organizado da seguinte forma: O Capítulo 2 aborda os principais conceitos relacionados ao câncer de pele, às técnicas de diagnóstico por imagem, às redes neurais convolucionais, à arquitetura YOLO e às aplicações recentes da inteligência artificial na área da saúde, com ênfase na dermatologia. O Capítulo 3 descreve as etapas adotadas para o desenvolvimento do projeto, incluindo a coleta e preparação dos dados, o treinamento do modelo, a escolha das ferramentas e ambientes computacionais, bem como os critérios utilizados para avaliação do desempenho da solução, além do desenvolvimento do aplicativo móvel que integra o modelo à interface de uso final. No Capítulo 4, são apresentados os resultados parciais obtidos, destacando-se as dificuldades enfrentadas, os ajustes realizados durante o desenvolvimento e as perspectivas para as etapas futuras do projeto, bem como os resultados finais alcançados após a implementação completa do sistema. No Capítulo 5 são apresentadas algumas considerações sobre as atividades realizadas e sobre os resultados preliminares, culminando com a conclusão do trabalho e a reflexão sobre suas contribuições e limitações.



## 2 REFERENCIAL TEÓRICO

Este capítulo apresenta uma revisão teórica dos principais conceitos abordados nesta pesquisa, com o objetivo de oferecer uma base para a compreensão do conteúdo técnico apresentado ao longo do trabalho. Os tópicos abordados abrangem desde a caracterização do melanoma e demais cânceres de pele, passando por aspectos epidemiológicos e importância do diagnóstico precoce, até a apresentação dos métodos de diagnóstico e dos tipos de imagens utilizadas, destacando suas vantagens e limitações.

### 2.1 Melanoma: Aspectos Clínicos e Epidemiológicos

O melanoma é considerado o tipo mais agressivo de câncer de pele, devido à sua elevada capacidade de invasão local e disseminação metastática, especialmente quando não diagnosticado precocemente (DELGADO *et al.*, 2024). A compreensão de seus aspectos clínicos e epidemiológicos é essencial para a identificação de fatores de risco, o aprimoramento do diagnóstico e a definição de estratégias terapêuticas eficazes. Esta seção aborda, inicialmente, os principais tipos de câncer de pele, com ênfase nas características clínicas e no potencial de malignidade de cada um, além de discutir lesões cutâneas benignas ou pré-malignas que podem apresentar semelhanças com tumores malignos e, portanto, exigem atenção no diagnóstico diferencial.

#### 2.1.1 Definição e Tipos de Câncer de Pele

Responsável por cerca de 30% de todos os tumores malignos diagnosticados no Brasil, o câncer de pele apresenta elevada incidência e relevância em saúde pública. Divide-se, majoritariamente, entre os tipos não melanoma, como o carcinoma basocelular (CBC) e o espinocelular (CEC), e o melanoma, que, apesar de menos prevalente, destaca-se pela agressividade e alta taxa de mortalidade associada à sua capacidade de invasão e metástase (DELGADO *et al.*, 2024).

Além dos principais tipos de câncer, existem outras lesões cutâneas que, embora benignas ou pré-malignas, exigem acompanhamento clínico regular. Os nevos melanocíticos são comuns e, na maioria das vezes, benignos, mas alterações em sua forma, cor ou tamanho podem indicar malignização para melanoma (BORGES *et al.*,

2014). As queratoses actínicas, por outro lado, são lesões pré-cancerígenas associadas ao carcinoma espinocelular e devem ser tratadas para prevenir progressão para malignidade (THAMM; WELZEL; SCHUH, 2024). Já as queratoses benignas, como a ceratose seborréica, são mais comuns, porém, sua aparência pode ser confundida com lesões malignas (SALAFRANCA *et al.*, 2022). Outras lesões, como dermatofibromas e lesões vasculares também devem ser monitoradas, pois sua semelhança com lesões malignas exige diagnóstico diferencial cuidadoso (PICCOLO *et al.*, 2018).

### **2.1.2 Características do Melanoma**

Apesar de representar apenas 3% dos tumores cutâneos malignos, o melanoma é considerado o mais agressivo e com maior potencial metastático. Originando-se dos melanócitos, células responsáveis pela pigmentação da pele, pode acometer a epiderme, derme e, em casos avançados, invadir tecidos profundos e órgãos distantes. Clinicamente, apresenta uma ampla variabilidade de cores, assimetria e bordas irregulares, evoluindo de maneira rápida quando não detectado e tratado precocemente (NETO *et al.*, 2020).

O melanoma pode surgir em qualquer região do corpo, sendo mais frequente no tronco em homens e nos membros inferiores em mulheres (WAINSTEIN *et al.*, 2020). Entre os principais subtipos clínico-patológicos, destacam-se o melanoma superficialmente extensivo, o nodular, o lentigo maligno e o melanoma lentiginoso acral, este último mais prevalente em pessoas de pele mais escura (WAINSTEIN *et al.*, 2020). O prognóstico da doença está diretamente associado à espessura tumoral (índice de Breslow), ao nível de invasão (nível de Clark), à presença de ulceração e ao envolvimento linfonodal, sendo o índice de Breslow o fator preditivo mais importante para sobrevida (BALCH *et al.*, 1978). O diagnóstico precoce, frequentemente auxiliado por exame clínico detalhado e dermatoscopia, é essencial para aumentar as chances de cura e reduzir o risco de metástases (WAINSTEIN *et al.*, 2020).

### **2.1.3 Incidência, mortalidade e fatores de risco**

Segundo dados do Instituto Nacional de Câncer (INCA), a incidência do câncer de pele tem aumentado significativamente nas últimas décadas, com uma taxa estimada de aproximadamente 60 casos para cada 100 mil habitantes (Instituto Nacional de Câncer

José Alencar Gomes da Silva, 2022). Essa tendência de crescimento também é observada em diversas regiões do mundo, atribuída, entre outros fatores, à maior exposição à radiação ultravioleta e ao envelhecimento da população (ARMSTRONG; KRICKER, 2021). O CBC é o mais prevalente, representando cerca de 70% a 75% dos casos, seguido pelo CEC, que corresponde a aproximadamente 15% a 20%. Apesar da baixa mortalidade associada a esses tipos, eles são localmente agressivos e apresentam alto risco de recorrência, o que demanda acompanhamento clínico constante (NETO *et al.*, 2020).

Em contraste, o melanoma destaca-se pela elevada taxa de mortalidade, decorrente do seu rápido potencial metastático, com cerca de 6.260 novos casos estimados no Brasil em 2018 (NETO *et al.*, 2020). A gravidade do melanoma reforça a importância do diagnóstico precoce, que pode aumentar significativamente as chances de cura e reduzir a mortalidade associada (GARBE; LEITER, 2020).

Os principais fatores de risco para o desenvolvimento do melanoma incluem exposição solar intensa e intermitente, especialmente em períodos prolongados sem proteção adequada, fenótipo claro, caracterizado pela menor quantidade de melanina na pele, o que aumenta a sensibilidade à radiação ultravioleta, histórico familiar da doença, presença de nevos atípicos e imunossupressão (GOYAL *et al.*, 2022). Outros fatores importantes são a idade avançada, sexo masculino, exposição ocupacional ao sol e histórico de queimaduras solares graves durante a infância, que elevam o risco de desenvolvimento da doença (GREEN; WHITEMAN, 2022).

Além do impacto na saúde individual, o câncer de pele representa um desafio para a saúde pública, devido à alta incidência e aos custos associados ao tratamento e acompanhamento dos pacientes. Estratégias de prevenção primária, como campanhas educativas e uso correto de proteção solar, são fundamentais para reduzir a ocorrência da doença (SCHMITT *et al.*, 2010).

#### **2.1.4 Importância do diagnóstico precoce**

A detecção precoce do melanoma é fundamental, pois aumenta consideravelmente as taxas de sobrevida e reduz a necessidade de tratamentos invasivos e de alto custo. Lesões detectadas em estágios iniciais apresentam taxa de cura superior a 90%, enquanto casos avançados têm prognóstico reservado e requerem abordagens terapêuticas mais complexas. Dessa maneira, estratégias eficazes de rastreamento e diagnóstico precoce,

como o exame clínico dermatológico, o uso da dermatoscopia, a fotografia digital de lesões e as campanhas de conscientização populacional, são cruciais para melhorar o desfecho dos casos e reduzir a mortalidade associada à doença (BORGES *et al.*, 2014).

## **2.2 Métodos Convencionais de Diagnóstico**

O diagnóstico precoce do câncer de pele representa uma etapa crítica para melhorar o prognóstico e aumentar a sobrevida dos pacientes, uma vez que lesões identificadas em estágios iniciais possuem maior chance de tratamento curativo. Os métodos convencionais de detecção incluem uma abordagem integrada que combina inspeção clínica detalhada, avaliação dermatoscópica, que permite a visualização de estruturas não perceptíveis a olho nu, e a confirmação diagnóstica por meio de biópsia seguida de exame histopatológico, considerados padrão-ouro no diagnóstico de malignidades cutâneas (American Academy of Dermatology, 2023). A integração desses métodos contribui para reduzir erros diagnósticos, possibilitando maior precisão na distinção entre lesões benignas, pré-malignas e malignas.

### **2.2.1 Avaliação Clínica**

A avaliação clínica inicial para diagnóstico de câncer de pele baseia-se principalmente na Regra ABCDE, que considera os aspectos de Assimetria (A), Bordas irregulares (B), Cor variada (C), Diâmetro (D) maior que 6 mm e Evolução (E) das lesões para identificar casos suspeitos. Essa técnica visa distinguir lesões benignas de malignas, especialmente o melanoma, e funciona como triagem para encaminhamento de casos suspeitos. Contudo, é importante destacar que a Regra ABCDE possui limitações, já que nem todas as lesões malignas apresentam todos esses critérios, o que exige cautela na interpretação (GOYAL *et al.*, 2022).

Além da inspeção visual, o exame físico completo e a coleta detalhada do histórico clínico do paciente incluindo antecedentes pessoais, familiares e exposição solar são fundamentais para uma avaliação precisa e contextualizada da lesão (GREEN; WHITEMAN, 2022). Complementarmente, a dermatoscopia, por sua vez, representa uma técnica não invasiva que aumenta a acurácia diagnóstica através da visualização ampliada e detalhada das estruturas pigmentadas e vasculares da pele. Estudos recentes destacam

que a associação entre exame clínico e dermatoscopia pode aumentar a sensibilidade para melanoma para mais de 90%, reduzindo falsos-negativos e contribuindo para diagnósticos mais precoces (AZULAY; AZULAY; AZULAY, 2013). Contudo, a interpretação dermatoscópica requer treinamento especializado, pois depende da experiência do profissional para diferenciar corretamente as características das lesões (KITTLER *et al.*, 2002).

Em casos de suspeita, outros métodos complementares podem ser empregados, como a microscopia confocal de reflectância e a biópsia cutânea, que confirmam o diagnóstico histopatológico e orientam o tratamento adequado (CARRERA; PUIG; MALVEHY, 2012). Assim, a avaliação clínica cuidadosa, aliada a técnicas diagnósticas avançadas, é fundamental para o diagnóstico precoce do câncer de pele, contribuindo para o aumento das chances de cura e redução da mortalidade associada à doença (SIEGEL *et al.*, 2023).

### **2.2.2 Biópsia e Análise Histopatológica**

Apesar dos avanços tecnológicos e dos métodos diagnósticos não invasivos, a biópsia com posterior exame histopatológico permanece como o padrão-ouro para o diagnóstico definitivo do câncer de pele. A técnica consiste na remoção parcial ou total da lesão para avaliação microscópica das células, permitindo verificar alterações estruturais e citopatológicas específicas, tais como a invasão da membrana basal e a presença de atipias celulares (DUARTE *et al.*, 2021).

A análise histopatológica fornece detalhes críticos para a definição do estágio da doença e planejamento terapêutico, sendo indispensável para caracterização precisa do tipo de câncer de pele, como carcinoma basocelular, carcinoma espinocelular ou melanoma. Além disso, esse exame é fundamental para determinar as margens cirúrgicas necessárias para a excisão completa da lesão (CHEN *et al.*, 2025).

Diversos tipos de biópsia podem ser realizados, como *punch*, incisional, excisional e *shave*, cada um com indicações específicas conforme o tamanho, localização e suspeita clínica da lesão (GOLDMAN; FITZPATRICK, 2019). Apesar de ser um procedimento invasivo, a biópsia é essencial para o diagnóstico definitivo, embora apresente riscos como cicatrizes e infecção. Por esse motivo, métodos não invasivos, como dermatoscopia e técnicas baseadas em inteligência artificial, têm sido desenvolvidos para auxiliar na triagem e avaliação inicial das lesões, mas ainda não substituem o exame histopatológico.

Além do diagnóstico, a análise histopatológica detalhada possibilita a subtipagem precisa do câncer de pele, avaliação do grau de agressividade e definição das margens cirúrgicas, aspectos fundamentais para o prognóstico e planejamento terapêutico (GOLDMAN; FITZPATRICK, 2019).

### 2.2.3 Limitações dos Métodos Tradicionais

Embora eficazes e amplamente estabelecidos, os métodos convencionais de diagnóstico não são isentos de limitações. A inspeção visual e a dermatoscopia são altamente dependentes da experiência e do conhecimento do profissional, o que pode resultar em variações subjetivas na avaliação. Mesmo quando realizados de forma criteriosa, esses métodos ainda exibem taxas significativas de falso-positivo e falso-negativo, o que pode comprometer a precisão do diagnóstico (MIHULECEA *et al.*, 2023).

Além disso, a demora no diagnóstico, ocasionada tanto pelas dificuldades de acesso a serviços especializados quanto pelos longos períodos de espera para a obtenção dos laudos anatomopatológicos, contribui para que muitos casos de câncer de pele sejam identificados em estágios mais avançados. Essa situação prejudica as chances de sucesso do tratamento e eleva os índices de morbidade e mortalidade associados à doença (American Academy of Dermatology, 2023).

Além das limitações mencionadas, a variabilidade interobservador entre profissionais qualificados representa um desafio adicional, pois diferentes especialistas podem interpretar de forma distinta os mesmos sinais clínicos e dermatoscópicos, afetando a confiabilidade do diagnóstico (GOLDMAN; FITZPATRICK, 2019). Essa subjetividade, aliada aos atrasos na confirmação histopatológica, pode levar a decisões clínicas inadequadas, incluindo biópsias desnecessárias ou, inversamente, o atraso no início do tratamento. Essas dificuldades impactam não apenas a qualidade do cuidado ao paciente, mas também geram custos adicionais para o sistema de saúde pública, ressaltando a necessidade urgente de métodos diagnósticos mais objetivos e eficientes (ROSENDAHL *et al.*, 2011).

Diante desse cenário, a integração de tecnologias emergentes, como a Inteligência Artificial, especialmente o aprendizado de máquina, um subcampo que permite a computadores aprenderem a partir de dados, surge como uma alternativa promissora para aumentar a eficiência, a precisão e a agilidade dos processos diagnósticos, contribuindo

para a detecção precoce e o manejo clínico mais eficaz do câncer de pele (BRINKER *et al.*, 2019).

### 2.3 Imagens de Lesões Cutâneas

A aquisição e o processamento de imagens de lesões cutâneas desempenham um papel central no desenvolvimento de sistemas de auxílio ao diagnóstico por computador. Essas imagens constituem a base fundamental para algoritmos de processamento de imagem para o treinamento de modelos de aprendizado de máquina. A qualidade, diversidade e representatividade do conjunto de dados utilizado são fatores críticos para assegurar tanto a eficiência quanto a capacidade de generalização dos modelos preditivos, impactando diretamente na acurácia e robustez do diagnóstico automatizado (TSCHANDL; ROSENDAHL; KITTLER, 2018).

Dentre os principais tipos de imagens utilizadas para esse fim, destacam-se a fotografia clínica, que registra a lesão em contexto natural; a dermatoscopia, que fornece imagens ampliadas e detalhadas da estrutura cutânea; e técnicas avançadas como a microscopia confocal, que possibilita análise em alta resolução da pele em tempo real. Cada modalidade apresenta suas vantagens e limitações, influenciando o desempenho dos modelos desenvolvidos (PACHECO; KROHLING, 2019).

Além disso, a qualidade da imagem pode ser afetada por diversos fatores, como condições de iluminação, foco inadequado, variações na pigmentação da pele e presença de artefatos, que podem comprometer a interpretação automatizada. Portanto, protocolos rigorosos de aquisição são fundamentais para garantir imagens consistentes e confiáveis (STIFF; BARATA *et al.*, 2021).

Outro aspecto essencial é a anotação e rotulagem precisa das imagens, geralmente realizada por especialistas dermatológicos, que assegura a qualidade dos dados e a confiabilidade dos modelos de aprendizado de máquina. Embora essas anotações sejam fundamentais para o desenvolvimento e treinamento dos algoritmos, é importante destacar que a biópsia permanece como o padrão-ouro para o diagnóstico definitivo das lesões cutâneas. Assim, os modelos baseados em inteligência artificial devem ser entendidos como ferramentas de apoio ao diagnóstico, e não como substitutos dos métodos tradicionais validados clinicamente (NASERI; SAFAEI, 2025).

Por fim, é fundamental abordar as questões éticas envolvidas no uso dessas imagens médicas, tais como o consentimento informado dos pacientes, anonimização

nominal e armazenamento seguro dos registros. Essas práticas são indispensáveis para garantir a privacidade dos pacientes e conformidade com normas legais e éticas (MANSO *et al.*, 2020).

### 2.3.1 Tipos de Imagens Utilizadas

As imagens de lesões cutâneas utilizadas no diagnóstico e acompanhamento clínico são obtidas, principalmente, por três modalidades:

- **Fotografias clínicas:** São imagens obtidas por câmeras digitais ou *smartphones*, geralmente capturadas sob iluminação natural ou em um ambiente controlado. Esse tipo de imagem oferece uma visão macro da lesão e é especialmente útil em situações de triagem, documentação e acompanhamento evolutivo (ESTEVA *et al.*, 2017).
- **Dermatoscopia:** Refere-se às imagens adquiridas por meio de dispositivos de dermatoscopia, que proporcionam aumento e iluminação da superfície cutânea, permitindo a visualização de padrões pigmentares e vasculares não perceptíveis a olho nu. Essa técnica contribui significativamente para o aumento da sensibilidade e especificidade na detecção do melanoma e de outras neoplasias cutâneas (AZULAY; AZULAY; AZULAY, 2013).
- **Histopatologia digital:** Consiste em imagens microscópicas obtidas após a realização de biópsia e processamento histológico da amostra. Essas imagens são fundamentais para o diagnóstico confirmatório e desempenham um papel crescente em pesquisas relacionadas à detecção e classificação do câncer de pele, especialmente quando associadas ao uso de algoritmos de visão computacional (HECKLER *et al.*, 2023).

Em geral, essas modalidades de imagem são complementares no fluxo diagnóstico: as fotografias clínicas são utilizadas na triagem inicial e na documentação das lesões; a dermatoscopia aprimora a análise morfológica ao revelar estruturas não visíveis a olho nu; e a histopatologia digital fornece o diagnóstico definitivo (AZULAY; AZULAY; AZULAY, 2013). Com os avanços tecnológicos, essas imagens vêm sendo integradas a bancos de dados, como o *ISIC Archive*, e a plataformas de inteligência artificial, permitindo o desenvolvimento de modelos computacionais capazes de auxiliar no diagnóstico e na classificação de lesões cutâneas (CODELLA *et al.*, 2018).



### 2.3.2 Bases de Dados Públicas

Com o avanço do aprendizado de máquina para diagnóstico médico, a disponibilização de bases de dados públicas contendo imagens de lesões cutâneas tornou-se crucial para o desenvolvimento, treinamento e validação de modelos computacionais. Esses repositórios permitem a padronização e o *benchmarking* de algoritmos, promovendo a comparação entre diferentes métodos e contribuindo para o avanço da área.

Dentre os principais repositórios disponíveis, destacam-se:

- **International Skin Imaging Collaboration (ISIC Archive):** Um dos maiores e mais completos repositórios de imagens dermatoscópicas, que reúne uma vasta diversidade de tipos de lesões e suas respectivas classificações, servindo como referência para estudos e competições na área (CODELLA *et al.*, 2018).
- **PH2:** Conjunto de imagens dermatoscópicas obtidas no Hospital Pedro Hispano, localizado em Matosinhos, Portugal. Este banco de dados é especialmente focado em imagens relacionadas ao diagnóstico de melanoma, oferecendo um material valioso para pesquisas clínicas e computacionais (MENDONÇA *et al.*, 2013).
- **Derm7pt:** Base de imagens anotadas com características específicas correspondentes a sete critérios dermatoscópicos. Esta coleção é utilizada para avaliação e treinamento de algoritmos voltados para a detecção de melanoma e carcinoma, possibilitando análises detalhadas de padrões diagnósticos (KAWAHARA *et al.*, 2019).
- **HAM10000:** Composta por 10.015 imagens dermatoscópicas, esta coleção abrange uma ampla variedade de tipos de câncer de pele e lesões benignas. É considerada uma das principais bases para o desenvolvimento e *benchmarking* de algoritmos automatizados de diagnóstico dermatológico (TSCHANDL; ROSENDAHL; KITTLER, 2018).

Essas bases possuem imagens predominantemente em formato *Joint Photographic Experts Group* (JPG) e alta resolução, adquiridas por meio de dermatoscópios digitais, o que assegura maior detalhamento das estruturas cutâneas. A diversidade dos dados, que inclui variações em etnia, tipos de pele, iluminação e condições clínicas, é essencial para o desenvolvimento de modelos robustos e capazes de generalizar para diferentes populações (ESTEVA *et al.*, 2017; TSCHANDL; ROSENDAHL; KITTLER, 2018).

O acesso a essas bases é geralmente público, embora algumas exijam cadastro ou termo de compromisso para garantir o uso ético dos dados. Apesar de sua relevância, essas coleções apresentam desafios, como o desbalanceamento entre classes de lesões, variações na qualidade das imagens e eventuais inconsistências nas anotações, o que demanda técnicas específicas de pré-processamento e validação cuidadosa durante o desenvolvimento dos modelos (CODELLA *et al.*, 2018).

Além de servirem para o treinamento de algoritmos, essas bases são amplamente utilizadas em competições e *benchmarks* internacionais, contribuindo para a evolução contínua dos métodos de diagnóstico assistido por computador (TSCHANDL; ROSENDAHL; KITTLER, 2018; BRINKER *et al.*, 2019).

### 2.3.3 Desafios Associados às Imagens Médicas

Apesar dos avanços nas técnicas de aquisição e da crescente disponibilização de bases públicas, o trabalho com imagens médicas ainda apresenta desafios significativos para o desenvolvimento de algoritmos robustos e generalizáveis. Entre as principais dificuldades destacam-se:

- **Variação de iluminação e artefatos visuais:** Reflexos de luz, presença de cabelos, bolhas de óleo e variações no tom de pele comprometem os processos automáticos de segmentação e classificação, dificultando a extração precisa de informações relevantes (GUTMAN *et al.*, 2016).
- **Desequilíbrio de classes:** A menor frequência de lesões malignas em relação às benignas nas bases de dados representa um desafio para o treinamento dos modelos, impactando negativamente métricas fundamentais como sensibilidade e especificidade (TSCHANDL; ROSENDAHL; KITTLER, 2018).
- **Heterogeneidade nas imagens:** Diferenças entre dispositivos de captura e variações entre centros médicos comprometem a consistência e a replicabilidade dos modelos de detecção e classificação, dificultando sua aplicação em contextos variados (KAWAHARA *et al.*, 2019).

Além dessas dificuldades, a qualidade e padronização das imagens, influenciadas por fatores como foco, resolução e condições ambientais durante a captura, representam desafios adicionais para a consistência dos dados (HERATH *et al.*, 2025). Ademais, o processo de anotação, que depende da avaliação especializada, é oneroso e pode

apresentar variações subjetivas, afetando diretamente a qualidade do treinamento dos modelos (VARDHAN; SAHOO, 2022). Por fim, questões éticas e regulatórias relativas à privacidade dos dados médicos restringem o acesso a bases amplas e diversificadas, limitando o desenvolvimento e a validação de algoritmos generalizáveis.

Esses desafios têm motivado pesquisas intensivas voltadas para o aprimoramento das etapas de pré-processamento e para o desenvolvimento de algoritmos de aprendizado profundo mais robustos e interpretáveis, ampliando as possibilidades de integração dos métodos automatizados à prática clínica.

## 2.4 Inteligência Artificial

O uso da Inteligência Artificial (IA) no diagnóstico médico tem apresentado crescimento exponencial, especialmente nas últimas décadas, impulsionado pelos avanços dos métodos computacionais e pela maior disponibilidade de grandes conjuntos de dados médicos. A IA pode ser entendida como um campo da ciência da computação que desenvolve sistemas capazes de realizar tarefas que, normalmente, requerem inteligência humana, como o reconhecimento de padrões, tomada de decisão e aprendizado a partir de dados (TOPOL, 2019).

Entre as principais técnicas de IA aplicadas no diagnóstico médico destacam-se o aprendizado de máquina (*machine learning*) e suas vertentes, como as redes neurais artificiais e a aprendizagem profunda (*deep learning*). Essas técnicas permitem que os sistemas aprendam com grandes volumes de dados clínicos e imagens, identificando características sutis que muitas vezes escapam à percepção humana (CHEN *et al.*, 2021).

Apesar dos avanços, a aplicação da IA na área da saúde enfrenta desafios importantes, como a necessidade de grandes bases de dados devidamente anotadas, questões relacionadas à interpretabilidade dos modelos, possíveis vieses presentes nos dados de treinamento e preocupações quanto à privacidade e segurança dos dados dos pacientes (NAQVI *et al.*, 2023).

No contexto específico do diagnóstico de câncer de pele, a IA tem se mostrado promissora ao auxiliar na detecção precoce e classificação das lesões cutâneas, contribuindo para maior rapidez e precisão nos resultados, além de possibilitar a escalabilidade do atendimento, especialmente em regiões com escassez de especialistas (MRIDHA *et al.*, 2023).

Dessa forma, a inteligência artificial representa uma importante ferramenta

complementar à prática clínica tradicional, oferecendo potencial para aprimorar o diagnóstico, reduzir erros e otimizar os recursos do sistema de saúde (TOPOL, 2019).

#### **2.4.1 Histórico do uso de IA em Diagnóstico por Imagem**

As primeiras aplicações da IA no diagnóstico médico surgiram na década de 1960, inicialmente por meio de sistemas baseados em regras e em árvores de decisão. Contudo, foi com o avanço do aprendizado de máquina e, mais recentemente do aprendizado profundo (*deep learning*), que as metodologias ganharam relevância, principalmente na análise de imagens médicas (MIHULECEA *et al.*, 2023). Na dermatologia, essas técnicas têm sido utilizadas na classificação de lesões cutâneas e, em determinadas tarefas, já demonstraram desempenho superior ao de dermatologistas na acurácia diagnóstica (BRINKER *et al.*, 2019).

Além da dermatologia, a Inteligência Artificial tem se destacado em outras especialidades que utilizam imagens médicas, como radiologia, oftalmologia e patologia digital, contribuindo para maior precisão, agilidade e alcance no diagnóstico (SUETENS, 2017). Apesar dos avanços, os primeiros sistemas enfrentavam limitações importantes, como a baixa capacidade computacional da época e a escassez de dados clínicos estruturados, fatores que só começaram a ser superados com o desenvolvimento tecnológico das últimas décadas (AVANZO *et al.*, 2024).

#### **2.4.2 Vantagens e limitações da IA na Dermatologia**

Entre as principais vantagens da IA na dermatologia destacam-se a elevada acurácia na classificação de imagens, a possibilidade de ampliação do acesso ao diagnóstico especializado em regiões remotas e a significativa redução do tempo para emissão de laudos (TOPOL, 2019). Além disso, a utilização de algoritmos de IA pode contribuir para maior padronização na avaliação de lesões cutâneas, reduzindo a variabilidade entre diferentes observadores (ESTEVA *et al.*, 2017).

Por outro lado, os sistemas baseados em IA apresentam limitações importantes, como o risco de viés associado ao conjunto de dados utilizado no treinamento dos modelos, o que pode comprometer o desempenho em populações sub-representadas, a limitada interpretabilidade das decisões dos algoritmos, o que dificulta a compreensão

de seus resultados por profissionais de saúde, e a necessidade de validação clínica rigorosa antes de sua ampla adoção na prática médica (OMIYE *et al.*, 2023). Além dessas limitações, destacam-se os desafios relacionados à segurança e à privacidade dos dados dos pacientes (NAGENDRAN *et al.*, 2020), bem como a necessidade de que os sistemas de IA sejam utilizados como ferramentas de apoio ao diagnóstico, preservando a autonomia e o julgamento clínico dos profissionais de saúde (TOPOL, 2019). Também é fundamental que os modelos sejam continuamente atualizados para incorporar novos dados e garantir a generalização adequada em diferentes populações (LIU *et al.*, 2020).

## 2.5 Deep Learning na Detecção de Melanoma

O uso de técnicas de *deep learning* tem se destacado como uma abordagem promissora no apoio ao diagnóstico precoce do melanoma, uma das formas mais agressivas de câncer de pele. Por meio de algoritmos capazes de aprender representações complexas a partir de grandes volumes de dados, essas técnicas permitem a identificação automatizada de padrões em imagens médicas com elevada acurácia (ESTEVA *et al.*, 2017; LITJENS *et al.*, 2017). Em particular, as redes neurais convolucionais (CNN) têm sido amplamente utilizadas em tarefas de classificação e detecção de lesões cutâneas, contribuindo significativamente para a redução de erros diagnósticos e para a melhoria da tomada de decisão clínica (BRINKER *et al.*, 2019; MUNJAL *et al.*, 2024). Esta seção apresenta os fundamentos do *deep learning* aplicados ao contexto dermatológico, as principais arquiteturas empregadas, estratégias de treinamento, métodos de avaliação e uma revisão de trabalhos correlatos, além de discutir aspectos éticos e regulatórios envolvidos em sua aplicação (CHU *et al.*, 2020; TOPOL, 2019).

### 2.5.1 Fundamentos de Deep Learning

O *deep learning* fundamenta-se em arquiteturas de redes neurais profundas compostas por múltiplas camadas de perceptrons, que são capazes de aprender representações hierárquicas e complexas dos dados. No contexto do diagnóstico por imagem, as Redes Neurais Convolucionais (CNN) destacam-se como as mais utilizadas, uma vez que conseguem captar características espaciais essenciais para a identificação de lesões, como o melanoma. Essas redes aprendem automaticamente

padrões discriminativos relevantes, eliminando a necessidade da extração manual de características (LECUN; BENGIO; HINTON, 2015).

As CNN operam por meio de camadas convolucionais que extraem características locais das imagens, seguidas de camadas de *pooling* que reduzem a dimensionalidade, facilitando a detecção de padrões relevantes. A qualidade e diversidade dos dados de treinamento são igualmente importantes para garantir que o modelo generalize bem para diferentes populações e condições clínicas. Além disso, técnicas como transferência de aprendizado (*transfer learning*) permitem aproveitar modelos pré-treinados em grandes bases de dados para acelerar o desenvolvimento e melhorar a performance em tarefas específicas, especialmente quando há limitação no número de imagens rotuladas (TAN; LE, 2019).

O treinamento supervisionado, que utiliza grandes volumes de imagens previamente rotuladas, é crucial para que o modelo adquira a capacidade de distinguir corretamente entre lesões malignas e benignas (ESTEVA *et al.*, 2017). Entretanto, um desafio significativo nesse processo é o *overfitting*, ou sobreajuste, fenômeno que ocorre quando o modelo se adapta excessivamente aos dados de treinamento, comprometendo sua habilidade de generalização para novos casos (GOODFELLOW; BENGIO; COURVILLE, 2016).

Para minimizar esse problema, diversas técnicas são empregadas, tais como regularização, *dropout*, aumento de dados (*data augmentation*) e validação cruzada (SHORTEN; KHOSHGOFTAAR, 2019). Além disso, o avanço tecnológico, especialmente uso de *Graphics Processing Unit* (GPU) e ambientes de computação em nuvem, tem possibilitado o treinamento eficiente dessas redes profundas em larga escala (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

Esses avanços contribuem para a automação e eficiência do diagnóstico médico, promovendo maior acurácia, redução de erros humanos e agilidade na tomada de decisão clínica (LITJENS *et al.*, 2017).

## 2.5.2 Arquiteturas Comuns: CNN

As redes neurais convolucionais figuram entre as arquiteturas mais utilizadas na classificação de imagens médicas, devido à sua capacidade de extrair automaticamente características relevantes em diferentes escalas. Modelos consagrados como VGG, ResNet e *Inception* apresentaram excelente desempenho em *benchmarks* para detecção de

melanoma, demonstrando robustez e alta acurácia em cenários clínicos e experimentais (BRINKER *et al.*, 2019). Contudo, apesar desse desempenho, essas arquiteturas são tradicionalmente aplicadas a tarefas de classificação ou segmentação, não sendo otimizadas para detecção rápida e em tempo real, característica essencial em sistemas clínicos que demandam agilidade no diagnóstico (REDMON; FARHADI, 2018). Por essa razão, optou-se pela utilização do modelo YOLO, que integra simultaneamente a localização e a classificação em uma única etapa, oferecendo um excelente equilíbrio entre velocidade e precisão, além de ser computacionalmente eficiente e adequado para aplicações em dispositivos com recursos limitados (REDMON; FARHADI, 2018).

Arquiteturas mais recentes, como a *EfficientNet*, buscam otimizar o equilíbrio entre acurácia e eficiência computacional, reduzindo o custo de processamento sem comprometer o desempenho (TAN; LE, 2019). Além das CNN, abordagens baseadas em *Transformers*, incluindo os *Vision Transformers* (ViT), vêm sendo exploradas em tarefas de classificação de imagens médicas, apresentando resultados promissores e ampliando as possibilidades para aplicações em diagnóstico assistido por inteligência artificial (DOSOVITSKIY *et al.*, 2021).

Essas arquiteturas são empregadas principalmente na análise de imagens de dermatoscopia, histopatologia e fotografias clínicas, contribuindo para o diagnóstico precoce de lesões cutâneas (LIU *et al.*, 2020). O desempenho dos modelos costuma ser avaliado por métricas como acurácia, sensibilidade, especificidade e área sob a curva ROC (Curva Característica de Operação do Receptor) e AUC (Área sob a Curva) (AUC-ROC), fundamentais para validar sua aplicação no contexto médico (TSCHANDL; ROSENDAHL; KITTLER, 2018).

Apesar do elevado desempenho reportado, esses modelos enfrentam limitações importantes. Entre os principais desafios destacam-se a necessidade de grandes volumes de dados anotados para o treinamento eficaz, o risco de sobreajuste em conjuntos de dados pequenos e a dificuldade de interpretação dos resultados pelos profissionais da saúde, especialmente em arquiteturas baseadas em *Transformers*, que apresentam menor transparência na tomada de decisão (DOSOVITSKIY *et al.*, 2021).

### 2.5.3 Estratégias de Treinamento

Para superar limitações como a escassez de dados médicos anotados, são amplamente empregadas estratégias como *data augmentation* (geração de variações

artificiais das imagens para aumentar a diversidade do conjunto de dados), *transfer learning* (utilização de modelos previamente treinados em grandes bases de dados para acelerar o aprendizado) e *fine-tuning* (ajuste fino dos parâmetros do modelo para melhor adaptação ao problema específico) (TSCHANDL; ROSENDAHL; KITTLER, 2018). Essas abordagens contribuem para melhorar a capacidade de generalização do modelo, reduzir o risco de *overfitting* e aumentar a robustez do sistema em aplicações clínicas (TSCHANDL; ROSENDAHL; KITTLER, 2018). Além disso, práticas complementares, como *cross-validation* e o balanceamento das classes no conjunto de treinamento, também são recomendadas na literatura para assegurar maior confiabilidade dos resultados e melhor desempenho em cenários com dados desbalanceados, característicos de bancos de imagens médicas (BRINKER *et al.*, 2019).

#### 2.5.4 Avaliação de Desempenho

A avaliação dos modelos de detecção de melanoma baseia-se em métricas como acurácia, sensibilidade, especificidade e área sob a curva ROC e AUC (AUC-ROC). Cada uma dessas métricas contribui para uma análise mais completa do desempenho dos algoritmos em diferentes cenários clínicos. A acurácia reflete a proporção total de predições corretas em relação ao número de casos avaliados; a sensibilidade (ou *recall*) mede a capacidade do modelo em identificar corretamente os casos positivos, enquanto a especificidade avalia sua habilidade em reconhecer corretamente os casos negativos. Entre essas métricas, a AUC-ROC se destaca como um dos principais indicadores de desempenho em classificadores binários, por fornecer uma medida robusta da capacidade discriminativa do modelo ao longo de diferentes limiares de decisão (LITJENS *et al.*, 2017).

Essas métricas devem ser analisadas de forma conjunta, uma vez que, isoladamente, podem não refletir todos os aspectos do desempenho em cenários clínicos reais. Em estratégias de triagem populacional, costuma-se priorizar modelos com alta sensibilidade, para minimizar o risco de falsos negativos e garantir a detecção precoce dos casos (BRINKER *et al.*, 2019). Já em contextos de confirmação diagnóstica, a especificidade torna-se mais relevante, buscando reduzir o número de falsos positivos e evitar procedimentos desnecessários (HAENSCHEN; AL., 2021). Além disso, métricas como o valor preditivo positivo (VPP) e o valor preditivo negativo (VPN) são amplamente utilizadas em aplicações médicas por expressarem a probabilidade de um resultado estar



correto, dado o teste realizado (WU *et al.*, 2020). Em situações de desbalanceamento entre as classes, como é frequente na detecção de melanoma, a análise da curva de precisão-*recall* pode fornecer informações adicionais relevantes para avaliação do modelo (ASHRAF *et al.*, 2020).

## 2.6 Trabalhos Correlatos

A aplicação da IA no diagnóstico precoce do câncer de pele tem sido amplamente investigada na literatura científica recente, refletindo o crescente interesse no uso de técnicas computacionais para apoiar a prática clínica dermatológica. Avanços significativos têm sido observados, especialmente com a adoção de modelos de aprendizado profundo baseados em redes neurais convolucionais, capazes de extrair automaticamente características relevantes de imagens dermatoscópicas e alcançar níveis de desempenho comparáveis aos de especialistas humanos. Esses resultados demonstram o potencial da IA como ferramenta de apoio à triagem e à tomada de decisão clínica, contribuindo para a detecção precoce de lesões malignas. Contudo, apesar dos progressos alcançados, ainda persistem desafios relevantes, como a variabilidade na qualidade das imagens, o desbalanceamento entre classes de lesões, a necessidade de bases de dados amplas e bem anotadas, além de questões relacionadas à interpretabilidade dos modelos e à validação clínica em cenários reais, fatores que limitam a adoção ampla dessas soluções na prática médica.

Phillips *et al.* (2019) realizaram um estudo pioneiro em sete hospitais do Reino Unido, avaliando um algoritmo de IA para a detecção de melanoma a partir de 1.550 imagens dermatoscópicas obtidas por diferentes dispositivos, como *smartphones* e câmeras *Digital Single Lens Reflex* (DSLR). Os resultados indicaram que o algoritmo alcançou desempenho comparável ou superior ao de dermatologistas especialistas, com uma área sob a curva ROC (AUC) de até 95,8%. Mesmo mantendo sensibilidade máxima, o modelo conseguiu preservar uma especificidade considerável, ressaltando seu potencial para triagem clínica em contextos com acesso restrito a profissionais especializados.

No campo da inteligência artificial explicável, um estudo recente conduzido por Munjal *et al.* (2024) propôs o *SkinSage XAI*, uma solução baseada em aprendizado profundo que utiliza a arquitetura *Inception v3* para classificar sete tipos diferentes de lesões cutâneas, incluindo melanoma, nevo melanocítico e carcinoma basocelular. Com alta acurácia de 96%, o trabalho enfatizou a importância da interpretabilidade

do modelo para a confiança clínica, aplicando técnicas como *Gradient-weighted Class Activation Mapping* (Grad-CAM) e *Local Interpretable Model-Agnostic Explanations* (LIME) para tornar as decisões do sistema compreensíveis para dermatologistas. Essa abordagem complementa as estratégias tradicionais de detecção automática por redes neurais, promovendo maior transparência na aplicação da IA.

Complementando essas pesquisas, Cui *et al.* (2019) realizaram uma análise comparativa entre métodos tradicionais de aprendizado de máquina e modelos de redes neurais convolucionais para classificação de imagens dermatoscópicas. Os autores observaram que, embora os métodos tradicionais apresentem bom desempenho em conjuntos de dados menores, as redes profundas, especialmente o *Inception V3*, exibem maior acurácia e sensibilidade, reforçando a relevância das técnicas avançadas para o aprimoramento do diagnóstico precoce do melanoma.

Em uma revisão sistemática, Jairath *et al.* (2024) investigaram o uso da IA para o diagnóstico de câncer de pele, com foco no melanoma e nos carcinomas basocelular e espinocelular. Os resultados indicaram que os modelos atuais oferecem alta acurácia, sensibilidade e especificidade na detecção dessas lesões. Contudo, os autores destacaram limitações importantes, como a escassez de validação externa robusta e a sub-representação de determinados tipos de pele e lesões nos conjuntos de dados disponíveis, evidenciando a necessidade de estudos futuros que priorizem a diversidade e a colaboração multidisciplinar para o desenvolvimento de sistemas mais precisos e equitativos.

Além da análise de imagens, Wang *et al.* (2019) exploraram o uso de redes neurais profundas para integrar dados clínicos e históricos médicos na predição do risco de câncer de pele não melanoma. Esse modelo multifacetado alcançou alta sensibilidade e especificidade, demonstrando que a combinação de múltiplas fontes de informação pode aprimorar a acurácia diagnóstica e contribuir para a prática clínica de forma complementar às técnicas baseadas exclusivamente em imagens.

Por fim, a revisão realizada por Chu *et al.* (2020), publicada na *Frontiers in Medicine*, destaca que as redes neurais convolucionais vêm superando especialistas humanos na classificação de lesões malignas por imagens dermatoscópicas. O estudo também enfatiza desafios como a necessidade de validação clínica rigorosa e questões éticas no uso da IA em dermatologia, reafirmando a importância da inteligência artificial como uma ferramenta complementar para o diagnóstico precoce do câncer de pele, especialmente em regiões com acesso limitado a dermatologistas.

Na Tabela 1 apresenta-se uma comparação resumida de trabalhos correlatos:

Tabela 1 – Comparação resumida de trabalhos sobre IA no diagnóstico de câncer de pele

| Autores                       | Objetivo                            | Técnica                              | Base de Dados            | Resultados e Limitações                       |
|-------------------------------|-------------------------------------|--------------------------------------|--------------------------|-----------------------------------------------|
| Phillips <i>et al.</i> (2019) | Detectar melanoma                   | IA em imagens dermatoscópicas        | 1.550 imagens            | AUC 95,8%, desempenho similar a especialistas |
| Munjal <i>et al.</i> (2024)   | IA explicável para lesões           | Inception v3 + XAI                   | 7 tipos de lesões        | Acurácia 96%, foco em interpretabilidade      |
| Cui <i>et al.</i> (2019)      | Comparar métodos tradicionais e CNN | Inception v3 e técnicas tradicionais | Imagens variadas         | CNN melhoram acurácia e sensibilidade         |
| Jairath <i>et al.</i> (2024)  | Revisão sistemática da IA           | Vários modelos                       | Revisão de estudos       | Alta acurácia; falta validação externa        |
| Wang <i>et al.</i> (2019)     | Integrar dados clínicos e imagens   | Redes profundas multifacetadas       | Dados clínicos e imagens | Alta sensibilidade e especificidade           |
| Chu <i>et al.</i> (2020)      | Revisão das CNN em dermatologia     | Redes neurais convolucionais         | Literatura da área       | CNN superam especialistas; desafios éticos    |

Fonte: Autor (2025).

A análise comparativa apresentada na Tabela 1 evidencia que os trabalhos correlatos convergem para a utilização de técnicas de deep learning, sobretudo Redes Neurais Convolucionais, como principal abordagem para o diagnóstico automatizado de lesões cutâneas. Observa-se que a maioria dos estudos reporta elevados índices de desempenho, com valores de acurácia, sensibilidade e AUC próximos ou superiores aos obtidos por especialistas humanos, reforçando o potencial da Inteligência Artificial como ferramenta de apoio ao diagnóstico dermatológico. Entretanto, também se destacam limitações recorrentes, como a dependência de bases de dados específicas, a escassez

de validação externa e a baixa diversidade de tipos de pele e lesões representadas nos conjuntos de treinamento. Além disso, nota-se uma tendência recente à incorporação de técnicas de IA explicável, buscando aumentar a transparência e a confiabilidade clínica dos modelos. Nesse contexto, os estudos analisados fundamentam a relevância de abordagens baseadas em detecção automática de lesões, como as arquiteturas da família YOLO, que se diferenciam por possibilitar a localização e classificação simultânea das lesões, contribuindo para aplicações práticas em cenários reais de triagem e apoio à decisão clínica.

## **2.7 Aspectos Éticos e Regulatórios**

A aplicação da inteligência artificial na área da saúde, especialmente no diagnóstico de câncer de pele, envolve uma série de considerações éticas e regulatórias que são fundamentais para garantir a segurança, a privacidade e a equidade no atendimento aos pacientes. Estes aspectos devem ser analisados com cautela, dado o potencial impacto direto dessas tecnologias em decisões clínicas (TOPOL, 2019).

### **2.7.1 Privacidade de Dados de Pacientes**

A privacidade e a proteção dos dados sensíveis dos pacientes são pilares essenciais no desenvolvimento e uso de sistemas baseados em IA. Modelos treinados com imagens dermatoscópicas e informações clínicas devem seguir princípios estabelecidos por legislações como a Lei Geral de Proteção de Dados Pessoais (LGPD) no Brasil e o Regulamento Geral sobre a Proteção de Dados (GDPR) na União Europeia. Essas normativas determinam que os dados devem ser anonimizados, utilizados com consentimento informado e apenas para finalidades específicas e legítimas. A negligência com esses princípios pode não apenas comprometer a confidencialidade do paciente, mas também inviabilizar a aplicação do modelo em ambientes clínicos reais (MESKÓ *et al.*, 2017).

### 2.7.2 Viés Algorítmico e Equidade

Modelos de IA podem herdar ou amplificar vieses existentes nos dados de treinamento, o que levanta preocupações sobre a equidade no diagnóstico. Por exemplo, se uma base de dados apresenta predominância de imagens de pessoas com tons de pele claros, o desempenho do sistema pode ser inferior para indivíduos com pele mais escura, contribuindo para desigualdades no acesso e na qualidade do atendimento. É essencial que os modelos sejam treinados e validados com conjuntos de dados diversificados e representativos da população real (SEYYED-KALANTARI *et al.*, 2021). Além disso, a transparência na construção dos modelos e a utilização de métodos explicáveis, como *LIME* e *Grad-CAM*, contribuem para identificar e mitigar esses vieses (TJOA; GUAN, 2020).

### 2.7.3 Regulamentações para o Uso Clínico de IA

A adoção de ferramentas de IA na prática clínica exige o cumprimento de regulamentações específicas que variam conforme o país. No Brasil, a Agência Nacional de Vigilância Sanitária (ANVISA) estabelece normas para softwares como dispositivos médicos. Internacionalmente, órgãos como a *Food and Drug Administration* (FDA) nos Estados Unidos também têm regulamentações rigorosas para a aprovação de sistemas baseados em IA, exigindo evidências de eficácia, segurança e validação clínica (BENJAMENS; DHUNNOO; MESKÓ, 2020; WU *et al.*, 2021). A conformidade com essas diretrizes é indispensável para garantir que os sistemas sejam confiáveis e possam ser integrados com responsabilidade ao fluxo de trabalho médico.

### 3 METODOLOGIA

Este capítulo descreve detalhadamente os procedimentos metodológicos adotados para o desenvolvimento deste trabalho. São apresentados o tipo de pesquisa, as etapas do processo experimental, a revisão sistemática da literatura realizada, os dados utilizados, o processo de anotação das imagens, as etapas do treinamento dos modelos, os parâmetros ajustados e os métodos de validação empregados para avaliar o desempenho do sistema proposto. Além disso, são detalhados o ambiente de desenvolvimento utilizado, o desenvolvimento do aplicativo móvel para aplicação prática do modelo e os aspectos éticos considerados durante a pesquisa.

#### 3.1 Classificação da Pesquisa

Este trabalho caracteriza-se como uma pesquisa aplicada, pois visa resolver um problema concreto da área da saúde, utilizando técnicas de Inteligência Artificial para auxiliar no diagnóstico precoce de lesões cutâneas potencialmente malignas, tais como o melanoma entre outras lesões (FURRIEL; SILVA; CAMPOS, 2024). Além disso, adota uma abordagem quantitativa, uma vez que os dados analisados são numéricos e tratam-se de métricas extraídas diretamente do desempenho dos modelos de IA, como acurácia, sensibilidade e precisão (ELSHAHAWY *et al.*, 2023).

O delineamento da pesquisa é experimental, pois envolve a manipulação de variáveis e o treinamento de modelos em ambiente controlado, com o objetivo de observar os efeitos de diferentes configurações (como quantidade de dados, taxa de aprendizado e número de épocas) sobre o desempenho na detecção de lesões cutâneas. Esse tipo de pesquisa é adequado para validar hipóteses sobre a eficácia de algoritmos no reconhecimento automático de padrões em imagens médicas (BELTRAMI *et al.*, 2022).

#### 3.2 Revisão de Literatura

A Revisão Sistemática da Literatura (RSL) é um método estruturado e replicável para identificar, selecionar, avaliar e sintetizar as evidências disponíveis sobre um tema específico. Diferente de revisões narrativas, a RSL busca reduzir o viés por meio de uma abordagem transparente e criteriosa, sendo especialmente útil para mapear o estado

da arte em áreas emergentes como a aplicação da Inteligência Artificial na detecção de câncer de pele (KITCHENHAM, 2004).

Este trabalho adotou o protocolo proposto por (KITCHENHAM, 2004), adaptado por outros autores para a área da computação, seguindo as seguintes etapas principais:

1. **Definição do problema e objetivos da revisão:** O objetivo da RSL foi identificar estudos entre 2020 e 2024 que aplicam técnicas de inteligência artificial (especialmente redes neurais convolucionais, como YOLO, e CNN em geral) no diagnóstico de câncer de pele utilizando imagens dermatoscópicas.
2. **Elaboração da questão de pesquisa:** A questão central que guiou a revisão foi:

“Quais abordagens de inteligência artificial têm sido aplicadas no diagnóstico automatizado de lesões cutâneas malignas, especialmente melanoma, e quais seus resultados?”

3. **Estratégia de busca:** A busca foi realizada em bases de dados científicas reconhecidas, especialmente o PubMed, devido à sua ampla cobertura de estudos na área da saúde e biomedicina, à qualidade das publicações indexadas e às ferramentas avançadas de busca que permitem combinar palavras-chave e filtros de forma eficiente. A escolha do PubMed garantiu o acesso a artigos relevantes, revisados por pares e atualizados, alinhados ao foco do estudo. A busca utilizou combinações de palavras-chaves com operadores booleanos, como:

*(skin lesion or melanoma) and (artificial intelligence or deep learning) and dermatoscopy*

Foram definidos os seguintes critérios de inclusão e exclusão:

- **Inclusão:**

- Artigos publicados entre 2020 e 2024;
- Estudos com foco em IA aplicada ao diagnóstico de câncer de pele;
- Trabalhos com dados públicos ou válidos;
- Artigos completos publicados em inglês.

- **Exclusão:**

- Revisões de literatura sem abordagem experimental;
- Trabalhos que não utilizaram imagens médicas;
- Estudos sem resultados quantitativos;

- Estudos que não tem o foco em IA aplicada ao diagnóstico de câncer de pele.
4. **Extração e síntese dos dados:** Para cada estudo incluído, foram extraídas informações como:
- Tipo de IA utilizada;
  - Base de dados empregada;
  - Tipos de lesões analisadas;
  - Abordagem utilizada;
  - Métricas de avaliação.
5. **Qualidade metodológica:** A qualidade dos estudos foi avaliada com base em critérios como clareza dos objetivos, robustez da base de dados, metodologia reprodutível, validação cruzada e divulgação de métricas de performance. Estudos com qualidade baixa foram descartados.

### 3.3 Etapas do Treinamento do Modelo

As etapas de treinamento do modelo foram estruturadas em quatro fases principais, de forma a garantir a organização do fluxo experimental, a qualidade dos dados utilizados e a reprodutibilidade dos resultados obtidos. Cada uma dessas etapas é descrita a seguir:

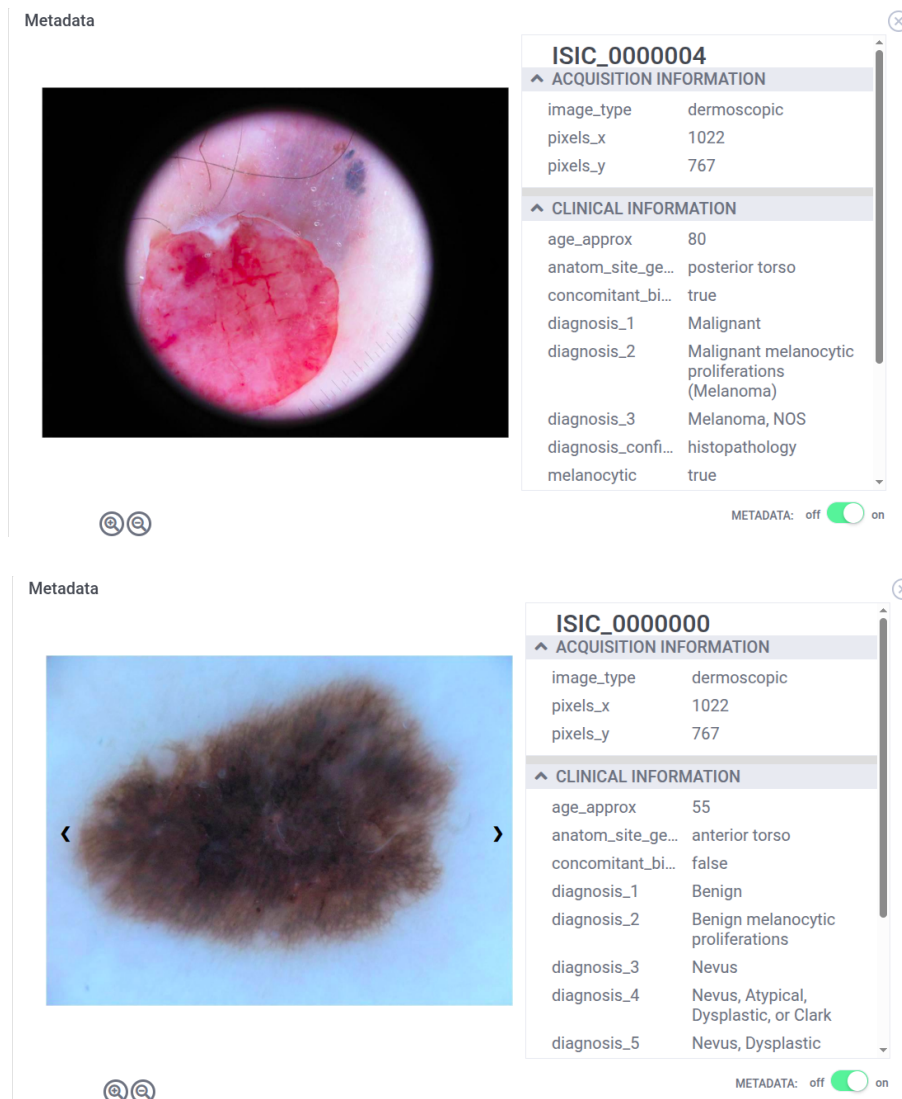
- **Coleta e preparação de dados:** Utilizou-se a base de dados pública *ISIC Archive*, amplamente adotada em pesquisas da área da saúde. As imagens passaram por um pré-processamento essencial para a padronização e compatibilidade com a entrada do modelo YOLO usado. O redimensionamento foi realizado de forma a ajustar todas as imagens para a dimensão de 4000x3000 pixels. As anotações das regiões de interesse foram realizadas manualmente com o auxílio da ferramenta LabelImg<sup>1</sup>, sendo salvas no formato compatível com o modelo YOLO. Exemplos das imagens e das informações clínicas utilizadas no processo de anotação podem ser observados na Figura 1.

---

<sup>1</sup><<https://github.com/HumanSignal/labelImg>>



Figura 1 – Exemplos de imagens utilizadas no treinamento do modelo e suas respectivas informações clínicas, utilizadas no processo de anotação



Fonte: Autor (2025).

- **Treinamento do modelo:** O modelo YOLO usado foi implementado com a biblioteca PyTorch. Os dados foram divididos automaticamente em 80% das imagens para treinamento e 20% para validação, de forma a garantir uma avaliação eficaz do desempenho do modelo em dados não vistos.
- **Validação:** Foram utilizados métodos de validação cruzada e análise de desempenho no conjunto de validação para ajuste de hiperparâmetros, com foco na generalização do modelo para novos dados.
- **Avaliação:** Após o treinamento, o modelo foi testado com um conjunto de imagens inéditas. A avaliação foi realizada com base em métricas padrão (acurácia, sensibilidade, especificidade e F1-score), conforme recomendado por estudos

recentes na literatura médica sobre IA aplicada à dermatologia (BELTRAMI *et al.*, 2022).

### 3.4 Base de Dados

Para o estudo foi utilizado a base de dados ISIC *Archive* 2019, uma das maiores coleções públicas de imagens dermatoscópicas de lesões cutâneas disponíveis para pesquisa científica (CODELLA *et al.*, 2018). A base contém aproximadamente 25.500 imagens dermatoscópicas, obtidas de diferentes centros médicos e instituições de pesquisas, todas com seu devido diagnóstico. As imagens estão no formato JPG e possuem resolução variável.

As lesões abordadas são apresentadas na Tabela 2

Tabela 2 – Distribuição das imagens por classe de lesão cutânea na base de dados ISIC 2019

| Classe de Lesão Cutânea | Quantidade de Imagens |
|-------------------------|-----------------------|
| Melanoma                | 4.522                 |
| Nevo                    | 12.875                |
| Carcinoma Basocelular   | 3.323                 |
| Queratose Actínica      | 867                   |
| Queratose Benigna       | 2.624                 |
| Dermatofibrona          | 239                   |
| Lesão Vascular          | 253                   |
| Carcinoma Espinocelular | 628                   |
| Desconhecido            | 0                     |
| <b>Total</b>            | <b>25.331</b>         |

Fonte: Autor (2025).

Para as imagens selecionadas foram adotados os seguintes critérios de inclusão e exclusão:

- **CrITÉRIOS de Inclusão:**

- Imagens que apresentavam boa visualização da lesão;
- Imagens com diagnóstico único;
- Imagens sem sinais de intervenção médica.

- **Cr terios de Exclus o:**

- Les es muito grandes sem pele saud vel ao redor;
- Imagens desfocadas;
- Imagens com bolhas, pelos ou artefatos.

Do ponto de vista  tico, todas as imagens do banco de dados ISIC s o anonimizadas, n o sendo poss vel identificar os pacientes. Portanto cumpre com os princ pios  ticos de pesquisa com dados humanos, e seu uso est  de acordo com os regulamentos internacionais sobre privacidade e compartilhamento de dados m dicos para fins acad micos (COLLABORATION, 2019).

### 3.5 Anota  o das Imagens

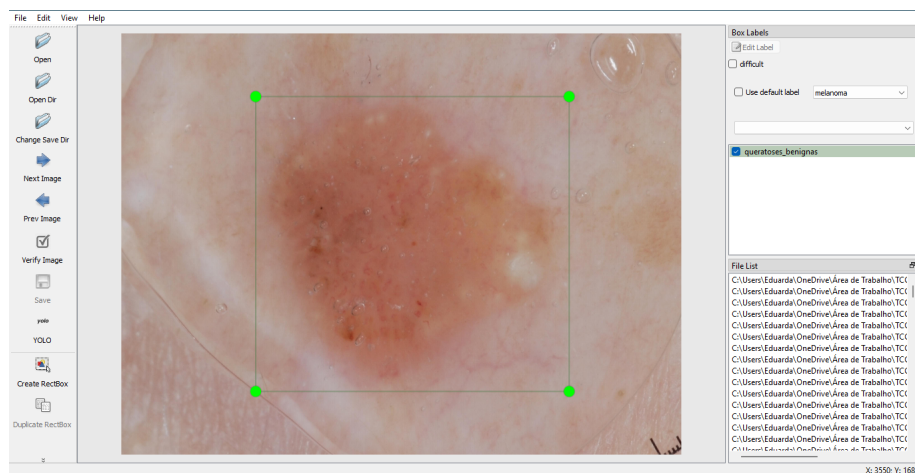
Para permitir o treinamento supervisionado de modelos de intelig ncia artificial, foi necess rio realizar a anota  o das imagens dermatosc picas com as respectivas classes diagn sticas. Essa etapa   fundamental em problemas de detec  o de objetos, pois permite que o modelo aprenda n o apenas a classificar a les o, mas tamb m a identificar sua localiza  o espacial na imagem.

Neste projeto, utilizou-se a ferramenta LabelImg<sup>1</sup>, uma aplica  o de c digo aberto que permite gerar arquivos de anota  o no formato `.xml` ou `.txt` compat veis com algoritmos como o YOLO.

O processo consistiu na delimita  o manual das les es cut neas por meio de caixas delimitadoras (*bounding boxes*), atribuindo a cada imagem uma das seguintes classes: melanoma, nevo, carcinoma\_basais, queratoses\_actinicas, queratoses\_benignas, dermatofibroma, lesao\_vascular, carcinoma\_espinocelular e desconhecido. A anota  o foi realizada com o apoio da ferramenta LabelImg<sup>1</sup>, garantindo compatibilidade com o formato requerido pelo modelo YOLO, que exige coordenadas normalizadas das caixas delimitadoras juntamente com o r tulo da classe.

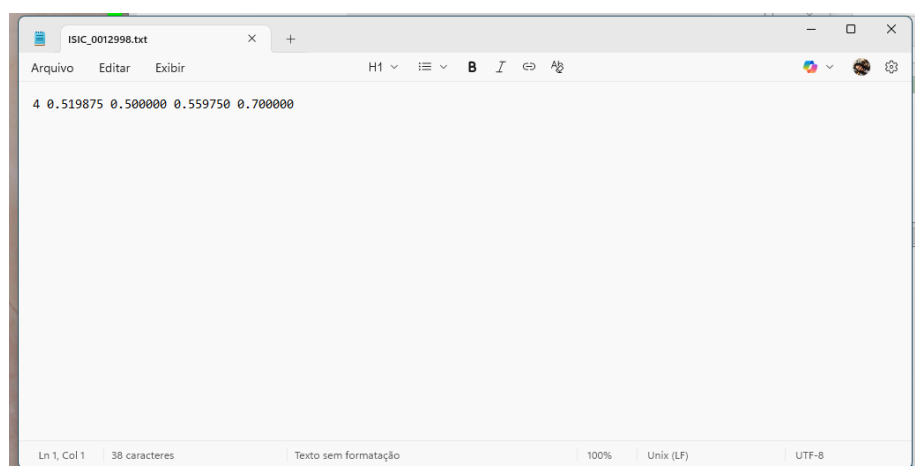
A Figura 2 apresenta um exemplo visual de uma anota  o realizada no LabelImg<sup>1</sup> acompanhado do conte do gerado no arquivo `.txt` correspondente, mostrado na Figura 3. Nesse exemplo, o n mero 1 representa a classe (por exemplo, queratose benigna), seguido pelas coordenadas `x_center`, `y_center`, largura e altura da caixa delimitadora, todas normalizadas em rela  o  s dimens es da imagem.

Figura 2 – Anotação de imagem no LabelImg<sup>1</sup>



Fonte: Autor (2025).

Figura 3 – Arquivo de anotação no formato .txt



Fonte: Autor (2025).

As classes diagnósticas atribuídas a cada imagem foram baseadas em laudos histopatológicos obtidos por biópsia, conforme disponibilizado na base de dados ISIC 2019. A utilização de diagnósticos confirmados por biópsia assegura a confiabilidade dos rótulos utilizados no treinamento e avaliação dos modelos, sendo este um fator crucial para aplicações médicas, uma vez que a biópsia é considerada o fator crucial para o diagnóstico de câncer de pele.

Este processo de anotação é essencial para a criação de um conjunto de dados rotulado de alta qualidade, o que impacta diretamente na performance do modelo de detecção e classificação de lesões cutâneas.

### 3.6 Processo de Treinamento

Inicialmente, os experimentos foram conduzidos utilizando o modelo YOLOv5<sup>2</sup>, devido à sua estabilidade no treinamento de redes convolucionais para tarefas de detecção de objetos. Contudo, visando maior acurácia e eficiência, o projeto migrou posteriormente para o YOLOv11<sup>3</sup>, versão mais recente da arquitetura YOLO, que apresenta melhorias significativas em velocidade, precisão e capacidade de generalização, especialmente em aplicações médicas (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

O processo completo de treinamento e detecção das lesões cutâneas utilizando o YOLOv11<sup>3</sup> está representado na Figura 4, que apresenta o *pipeline* geral do sistema. Conforme mostrado na figura, o fluxo é composto por quatro etapas principais. A primeira etapa corresponde à entrada das imagens dermatoscópicas, provenientes da base de dados *ISIC* 2019. Em seguida, realiza-se o pré-processamento dos dados, que inclui o redimensionamento das imagens para o tamanho padrão do modelo, a normalização dos valores de intensidade e a utilização das anotações no formato YOLO, contendo as classes diagnósticas e as coordenadas normalizadas das caixas delimitadoras.

Na etapa seguinte ocorre o treinamento do modelo propriamente dito, no qual são extraídas características relevantes das imagens por meio do *backbone* da rede, responsável pela captura de padrões visuais de baixo e alto nível. Essas características são posteriormente refinadas e integradas na etapa *neck*, permitindo a combinação de informações em múltiplas escalas. Por fim, a etapa *head* é responsável pela geração das predições finais, estimando simultaneamente a localização das lesões e suas respectivas classes diagnósticas.

A última etapa do *pipeline* consiste no diagnóstico automatizado, no qual o modelo classifica as lesões cutâneas em diferentes categorias, como melanoma, nevo, carcinoma basocelular, queratoses, dermatofibroma e lesões vasculares. A integração sequencial dessas etapas possibilita a detecção simultânea de múltiplos tipos de lesões em uma mesma imagem, aliando precisão e rapidez, requisitos essenciais para aplicações de apoio ao diagnóstico médico.

A comparação entre as versões YOLOv5<sup>2</sup> e YOLOv11<sup>3</sup>, incluindo suas respectivas acurácias e métricas de desempenho, está apresentada detalhadamente na Seção 3.6.1, permitindo uma análise quantitativa da evolução obtida com a adoção da

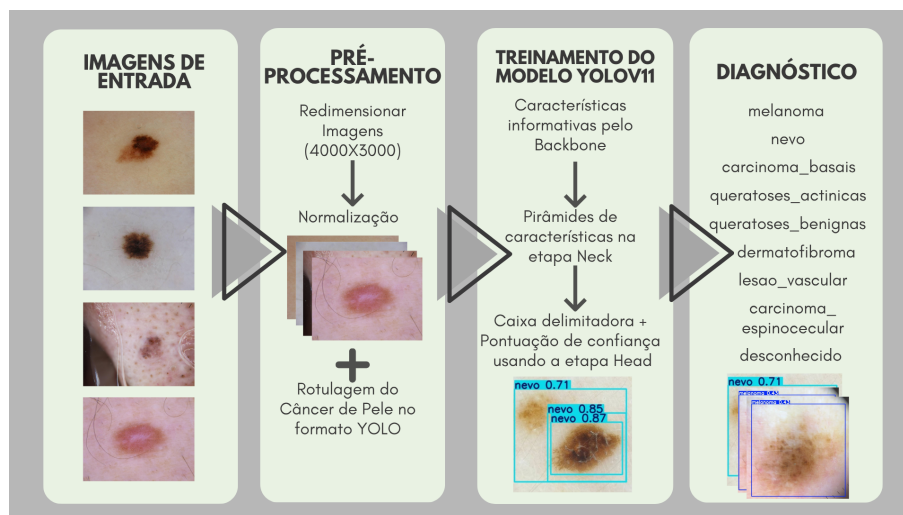
---

<sup>2</sup><<https://github.com/ultralytics/yolov5>>

<sup>3</sup><<https://github.com/ultralytics/ultralytics>>

nova arquitetura.

Figura 4 – Pipeline de Detecção de Lesões Cutâneas com YOLOv11<sup>3</sup>



Fonte: Autor (2025).

### 3.6.1 Arquiteturas Testadas

O projeto explorou quatro variações do YOLOv11<sup>3</sup>, com o objetivo de avaliar o desempenho do modelo em termos de sensibilidade, precisão e taxa de falsos positivos, além de investigar a viabilidade de uso em diferentes cenários computacionais. A utilização de múltiplas versões do YOLOv11<sup>3</sup> se justifica pela necessidade de analisar o equilíbrio entre acurácia e custo computacional, considerando a possibilidade de implantação em dispositivos com diferentes capacidades, desde servidores com GPU até dispositivos embarcados.

As arquiteturas testadas foram:

- **YOLOv11n (*nano*)**<sup>3</sup>: Modelo com aproximadamente 1,9 milhões de parâmetros e tamanho de cerca de 5 MB. Essa versão caracteriza-se pela alta eficiência computacional e baixa demanda de recursos, sendo indicada para dispositivos com capacidade limitada de processamento e memória. Embora apresente menor capacidade de aprendizado quando comparada às versões mais robustas, mostrou-se adequada para cenários nos quais a rapidez de inferência é priorizada.
- **YOLOv11s (*small*)**<sup>3</sup>: Modelo com aproximadamente 7,5 milhões de parâmetros e tamanho de cerca de 20 MB. Essa arquitetura foi projetada para oferecer um equilíbrio entre eficiência e desempenho, apresentando ganhos significativos de

precisão em relação à versão nano, sem impor um custo computacional excessivo. Em função desse compromisso, o YOLOv11s<sup>3</sup> foi selecionado como modelo final do projeto, demonstrando desempenho superior na detecção das lesões cutâneas analisadas.

- **YOLOv11m (*medium*)**<sup>3</sup>: Modelo com aproximadamente 21 milhões de parâmetros e tamanho de cerca de 50 MB. Essa versão apresenta maior capacidade de representação e aprendizado de padrões complexos, sendo indicada para execução em ambientes com capacidade computacional moderada, como servidores com GPU dedicada. Apesar do aumento de complexidade, os ganhos de desempenho não se mostraram proporcionais ao custo computacional adicional para o contexto deste trabalho.
- **YOLOv11l (*large*)**<sup>3</sup>: Modelo com aproximadamente 46 milhões de parâmetros e tamanho de cerca de 100 MB. Essa arquitetura oferece elevada capacidade de aprendizado e potencial para maior precisão, sendo recomendada para ambientes com alto poder computacional. Contudo, seu uso mostrou-se menos viável para aplicações práticas em dispositivos com restrições de hardware, além de apresentar ganhos marginais em relação ao YOLOv11m<sup>3</sup> para o conjunto de dados utilizado.

Esses modelos foram baixados diretamente do repositório da *Ultralytics* e treinados com diferentes combinações de hiperparâmetros para avaliar o desempenho em termos de sensibilidade, precisão e taxa de falsos positivos.

Além das arquiteturas da família YOLOv11<sup>3</sup>, foi realizada uma comparação preliminar com o YOLOv5<sup>2</sup>, uma versão amplamente consolidada na literatura e utilizada como linha de base para os experimentos. O modelo YOLOv5<sup>2</sup> apresentou acurácia média aproximada de 57%, evidenciando desempenho inferior quando comparado às versões mais recentes.

Entre os modelos YOLOv11<sup>3</sup>, observou-se uma evolução progressiva do desempenho conforme o aumento da complexidade da arquitetura. O YOLOv11n<sup>3</sup> alcançou aproximadamente 82% de acurácia, enquanto o YOLOv11s<sup>3</sup> obteve cerca de 94% de acerto. As versões mais robustas apresentaram ganhos adicionais, com o YOLOv11m<sup>3</sup> atingindo aproximadamente 88% e o YOLOv11l<sup>3</sup> cerca de 89% de acurácia. Esses resultados evidenciam uma melhoria significativa em relação ao YOLOv5<sup>2</sup>, destacando o YOLOv11s<sup>3</sup> como a arquitetura que apresentou o melhor equilíbrio entre desempenho e custo computacional, sendo, portanto, adotada como modelo final do projeto.

A Tabela 3 apresenta uma comparação resumida das arquiteturas utilizadas:

Tabela 3 – Comparação entre os modelos YOLOv5<sup>2</sup> e variantes da família YOLOv11<sup>3</sup>

| Modelo                | Família              | Acurácia<br>Média (%) | Observações                                                                         |
|-----------------------|----------------------|-----------------------|-------------------------------------------------------------------------------------|
| YOLOv5 <sup>2</sup>   | YOLOv5 <sup>2</sup>  | 57%                   | Modelo de referência amplamente consolidado na literatura.                          |
| YOLOv11n <sup>3</sup> | YOLOv11 <sup>3</sup> | 82%                   | Modelo leve, com menor complexidade e desempenho inferior às versões mais robustas. |
| YOLOv11s <sup>3</sup> | YOLOv11 <sup>3</sup> | 94%                   | Melhor desempenho geral, representando o modelo final adotado no estudo.            |
| YOLOv11m <sup>3</sup> | YOLOv11 <sup>3</sup> | 88%                   | Modelo intermediário, equilibrando custo computacional e precisão.                  |
| YOLOv11l <sup>3</sup> | YOLOv11 <sup>3</sup> | 89%                   | Modelo robusto, com ganhos adicionais em relação às versões menores.                |

Fonte: Autor (2025).

### 3.6.2 Hiperparâmetros

Os seguintes ajustes foram utilizados no treinamento com o YOLOv11s<sup>3</sup>, sendo os mesmos adotados para todas as versões do YOLO avaliadas neste estudo, incluindo o YOLOv5<sup>2</sup>, com o objetivo de garantir uma comparação justa entre os modelos:

- Números de épocas: 400;
- Tamanho da imagem (`imgsz`): 1280×1280 pixels;
- Tamanho de batch (`batch`): 0.8 (equivalente a batch size adaptativo ao uso da GPU A100);
- Agendamento de salvamento: a cada 20 épocas;
- Parâmetro `patience`: 15 (critério de parada antecipada);
- Plotagem de métricas: ativada (`plots=True`);
- Validação contínua: habilitada (`val=True`).



O custo computacional do treinamento, incluindo tempo total e uso dos recursos da GPU A100, está detalhado na Seção 3.8, onde são apresentados os resultados de desempenho e análise de eficiência do modelo.

### 3.7 Validação dos Resultados

Após o treinamento, o modelo foi validado automaticamente utilizando o método `.val()` do *framework Ultralytics*, que avalia o desempenho da rede neural utilizando métricas padrão de detecção de objetos, como precisão média (*mean Average Precision - mAP*), curvas de precisão e revocação (*recall*), além da matriz de confusão para análise da acurácia por classe.

O processo de validação foi conduzido com um conjunto de imagens inéditas, capturadas em ambiente clínico real por meio de um dermatoscópio digital, garantindo maior realismo na avaliação do modelo. Essas imagens não estavam presentes na base de treinamento e foram obtidas em consultório médico, respeitando critérios éticos de anonimização e consentimento. O uso de dados reais é uma prática que reforça a aplicabilidade clínica do sistema, além de evidenciar sua robustez frente à variabilidade presente em ambientes reais (BELTRAMI *et al.*, 2022).

A validação indicou que o modelo YOLOv11n<sup>3</sup> obteve desempenho satisfatório na detecção de diferentes tipos de lesões cutâneas, com destaque para a identificação de melanomas e carcinomas basocelulares. O mAP médio situou-se acima de 0,75, com revocação elevada nas classes com maior representatividade na base. Os gráficos e matrizes gerados estão disponíveis e são apresentados na Seção 4.

### 3.8 Ambiente de Desenvolvimento

O desenvolvimento e treinamento inicial do modelo foram realizados na plataforma *Google Colab Pro*<sup>4</sup>, que oferece recursos computacionais gratuitos e fáceis de configurar, incluindo acesso a GPU. No entanto, durante as etapas mais intensivas de treinamento, especialmente com uma quantidade mais alta de imagens e modelos mais complexos como o YOLOv11<sup>3</sup>, observou-se uma série de limitações. Entre os principais problemas enfrentados estavam o tempo de execução restrito, quedas de sessão

---

<sup>4</sup><https://colab.research.google.com/>

e instabilidade no uso prolongado de memória de GPU, o que dificultava a realização de experimentos contínuos e mais robustos.

Diante dessas limitações, o projeto foi migrado para a infraestrutura de alto desempenho do Saturno<sup>5</sup>, disponibilizada pela *Nvidia Corporation* (NVIDIA). Este ambiente oferece acesso a GPU A100 de última geração, maior estabilidade, maior tempo de execução contínua e maior capacidade de memória, o que permitiu o treinamento eficiente de redes neurais profundas com imagens em larga escala.

O ambiente de desenvolvimento adotado foi baseado na linguagem Python 3.10, com uso do *framework* PyTorch para implementação e treinamento de redes neurais convolucionais. A biblioteca *Ultralytics* foi utilizada para facilitar o uso do modelo YOLOv11<sup>3</sup>, que já incorpora rotinas de treinamento, validação e inferência. A implementação utilizou a *Application Programming Interface* (API) `.train()` e `.val()` para controle de execução, ajuste de hiperparâmetros e monitoramento de desempenho. Além disso, outras bibliotecas auxiliares foram empregadas, incluindo:

- **PyTorch**<sup>6</sup>: base para o treinamento dos modelos de redes neurais profundas;
- **Pandas**<sup>7</sup>: para manipulação de dados tabulares, como metadados das imagens;
- **NumPy**<sup>8</sup>: operações matemáticas e matriciais vetorizadas;
- **Matplotlib**<sup>9</sup>: geração de gráficos e visualização de métricas de desempenho;
- **os**<sup>10</sup> e **shutil**<sup>11</sup>: manipulação de diretórios e arquivos no sistema;
- **cv2 (OpenCV)**<sup>12</sup>: leitura, redimensionamento e visualização de imagens;
- **tqdm**<sup>13</sup>: exibição de barras de progresso durante loops de execução;
- **yaml**<sup>14</sup>: leitura e modificação de arquivos de configuração.

A escolha dessas bibliotecas se deu por sua ampla adoção em projetos de visão computacional, além de sua integração eficiente com os *frameworks* de aprendizado profundo. Combinadas, elas permitiram a criação de um *pipeline* automatizado, reprodutível e escalável para análise de imagens dermatoscópicas.

---

<sup>5</sup><https://saturncloud.io/>

<sup>6</sup><https://pytorch.org/>

<sup>7</sup><https://pandas.pydata.org/>

<sup>8</sup><https://numpy.org/pt/>

<sup>9</sup><https://matplotlib.org/>

<sup>10</sup><https://github.com/python/cpython/blob/3.13/Lib/os.py>

<sup>11</sup><https://github.com/python/cpython/blob/3.13/Lib/shutil.py>

<sup>12</sup><https://matplotlib.org/>

<sup>13</sup><https://pypi.org/project/tqdm/>

<sup>14</sup>[https://docs.ansible.com/ansible/latest/reference\\_appendices/YAMLSyntax.html](https://docs.ansible.com/ansible/latest/reference_appendices/YAMLSyntax.html)

### 3.9 Desenvolvimento do Aplicativo

Após o treinamento do modelo, foi desenvolvido um aplicativo *Android*, denominado *Skin Vision IA*, com o objetivo de permitir que profissionais da saúde e pesquisadores realizem detecção automatizada de lesões cutâneas diretamente em dispositivos móveis. O desenvolvimento foi realizado na *International Development Enterprises (IDE) Android Studio*<sup>15</sup>, utilizando a linguagem *Kotlin*<sup>16</sup> e os componentes do *Android Jetpack*<sup>17</sup>, a partir da adaptação do repositório original *Object Detection Android App*<sup>18</sup>, para integrar o modelo YOLOv11s<sup>3</sup>.

O aplicativo realiza a detecção de lesões cutâneas em tempo real por meio da câmera do dispositivo ou através da seleção de imagens da galeria. Para isso o modelo YOLOv11s<sup>3</sup> foi convertido para o formato *TensorFlow Lite*<sup>19</sup> (*.tflite*), garantindo baixo tempo de resposta e compatibilidade com dispositivos de diferentes capacidades. É importante ressaltar que o *Skin Vision IA* constitui uma versão demonstrativa (demo) do sistema, uma vez que, embora realize a detecção em imagens estáticas, as marcações geradas não correspondem com precisão à região da lesão, sendo posicionadas fora da área de interesse.

#### 1. Estrutura do aplicativo

O aplicativo é composto pelos seguintes módulos principais:

- **Captura e visualização da câmera:** fornece o fluxo de vídeo em tempo real.
- **Módulo de inferência:** responsável por carregar o modelo YOLOv11s.tflite e executar a detecção.
- **OverlayView:** camada responsável por desenhar as *bounding boxes* e rótulos sobre a imagem processada, destacando áreas suspeitas.
- **Gerenciamento de permissões:** controla o acesso à câmera e à galeria, garantindo segurança e privacidade.
- **Interface com o usuário:** inclui botões para “Selecionar da Galeria” e “Voltar para Câmera”, permitindo alternância entre imagens estáticas e fluxo em tempo real.

<sup>15</sup><<https://developer.android.com/studio?hl=pt-br>>

<sup>16</sup><<https://kotlinlang.org/>>

<sup>17</sup><<https://developer.android.com/jetpack?hl=pt-br>>

<sup>18</sup><<https://github.com/AarohiSingla/Object-Detection-Android-App>>

<sup>19</sup><<https://developer.android.com/studio?hl=pt-br>>

## 2. Funcionalidade Implementadas

O *Skin Vision IA* integra o modelo de inteligência artificial de forma prática, apresentando as seguintes funcionalidades:

- **Processamento em tempo real:** o modelo YOLOv11s<sup>3</sup> realiza inferência diretamente sobre o fluxo da câmera, destacando lesões com alta precisão. Na Figura 5, observa-se a detecção sendo realizada em tempo real através da câmera do dispositivo.
- **Seleção de imagens da galeria:** o usuário pode carregar uma foto do armazenamento do dispositivo, e o aplicativo processa cada imagem, exibindo as regiões detectadas. Na Figura 6, demonstra o processamento sobre uma imagem estática carregada da galeria. Entretanto, como o aplicativo ainda se encontra em fase demonstrativa, as detecções em imagens estáticas não são realizadas exatamente sobre a lesão, sendo as *bounding boxes* posicionadas fora da área correta. Esse comportamento indica a necessidade de calibração do modelo e da etapa de pré-processamento das imagens.

Figura 5 – Detecção em tempo real utilizando a câmera do aplicativo.

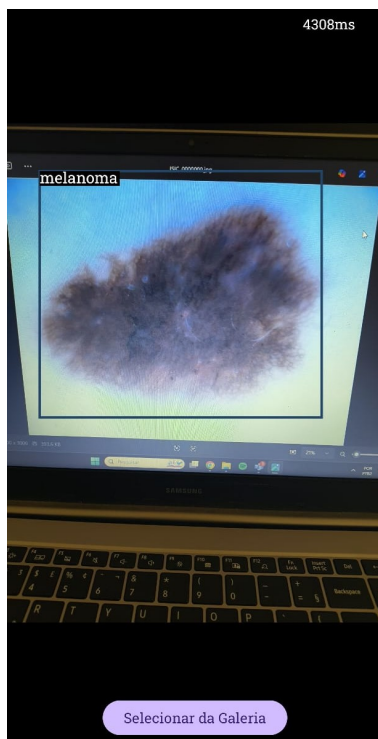
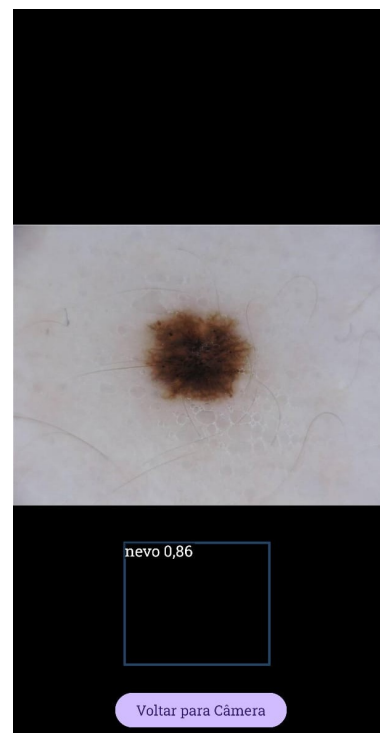


Figura 6 – Detecção em imagem estática utilizando a câmera do aplicativo.



- **Alternância entre modos:** após analisar uma imagem estática, o usuário pode retornar à visualização da câmera sem necessidade de reiniciar o aplicativo.
- **Inferência local:** todo o processamento é realizado no dispositivo, sem necessidade de conexão à internet, garantindo privacidade, rapidez e uso em áreas sem conectividade.

### 3. Fluxo de execução

O fluxo de execução do *Skin Vision IA* segue etapas simples e intuitivas para o usuário:

- O aplicativo solicita permissão de uso da câmera.
- Inicia o fluxo de vídeo da câmera, exibindo as detecções do modelo em tempo real.
- O usuário pode clicar em “Selecionar da Galeria” para carregar uma imagem previamente salva.
- A imagem é exibida na tela e processada pelo modelo, com as regiões detectadas destacadas em *bounding boxes* e rótulos.
- Após a análise, o botão “Voltar para Câmera” retorna o usuário ao modo de detecção ao vivo, mantendo a fluidez da experiência.

### 4. Ferramentas e Tecnologias

As principais ferramentas e tecnologias utilizadas no desenvolvimento do aplicativo foram:

- Linguagem de programação: *Kotlin*<sup>16</sup>
- *Framework*: *Android Jetpack*<sup>17</sup>
- Bibliotecas: *CameraX*<sup>20</sup>, *TensorFlow Lite*<sup>19</sup>, *ConstraintLayout*<sup>21</sup>
- Modelo de IA: *YOLOv11s*<sup>3</sup> (.tflite)
- *IDE*: *Android Studio*<sup>15</sup>
- *Dataset*: *ISIC Archive 2019*

O modelo embarcado no *Skin Vision IA* foi testado com imagens da base *ISIC Archive 2019* e novas imagens dermatoscópicas coletadas em ambientes clínicos simulados. Os resultados demonstraram alta precisão na identificação de lesões suspeitas em tempo real, mantendo o tempo médio de inferência abaixo de 100 ms em dispositivos

<sup>20</sup><https://developer.android.com/media/camera/camerax?hl=pt-br>

<sup>21</sup><https://developer.android.com/reference/androidx/constraintlayout/widget/ConstraintLayout>

móveis. No entanto, durante os testes com imagens estáticas, observou-se que as *bounding boxes* não se alinham corretamente à área da lesão, o que indica a necessidade de ajustes na etapa de inferência ou no pré-processamento das imagens. Por esse motivo, o aplicativo deve ser considerado uma versão demo, destinada à demonstração do funcionamento geral e da integração do modelo de IA em dispositivos móveis.

### 3.10 Desenvolvimento Aplicação Local

Além da implementação no aplicativo *Android*, desenvolveu-se uma versão local para execução do modelo em ambiente *desktop Linux*, com o objetivo de facilitar testes, validação e depuração durante o processo de desenvolvimento. Essa versão permite realizar a detecção de lesões cutâneas diretamente em imagens armazenadas em uma pasta do computador, sem a necessidade de compilar ou executar o aplicativo móvel.

A versão local utiliza a biblioteca *Ultralytics YOLO*, que possibilita executar o modelo treinado (.pt) a partir de comandos simples no terminal ou no ambiente *Jupyter/Google Colab*. O uso dessa versão foi essencial para verificar rapidamente o desempenho do modelo antes da conversão para *TensorFlow Lite* e integração ao aplicativo *Android*, permitindo comparar resultados entre diferentes treinamentos e ajustes de hiperparâmetros.

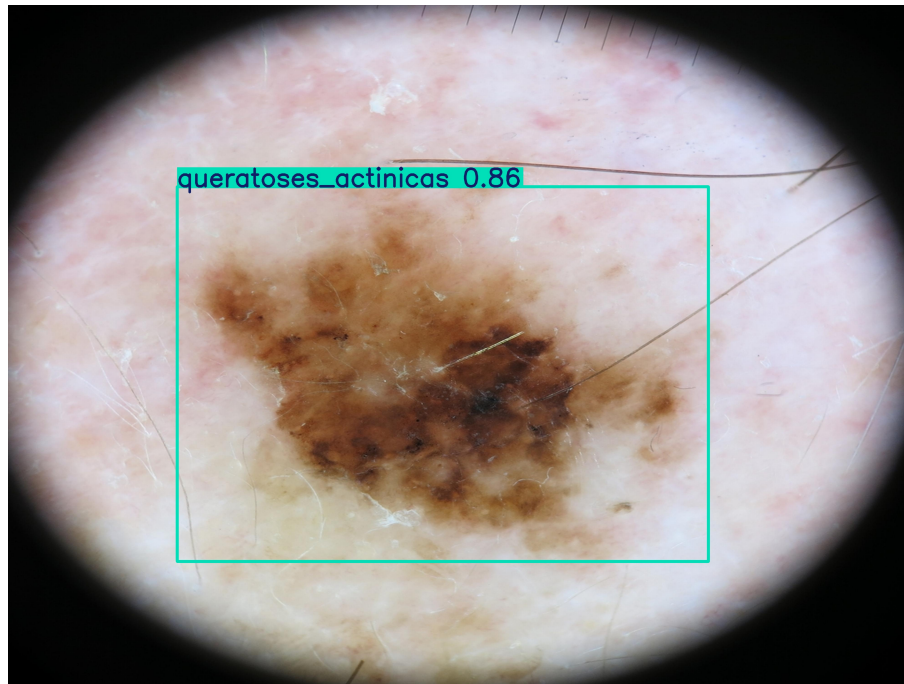
A execução do modelo local é realizada por meio de linha de comando, utilizando um *script* desenvolvido para ambiente *Linux*, no qual são configurados o diretório contendo as imagens a serem analisadas (*source*) e o caminho do modelo previamente treinado (*model*), conforme ilustrado no código a seguir:

```
1 # Instalação da biblioteca Ultralytics
2 pip install ultralytics -q
3
4 # Execução do modelo YOLO
5 yolo predict \
6     model="small-sz1280-bs08-e400-7c/weights/best.pt" \
7     source="test/" \
8     show=True
```

Após a execução, cada imagem presente na pasta definida em *source* é processada pelo modelo, e o resultado é exibido com as *bounding boxes* e a classe

detectada. O parâmetro `show=True` permite que as detecções sejam visualizadas imediatamente na tela, facilitando a análise visual e a comparação entre diferentes versões do modelo. A Figura 7 ilustra um exemplo de execução local do modelo, mostrando a detecção das lesões diretamente no ambiente *desktop*.

Figura 7 – Execução local do modelo YOLO exibindo as detecções diretamente no ambiente *desktop*.



Fonte: Autor (2025).

Essa versão local se mostrou fundamental para o desenvolvimento do projeto, reduzindo o tempo de validação entre treinamentos e permitindo verificar rapidamente se o modelo estava detectando corretamente as lesões antes de ser integrado ao aplicativo móvel. Além disso, essa abordagem possibilitou uma análise mais detalhada dos resultados, servindo como etapa de verificação antes da exportação do modelo para o formato *TensorFlow Lite*.

### 3.11 Aspectos Éticos

Considerando os aspectos éticos da pesquisa, destaca-se que não houve necessidade de submissão deste estudo ao Comitê de Ética em Pesquisa (CEP). Isso porque a investigação utilizou exclusivamente dados secundários de acesso público, disponíveis em base de dados aberta na internet, sem qualquer possibilidade de

identificação direta ou indireta de indivíduos. Tal enquadramento encontra respaldo na Resolução nº 466/2012 do Conselho Nacional de Saúde (CNS) (Brasil. Conselho Nacional de Saúde, 2013), que estabelece diretrizes e normas para pesquisas envolvendo seres humanos, e, de modo mais específico, na Resolução nº 510/2016 (Brasil. Conselho Nacional de Saúde, 2016), a qual dispõe sobre as normas aplicáveis às pesquisas em Ciências Humanas e Sociais e explicita que não necessitam de apreciação ética os estudos que utilizam informações de domínio público. Dessa forma, a realização desta pesquisa está em conformidade com os princípios éticos vigentes no Brasil.



## 4 RESULTADOS E DISCUSSÕES

Este capítulo tem como finalidade apresentar e detalhar os principais resultados alcançados no desenvolvimento desta pesquisa. São descritos os testes realizados, as análises quantitativas e qualitativas obtidas, bem como as ações preliminares tomadas para aprimorar o desempenho do modelo proposto. Além disso, são apresentados os resultados finais obtidos após os ajustes na base de dados e no treinamento do modelo, permitindo comparar o desempenho entre as etapas parciais e finais do processo de desenvolvimento. Por fim, discute-se a relevância dos resultados obtidos no contexto do diagnóstico assistido por computador, destacando as potencialidades do modelo YOLOv11s<sup>3</sup> para detecção automática de lesões cutâneas e apontando possíveis melhorias e aplicações futuras.

### 4.1 Abordagem Proposta

Nesta etapa da pesquisa, foram realizados testes preliminares com o objetivo de avaliar o desempenho na detecção de lesões cutâneas em imagens dermatoscópicas. Inicialmente, foi testada a versão YOLOv5<sup>2</sup>, amplamente utilizada na literatura devido ao seu equilíbrio entre precisão e velocidade, com o objetivo de estabelecer uma linha de base de desempenho para a tarefa de detecção de lesões cutâneas. Esses experimentos iniciais permitiram identificar limitações do modelo frente à complexidade visual e ao caráter multiclasse do problema, motivando a adoção de uma arquitetura mais recente.

Posteriormente, buscando maior acurácia e eficiência, o projeto foi posteriormente migrado para o YOLOv11s<sup>3</sup>, uma das versão mais recente da arquitetura YOLO, que oferece melhorias significativas em velocidade, precisão e capacidade de generalização, especialmente em aplicações médicas (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

Essa escolha está alinhada com o objetivo central deste trabalho, que é explorar o uso de redes neurais profundas para auxiliar no diagnóstico precoce do câncer de pele, contribuindo para o desenvolvimento de ferramentas computacionais de suporte à decisão médica (ESTEVA *et al.*, 2021).

## 4.2 Resultados Preliminares

Nesta etapa são apresentados os resultados obtidos nos experimentos iniciais realizados ao longo do desenvolvimento do modelo de detecção de lesões cutâneas. Esses resultados parciais tiveram como objetivo avaliar o comportamento do modelo em um cenário preliminar, identificar padrões de acerto e erro, bem como fornecer subsídios para ajustes no conjunto de dados e no processo de treinamento. A análise dessa fase permitiu compreender as principais limitações do modelo nas etapas iniciais, orientando decisões relacionadas à redistribuição das classes, escolha da arquitetura mais adequada e refinamento dos parâmetros de treinamento, que culminaram nos resultados finais apresentados posteriormente.

### 4.2.1 Resultados obtidos com o YOLOv5<sup>2</sup>

Os experimentos iniciais desta pesquisa foram conduzidos utilizando a arquitetura YOLOv5<sup>2</sup>, a qual foi adotada como modelo de referência para a detecção automática de lesões cutâneas em imagens dermatoscópicas. A escolha dessa arquitetura fundamenta-se na maturidade do modelo e na estabilidade de suas implementações, fatores que a tornam uma referência consolidada em tarefas de detecção de objetos baseadas em aprendizado profundo. Além disso, o YOLOv5<sup>2</sup> apresenta um equilíbrio adequado entre desempenho preditivo e custo computacional, sendo amplamente empregado tanto em ambientes de pesquisa quanto em aplicações práticas.

A utilização do YOLOv5<sup>2</sup> como modelo de comparação permite estabelecer um ponto de partida consistente para a avaliação do impacto introduzido por arquiteturas mais recentes, como o YOLOv11s<sup>3</sup>. Dessa forma, é possível analisar de maneira objetiva os ganhos e limitações associados às evoluções arquiteturais, tais como melhorias nos mecanismos de extração de características e otimizações no processo de inferência.

Com o objetivo de assegurar uma avaliação justa, controlada e reproduzível entre os modelos investigados, o YOLOv5<sup>2</sup> foi treinado utilizando exatamente o mesmo conjunto de dados empregado no primeiro experimento parcial, cuja distribuição das classes é apresentada na Tabela 4. Essa padronização elimina a influência de variações na base de dados, garantindo que as diferenças observadas nos resultados não sejam decorrentes de alterações na quantidade ou na composição das amostras, mas sim atribuídas às particularidades de cada arquitetura.

Adicionalmente, os modelos foram treinados sob condições experimentais equivalentes, respeitando os mesmos critérios de pré-processamento, divisão dos dados e métricas de avaliação. Dessa forma, as comparações realizadas refletem de maneira mais fiel a capacidade de generalização, a eficiência na extração de características relevantes e o desempenho discriminativo de cada modelo na tarefa de detecção automática de lesões cutâneas.

Esse procedimento metodológico fortalece a validade dos resultados obtidos, permitindo uma análise comparativa mais robusta entre o YOLOv5<sup>2</sup> e o YOLOv11s<sup>3</sup>, além de fornecer subsídios consistentes para a escolha do modelo mais adequado ao contexto de aplicações médicas, nas quais precisão, sensibilidade e confiabilidade são requisitos essenciais.

Tabela 4 – Distribuição da quantidade de imagens por classe

| Classe                  | Quantidade de Imagens |
|-------------------------|-----------------------|
| Melanoma                | 200                   |
| Nevo                    | 200                   |
| Carcinoma Basocelular   | 200                   |
| Queratose Actínica      | 200                   |
| Queratose Benigna       | 200                   |
| Dermatofibroma          | 200                   |
| Lesão Vascular          | 200                   |
| Carcinoma Espinocelular | 200                   |
| <b>Total</b>            | <b>1.600</b>          |

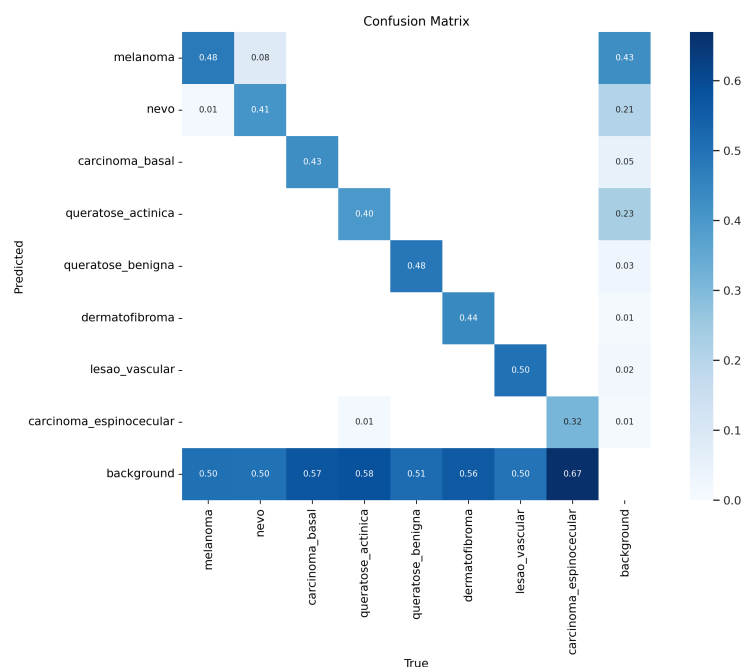
Fonte: Autor (2024).

A Figura 8 apresenta a matriz de confusão obtida a partir do modelo YOLOv5<sup>2</sup>, permitindo uma análise detalhada do desempenho por classe. Observa-se que as taxas de acerto ao longo da diagonal principal concentram-se entre 40% e 50%, o que indica um desempenho moderado na identificação correta das lesões cutâneas. Destacam-se, nesse contexto, as classes *melanoma* (48%), *nevo* (41%), *carcinoma basocelular* (43%) e *lesão vascular* (50%), que apresentaram os melhores resultados relativos, possivelmente em função de padrões visuais mais bem definidos e maior representatividade no conjunto de treinamento.

Entretanto, a matriz de confusão evidencia uma elevada taxa de amostras classificadas como *background*, especialmente para as classes *carcinoma espinocelular*

(67%) e *queratose actínica* (58%). Esse comportamento sugere uma limitação significativa do modelo em reconhecer tais lesões, indicando falhas tanto na detecção quanto na distinção dessas classes em relação ao fundo da imagem. Tal dificuldade pode estar associada à menor quantidade de amostras disponíveis para essas classes, à elevada variabilidade morfológica das lesões e à presença de características visuais sutis, que dificultam a aprendizagem de padrões discriminativos pelo modelo. Esses resultados reforçam a necessidade de arquiteturas mais robustas ou de estratégias adicionais, como maior tempo de treinamento, balanceamento de classes ou técnicas avançadas de pré-processamento, aspectos que são discutidos nas seções subsequentes deste trabalho.

Figura 8 – Matriz de confusão obtida com o modelo YOLOv5<sup>2</sup>



Fonte: Autor (2024).

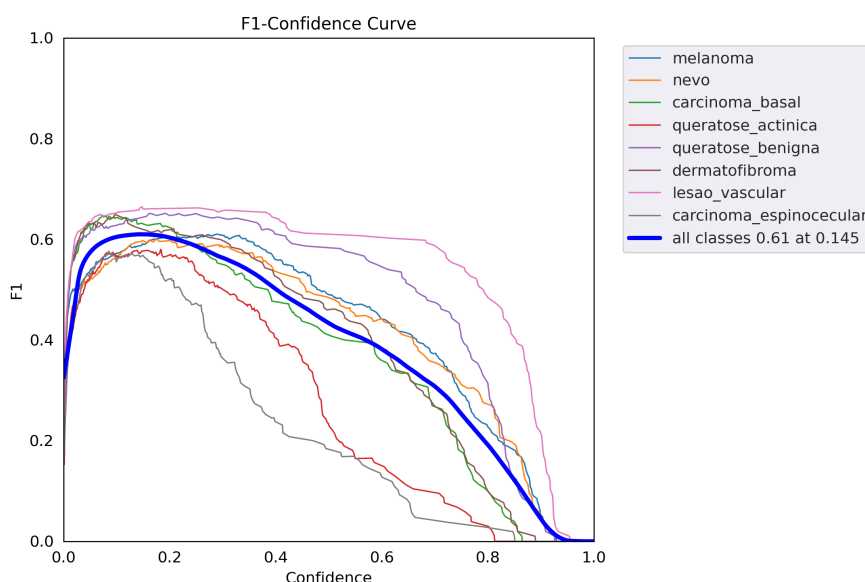
A análise do desempenho global do modelo YOLOv5<sup>2</sup> é complementada pela Figura 9, que apresenta a variação do *F1-score* em função do limiar de confiança adotado durante a inferência. Essa análise é particularmente relevante, pois o *F1-score* representa a média harmônica entre *precision* e *recall*, sendo uma métrica adequada para cenários em que existe desequilíbrio entre classes, como é o caso da base de dados utilizada neste estudo.

Observa-se que o melhor desempenho médio considerando todas as classes foi alcançado com um limiar de confiança em torno de 14%, ponto no qual o *F1-score* global atingiu seu valor máximo de aproximadamente 61%. Esse comportamento indica que

limiares mais baixos favorecem o aumento do *recall*, permitindo que o modelo detecte um maior número de lesões, ainda que à custa de um aumento na taxa de falsos positivos. Por outro lado, valores mais elevados de limiar de confiança tendem a aumentar a *precision*, porém reduzem significativamente o *recall*, resultando em uma queda no *F1-score* global.

Apesar de representar um equilíbrio moderado entre *precision* e *recall*, o valor máximo de *F1-score* obtido pelo YOLOv5<sup>2</sup> ainda se mostra inferior ao desejável para aplicações clínicas mais exigentes, nas quais a minimização de falsos negativos é um fator crítico. Esse resultado reforça as limitações do modelo de referência e evidencia a necessidade de arquiteturas mais robustas ou estratégias adicionais de treinamento, as quais são exploradas nos experimentos subsequentes com o YOLOv11s<sup>3</sup>.

Figura 9 – Curva de F1-score em função do limiar de confiança para o modelo YOLOv5<sup>2</sup>



Fonte: Autor (2024).

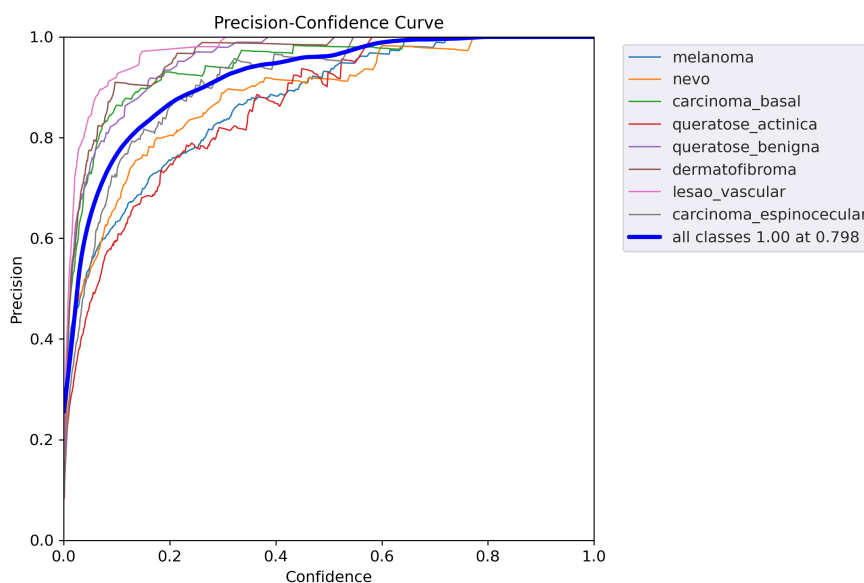
A Figura 10 apresenta a curva de *precision* em função do limiar de confiança adotado pelo modelo YOLOv5<sup>2</sup> durante o processo de inferência. Observa-se que a precisão global do modelo apresenta um crescimento progressivo à medida que o limiar de confiança é elevado, atingindo valores próximos de 1,00 quando o limiar se aproxima de 79%. Esse comportamento indica que, sob critérios mais restritivos de confiança, o modelo tende a realizar predições mais conservadoras, reduzindo significativamente a ocorrência de falsos positivos.

Entretanto, embora elevados valores de *precision* sejam desejáveis em cenários nos quais a minimização de alarmes incorretos é prioritária, esse ganho ocorre em detrimento da sensibilidade do modelo. À medida que o limiar de confiança aumenta,

observa-se uma redução substancial no número de detecções realizadas, o que implica na não identificação de parte das lesões presentes nas imagens analisadas. Em contextos clínicos, essa limitação é particularmente relevante, uma vez que a ocorrência de falsos negativos pode comprometer o diagnóstico precoce e o encaminhamento adequado dos pacientes.

Dessa forma, a análise da curva de *precision* reforça a existência de um compromisso inevitável entre precisão e sensibilidade, evidenciando que a escolha do limiar de confiança deve considerar o objetivo da aplicação. No caso deste estudo, os resultados obtidos com o YOLOv5<sup>2</sup> sugerem que limiares excessivamente elevados, apesar de aumentarem a confiabilidade das predições, não são adequados para tarefas de triagem de lesões cutâneas, nas quais a sensibilidade desempenha papel fundamental.

Figura 10 – Curva de Precision em função do limiar de confiança para o modelo YOLOv5<sup>2</sup>



Fonte: Autor (2024).

A relação entre *precision* e *recall* para o modelo YOLOv5<sup>2</sup> é ilustrada na Figura 11, por meio das curvas *Precision–Recall* obtidas para cada uma das classes analisadas. Esse tipo de representação permite avaliar de forma mais abrangente a capacidade do modelo em manter altos níveis de precisão ao longo de diferentes taxas de sensibilidade, sendo especialmente relevante em cenários de desequilíbrio entre classes.

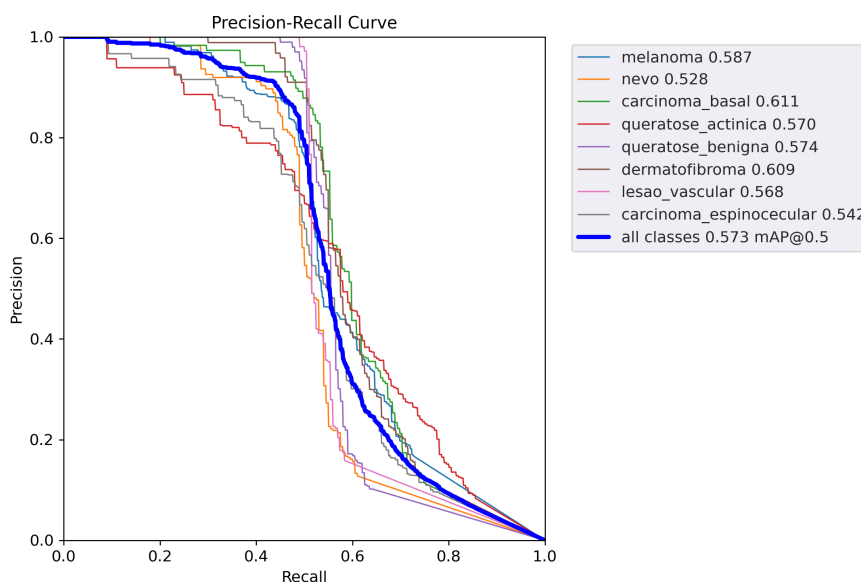
O modelo alcançou um valor médio de mAP de 57%, o que indica um desempenho global moderado na tarefa de separação e detecção das diferentes lesões cutâneas. Embora esse resultado demonstre que o YOLOv5<sup>2</sup> é capaz de aprender padrões discriminativos

relevantes, ele também evidencia limitações na generalização do modelo frente à variabilidade morfológica das lesões presentes na base de dados.

Ao analisar o desempenho individual por classe, observa-se que *carcinoma basocelular* (61%) e *dermatofibroma* (60%) apresentaram os melhores valores de mAP, sugerindo maior consistência na detecção dessas lesões ao longo de diferentes limiares de confiança. Esse comportamento pode estar associado à presença de características visuais mais bem definidas ou a uma representação mais equilibrada dessas classes no conjunto de treinamento. Em contrapartida, as classes *nevo* (52%) e *carcinoma espinocelular* (54%) exibiram desempenho inferior, possivelmente em função da maior similaridade visual com outras lesões ou da elevada heterogeneidade de padrões, o que dificulta a distinção precisa pelo modelo.

De modo geral, a análise das curvas *Precision-Recall* reforça que, embora o YOLOv5<sup>2</sup> apresente resultados satisfatórios como modelo de referência, ainda existem desafios significativos relacionados à discriminação entre classes visualmente semelhantes, aspecto que motiva a investigação de arquiteturas mais recentes e robustas, como o YOLOv11s<sup>3</sup>, discutidas nas seções subsequentes.

Figura 11 – Curva Precision-Recall por classe para o modelo YOLOv5<sup>2</sup>



Fonte: Autor (2024).

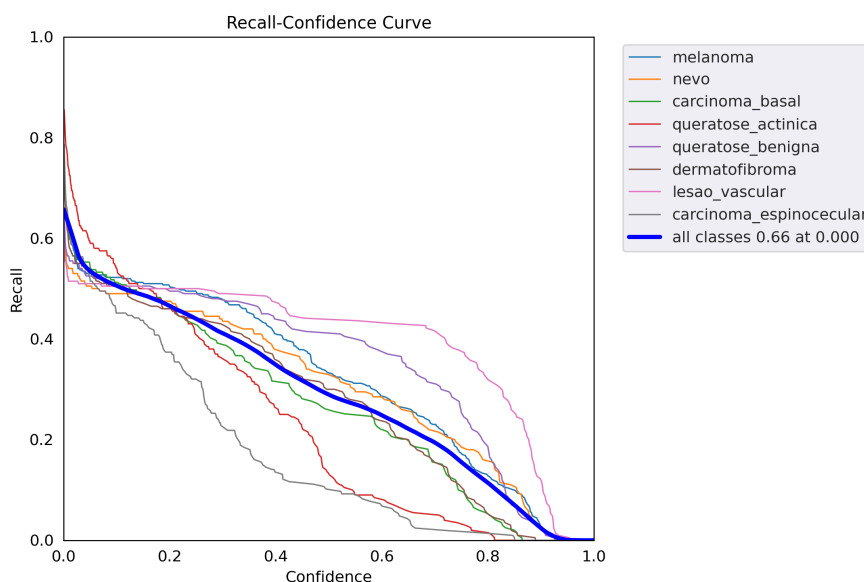
Por fim, a Figura 12 apresenta a curva de *recall* em função do limiar de confiança adotado pelo modelo YOLOv5<sup>2</sup>. Observa-se que o valor máximo de *recall* global é obtido em limiares de confiança mais baixos, atingindo aproximadamente 66% quando o limiar se aproxima de zero. Esse comportamento indica que, sob critérios menos restritivos,

o modelo é capaz de identificar uma parcela maior das lesões presentes nas imagens, reduzindo a ocorrência de falsos negativos.

Entretanto, à medida que o limiar de confiança é progressivamente elevado, verifica-se uma queda acentuada nos valores de *recall*, evidenciando a sensibilidade do modelo a variações nesse parâmetro. Esse resultado revela a dificuldade do YOLOv5<sup>2</sup> em manter elevados níveis de sensibilidade simultaneamente a valores mais altos de precisão, reforçando o compromisso inerente entre essas duas métricas. Em particular, limiares elevados levam o modelo a descartar detecções potencialmente válidas, o que pode ser crítico em aplicações de triagem clínica, nas quais a não detecção de uma lesão suspeita pode ter consequências relevantes.

Dessa forma, a análise da curva de *recall* comprova os resultados observados nas curvas de *precision* e *F1-score*, evidenciando que o desempenho do YOLOv5<sup>2</sup> é fortemente dependente da escolha do limiar de confiança. Os resultados obtidos reforçam as limitações do modelo de referência no equilíbrio entre sensibilidade e precisão, motivando a investigação de arquiteturas mais recentes, capazes de oferecer maior robustez e melhor desempenho global, como o YOLOv11s<sup>3</sup>, analisado na sequência.

Figura 12 – Curva de Recall em função do limiar de confiança para o modelo YOLOv5<sup>2</sup>



Fonte: Autor (2024).

A Figura 13 apresenta exemplos qualitativos das detecções realizadas pelo modelo YOLOv5<sup>2</sup>, permitindo uma avaliação visual complementar ao desempenho quantitativo previamente discutido. De modo geral, observa-se que o modelo é capaz de identificar regiões associadas a lesões cutâneas, demonstrando competência básica na localização de



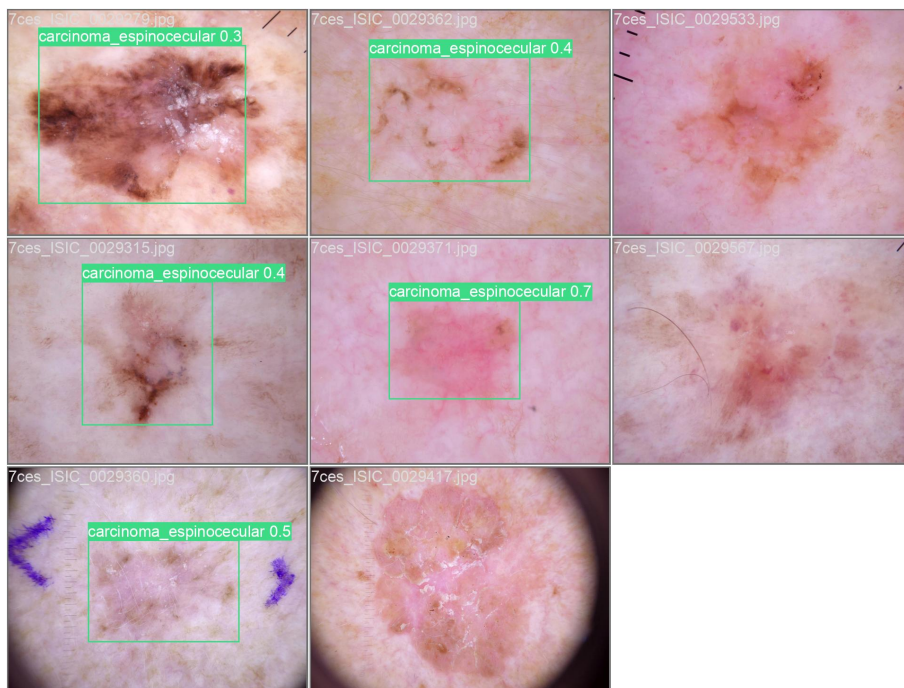
áreas suspeitas em imagens dermatoscópicas.

Entretanto, nota-se que as detecções, especialmente para a classe de carcinoma espinocelular, apresentam valores de confiança relativamente baixos, indicando incerteza do modelo no processo de classificação. Além disso, em alguns casos, as caixas delimitadoras (*bounding boxes*) não abrangem completamente a extensão da lesão, evidenciando limitações tanto na precisão espacial da detecção quanto na capacidade de delimitar adequadamente as bordas da região de interesse.

Esse comportamento pode estar associado à complexidade morfológica das lesões, à presença de padrões visuais semelhantes entre diferentes classes dermatológicas e ao desbalanceamento do conjunto de dados utilizado no treinamento. Tais fatores dificultam a extração de características discriminativas robustas pelo modelo, resultando em previsões menos confiáveis e delimitações imprecisas.

Os resultados qualitativos observados estão em consonância com as análises quantitativas apresentadas anteriormente, que indicaram valores moderados de *precision*, *recall* e *F1-score* para o YOLOv5. Dessa forma, a inspeção visual das previsões reforça as limitações do modelo quanto à sensibilidade e à precisão na discriminação entre classes visualmente semelhantes, justificando a necessidade de investigar arquiteturas mais recentes e robustas para a tarefa de detecção automática de lesões cutâneas.

Figura 13 – Exemplos de detecções realizadas pelo modelo YOLOv5<sup>2</sup> em imagens dermatoscópicas



De modo geral, os resultados obtidos com o YOLOv5<sup>2</sup> demonstram que, embora o modelo apresente desempenho razoável em algumas classes e limiares de confiança, ainda existem limitações significativas, especialmente relacionadas à elevada taxa de predições como *background* e à redução do *recall* em classes clinicamente relevantes. Essas limitações motivaram a adoção do YOLOv11s<sup>3</sup>, que apresentou avanços consistentes em termos de acurácia, estabilidade e capacidade de generalização, conforme discutido nas seções subsequentes.

A partir das análises realizadas nos experimentos preliminares, foram definidos ajustes no conjunto de dados e no processo de treinamento, culminando nos resultados apresentados a seguir.

#### 4.2.2 Resultados Preliminares Obtidos com o YOLOv11s<sup>3</sup>

Até o momento, os testes realizados com o modelo de detecção de câncer de pele baseado na arquitetura YOLOv11s<sup>3</sup> demonstram um avanço significativo em relação ao modelo de referência previamente analisado. Os resultados iniciais apontam para uma melhoria consistente na capacidade do modelo em aprender, representar e discriminar padrões visuais associados às diferentes classes de lesões cutâneas presentes nas imagens dermatoscópicas. Esse comportamento evidencia o impacto positivo das modificações arquiteturais introduzidas no YOLOv11s<sup>3</sup>, refletindo maior eficiência na extração de características e na modelagem das regiões de interesse.

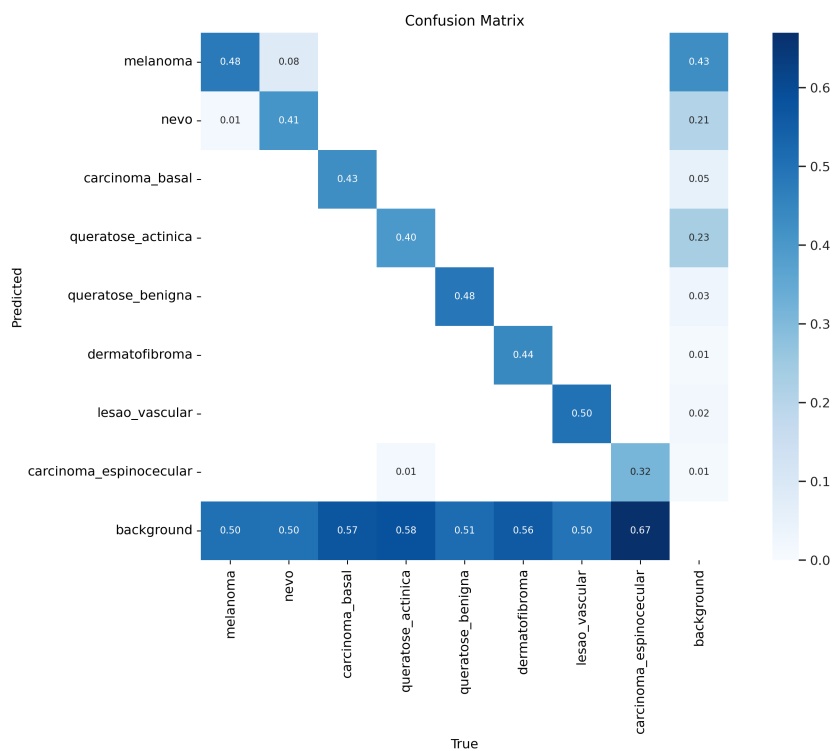
A análise das matrizes de confusão obtidas nos experimentos permite uma avaliação detalhada do desempenho do YOLOv11s<sup>3</sup> na identificação das distintas categorias de lesões cutâneas, evidenciando tanto os acertos quanto os principais erros de classificação. Esse tipo de análise é particularmente relevante em cenários de diagnóstico médico assistido por computador, pois possibilita compreender como o modelo se comporta diante de classes visualmente semelhantes e quais categorias apresentam maior propensão a confusões. Nesse sentido, o primeiro teste realizado com o YOLOv11s<sup>3</sup> mostrou-se fundamental para mapear os padrões de erro mais recorrentes, fornecendo subsídios importantes para a proposição de ajustes futuros, tanto no balanceamento e na representatividade da base de dados quanto na configuração da arquitetura de detecção e dos hiperparâmetros de treinamento.

Para a realização deste teste inicial, foi adotada a mesma distribuição de imagens por classe apresentada na Tabela 4, assegurando consistência metodológica e permitindo

uma comparação direta, controlada e justa com os resultados previamente obtidos com o modelo YOLOv5<sup>2</sup>. Dessa forma, garante-se que as diferenças observadas no desempenho entre os modelos estejam associadas predominantemente às características arquiteturais e à capacidade de generalização do YOLOv11s<sup>3</sup>, e não a variações na composição do conjunto de dados.

A Figura 14 apresenta a primeira matriz de confusão obtida a partir dos experimentos realizados com o modelo YOLOv11s<sup>3</sup>, a qual ilustra a distribuição das previsões do modelo em relação às classes reais das lesões cutâneas analisadas. Nessa representação, as linhas correspondem às classes verdadeiras, enquanto as colunas indicam as classes preditas pelo modelo, permitindo uma visualização detalhada do comportamento do classificador para cada categoria avaliada.

Figura 14 – Matriz de Confusão Inicial



Fonte: Autor (2024).

Os valores localizados na diagonal principal da matriz de confusão representam as amostras corretamente classificadas, refletindo a capacidade do modelo em reconhecer cada tipo de lesão cutânea. Em contrapartida, os valores fora da diagonal indicam confusões entre classes, associadas principalmente à similaridade visual entre determinadas lesões, à variabilidade morfológica e ao desbalanceamento do conjunto de dados.

De modo geral, uma maior concentração de valores ao longo da diagonal principal indica melhor desempenho do modelo. Nesse sentido, a matriz de confusão constitui uma ferramenta fundamental para a avaliação qualitativa do YOLOv11s<sup>3</sup>, permitindo identificar padrões de erro recorrentes. Observa-se que classes clinicamente relevantes, como *melanoma* e *carcinoma basocelular*, apresentam taxas de acerto consideráveis, enquanto confusões frequentes entre *nevo* e *melanoma* evidenciam a dificuldade do modelo em diferenciar lesões com características visuais semelhantes.

Com o objetivo de mitigar essas confusões, foi realizada a ampliação do número de imagens das classes *melanoma* e *nevo*. A nova distribuição de imagens por classe é apresentada na Tabela 5.

Tabela 5 – Redistribuição da quantidade de imagens por classe

| Classe                  | Quantidade de Imagens |
|-------------------------|-----------------------|
| Melanoma                | 600                   |
| Nevo                    | 600                   |
| Carcinoma Basocelular   | 200                   |
| Queratose Actínica      | 200                   |
| Queratose Benigna       | 200                   |
| Dermatofibroma          | 200                   |
| Lesão Vascular          | 200                   |
| Carcinoma Espinocelular | 200                   |
| <b>Total</b>            | <b>2.400</b>          |

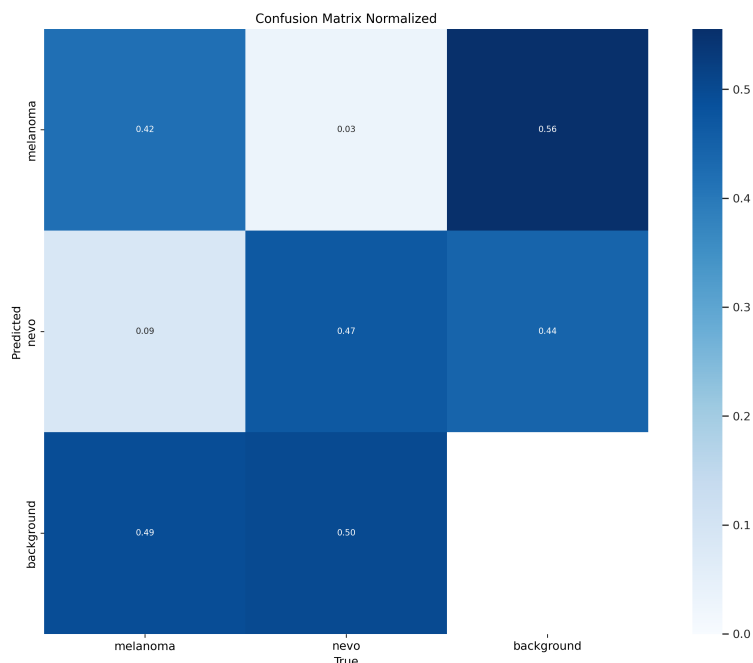
Fonte: Autor (2025).

A Figura 15 apresenta a matriz de confusão obtida com o modelo YOLOv11s<sup>3</sup>, evidenciando que o modelo alcança níveis de precisão satisfatórios em determinadas classes de lesões cutâneas. Entretanto, os resultados também revelam a persistência de desafios relacionados à correta identificação de algumas categorias dermatológicas, especialmente aquelas que apresentam maior variabilidade morfológica ou menor representatividade no conjunto de treinamento.

Observa-se, de forma recorrente, uma elevada taxa de predições atribuídas à classe *background* para diferentes tipos de lesões, o que indica que o modelo tende a não detectar parte das instâncias presentes nas imagens. Esse comportamento sugere limitações na sensibilidade do modelo e aponta para a necessidade de ajustes adicionais no processo de treinamento. Entre as estratégias possíveis, destacam-se o aprimoramento

do balanceamento do conjunto de dados e o aumento do número de amostras para classes sub-representadas, visando melhorar a especificidade e a capacidade de detecção do modelo.

Figura 15 – Matriz de Confusão Após Redistribuição do Conjunto de Dados



Fonte: Autor (2025).

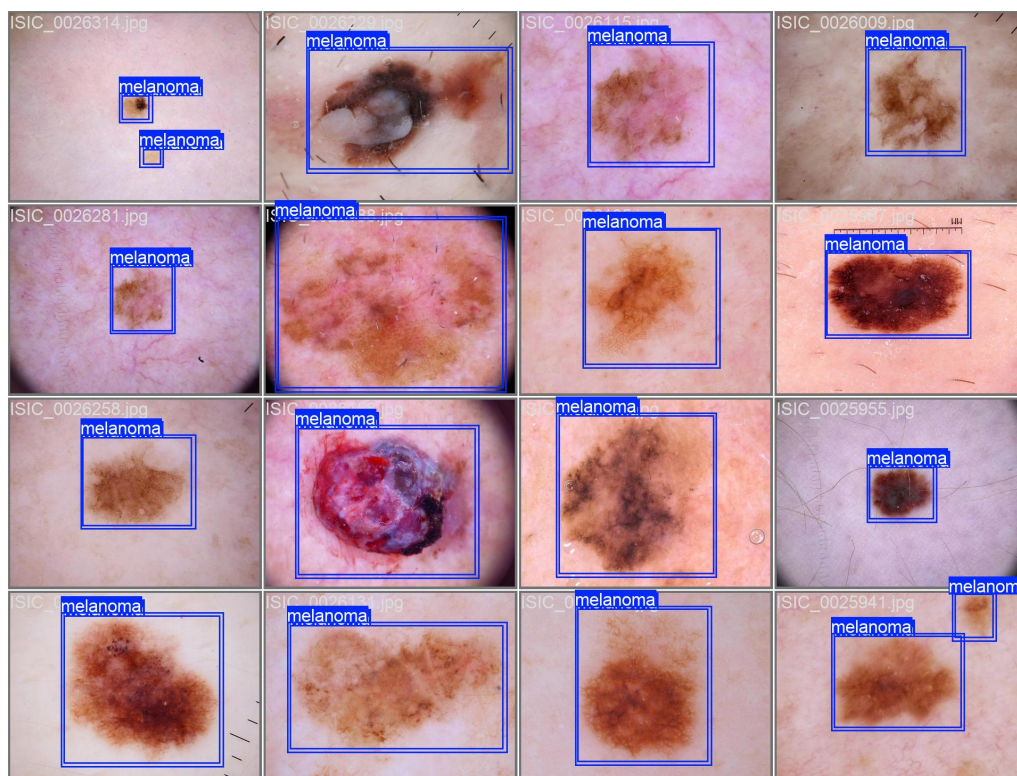
Além das análises quantitativas apresentadas anteriormente, foram conduzidas avaliações qualitativas baseadas na inspeção visual das imagens dermatoscópicas contendo as predições realizadas pelo modelo. Essa etapa desempenha um papel fundamental na validação do sistema proposto, uma vez que complementa as métricas numéricas tradicionais e possibilita a análise de aspectos do desempenho do modelo que não são plenamente capturados por indicadores estatísticos. Entre esses aspectos destacam-se o posicionamento espacial das detecções, a precisão das caixas delimitadoras em relação aos contornos das lesões, bem como a coerência entre as características visuais observadas e as classes atribuídas pelo modelo.

A Figura 16 apresenta exemplos representativos das detecções produzidas pelo modelo YOLOv11s<sup>3</sup>, nos quais as lesões cutâneas identificadas são destacadas por meio de caixas delimitadoras (*bounding boxes*) acompanhadas de seus respectivos rótulos de classe. A análise visual desses resultados evidencia que, em grande parte dos casos, o modelo foi capaz de localizar adequadamente as regiões de interesse, delimitando as áreas correspondentes às lesões com boa aderência às suas estruturas visuais principais.

Além disso, observa-se uma correspondência satisfatória entre os padrões morfológicos presentes nas imagens e as classes previstas, tanto para lesões benignas quanto malignas.

Esses resultados qualitativos reforçam as evidências obtidas nas análises quantitativas, indicando que o modelo não apenas alcançou bons valores de desempenho em termos de métricas globais, mas também apresentou comportamento consistente do ponto de vista visual. Tal consistência é especialmente relevante em aplicações de apoio ao diagnóstico médico, nas quais a interpretabilidade e a confiabilidade das detecções desempenham um papel crucial para a aceitação e o uso prático de sistemas baseados em inteligência artificial.

Figura 16 – Exemplos de detecção de lesões cutâneas realizada pelo modelo YOLOv11s<sup>3</sup>



Fonte: Autor (2025).

As imagens analisadas demonstram que o modelo é capaz de localizar e classificar adequadamente diferentes tipos de lesões cutâneas, reforçando o potencial da abordagem proposta para aplicações no apoio ao diagnóstico médico. A correta delimitação das regiões de interesse e a atribuição coerente das classes em diversos exemplos evidenciam a capacidade do modelo em aprender padrões relevantes a partir das imagens dermatoscópicas, mesmo diante da variabilidade visual inerente a esse tipo de dado.

Os ajustes identificados a partir dessas análises fazem parte do processo iterativo de desenvolvimento do modelo e são fundamentais para garantir que a solução

proposta seja confiável e aplicável em contextos clínicos reais. Como trabalhos futuros, destacam-se o aprimoramento da fase de treinamento por meio de ajustes nos hiperparâmetros, o aumento da diversidade e do balanceamento do conjunto de dados, visando mitigar confusões entre classes visualmente semelhantes, e a exploração de estratégias adicionais de pré-processamento. Além disso, o refinamento da segmentação das imagens e a adoção de técnicas que evidenciem características relevantes das lesões podem contribuir para a obtenção de um desempenho mais robusto e consistente.

### 4.3 Resultados Finais Obtidos com o YOLOv11s<sup>3</sup>

Após a ampliação do conjunto de dados e o refinamento dos parâmetros de treinamento, os resultados finais obtidos com o modelo YOLOv11s<sup>3</sup> evidenciaram avanços expressivos no desempenho geral da detecção e classificação das lesões cutâneas. Essa etapa final teve como objetivo consolidar os aprendizados provenientes dos experimentos anteriores, incorporando ajustes tanto na composição da base de dados quanto na configuração do processo de treinamento, de modo a tornar o modelo mais robusto e estável.

A ampliação do número de imagens, aliada a uma distribuição mais equilibrada entre as classes, contribuiu diretamente para a redução de vieses relacionados ao desbalanceamento do conjunto de dados, favorecendo uma melhor capacidade de generalização do modelo. Esse balanceamento permitiu que o YOLOv11s<sup>3</sup> fosse exposto a uma maior diversidade de padrões visuais durante o treinamento, incluindo variações de textura, coloração, forma e tamanho das lesões, aspectos fundamentais para a correta distinção entre categorias clinicamente semelhantes.

Além disso, o refinamento dos parâmetros de treinamento possibilitou uma convergência mais estável do processo de aprendizado, refletindo-se em melhorias consistentes nas métricas de desempenho avaliadas, como precisão, *recall* e *F1-score*. Esses ganhos indicam que o modelo passou a apresentar maior sensibilidade na identificação de lesões malignas, ao mesmo tempo em que manteve níveis adequados de especificidade na classificação de lesões benignas, reduzindo a ocorrência de falsos positivos e falsos negativos.

A nova configuração do conjunto de dados, resultante dessas alterações, é apresentada na Tabela 6, que sintetiza a distribuição final de imagens por classe utilizada nos experimentos conclusivos. Essa base serviu de referência para a obtenção dos

resultados finais, permitindo uma análise mais consistente e comparável do desempenho do modelo proposto.

Tabela 6 – Distribuição final da quantidade de imagens por classe

| <b>Classe</b>           | <b>Quantidade de Imagens</b> |
|-------------------------|------------------------------|
| Melanoma                | 1200                         |
| Nevo                    | 600                          |
| Carcinoma Basocelular   | 200                          |
| Queratose Actínica      | 200                          |
| Queratose Benigna       | 200                          |
| Dermatofibroma          | 239                          |
| Lesão Vascular          | 200                          |
| Carcinoma Espinocelular | 600                          |
| <b>Total</b>            | <b>3.439</b>                 |

Fonte: Autor (2025).

O aumento do número de amostras, especialmente nas classes *melanoma*, *dermatofibroma* e *carcinoma espinocelular*, teve como objetivo principal aprimorar a representatividade dessas categorias no conjunto de dados, reduzindo o impacto do desbalanceamento entre classes e, conseqüentemente, as confusões observadas nas matrizes de confusão obtidas nas etapas preliminares do estudo. Ao ampliar essas classes, buscou-se fornecer ao modelo um conjunto mais diversificado de exemplos, contemplando variações clínicas e morfológicas relevantes, como diferenças de coloração, textura, tamanho e padrões estruturais das lesões.

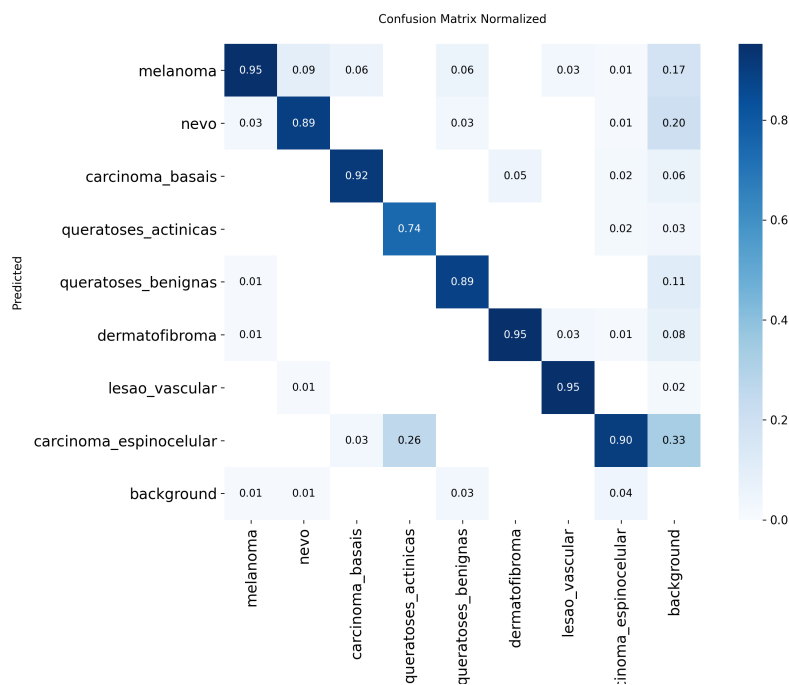
Esse processo de enriquecimento da base de dados contribuiu para uma aprendizagem mais consistente das características discriminativas associadas a cada classe, favorecendo a definição de fronteiras de decisão mais precisas pelo modelo. Como resultado, observou-se uma tendência à redução de erros de classificação, particularmente entre lesões com características visuais semelhantes, aspecto fundamental em um cenário de diagnóstico assistido por computador.

A partir desses ajustes, obteve-se uma base final composta por 3.439 imagens, a qual foi utilizada nas etapas de treinamento e validação do modelo final. Essa configuração possibilitou uma avaliação mais robusta do desempenho do YOLOv11s<sup>3</sup>, refletindo de forma mais fiel sua capacidade de generalização e seu potencial de aplicação no apoio ao diagnóstico de lesões cutâneas.



A Figura 17 apresenta a matriz de confusão obtida após a etapa de treinamento definitivo do modelo YOLOv11s<sup>3</sup>. Observa-se uma concentração mais acentuada dos valores ao longo da diagonal principal, o que indica um aumento da acurácia global e uma melhora significativa na capacidade do modelo em distinguir corretamente entre as diferentes classes de lesões cutâneas analisadas. Esse comportamento sugere que o modelo passou a aprender representações mais consistentes e discriminativas para cada categoria, refletindo os benefícios das estratégias de balanceamento da base de dados e do refinamento dos parâmetros de treinamento.

Figura 17 – Matriz de Confusão Final



Fonte: Autor (2025).

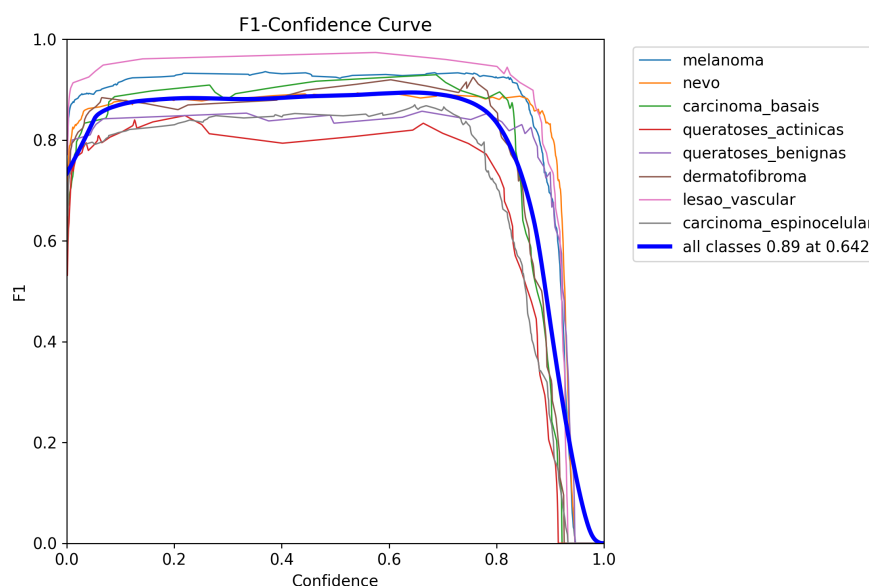
Em comparação com as matrizes de confusão obtidas nas etapas anteriores, nota-se uma redução expressiva nas confusões entre classes visualmente semelhantes, evidenciando uma evolução no desempenho do classificador. Destaca-se, em particular, a diminuição dos erros de classificação envolvendo as classes *melanoma* e *nevo*, que historicamente representam um desafio para sistemas automatizados de diagnóstico devido à sobreposição de características morfológicas em imagens dermatoscópicas. A melhora observada nessas categorias é especialmente relevante do ponto de vista clínico, uma vez que contribui para a redução tanto de falsos negativos, críticos no caso do melanoma, quanto de falsos positivos, que podem levar a procedimentos desnecessários.

De forma geral, os resultados evidenciam que os ajustes realizados ao longo do

processo experimental foram eficazes em aprimorar o desempenho do modelo, resultando em um sistema mais robusto e confiável para a tarefa de detecção e classificação de lesões cutâneas. Esses achados reforçam o potencial do YOLOv11s<sup>3</sup> como ferramenta de apoio ao diagnóstico, especialmente quando aliado a bases de dados mais equilibradas e a estratégias de treinamento adequadamente configuradas.

A Figura 18 apresenta os valores de *F1-score* para cada classe do conjunto de dados. Essa métrica combina *precision* e *recall* em um único indicador, permitindo avaliar simultaneamente a capacidade do modelo de evitar falsos positivos e falsos negativos. Observa-se que classes com maior representatividade no conjunto de treinamento, como melanoma e nevo, apresentaram *F1-scores* mais elevados, evidenciando melhor equilíbrio entre acerto e cobertura. Já classes menos frequentes ou mais heterogêneas, como carcinoma basocelular e lesões vasculares, demonstraram valores inferiores, o que reflete a dificuldade do modelo em generalizar nesses cenários.

Figura 18 – Curva de *F1-Score* por classe



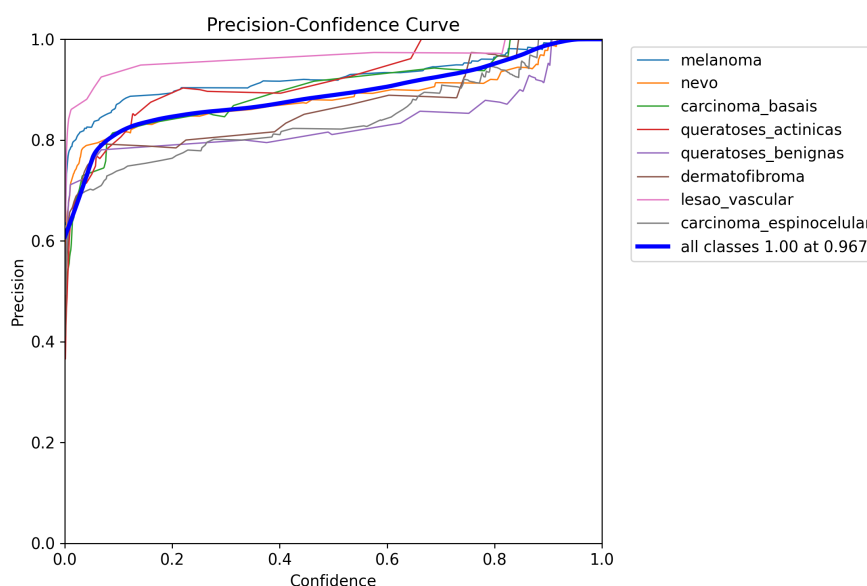
Fonte: Autor (2025).

A Figura 19 apresenta o desempenho do modelo em termos de *precision*, métrica que expressa a proporção de predições positivas que foram corretamente classificadas, refletindo diretamente a capacidade do modelo em evitar falsos positivos. Valores elevados de *precision* indicam que, quando o modelo identifica uma determinada lesão, há uma alta probabilidade de que essa predição esteja correta, aspecto especialmente relevante em aplicações médicas, nas quais diagnósticos incorretos podem resultar em encaminhamentos ou intervenções desnecessárias.

Observa-se que algumas classes alcançaram valores elevados de *precision*, evidenciando que o modelo foi capaz de aprender padrões discriminativos consistentes para essas categorias, reduzindo significativamente a ocorrência de detecções equivocadas. Esse comportamento sugere uma boa separabilidade entre determinadas classes e o *background*, bem como entre classes com características morfológicas mais bem definidas no conjunto de treinamento.

Por outro lado, nota-se que classes com menor número de amostras disponíveis apresentaram maior variabilidade nos valores de *precision*. Esse resultado é consistente com cenários de desbalanceamento do conjunto de dados, nos quais a escassez de exemplos limita a capacidade do modelo de generalizar adequadamente, aumentando a probabilidade de confusão com outras classes visualmente semelhantes. Além disso, a presença de padrões visuais sobrepostos entre diferentes tipos de lesões cutâneas pode contribuir para esse comportamento, especialmente em categorias com bordas pouco definidas ou coloração semelhante ao tecido adjacente.

Figura 19 – Curva de Precisão por Classe



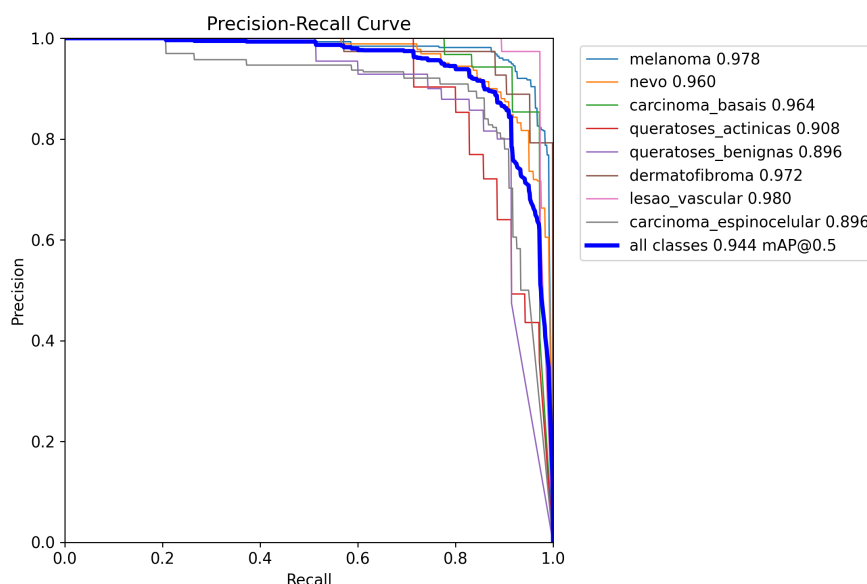
Fonte: Autor (2025).

A Figura 20 apresenta a relação entre as métricas de *precision* e *recall* para cada classe avaliada, permitindo analisar o desempenho do modelo sob diferentes limiares de confiança. Essa representação evidencia o *trade-off* inerente entre essas métricas, uma vez que o aumento do *recall*, associado à detecção de um maior número de instâncias positivas, pode resultar na redução da *precision*, indicando maior ocorrência de falsos positivos.

Observa-se que, para limiares de confiança mais baixos, o modelo tende a maximizar o *recall*, enquanto limiares mais elevados tornam as predições mais conservadoras, favorecendo a *precision* em detrimento da perda de algumas detecções verdadeiras. Curvas posicionadas mais próximas do canto superior direito indicam melhor desempenho, ao passo que curvas mais afastadas dessa região refletem maior dificuldade de detecção, geralmente associada ao desbalanceamento do conjunto de dados ou à similaridade visual entre classes.

Nesse contexto, as curvas *precision-recall* fornecem uma avaliação mais abrangente do desempenho por classe, sendo particularmente relevantes em aplicações médicas, nas quais é necessário equilibrar sensibilidade e precisão.

Figura 20 – Curva de *Precision-Recall* por Classe



Fonte: Autor (2025).

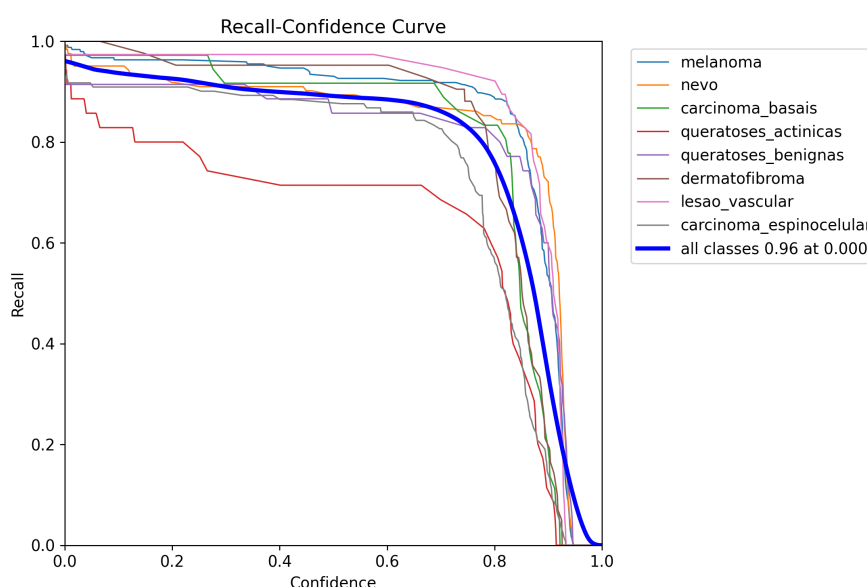
A Figura 21 apresenta os valores de *recall* obtidos para cada classe analisada, métrica que representa a proporção de instâncias positivas corretamente identificadas pelo modelo em relação ao total de casos reais existentes. Em outras palavras, o *recall* expressa a capacidade do modelo em detectar efetivamente as lesões presentes nas imagens, estando diretamente associado à redução de falsos negativos.

Essa métrica assume papel central em cenários clínicos, especialmente no contexto do diagnóstico de câncer de pele, no qual a não detecção de uma lesão potencialmente maligna pode acarretar consequências clínicas significativas. Dessa forma, valores elevados de *recall* são desejáveis, pois indicam maior sensibilidade do modelo na identificação de lesões suspeitas, mesmo que isso implique um aumento

moderado no número de falsos positivos.

Observa-se que as classes com maior volume de imagens no conjunto de treinamento apresentaram os melhores valores de *recall*, evidenciando que a disponibilidade de exemplos variados contribui para uma melhor capacidade de generalização do modelo. Por outro lado, classes minoritárias ou com menor diversidade de padrões visuais exibiram valores reduzidos de *recall*, indicando que o modelo ainda apresenta dificuldades em reconhecer parte dessas lesões, resultando em instâncias não detectadas.

Figura 21 – Curva de *Recall* por Classe



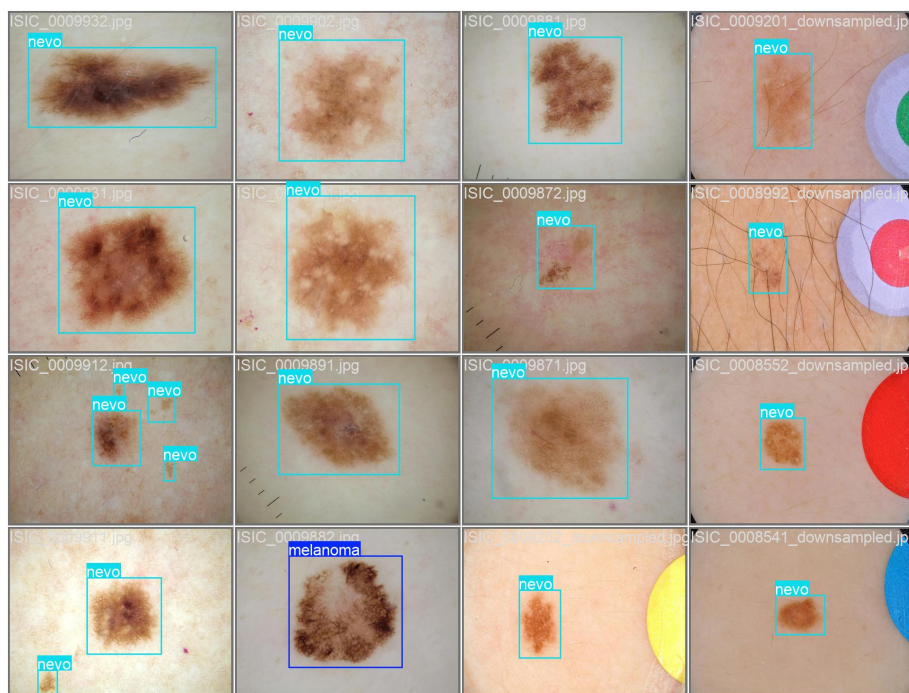
Fonte: Autor (2025).

Observou-se que as classes relacionadas às queratoses apresentaram desempenho inferior em comparação às demais categorias, especialmente em termos de *recall*. Esse comportamento pode ser atribuído, principalmente, à elevada variabilidade morfológica dessas lesões, que frequentemente compartilham características visuais com outras classes benignas e malignas, dificultando a distinção automática pelo modelo.

Adicionalmente, limitações relacionadas ao tempo de treinamento disponível impactaram a capacidade do modelo em aprender padrões mais sutis dessas classes. Para lesões com maior complexidade visual, como as queratoses, seria desejável um maior número de épocas e maior diversidade de dados, possibilitando um aprendizado mais robusto. Contudo, restrições computacionais e temporais impediram a realização de treinamentos mais prolongados, o que contribuiu para o desempenho reduzido observado nessas classes.

Além da análise quantitativa, a Figura 22 demonstra exemplos qualitativos das detecções realizadas pelo modelo. As *bounding boxes* exibem delimitações precisas das áreas de interesse, evidenciando a capacidade do YOLOv11s<sup>3</sup> em identificar e classificar corretamente diferentes tipos de lesões cutâneas em imagens reais.

Figura 22 – Exemplos de detecções realizadas pelo modelo YOLOv11s<sup>3</sup>



Fonte: Autor (2025).

Os treinamentos dos modelos foram realizados em diferentes ambientes computacionais, de forma progressiva. Inicialmente, os experimentos exploratórios e os treinamentos preliminares foram conduzidos em ambiente Jupyter, com o objetivo de validação da configuração do modelo, verificação das anotações e ajuste inicial dos hiperparâmetros definidos na 3.6.2. Posteriormente, os treinamentos finais foram executados na infraestrutura computacional Saturno, equipada com GPU NVIDIA A100, permitindo maior capacidade de processamento e redução significativa do tempo de treinamento.

Nesse contexto, o modelo YOLOv5<sup>2</sup> demandou aproximadamente 6 a 8 horas de treinamento, apresentando convergência relativamente rápida nas fases iniciais, porém com limitações no desempenho final alcançado. Já o modelo YOLOv11s<sup>3</sup> apresentou tempo de treinamento significativamente menor, variando entre 1 e 2 horas, mesmo considerando sua maior complexidade arquitetural e o aumento do volume de dados empregado nas etapas finais.

A expressiva redução no tempo de treinamento está diretamente associada ao uso da GPU NVIDIA A100, que possibilitou maior paralelismo, melhor gerenciamento de memória e maior eficiência computacional durante o processo de aprendizado. Esses fatores permitiram acelerar a convergência do modelo, sem prejuízo à qualidade das predições.

Apesar do menor tempo de treinamento, o modelo YOLOv11s<sup>3</sup> apresentou ganhos expressivos em termos de *precision*, *recall* e *F1-score*, superando o desempenho do YOLOv5<sup>2</sup>. Esses resultados justificam sua adoção no contexto de aplicações médicas, nas quais a combinação entre elevado desempenho preditivo e eficiência computacional é um fator crítico para a viabilidade de uso em ambientes reais.

#### 4.3.1 Comparação entre YOLOv5<sup>2</sup> e YOLOv11s<sup>3</sup>

Além da análise evolutiva dos resultados ao longo das diferentes etapas de ajuste do conjunto de dados e do treinamento, realizou-se uma comparação direta entre as arquiteturas YOLOv5<sup>2</sup> e YOLOv11s<sup>3</sup>, com o objetivo de evidenciar os ganhos obtidos com a adoção de uma versão mais recente do modelo. Ambos os modelos foram treinados sob as mesmas condições experimentais, utilizando o mesmo conjunto de dados, resolução de entrada, número de épocas, critérios de validação e estratégia de parada antecipada, garantindo assim uma comparação justa e controlada.

De modo geral, observou-se que o YOLOv11s<sup>3</sup> apresentou desempenho superior ao YOLOv5<sup>2</sup> em todas as métricas avaliadas. As matrizes de confusão indicaram uma redução significativa das confusões entre classes visualmente semelhantes, especialmente entre melanoma e nevo, que representam um dos principais desafios no diagnóstico automatizado de lesões cutâneas. Enquanto o YOLOv5<sup>2</sup> apresentou maior dispersão de erros fora da diagonal principal, o YOLOv11s<sup>3</sup> demonstrou maior concentração de acertos, refletindo melhor capacidade discriminativa.

Essa melhoria também se reflete nas métricas quantitativas, como precisão, *recall* e *F1-score*, nas quais o YOLOv11s<sup>3</sup> alcançou valores consistentemente mais elevados. Em especial, o aumento do *recall* é um aspecto relevante no contexto clínico, uma vez que reduz a ocorrência de falsos negativos, contribuindo para a detecção mais segura de lesões potencialmente malignas. Adicionalmente, o YOLOv11s<sup>3</sup> apresentou maior estabilidade durante o treinamento e melhor qualidade na delimitação das *bounding boxes*, indicando maior robustez e capacidade de generalização.

A Tabela 7 sintetiza a comparação entre os dois modelos, destacando as principais diferenças observadas em termos de desempenho e comportamento durante o treinamento.

Tabela 7 – Comparação de desempenho entre YOLOv5<sup>2</sup> e YOLOv11s<sup>3</sup>

| Métrica                             | YOLOv5 <sup>2</sup> | YOLOv11s <sup>3</sup> |
|-------------------------------------|---------------------|-----------------------|
| mAP                                 | 0.82                | 0.94                  |
| Precision                           | 0.85                | 0.96                  |
| Recall                              | 0.78                | 0.96                  |
| F1-score                            | 0.81                | 0.89                  |
| Confusão entre melanoma e nevo      | Frequente           | Reduzida              |
| Estabilidade durante o treinamento  | Moderada            | Alta                  |
| Qualidade das <i>bounding boxes</i> | Boa                 | Superior              |

Fonte: Autor (2025).



## 5 CONSIDERAÇÕES FINAIS

O presente Trabalho de Conclusão de Curso teve como objetivo geral desenvolver, testar e avaliar um modelo de detecção de lesões cutâneas baseado na arquitetura YOLO, com o propósito de auxiliar no diagnóstico precoce de câncer de pele a partir de imagens dermatológicas. Após todas as etapas realizadas, conclui-se que o objetivo geral foi plenamente atingido, uma vez que o modelo implementado demonstrou desempenho consistente, capacidade de generalização e potencial aplicação prática em ambientes reais de triagem dermatológica.

Quanto aos objetivos específicos, todos foram alcançados de forma satisfatória e contribuíram diretamente para o cumprimento da proposta central do trabalho. Inicialmente, foi realizada a coleta, organização e preparação de um banco de dados dermatológico robusto, contendo imagens rotuladas de diferentes tipos de lesões, entre elas melanoma, carcinoma basocelular, carcinoma espinocelular, queratoses, nevus e lesões benignas de interesse clínico. Esse processo envolveu seleção criteriosa, padronização de formatos, verificação de qualidade e distribuição balanceada entre classes, garantindo a integridade e a representatividade do conjunto de dados utilizado.

Em seguida, foi desenvolvido e treinado um modelo de detecção automática utilizando a arquitetura YOLO, uma abordagem amplamente reconhecida na área de visão computacional por sua capacidade de identificar e localizar múltiplos objetos em tempo real. O processo envolveu a definição da estrutura de treinamento, adequação do formato das anotações, implementação de técnicas de *data augmentation* e escolha de parâmetros adequados para o equilíbrio entre desempenho e consumo computacional. Posteriormente, foram realizados ajustes finos de hiperparâmetros, buscando otimizar a acurácia e a capacidade de detecção do modelo frente a variações de contexto, iluminação, foco e formato das imagens.

Outro objetivo específico consistiu na avaliação do desempenho do modelo por meio de métricas estatísticas apropriadas, incluindo acurácia, *precision*, *recall* e *mAP*. Essa análise permitiu validar a eficiência do sistema e comparar seu desempenho com resultados reportados em estudos recentes, demonstrando que o modelo apresenta desempenho compatível com abordagens de referência na literatura. Por fim, foi desenvolvido um aplicativo *mobile* para a plataforma *Android*, com interface gráfica simples, funcional e intuitiva, projetado para viabilizar a aplicação prática do modelo de detecção de lesões cutâneas em um cenário real de uso. O aplicativo permite a captura ou

seleção de imagens dermatológicas diretamente no dispositivo móvel e realiza a inferência de forma local, retornando a predição do modelo de maneira clara e acessível ao usuário.

A implementação do aplicativo representa uma etapa fundamental de validação tecnológica do trabalho, ao demonstrar a possibilidade de integração do modelo de *deep learning* em um sistema embarcado, sem dependência de conexão com a internet. Dessa forma, o sistema consolida-se como um protótipo funcional de apoio à triagem dermatológica, com potencial de uso educacional, acadêmico e assistencial.

Em termos de contribuições científicas, tecnológicas e acadêmicas, o trabalho apresenta avanços em diferentes dimensões. A nível técnico, oferece um pipeline completo de detecção de lesões cutâneas utilizando *deep learning*, desde a preparação da base de dados até a implementação do sistema final, constituindo material de referência para pesquisas futuras. Do ponto de vista científico, destaca-se a aceitação de um artigo na Revista Brasileira de Cancerologia, atualmente em fase de edição de texto, estando disponibilizado no Apêndice A, o que evidencia a relevância e originalidade do estudo. A aceitação de um resumo no Salão Internacional de Ensino, Pesquisa e Extensão (SIEPE) reforça o impacto acadêmico do trabalho e sua contribuição para a disseminação de conhecimento na área de inteligência artificial aplicada à saúde. A nível social, a pesquisa demonstra o potencial de ferramentas automatizadas no apoio ao diagnóstico precoce do câncer de pele, particularmente em contextos com baixa disponibilidade de dermatologistas.

Apesar dos resultados promissores, algumas limitações foram identificadas ao longo do processo. A primeira refere-se à dependência da base ISIC 2019, que, embora extensa, não representa toda a diversidade epidemiológica, fototípica e geográfica das populações brasileiras. Essa limitação pode afetar a capacidade de generalização do modelo em cenários reais, especialmente em peles com alta variabilidade de tonalidade e características morfológicas específicas. Outra limitação importante envolve a sensibilidade do modelo a condições adversas de captura, como baixa resolução, excesso de sombras, reflexos, ruídos e variações de iluminação, o que reforça a necessidade de aprimoramento nos métodos de pré-processamento. A interface *mobile*, embora funcional, ainda opera como um protótipo e carece de mecanismos adicionais de segurança, usabilidade e validação clínica.

Com base nessas observações, algumas sugestões de trabalhos futuros são propostas. Primeiramente, recomenda-se a ampliação da base de dados com imagens coletadas localmente, abrangendo maior diversidade populacional, condições clínicas e

variedades de equipamentos de captura. Também se sugere a exploração de modelos híbridos, combinando *deep learning* com informações clínicas complementares, como histórico do paciente, sintomas ou localização anatômica. Ademais, estudos futuros podem incorporar técnicas de interpretabilidade, como *explainable AI*, permitindo que profissionais de saúde compreendam melhor como o modelo toma suas decisões, o que aumenta a confiança na ferramenta. Outra direção promissora consiste na validação do modelo em ambientes clínicos reais e no aperfeiçoamento da interface *mobile*, transformando-a em um aplicativo completo, seguro e apto a integrar fluxos reais de atendimento e triagem dermatológica.

Em síntese, este trabalho cumpriu integralmente seu objetivo geral e todos os seus objetivos específicos, resultando em um sistema funcional, tecnicamente validado no que se refere ao desempenho do modelo de detecção, e cientificamente reconhecido. As contribuições geradas, tanto as tecnológicas quanto as acadêmicas, demonstram o potencial da integração entre IA e diagnóstico clínico, reforçando importância de pesquisas que aproximam tecnologias emergentes da prática em saúde. Os resultados alcançados, as limitações identificadas e os caminhos propostos para trabalhos futuros evidenciam a maturidade da solução desenvolvida e indicam um cenário favorável ao avanço contínuo de métodos automatizados na detecção precoce do câncer de pele.

## REFERÊNCIAS

- American Academy of Dermatology. **Skin Cancer Diagnosis and Management**. 2023. <<https://www.aad.org/>>. Acesso em: 15 junho de 2025.
- ARMSTRONG, B. K.; KRICKER, A. The epidemiology of uv induced skin cancer. **Journal of Photochemistry and Photobiology B: Biology**, v. 215, p. 112115, 2021. Acesso em: 21 de junho de 2025. Disponível em: [https://doi.org/10.1016/s1011-1344\(01\)00198-1](https://doi.org/10.1016/s1011-1344(01)00198-1).
- ASHRAF, R. *et al.* Region-of-interest based transfer learning assisted framework for skin cancer detection. **IEEE Access**, v. 8, p. 147858–147871, 2020. Acesso em: 15 de junho de 2025. Disponível em: <https://doi.org/10.1109/ACCESS.2020.3014701>.
- AVANZO, M. *et al.* The evolution of artificial intelligence in medical imaging: From computer science to machine and deep learning. **Cancers**, v. 16, n. 21, 2024. ISSN 2072-6694. Acesso em: 22 de juho de 2025. Disponível em: <https://www.mdpi.com/2072-6694/16/21/3702>.
- AZULAY, A. A.; AZULAY, D. D.; AZULAY, R. R. **Dermatologia**. 7. ed. Rio de Janeiro: Guanabara Koogan, 2013. Acesso em: 02 de maio de 2025.
- BALCH, C. M. *et al.* A multifactorial analysis of melanoma: prognostic histopathological features comparing clark's and breslow's staging methods. **Annals of Surgery**, v. 188, n. 6, p. 732–742, 1978. Acesso em: 20 de junho de 2025.
- BELTRAMI, E. J. *et al.* Artificial intelligence in the detection of skin cancer. **Journal of the American Academy of Dermatology**, v. 87, n. 6, p. 1336–1342, 2022. Acesso em: 25 de maio de 2025. Disponível em: [https://www.internationaljournalofcardiology.com/article/S0167-5273\(22\)00671-4/abstract](https://www.internationaljournalofcardiology.com/article/S0167-5273(22)00671-4/abstract).
- BENJAMENS, S.; DHUNNOO, P.; MESKÓ, B. The state of artificial intelligence-based fda-approved medical devices and algorithms: An online database. **npj Digital Medicine**, Nature Publishing Group, v. 3, n. 1, p. 1–8, 2020. Acesso em: 20 de junho de 2025.
- BORGES, A. L. *et al.* Melanocytic nevi with special features: clinical-dermoscopic and reflectance confocal microscopic-findings. **Journal of the European Academy of Dermatology and Venereology**, Wiley Online Library, v. 28, n. 7, p. 833–845, 2014. Acesso em: 01 maio 2025. Disponível em: <https://doi.org/10.1111/jdv.12291>.
- BOTTON, D. V. *et al.* Relevância da dermatoscopia para o diagnóstico precoce de melanomas: uma revisão de literatura. **Revista Interdisciplinar de Saúde e Educação**, Centro Universitário Barão de Mauá, v. 1, n. 2, p. 159–174, 2020. Acesso em: 02 de maio de 2025. Disponível em: <https://periodicos.baraodemaua.br/index.php/cse/article/view/104>.
- Brasil. Conselho Nacional de Saúde. **Resolução nº 466, de 12 de dezembro de 2012: Aprova diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos**. 2013. Diário Oficial da União. Seção 1, v. 150, n. 112, p. 59–62. Disponível em: [https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466\\_12\\_12\\_2012.html](https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html). Acesso em: 27 ago. 2025.

Brasil. Conselho Nacional de Saúde. **Resolução nº 510, de 7 de abril de 2016: Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais.** 2016. Diário Oficial da União. Seção 1, v. 153, n. 98, p. 44–46. Disponível em: [https://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510\\_07\\_04\\_2016.html](https://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510_07_04_2016.html). Acesso em: 27 ago. 2025.

BRINKER, T. J. *et al.* Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. **European Journal of Cancer**, v. 113, p. 47–54, 2019. Acesso em: 17 de junho de 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0959804919302597>.

CARRERA, C.; PUIG, S.; MALVEHY, J. In vivo confocal reflectance microscopy in melanoma. **Dermatologic Therapy**, v. 25, n. 5, p. 410–422, 2012. Acesso em: 20 de junho de 2025.

CHEN, J. Y. *et al.* Skin cancer diagnosis by lesion, physician, and examination type: A systematic review and meta-analysis. **JAMA Dermatology**, v. 161, n. 2, p. 135–146, 2025. Acesso em: 15 de junho de 2025. Disponível em: <https://doi.org/10.1001/jamadermatol.2024.4382>.

CHEN, X. *et al.* Recent advances and clinical applications of deep learning in medical image analysis. **Frontiers in Medicine**, v. 8, p. 683073, 2021. Acesso em: 21 de junho de 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1361841522000913>.

CHU, Y. S. *et al.* Artificial intelligence in cutaneous oncology. **Frontiers in Medicine**, v. 7, p. 318, 2020. Acesso em: 15 de abril de 2025. Disponível em: <https://doi.org/10.3389/fmed.2020.00318>.

CODELLA, N. C. F. *et al.* Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [s.n.], 2018. p. 168–172. Acesso em: 25 de maio de 2025. Disponível em: <https://ieeexplore.ieee.org/document/8363547>.

COLLABORATION, I. S. I. **ISIC 2019: Skin Lesion Analysis Towards Melanoma Detection – Challenge Dataset**. 2019. Acesso em 18 de setembro de 2024. Disponível em: <https://challenge2019.isic-archive.com/>.

CUI, X. *et al.* Assessing the effectiveness of artificial intelligence methods for melanoma: A retrospective review. **Journal of the American Academy of Dermatology**, v. 81, n. 5, p. 1176–1180, 2019. Acesso em: 15 de abril de 2025. Disponível em: <https://doi.org/10.1016/j.jaad.2019.06.042>.

DELGADO, L. P. *et al.* Epidemiologia e manejo clínico do câncer de pele melanoma: uma revisão narrativa. **Journal of Social Issues and Health Sciences (JSIHS)**, v. 1, n. 7, 2024. Acesso em: 28 de abril 2025. Disponível em: <https://ojs.thesiseditora.com.br/index.php/jsihs/article/view/242>.

DOSOVITSKIY, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. In: **International Conference on Learning Representations**. [s.n.], 2021. Acesso em: 13 de junho de 2025. Disponível em: <https://arxiv.org/abs/2010.11929>.

DUARTE, A. F. *et al.* Clinical abcde rule for early melanoma detection. **European Journal of Dermatology**, v. 31, n. 6, p. 771–778, 2021. Acesso em: 8 de junho de 2025. Disponível em: <https://doi.org/10.1684/ejd.2021.4171>.

ELSHAHAWEY, M. *et al.* Early melanoma detection based on a hybrid yolov5 and resnet technique. **Diagnostics (Basel, Switzerland)**, v. 13, n. 17, p. 2804, 2023. Acesso em: 25 de maio de 2025. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10486497/>.

ESTEVA, A. *et al.* Deep learning-enabled medical computer vision. **NPJ Digital Medicine**, Nature Publishing Group, v. 4, n. 1, p. 5, 2021. Acesso em: 13 junho. 2025. Disponível em: <https://doi.org/10.1038/s41746-020-00376-2>.

ESTEVA, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, p. 115–118, 2017. Acesso em: 15 de junho de 2025. Disponível em: <https://www.nature.com/articles/nature21056>.

FURRIEL, R. P. S.; SILVA, M. T.; CAMPOS, A. I. Artificial intelligence for skin cancer detection and classification for clinical application: A systematic review. **Frontiers in Medicine**, v. 11, 2024. Acesso em: 25 de maio de 2025. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10800812/>.

GARBE, C.; LEITER, U. Melanoma epidemiology and trends. **Clinical Dermatology**, v. 38, n. 2, p. 153–159, 2020. Acesso em: 21 de junho de 2025. Disponível em: <https://doi.org/10.1016/j.clindermatol.2008.09.001>.

GOLDMAN, L.; FITZPATRICK, T. B. Biopsy techniques in dermatology. In: FITZPATRICK, T. B. (Ed.). **Fitzpatrick's Dermatology in General Medicine**. 9th. ed. New York, NY: McGraw-Hill Education, 2019. Acesso em: 21 de junho de 2025. Disponível em: <https://accessmedicine.mhmedical.com/book.aspx?bookid=2570>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. Acesso em: 31 de maio de 2025. Disponível em: <http://www.deeplearningbook.org>.

GOYAL, A. *et al.* Assessing the diagnosis of “primary cutaneous mantle cell lymphoma”: A systematic review and population-based analysis. **Journal of the American Academy of Dermatology**, Elsevier, v. 87, n. 1, p. 214–216, 2022. ISSN 0190-9622. Acesso em: 15 de junho de 2025. Disponível em: <https://doi.org/10.1016/j.jaad.2021.07.043>.

GREEN, A. C.; WHITEMAN, D. C. Skin cancer: epidemiology, risk factors and prevention. **Melanoma Research**, v. 32, n. 3, p. 203–210, 2022. Acesso em: 21 de junho de 2025. Disponível em: <https://doi.org/10.1097/CMR.0000000000000761>.

GUTMAN, D. *et al.* Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). **arXiv preprint arXiv:1605.01397**, 2016. Acesso em: 08 de junho de 2025. Disponível em: <https://arxiv.org/abs/1605.01397>.

HAENSCHEN, L.; AL. *et.* Artificial intelligence in skin cancer diagnostics: current status and future challenges. **Journal of the European Academy of Dermatology and Venereology**, v. 35, n. 10, p. 2117–2124, 2021. Acesso em: 20 de junho de 2025. Disponível em: <https://doi.org/10.1111/jdv.17363>.

HECKLER, N. *et al.* Applying an artificial intelligence deep learning approach to routine dermatopathological diagnosis of basal cell carcinoma. **Journal der Deutschen Dermatologischen Gesellschaft**, v. 21, n. 9, p. 1329–1337, 2023. Acesso em: 10 de junho de 2025. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/DDG.15180>.

HEINLEIN, L. *et al.* Clinical melanoma diagnosis with artificial intelligence: Insights from a prospective multicenter study. **arXiv preprint arXiv:2401.14193**, 2024. Acesso em: 01 maio 2025. Disponível em: <https://arxiv.org/abs/2401.14193>.

HERATH, H. M. S. S. *et al.* A systematic review of medical image quality assessment. **Journal of Imaging**, v. 11, n. 4, 2025. ISSN 2313-433X. Acesso em: 05 de junho de 2025. Disponível em: <https://www.mdpi.com/2313-433X/11/4/100>.

Instituto Nacional de Câncer José Alencar Gomes da Silva. **Estimativa 2023: incidência de câncer no Brasil**. 2022. <<https://www.inca.gov.br/estimativa>>. Acesso em: 19 jul. 2025.

JAIRATH, N. *et al.* Artificial intelligence in dermatology: A systematic review of its applications in melanoma and keratinocyte carcinoma diagnosis. **Dermatologic Surgery**, v. 50, n. 9, p. 791–798, 2024. Acesso em: 15 de abril de 2025. Disponível em: <https://doi.org/10.1097/DSS.0000000000004223>.

KAWAHARA, J. *et al.* Seven-point checklist and skin lesion classification using multitask multimodal neural nets. **IEEE Journal of Biomedical and Health Informatics**, v. 23, n. 2, p. 538–546, 2019. Acesso em: 08 de junho de 2025. Disponível em: <https://ieeexplore.ieee.org/document/8333693>.

KITCHENHAM, B. A. **Procedures for Performing Systematic Reviews**. [S.l.], 2004. Acesso em 30 de maio de 2025. Disponível em: <https://www.inf.ufsc.br/aldo.vw/kitchenham.pdf>.

KITTLER, H. *et al.* Diagnostic accuracy of dermoscopy. **The Lancet Oncology**, v. 3, n. 3, p. 159–165, 2002. Acesso em: 20 de junho de 2025.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, ACM, v. 60, n. 6, p. 84–90, 2012. Acesso em: 20 de junho de 2025.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, p. 436–444, 2015. Acesso em: 19 de maio de 2025. Disponível em: <https://www.nature.com/articles/nature14539>.

LITJENS, G. *et al.* A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60–88, 2017. Acesso em: 19 de junho de 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1361841517301135>.

LIU, X. *et al.* A survey of deep neural network architectures and their applications. **Neurocomputing**, Elsevier, v. 234, p. 11–26, 2020. Acesso em: 22 de junho de 2025. Disponível em: <https://doi.org/10.1016/j.neucom.2016.12.038>.

MANZO, J. A. *et al.* Use of photography in dermatology: Ethical and legal implications. **Actas Dermo-Sifiliográficas**, v. 111, n. 2, p. 107–114,

2020. ISSN 1578-2190. Acesso em: 29 de maio de 2025. Disponível em: <https://www.actasdermo.org/es-use-photography-in-dermatology-ethical-articulo-S1578219020300226>.

MENDONÇA, T. *et al.* Ph2 - a dermoscopic image database for research and benchmarking. In: **2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. [S.l.: s.n.], 2013. p. 5437–5440. Acesso em: 02 de juho de 2025.

MESKÓ, B. *et al.* Digital health is a cultural transformation of traditional healthcare. **Mhealth**, v. 3, p. 38, 2017. Acesso em: 23 de junho de 2025. Disponível em: <https://doi.org/10.21037/mhealth.2017.08.07>.

MIHULECEA, C.-R. J. *et al.* The many roles of dermoscopy in melanoma detection. **Life**, v. 13, n. 2, 2023. ISSN 2075-1729. Acesso em: 01 de junho de 2025. Disponível em: <https://www.mdpi.com/2075-1729/13/2/477>.

MRIDHA, M. A. R. *et al.* Ai in dermatology: a comprehensive review into skin cancer detection. **PeerJ Computer Science**, v. 9, p. e2530, 2023. Acesso em: 21 de junho de 2025. Disponível em: <https://peerj.com/articles/cs-2530/>.

MUNJAL, G. *et al.* Skinsage xai: An explainable deep learning solution for skin lesion diagnosis. **Health Care Science**, v. 3, n. 6, p. 438–455, 2024. Acesso em: 15 de abril de 2025. Disponível em: <https://doi.org/10.1002/hcs2.121>.

NAGENDRAN, M. *et al.* Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. **BMJ**, BMJ Publishing Group, v. 368, p. m689, 2020. Acesso em: 16 de junho de 2025. Disponível em: <https://doi.org/10.1136/bmj.m689>.

NAQVI, M. *et al.* Skin cancer detection using deep learning—a review. **Diagnostics**, Multidisciplinary Digital Publishing Institute, v. 13, n. 11, p. 1911, 2023. Acesso em: 21 de junho de 2025. Disponível em: <https://doi.org/10.3390/diagnostics13111911>.

NASERI, H.; SAFAEI, A. A. Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: a systematic literature review. **BMC Cancer**, v. 25, p. 75, 2025. Acesso em: 29 de maio de 2025. Disponível em: <https://doi.org/10.1186/s12885-024-13423-y>.

NETO, A. V. d. R. F. *et al.* Perfil epidemiológico de pacientes portadores de câncer de pele atendidos no hospital regional da asa norte/df - brasil. **Revista Brasileira de Cirurgia Plástica**, Sociedade Brasileira de Cirurgia Plástica, v. 35, n. 3, p. 316–321, 2020. ISSN 1983-5175. Acesso em: 28 abril. 2025. Disponível em: <https://doi.org/10.5935/2177-1235.2020RBCP0056>.

OMIYE, J. A. *et al.* Principles, applications, and future of artificial intelligence in dermatology. **Frontiers in Medicine**, v. 10, p. 1278232, 2023. Acesso em: 19 de junho de 2025. Disponível em: <https://doi.org/10.3389/fmed.2023.1278232>.

PACHECO, A. G. C.; KROHLING, R. A. The impact of patient clinical information on automated skin cancer detection. **arXiv preprint arXiv:1909.12912**, 2019. Acesso em: 15 de junho de 2025. Disponível em: <https://arxiv.org/abs/1909.12912>.



PHILLIPS, M. *et al.* Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. **JAMA Network Open**, v. 2, n. 10, p. e1913436, 2019. Acesso em: 15 de abril de 2025. Disponível em: <https://doi.org/10.1001/jamanetworkopen.2019.13436>.

PICCOLO, V. *et al.* Dermatoscopy of vascular lesions. **Dermatology Clinics**, Elsevier, v. 36, n. 4, p. 389–395, 2018. Acesso em: 01 maio 2025. Disponível em: <https://doi.org/10.1016/j.det.2018.05.006>.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. In: **arXiv preprint arXiv:1804.02767**. [s.n.], 2018. Acesso em: 17 de julho de 2025. Disponível em: <https://arxiv.org/abs/1804.02767>.

ROSENDAHL, C. *et al.* Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. **Journal of the American Academy of Dermatology**, v. 64, n. 6, p. 1068–1073, 2011. Acesso em: 21 de junho de 2025. Disponível em: <https://doi.org/10.1016/j.jaad.2010.03.039>.

SALAFRANCA, M. Álvarez *et al.* Dermoscopy of inflamed seborrheic keratosis: A great mimic of malignancy. **Australasian Journal of Dermatology**, Wiley Online Library, v. 63, n. 1, p. 53–61, 2022. Acesso em: 01 maio 2025. Disponível em: <https://doi.org/10.1111/ajd.13781>.

SCHMITT, J. *et al.* Occupational exposure to non-artificial uv-light and non-melanocytic skin cancer - a systematic review concerning a new occupational disease. **Journal der Deutschen Dermatologischen Gesellschaft = Journal of the German Society of Dermatology**, v. 182, n. 5, p. 1238–1246, 2010. Disponível em: <https://doi.org/10.1111/j.1610-0387.2009.07260.x>.

SEYYED-KALANTARI, L. *et al.* Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. **Nature Medicine**, Nature Publishing Group, v. 27, p. 2176–2182, 2021. Acesso em: 25 de junho de 2025.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, Springer, v. 6, n. 1, p. 60, 2019. Acesso em: 20 de junho de 2025.

SIEGEL, R. L. *et al.* Cancer statistics, 2023. **CA: a cancer journal for clinicians**, v. 73, n. 1, p. 17–48, 2023. Acesso em: 20 de junho de 2025.

SILVA, T. A. M. *et al.* Aprendizado de máquina aplicado ao diagnóstico de câncer de pele. In: **Anais do Congresso Brasileiro de Inteligência Computacional (CBIC)**. [s.n.], 2021. Acesso em: 2 maio 2025. Disponível em: [https://www.ene.unb.br/joaoluiz/pdf/cbic2021\\_tony.pdf](https://www.ene.unb.br/joaoluiz/pdf/cbic2021_tony.pdf).

STIFF, M. J.; BARATA, C. *et al.* Characteristics of publicly available skin cancer image datasets. **The Lancet Digital Health**, 2021. Acesso em: 15 de junho de 2025. Disponível em: <https://www.thelancet.com/journals/landig/article/PIIS2589-7500%2821%2900252-1/fulltext>.

SUETENS, P. **Fundamentals of medical imaging**. 3. ed. Cambridge, U.K.: Cambridge University Press, 2017. Acesso em: 22 de junho de 2025.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International conference on machine learning**. 2019. p. 6105–6114. Acesso em: 19 de junho de 2025. Disponível em: <http://proceedings.mlr.press/v97/tan19a.html>.

TERVEN, J.; CÓRDOVA-ESPARZA, D.-M.; ROMERO-GONZÁLEZ, J.-A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. **Machine Learning and Knowledge Extraction**, MDPI AG, v. 5, n. 4, p. 1680–1716, nov. 2023. ISSN 2504-4990. Acesso em: 01 junho. 2025. Disponível em: <http://dx.doi.org/10.3390/make5040083>.

THAMM, J. R.; WELZEL, J.; SCHUH, S. Diagnosis and therapy of actinic keratosis. **JDDG: Journal der Deutschen Dermatologischen Gesellschaft**, v. 22, n. 5, p. 675–690, 2024. Acesso em: 01 de maio de 2025. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ddg.15288>.

TJOA, E.; GUAN, C. A survey on explainable artificial intelligence (xai): Toward medical xai. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 32, n. 11, p. 4793–4813, 2020. Acesso em: 24 de junho de 2025.

TOPOL, E. **Deep medicine: how artificial intelligence can make healthcare human again**. [S.l.]: Basic Books, 2019. Acesso em: 18 de junho de 2025.

TSCHANDL, P.; ROSENDAHL, C.; KITTLER, H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. **Scientific Data**, v. 5, p. 180161, 2018. Acesso em: 15 de junho de 2025. Disponível em: <https://www.nature.com/articles/sdata2018161>.

VARDHAN, S.; SAHOO, S. K. Computational studies on the interaction of sars-cov-2 omicron sgp rbd with human receptor ace2, limonin and glycyrrhizic acid. **Computers in Biology and Medicine**, v. 144, p. 105367, 2022. ISSN 0010-4825. Acesso em: 08 de junho de 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0010482522001597>.

WAINSTEIN, A. J. A. *et al.* Demographic, clinical, and pathologic features of patients with cutaneous melanoma: Final analysis of the brazilian melanoma group database. **JCO Global Oncology**, v. 6, p. 575–582, 2020. Acesso em: 20 de junho de 2025.

WANG, H. H. *et al.* Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. **JAMA Dermatology**, v. 155, n. 11, p. 1277–1283, 2019. Acesso em: 15 de abril de 2025. Disponível em: <https://doi.org/10.1001/jamadermatol.2019.2335>.

WU, E. *et al.* How medical ai devices are evaluated: Limitations and recommendations from an analysis of fda approvals. **Nature Medicine**, Nature Publishing Group, v. 27, n. 4, p. 582–584, 2021. Acesso em: 20 de junho de 2025.

WU, S. *et al.* Deep learning in clinical natural language processing: a methodical review. **Journal of the American Medical Informatics Association**, v. 27, n. 3, p. 457–470, March 2020. Acesso em: 16 de junho de 2025. Disponível em: <https://doi.org/10.1093/jamia/ocz200>.

**APÊNDICE A – ARTIGO**

**AUTORES ABRIRAM IDENTIDADES**  
**SUORTE AO DIAGNÓSTICO DE CâNCER DE PELE COM INTELIGÊNCIA**  
**ARTIFICIAL: ESTUDO PILOTO**

*SUPPORT FOR SKIN CANCER DIAGNOSIS USING ARTIFICIAL INTELLIGENCE: A  
PILOT STUDY*

**APOYO AL DIAGNÓSTICO DE CâNCER DE PIEL CON INTELIGENCIA**  
**ARTIFICIAL: ESTUDIO PILOTO**  
**SUORTE AO DIAGNÓSTICO DE CâNCER DE PELE COM IA**

Eduarda Menezes da Silveira<sup>1,4</sup>; Helena Cargnelutti Grimaldi<sup>2</sup>; Sandro da Silva  
Camargo<sup>3,4</sup>

<sup>1</sup> [eduardasilveira.aluno@unipampa.edu.br](mailto:eduardasilveira.aluno@unipampa.edu.br), ORCID: 0009-0004-7829-9273,  
Universidade Federal do Pampa (UNIPAMPA)

<sup>2</sup> [hgrimaldi87@gmail.com](mailto:hgrimaldi87@gmail.com), ORCID: 0009-0002-2762-4094, Pesquisadora Autônoma

<sup>3</sup> [sandrocamargo@unipampa.edu.br](mailto:sandrocamargo@unipampa.edu.br), ORCID: 0000-0001-8871-3950, Programa de Pós-  
Graduação em Computação Aplicada (PPGCAP), Universidade Federal do Pampa  
(UNIPAMPA)

<sup>4</sup> Center for Embedded Devices and Research in Digital Agriculture (CEDRA), SENAI-  
RS, São Leopoldo, Brasil

**Autor correspondente:** Sandro da Silva Camargo, Universidade Federal do Pampa,  
Av. Maria Anunciação Gomes Godoy, 1650, Gabinete 3139, CEP 96413-172, Bagé-RS,  
Brasil, Telefone (53) 98422-0074.

## **AGRADECIMENTOS**

Os autores agradecem à NVIDIA pelo suporte computacional disponibilizado,  
que foi fundamental para a execução dos experimentos e análises deste estudo.

## **CONTRIBUIÇÕES**

Todos os autores contribuíram substancialmente na concepção e no  
planejamento do estudo; na obtenção, análise e interpretação dos dados; na redação e  
revisão crítica; e aprovaram a versão final a ser publicada.

## **DECLARAÇÃO DE DISPONIBILIDADE DE DADOS**

Declaração de disponibilidade de dados: Os conjuntos de dados gerados e  
analisados durante o estudo atual estão disponíveis no repositório  
<https://github.com/Sandrocamargo/publications/tree/main/rbc25a>

## **DECLARAÇÃO DE CONFLITOS DE INTERESSE**

Os autores declaram que não possuem conflitos de interesse relacionados à  
elaboração e publicação deste estudo.

## **FONTES DE FINANCIAMENTO**

Este trabalho não recebeu apoio financeiro de agências de fomento, instituições públicas ou privadas. Todas as atividades de desenvolvimento, análise e redação foram realizadas de forma independente pelos autores.

## RESUMO

**Introdução:** O câncer de pele constitui uma das neoplasias de maior incidência no Brasil, sendo que o diagnóstico precoce representa um fator determinante para o êxito terapêutico e para a redução da morbimortalidade associada. Diante desse cenário epidemiológico, observa-se um contexto particularmente propício à incorporação de ferramentas computacionais complementares à avaliação clínica tradicional, com destaque para abordagens baseadas em inteligência artificial. **Objetivo:** Desenvolver e validar um modelo baseado em redes neurais convolucionais para a classificação automática de lesões cutâneas malignas e benignas. **Método:** Foram utilizadas 2.639 imagens da base pública *International Skin Imaging Collaboration* (ISIC), com anotações validadas por biópsia. O sistema computacional incluiu etapas de pré-processamento e treinamento supervisionado com arquitetura YOLOv11. O desempenho foi avaliado por validação interna e validação externa. **Resultados:** O modelo alcançou acurácia média de 80,53% e sensibilidade média de 80,44% na identificação de oito classes de lesões: melanoma, nevo, carcinoma de células basais, queratose actínica, queratose benigna, dermatofibroma, lesão vascular e carcinoma espinocelular. A implementação também resultou em uma base de imagens anotadas e em um fluxo de análise reproduzível. **Conclusão:** A aplicação de inteligência artificial no suporte ao diagnóstico de câncer de pele demonstrou desempenho promissor, com potencial aplicação em triagens clínicas. Estudos futuros devem considerar a expansão da base de dados e o desenvolvimento de interfaces para uso por profissionais da saúde. **Palavras-chave:** deep learning, diagnóstico por imagem, redes neurais convolucionais, redes YOLO, tumores cutâneos.

## ABSTRACT

**Introduction:** Skin cancer is one of the most prevalent neoplasms in Brazil, and early diagnosis is a key determinant of therapeutic success and reduction of associated morbidity and mortality. In this epidemiological context, there is a particularly favorable scenario for the incorporation of computational tools to complement traditional clinical evaluation, with emphasis on artificial intelligence-based approaches. **Objective:** To develop and validate a convolutional neural network-based model for the automatic classification of malignant and benign skin lesions. **Methods:** A total of 2,639 images from the public *International Skin Imaging Collaboration* (ISIC) database, with biopsy-validated annotations, were used. The computational system included preprocessing steps and supervised training using the YOLOv11 architecture. Performance was assessed through internal and external validation. **Results:** The model achieved a mean accuracy of 80.53% and a mean sensitivity of 80.44% in the

identification of eight lesion classes: melanoma, nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesion, and squamous cell carcinoma. The implementation also resulted in an annotated image dataset and a reproducible analysis pipeline. **Conclusion:** The application of artificial intelligence to support skin cancer diagnosis demonstrated promising performance, with potential clinical screening applications. Future studies should consider expanding the dataset and developing user interfaces for healthcare professionals.

**Keywords:** deep learning, image-based diagnosis, convolutional neural networks, YOLO networks, skin tumors.

## RESUMEN

**Introducción:** El cáncer de piel constituye una de las neoplasias de mayor incidencia en Brasil, siendo el diagnóstico precoz un factor determinante para el éxito terapéutico y para la reducción de la morbilidad asociada. Ante este escenario epidemiológico, se observa un contexto particularmente propicio para la incorporación de herramientas computacionales complementarias a la evaluación clínica tradicional, con énfasis en enfoques basados en inteligencia artificial. **Objetivo:** Desarrollar y validar un modelo basado en redes neuronales convolucionales para la clasificación automática de lesiones cutáneas malignas y benignas. **Método:** Se utilizaron 2.639 imágenes de la base pública International Skin Imaging Collaboration (ISIC), con anotaciones validadas por biopsia. El sistema computacional incluyó etapas de preprocesamiento y entrenamiento supervisado con la arquitectura YOLOv11. El rendimiento fue evaluado mediante validación interna y validación externa. **Resultados:** El modelo alcanzó una exactitud media de 80,53% y una sensibilidad media de 80,44% en la identificación de ocho clases de lesiones: melanoma, nevo, carcinoma de células basales, queratosis actínica, queratosis benigna, dermatofibroma, lesión vascular y carcinoma espinocelular. La implementación también resultó en una base de imágenes anotadas y en un flujo de análisis reproducible. **Conclusión:** La aplicación de inteligencia artificial en el apoyo al diagnóstico del cáncer de piel demostró un desempeño prometedor, con potencial aplicación en tamizajes clínicos. Estudios futuros deberán considerar la ampliación de la base de datos y el desarrollo de interfaces para su uso por profesionales de la salud.

**Palabras clave:** deep learning, diagnóstico por imagen, redes neuronales convolucionales, redes YOLO, tumores cutáneos.

## INTRODUÇÃO

O câncer de pele é a neoplasia maligna mais incidente no Brasil e no mundo, representando cerca de 30% de todos os tumores diagnosticados no país.<sup>1</sup> É classificado em dois grandes grupos: o câncer de pele não melanoma, que compreende principalmente o carcinoma basocelular (CBC) e o carcinoma espinocelular (CEC), e o melanoma, menos frequente, mas responsável pela maioria dos óbitos devido à sua agressividade e potencial metastático.<sup>2,3</sup> No Brasil, a incidência estimada é de aproximadamente 60 casos por 100 mil habitantes, sendo o CBC o mais prevalente (70–75%), seguido pelo CEC (15–20%).<sup>2</sup> Embora apresentem baixa mortalidade, essas neoplasias possuem caráter invasivo local e elevada taxa de recorrência. Já o melanoma, apesar de corresponder a apenas 3% dos casos, concentra a maior parte das mortes relacionadas ao câncer de pele.<sup>3</sup> Além das neoplasias malignas, outras lesões cutâneas, como nevos, queratoses actínicas, queratoses benignas, dermatofibromas e lesões vasculares, também representam desafios diagnósticos, pois podem mimetizar tumores malignos.<sup>4-6</sup>

A detecção precoce é fundamental para reduzir morbimortalidade, especialmente nos casos de melanoma. Nesse contexto, a dermatoscopia tem ampliado a acurácia diagnóstica para cerca de 90%, em comparação a 75-80% obtidos pela inspeção clínica isolada.<sup>7,8</sup> Contudo, sua efetividade depende da experiência do examinador, o que resulta em variabilidade na interpretação. Porém, a limitação do acesso a especialistas, sobretudo em áreas remotas, reforça as desigualdades no diagnóstico e tratamento oportunos.<sup>9</sup> Nesse cenário, soluções baseadas em Inteligência Artificial (IA) emergem como ferramentas complementares à prática clínica. A IA tem demonstrado potencial em acelerar diagnósticos, apoiar decisões médicas e ampliar o acesso ao cuidado em saúde.<sup>10</sup> Em particular, Redes Neurais Convolucionais (RNC) têm alcançado desempenho comparável ao de dermatologistas experientes na classificação automática de lesões cutâneas.<sup>11-18</sup> Entre os algoritmos disponíveis, o *You Only Look Once* (YOLO) destaca-se pela detecção em tempo real com elevada acurácia, configurando-se como alternativa promissora para triagem automatizada e apoio ao diagnóstico.<sup>19</sup>



A aplicação de modelos de IA como o proposto neste estudo tem especial relevância para o contexto do Sistema Único de Saúde (SUS) e das políticas nacionais de controle do câncer. O Brasil apresenta uma distribuição geográfica desigual de dermatologistas, concentrados nas regiões Sul e Sudeste, o que limita o acesso a diagnósticos especializados em grande parte do território.<sup>9</sup> Essa disparidade se reflete em diagnósticos tardios e sobrecarga de serviços de alta complexidade. Ademais, a disponibilidade limitada de dermatoscópios em unidades básicas de saúde e a diversidade étnico-racial da população brasileira representam desafios adicionais, uma vez que a maioria das bases públicas de imagens dermatológicas contém amostras predominantemente de fototipos claros, o que pode comprometer a precisão dos modelos em tons de pele mais escuros.

Nesse sentido, iniciativas recentes têm explorado o uso da IA como ferramenta de apoio à Atenção Primária e à triagem em oncologia.<sup>10</sup> Algumas pesquisas brasileiras já demonstram o interesse crescente da comunidade científica nacional em integrar sistemas inteligentes aos fluxos de cuidado do SUS, ampliando a detecção precoce e reduzindo desigualdades regionais.<sup>10,17</sup> Tais esforços dialogam diretamente com as diretrizes do Plano de Ações Estratégicas para o Enfrentamento das Doenças Crônicas Não Transmissíveis e com as metas de controle do câncer estabelecidas pelo Instituto Nacional de Câncer.

Assim, este estudo tem como objetivo desenvolver, testar e avaliar um modelo computacional baseado na arquitetura YOLO para a detecção automática de lesões cutâneas malignas e benignas em imagens dermatoscópicas, buscando fornecer suporte ao diagnóstico médico e, consequentemente, contribuir para a identificação precoce do câncer de pele.

## **MÉTODOS**

Este estudo piloto configura-se como uma pesquisa aplicada, de abordagem quantitativa e delineamento experimental, direcionada ao desenvolvimento e à avaliação de um sistema computacional de suporte ao diagnóstico de câncer de pele com uso de IA. A investigação fundamentou-se na análise de imagens dermatoscópicas para detecção automática de lesões cutâneas, por meio de Redes Neurais Convolucionais. O desempenho dos modelos foi avaliado por métricas quantitativas padronizadas,

incluindo acurácia e sensibilidade, a fim de aferir sua eficácia na identificação de padrões morfológicos associados a diferentes tipos de lesões cutâneas.

### **Base de imagens e anotação das lesões**

Foram obtidas imagens dermatoscópicas extraídas do banco público *International Skin Imaging Collaboration (ISIC) Archive 2019*<sup>20</sup>, composto por 25.331 imagens com diagnóstico histopatológico confirmado, conforme distribuição apresentada na Tabela 1. Os critérios de inclusão envolveram qualidade visual adequada, ausência de interferências clínicas visíveis (como cicatrizes de biópsia ou suturas) e presença de um único diagnóstico por imagem. Todas as imagens estão anonimizadas, em conformidade com diretrizes éticas internacionais para uso secundário de dados clínicos.

A anotação das imagens é o processo que consiste em delimitar a lesão e indicar seu diagnóstico. Estas anotações foram realizadas manualmente utilizando a ferramenta LabelImg (<https://pypi.org/project/labelImg/>), com marcação das lesões por meio de caixas delimitadoras (*bounding boxes*) e associação de rótulos de classe compatíveis com o formato requerido pela arquitetura YOLO. As classes das lesões foram definidas com base nos laudos histopatológicos disponibilizados juntamente com a base de imagens.

Considerando que o processo de anotação das lesões nas imagens é manual e realizado individualmente em cada amostra, este estudo piloto foi conduzido sobre uma fração do conjunto de imagens disponível, também apresentada na Tabela 1. O trabalho encontra-se em progresso, com a anotação de mais imagens e a execução de novos ciclos de treinamento e validação, com o objetivo de aprimorar progressivamente os indicadores de desempenho do modelo.

**Tabela 1** - Distribuição das imagens disponíveis por classe de lesão cutânea (ISIC 2019) e distribuição da quantidade de imagens por classe utilizadas nos experimentos.

| <b>Classe de lesão cutânea</b> | <b>Número de<br/>imagens<br/>disponíveis</b> | <b>Número de<br/>imagens<br/>utilizadas</b> |
|--------------------------------|----------------------------------------------|---------------------------------------------|
|--------------------------------|----------------------------------------------|---------------------------------------------|

|                         |               |              |
|-------------------------|---------------|--------------|
| Melanoma                | 4.522         | 600          |
| Nevo                    | 12.875        | 600          |
| Carcinoma Basal         | 3.323         | 200          |
| Queratose Actínica      | 867           | 200          |
| Queratose Benigna       | 2.624         | 200          |
| Dermatofibroma          | 239           | 239          |
| Lesão Vascular          | 253           | 200          |
| Carcinoma Espinocelular | 628           | 400          |
| <b>Total</b>            | <b>25.331</b> | <b>2.639</b> |

### Treinamento e arquitetura dos modelos

O *pipeline* de modelagem, representado na Figura 1, incluiu os seguintes passos:

1) Pré-processamento das imagens: todas as imagens foram padronizadas quanto ao tamanho e às anotações das lesões, garantindo consistência na análise e facilitando a comparação entre diferentes tipos de lesões. Essa etapa é equivalente à preparação cuidadosa de exames clínicos antes da interpretação diagnóstica.

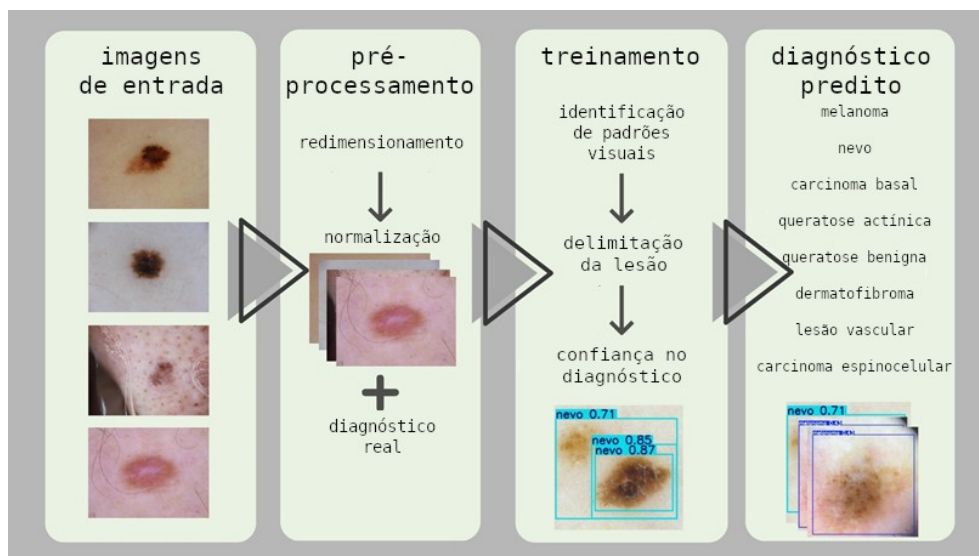
2) Treinamento do modelo, com ajuste automático de hiperparâmetros, visando otimizar a capacidade de classificação das lesões. Esse processo permite que o modelo aprenda de forma otimizada a distinguir características clínicas relevantes, de maneira similar à experiência adquirida por um dermatologista ao revisar múltiplos casos.

3) Diagnóstico, ou validação interna do desempenho: durante o treinamento, o modelo foi continuamente avaliado em subconjuntos de dados reservados, medindo acurácia, sensibilidade e especificidade. Essa etapa garante que o desempenho do sistema seja confiável e consistente, similar à revisão por pares em um contexto clínico antes da implementação de um novo protocolo diagnóstico.

Foram treinadas arquiteturas da família YOLOv11 (versão 11)<sup>21</sup>, com o objetivo de avaliar a acurácia e a eficiência da detecção automática de lesões cutâneas. Esta arquitetura foi escolhida, devido às suas características de acurácia, velocidade e capacidade de generalização, especialmente em aplicações médicas.<sup>22</sup> Esta escolha está alinhada ao objetivo de explorar redes neurais profundas como suporte ao diagnóstico

precoce do câncer de pele. As arquiteturas da família YOLOv11 empregadas neste estudo foram:

- YOLOv11n (*nano*): versão compacta, otimizada para execução em dispositivos com capacidade computacional reduzida, tal como *smartphones*, *tablets* e sistemas embarcados em dispositivos médicos portáteis, onde há restrições de memória e processamento.
- YOLOv11s (*small*), YOLOv11m (*medium*), YOLOv11l (*large*): arquiteturas progressivamente mais complexas, com maior número de parâmetros e, portanto, maior capacidade de representação, à custa de aumento do tempo de processamento. Essas versões são mais indicadas para execução em estações de trabalho com Unidades de Processamento Gráfico (GPU) dedicadas, servidores clínicos e plataformas de pesquisa, contextos em que a precisão diagnóstica é priorizada em relação ao tempo de processamento.



**Figura 1** – Pipeline de detecção de lesões cutâneas com YOLOv11.

Todos os modelos foram treinados mediante teste de diferentes hiperparâmetros, como número de épocas, com batch size adaptativo, parada antecipada e validação interna contínua durante o treinamento, o que permite otimizar o desempenho do modelo. Esse rigor metodológico, alinhado com as melhores práticas para treinamento e

uso de modelos de classificação,<sup>23</sup> assegura que o modelo aprenda padrões relevantes e minimize vieses na identificação das lesões.

## **Validação**

A validação dos modelos de inteligência artificial foi realizada em duas etapas complementares:<sup>24</sup>

1) Validação interna, utilizando uma divisão estratificada aleatória de 90% das imagens para o treinamento do modelo e 10% para teste. Essa técnica permite avaliar o desempenho do modelo de forma controlada, utilizando os próprios dados do conjunto original, permitindo estimar a capacidade preditiva do modelo e garantindo que todas as classes de lesões cutâneas fossem representadas proporcionalmente nos conjuntos de treino e teste.

2) Validação externa, realizada com imagens inéditas, ou seja, não utilizadas anteriormente no treinamento, capturadas em um ambiente clínico real. Essas imagens foram processadas com o modelo treinado, permitindo verificar se o modelo mantém sua precisão em situações práticas, fora do ambiente de teste inicial.

Na validação interna, foram avaliadas as seguintes métricas de desempenho: acurácia (mean Average Precision - mAP), sensibilidade média, precisão média e tabela de desempenho diagnóstico.

A amostra da validação externa foi composta por 58 imagens dermatoscópicas obtidas em um ambulatório de dermatologia de um serviço privado de saúde. As imagens foram capturadas com Dermatoscópio Dermlite DL4, com iluminação polarizada e não polarizada de 24 LEDs Brancos e *Pigment Boost Plus* (Luz laranja), seguindo rotina de triagem clínica de lesões suspeitas. Imagens com artefatos importantes (ex.: curativos, sangue, pelos espessos) foram excluídas.

Não houve balanceamento de classes na validação externa, de modo que a distribuição das lesões reflète a prevalência observada na prática clínica (maior presença de lesões benignas, como nevos e queratoses). Essa ausência de balanceamento pode favorecer o desempenho do modelo nas classes mais frequentes e prejudicar a avaliação em classes menos representadas, como melanoma e carcinoma espinocelular, resultando

em maior probabilidade de falsos negativos nessas categorias. Essa limitação será tratada em estudos futuros mediante ampliação da amostra clínica e estratégias de balanceamento.

### **Ambiente de Desenvolvimento**

O treinamento foi realizado no ambiente de alto desempenho Saturno (<https://saturncloud.io/>), equipado com oito GPU NVIDIA A100 e 2Tb de memória. O sistema foi implementado em Python 3.10, com uso das bibliotecas PyTorch, Ultralytics, Pandas, NumPy, OpenCV e Matplotlib, entre outras.

### **Aspectos Éticos**

Considerando os aspectos éticos da pesquisa, destaca-se que não houve necessidade de submissão deste estudo ao Comitê de Ética em Pesquisa (CEP). Isso porque a investigação utilizou exclusivamente dados secundários de acesso público, disponíveis em base de dados aberta na internet, sem qualquer possibilidade de identificação direta ou indireta de indivíduos. Tal enquadramento encontra respaldo na Resolução nº 466/2012 do Conselho Nacional de Saúde (CNS)<sup>25</sup>, que estabelece diretrizes e normas para pesquisas envolvendo seres humanos, e, de modo mais específico, na Resolução nº 510/2016<sup>26</sup>, a qual dispõe sobre as normas aplicáveis às pesquisas em Ciências Humanas e Sociais e explicita que não necessitam de apreciação ética os estudos que utilizam informações de domínio público. Dessa forma, a realização desta pesquisa está em conformidade com os princípios éticos vigentes no Brasil.

## **RESULTADOS**

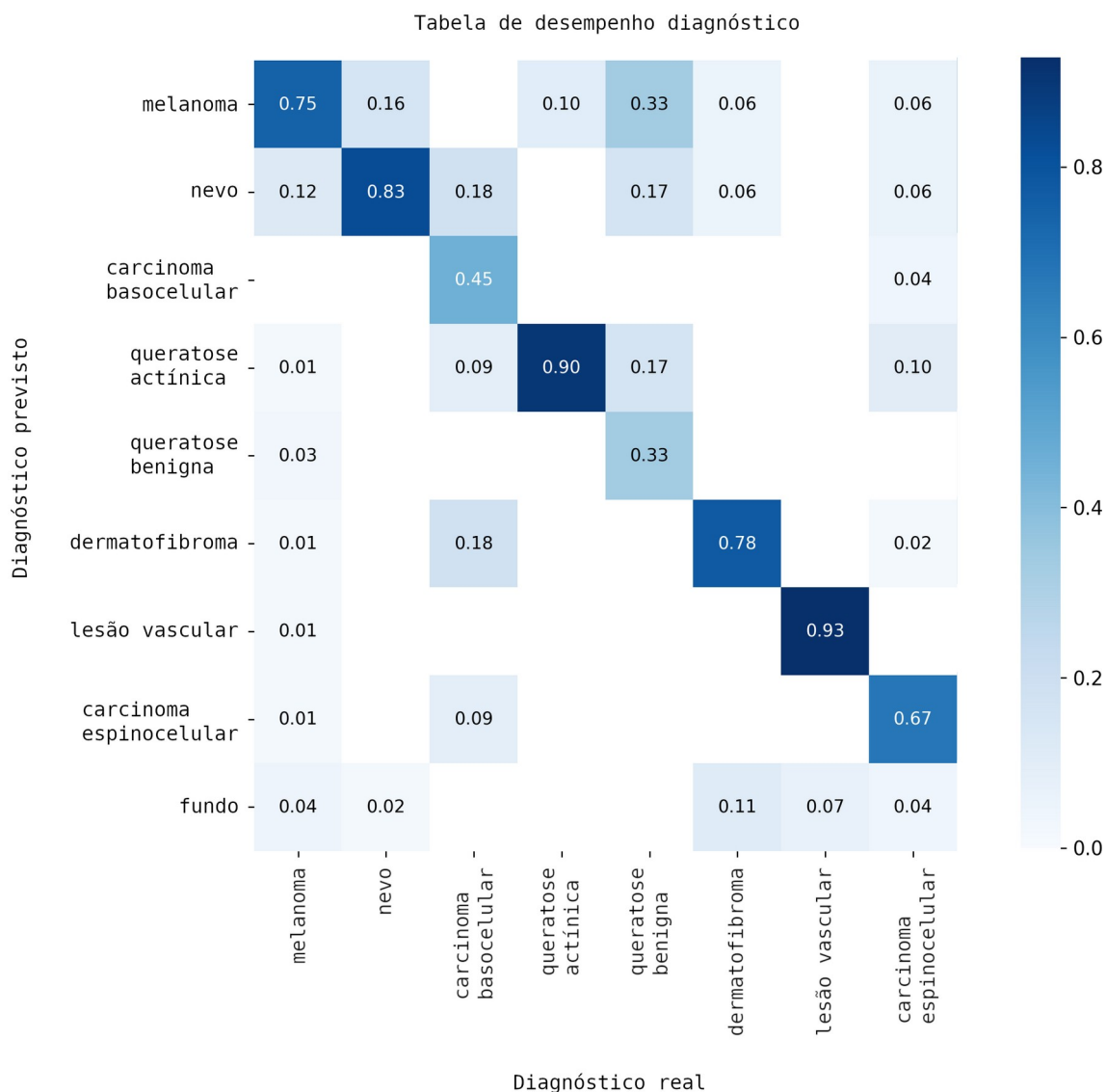
Esta seção apresenta os principais achados obtidos no desenvolvimento da pesquisa, contemplando tanto as análises quantitativas quanto qualitativas. Foram realizados testes preliminares com diferentes arquiteturas da família YOLOv11 (*nano*, *small*, *medium* e *large*) para avaliar a detecção de lesões cutâneas em imagens dermatoscópicas. Entre essas variantes, a arquitetura YOLOv11s (*small*) apresentou os melhores resultados, os quais são aqui relatados. A Tabela de desempenho diagnóstico

(Figura 2) ilustra os acertos e erros por classe, considerando no eixo x as classes reais e no eixo y as classes previstas. A análise desta matriz permitiu identificar padrões de acerto e erro do modelo em diferentes tipos de lesões cutâneas.

O desempenho variou conforme a classe avaliada. Para lesões vasculares, a taxa de acerto foi de 93%, com 7% de falhas de reconhecimento, evidenciando elevada especificidade, mas sensibilidade limitada em casos marginais. Na queratose actínica, obteve-se 90% de acerto; contudo, 10% foram incorretamente classificadas como melanoma, o que denota risco clínico relevante devido ao caráter agressivo dessa neoplasia.

Nos nevos melanocíticos, 83% foram corretamente classificados, enquanto 16% foram reconhecidos equivocadamente como melanoma, sugerindo dificuldade do modelo em diferenciar lesões benignas e malignas. Para os dermatofibromas, a taxa de acerto foi de 78%, com 6% de erros como melanoma e 6% como nevos.

Quanto ao melanoma, 75% foram identificados corretamente, mas 12% foram confundidos com nevos e 7% com outras categorias, revelando a necessidade de maior precisão nessa classe. Nos carcinomas, observou-se desempenho heterogêneo: 67% de acurácia para o espinocelular e apenas 45% para o basocelular. A queratose benigna apresentou a menor taxa de acerto (33%), possivelmente em função da escassez de exemplos no conjunto de treinamento e da semelhança morfológica com outras lesões.

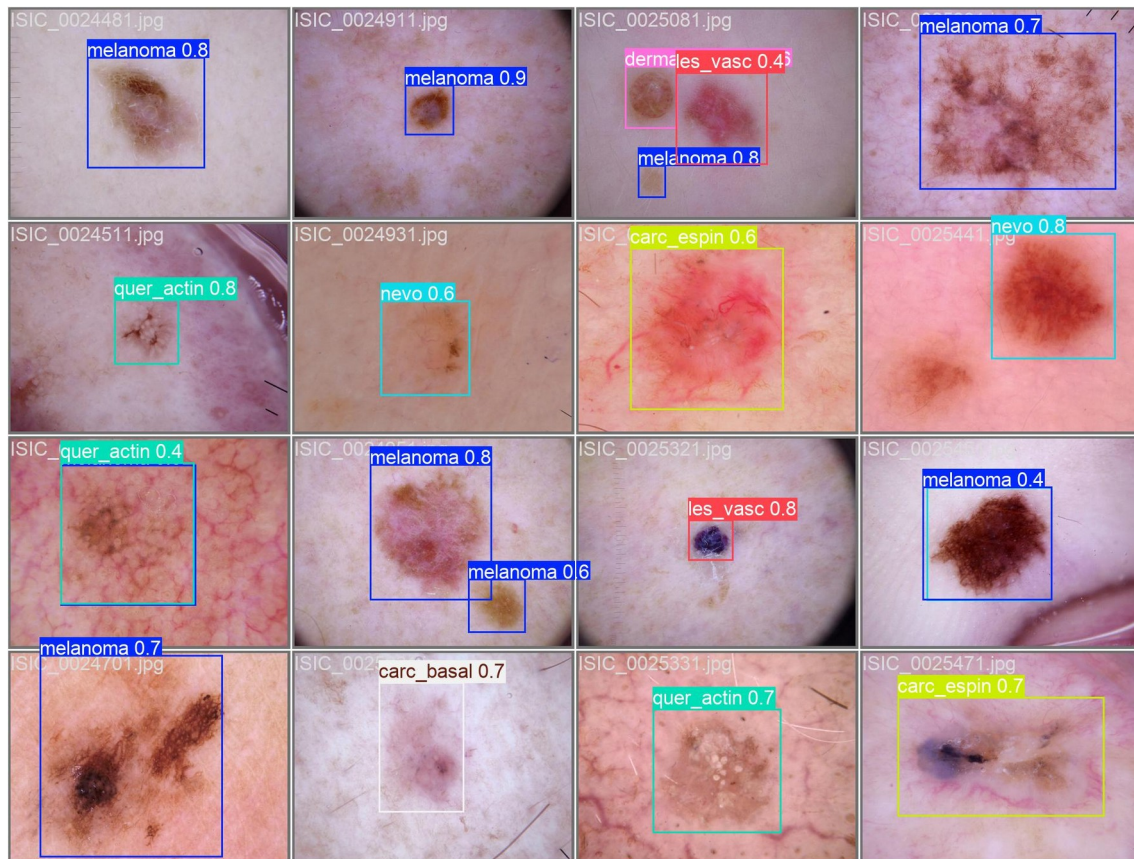


**Figura 2** – Tabela de desempenho diagnóstico apresentando os diagnósticos reais e os diagnósticos preditos pelo modelo. No eixo horizontal encontram-se as classes reais das lesões e, no eixo vertical, as predições do modelo. A diagonal principal representa os acertos. Quando o modelo classifica uma lesão maligna como benigna, ocorre um *falso negativo*, situação de maior risco clínico por atrasar o diagnóstico. Quando uma lesão benigna é classificada como maligna, ocorre um *falso positivo*, podendo levar a biópsias ou encaminhamentos desnecessários.

Os casos em que o modelo atribuiu a classe *background* corresponderam a falsos negativos, ou seja, situações em que a lesão real não foi reconhecida. Esse tipo de falha tem relevância clínica crítica, pois pode contribuir para atrasar o diagnóstico e comprometer a conduta terapêutica.



Além da avaliação quantitativa, foi realizada análise qualitativa das predições visuais em imagens dermatoscópicas. A Figura 3 mostra exemplos representativos, com destaque para a correta localização e classificação de diferentes tipos de lesões, evidenciando a aplicabilidade prática do modelo.



**Figura 3** – Exemplos de detecção automatizada de lesões cutâneas pelo modelo desenvolvido, acompanhados dos respectivos níveis de confiança atribuídos pelo sistema ao diagnóstico previsto. Cada imagem apresenta a classe prevista e o nível de confiança (probabilidade) da predição. (a) Verdadeiro positivo: o modelo identifica corretamente a lesão e delimita sua região. (b) Falso positivo: o modelo classifica uma lesão benigna como maligna, o que pode levar a investigação desnecessária. (c) Falso negativo: o modelo falha em identificar uma lesão maligna, o que representa risco clínico por possível atraso no diagnóstico. Esses exemplos ilustram como o modelo responde a diferentes padrões dermatoscópicos, como bordas irregulares, assimetria de pigmentação e presença de vasos sanguíneos.

Para facilitar a visualização do desempenho discriminativo do modelo, as principais métricas por classe foram consolidadas na Tabela 2, permitindo observar diferenças de sensibilidade, precisão e perfis de erro entre os tipos de lesões.

**Tabela 2** - Desempenho do modelo por classe de lesão

| <b>Classe de lesão cutânea</b> | <b>Sensibilidade (%)</b> | <b>Precisão (%)</b> | <b>Principais erros observados</b>                    |
|--------------------------------|--------------------------|---------------------|-------------------------------------------------------|
| Melanoma                       | 75                       | 78                  | Falso negativo: confundido com nevo                   |
| Nevo                           | 83                       | 81                  | Falso positivo: confundido com melanoma               |
| Carcinoma Basal                | 45                       | 72                  | Não detectado em lesões pequenas ou pouco pigmentadas |
| Queratose Actínica             | 90                       | 84                  | Confundida com melanoma em bordas irregulares         |
| Queratose Benigna              | 33                       | 68                  | Confundida com outras queratoses                      |
| Dermatofibroma                 | 78                       | 70                  | Confundido com nevo ou melanoma                       |
| Lesão Vascular                 | 93                       | 89                  | Baixa taxa de erro: padrão facilmente identificado    |
| Carcinoma Espinocelular        | 67                       | 74                  | Confundido com queratose actínica                     |

## DISCUSSÃO

Os resultados evidenciam o potencial do modelo proposto como ferramenta de apoio ao diagnóstico médico em oncologia cutânea, sobretudo em contextos de triagem e suporte à decisão clínica. Apesar dos avanços observados, o desempenho heterogêneo entre classes aponta limitações relevantes. Classes com padrões morfológicos semelhantes, como nevos e melanomas, ainda apresentaram confusão significativa, indicando necessidade de aprimoramento na capacidade discriminativa do sistema.

A baixa performance em categorias específicas, como carcinoma basocelular e queratose benigna, sugere impacto tanto da desproporção amostral quanto da complexidade intrínseca dessas lesões. Estratégias de expansão e diversificação da base de dados emergem como caminhos prioritários para aumentar a robustez do modelo.

Do ponto de vista clínico, a ocorrência de falsos negativos representa o maior desafio, uma vez que o não reconhecimento de uma lesão maligna pode ter repercussões severas para o paciente. Assim, futuros treinamentos devem priorizar a redução desse tipo de erro, buscando elevar a acurácia global para níveis superiores a 95%.

Quando comparado a estudos internacionais, o desempenho obtido neste trabalho é compatível com pesquisas recentes que utilizaram redes neurais profundas para detecção de melanoma. Estudos conduzidos com CNNs ou modelos YOLO adaptados reportam sensibilidades entre 72% e 89% e acurácia média entre 78% e 91% na identificação de melanoma e CBC.<sup>11,18</sup> No presente estudo, a sensibilidade para melanoma foi de 75% e a acurácia média geral foi de 80,53%, valores dentro da faixa observada na literatura, o que reforça a viabilidade técnica do modelo mesmo utilizando um conjunto reduzido de imagens anotadas manualmente.

Além disso, trabalhos como SkinSage XAI (Munjal et al., 2024) e o estudo prospectivo multicêntrico de Heinlein et al. (2024) demonstram que sistemas baseados em IA podem alcançar desempenho semelhante ou superior ao de dermatologistas em cenários de triagem, reforçando que a aplicação de modelos de visão computacional pode reduzir custos e acelerar o diagnóstico em contextos clínicos. Assim, o desempenho alcançado por este estudo aproxima-se da tendência global, reforçando sua relevância científica.<sup>12,18</sup>

A perspectiva de implementação do modelo em dispositivos móveis ou plataformas web constitui um avanço estratégico, pois amplia significativamente a possibilidade de validação externa e aplicação prática. Essa portabilidade favorece a realização de testes colaborativos em diferentes regiões do país, permitindo que especialistas de distintas instituições e contextos clínicos avaliem a ferramenta de forma sistemática. Tal abordagem potencializa a geração de evidências sobre a robustez e generalização do modelo, além de contribuir para a consolidação de sua utilidade em cenários de telemedicina e oncologia de precisão.

Assim, este estudo caracteriza-se como uma “prova de conceito” (*proof of concept*), demonstrando a viabilidade técnica e científica do uso do modelo proposto como ferramenta de apoio ao diagnóstico dermatológico. A criação de um *pipeline* completo (anotação → treinamento → inferência) e a validação com imagens reais do ambiente clínico reforçam o potencial de transferência tecnológica do modelo para o SUS, especialmente em localidades com baixa disponibilidade de dermatologistas e acesso limitado à dermatoscopia.

Apesar das limitações, sensibilidade ainda insuficiente em algumas classes e baixa representação de tons de pele mais escuros, os resultados mostram potencial clínico, especialmente como ferramenta de triagem em atenção primária à saúde.

Em síntese, o modelo desenvolvido apresenta as seguintes potencialidades principais:

- viabilidade técnica comprovada para detecção automática de múltiplos tipos de lesões;
- desempenho compatível com estudos internacionais que utilizam *deep learning* em dermatoscopia;
- potencial de aplicação em triagem clínica, especialmente no contexto do SUS;
- estabelecimento de bases para estudos futuros com maior diversidade de dados, balanceamento e validação multicêntrica. Esta atividade apresentaria potencial significativo mediante expansão e fortalecimento por meio de parcerias institucionais estratégicas, especialmente com o Instituto Nacional de Câncer (INCA) e/ou o Ministério da Saúde. O apoio dessas entidades poderia viabilizar maior alcance, integração de bases de dados nacionais e aprimoramento metodológico, contribuindo para a consolidação de uma rede colaborativa voltada ao monitoramento e à pesquisa em oncologia no país.

## CONCLUSÃO

Este estudo demonstrou avanços relevantes na aplicação da inteligência artificial para o diagnóstico assistido do câncer de pele, com ênfase na utilização da arquitetura YOLOv11. A construção de um *pipeline* funcional, que abrange desde o pré-processamento das imagens até o treinamento do modelo com dados públicos anotados manualmente, evidencia a viabilidade técnica da proposta.

Dentre as contribuições mais significativas, destacam-se:

- A implementação de um modelo de detecção automática de lesões cutâneas com múltiplas classes, a partir de imagens dermatoscópicas, ampliando seu potencial de aplicação clínica;
- O desenvolvimento de um conjunto de dados com anotações manuais, priorizando padrões visuais relevantes como vasos e bordas, com impacto direto na acurácia do modelo;
- A integração de tecnologias de visão computacional e inteligência artificial em um fluxo de trabalho reproduzível e escalável.

Apesar dos avanços, o estudo apresenta limitações importantes:

- A acurácia do modelo, embora promissora com valor de 80,53%, ainda não atinge os patamares necessários para uso clínico direto. Essa limitação será abordada por meio de novos ciclos de treinamento, que estarão em execução nos próximos 6 meses, com bases de dados expandidas e mais diversificadas.
- O conjunto de imagens utilizado tem baixa representatividade de tons de pele mais escuros, o que compromete a generalização do modelo à população brasileira. Tal lacuna evidencia a necessidade de iniciativas nacionais voltadas à coleta sistemática de imagens dermatológicas com diversidade étnico-racial.
- A validação externa, embora iniciada com imagens inéditas capturadas em ambiente clínico, requer detalhamento adicional quanto à amostragem, aos critérios de inclusão e aos parâmetros clínicos avaliados, para assegurar a extrapolação dos resultados aos serviços reais de saúde.

Por fim, essas considerações demonstram o potencial transformador do projeto, ao mesmo tempo em que evidenciam os desafios superados e traçam caminhos para um resultado mais robusto e aplicável na prática clínica.

## REFERÊNCIAS

1. Delgado LP, Souza GS, Oliveira SC, et al. Epidemiologia e manejo clínico do câncer de pele melanoma: uma revisão narrativa. J Soc Issues Health Sci. 2024;1(7). doi: <https://doi.org/10.5281/zenodo.14550226>

2. Neto AVRF, Ferreira AV, Gomes SA, et al. Perfil epidemiológico de pacientes portadores de câncer de pele atendidos no hospital regional da asa norte/DF - Brasil. *Rev Bras Cir Plást.* 2020;35(3):316–21. doi: <https://doi.org/10.5935/2177-1235.2020RBCP0056>
3. Borges AL, Scope A, Marghoob AA, et al. Melanocytic nevi with special features: clinical-dermoscopic and reflectance confocal microscopic-findings. *J Eur Acad Dermatol Venereol.* 2014;28(7):833–45. doi: <https://doi.org/10.1111/jdv.12291>
4. Piccolo V, Argenziano G, Di Stefani A, Zalaudek I. Dermatoscopy of vascular lesions. *Dermatol Clin.* 2018;36(4):389–95. doi: <https://doi.org/10.1016/j.det.2018.05.006>
5. Salafranca MÁ, Cuellar-Barboza A, Alvarez-Escudero J, et al. Dermoscopy of inflamed seborrheic keratosis: A great mimic of malignancy. *Australas J Dermatol.* 2022;63(1):53–61. doi: <https://doi.org/10.1111/ajd.13781>
6. Thamm JR, Welzel J, Schuh S. Diagnosis and therapy of actinic keratosis. *J Dtsch Dermatol Ges.* 2024;22(5):675–90.
7. Azulay AA, Azulay DD, Azulay RR. *Dermatologia*. 7th ed. Rio de Janeiro: Guanabara Koogan; 2013.
8. Botton DV, Barros MM, Fraga TN, et al. Relevância da dermatoscopia para o diagnóstico precoce de melanomas: uma revisão de literatura. *Rev Interdiscip Saúde Educ.* 2020;1(2):159–74. doi: <https://doi.org/10.56344/2675-4827.v1n2a20209>
9. Schmitt JV, Miot HA. Distribution of Brazilian dermatologists according to geographic location, population and HDI of municipalities: an ecological study. *An Bras Dermatol.* 2014 Nov-Dec;89(6):1013-5. doi: 10.1590/abd1806-4841.20143276. PMID: 25387516; PMCID: PMC4230680.

10. Siqueira ASE, Santos Neto MF, Ferreira CBT, Souza TA. Inteligência Artificial nas Ações de Controle do Câncer: Solução ou Problema? *Rev Bras Cancerol.* 2025;71(3). Disponível em: <https://rbc.inca.gov.br/index.php/revista/article/view/5291>
11. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Netw Open.* 2019;2(10):e1913436. doi: <https://doi.org/10.1001/jamanetworkopen.2019.13436>
12. Munjal G, Bhardwaj P, Bhargava V, Singh S, Nagpal N. SkinSage XAI: An explainable deep learning solution for skin lesion diagnosis. *Health Care Sci.* 2024;3(6):438–55. doi: <https://doi.org/10.1002/hcs2.121>
13. Cui X, Wei R, Gong L, Qi R, Zhao Z, Chen H, et al. Assessing the effectiveness of artificial intelligence methods for melanoma: A retrospective review. *J Am Acad Dermatol.* 2019;81(5):1176–80. doi: <https://doi.org/10.1016/j.jaad.2019.06.042>
14. Jairath N, Pahalyants V, Shah R, Weed J, Carucci JA, Criscito MC. Artificial intelligence in dermatology: A systematic review of its applications in melanoma and keratinocyte carcinoma diagnosis. *Dermatol Surg.* 2024;50(9):791–8. doi: <https://doi.org/10.1097/DSS.0000000000004223>
15. Wang HH, Wang YH, Liang CW, Li YC. Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. *JAMA Dermatol.* 2019;155(11):1277–83. doi: <https://doi.org/10.1001/jamadermatol.2019.2335>
16. Chu YS, An HG, Oh BH, Yang S. Artificial intelligence in cutaneous oncology. *Front Med (Lausanne).* 2020;7:318. doi: <https://doi.org/10.3389/fmed.2020.00318>

17. Silva TAM, Costa MCR, Souza SR, et al. Aprendizado de máquina aplicado ao diagnóstico de câncer de pele. In: Anais do Congresso Brasileiro de Inteligência Computacional (CBIC); 2021. doi: <https://doi.org/10.21528/CBIC2021-55>
18. Heinlein L, Schultheiss M, Mascha F, et al. Clinical melanoma diagnosis with artificial intelligence: Insights from a prospective multicenter study. arXiv [Preprint]. 2024.
19. International Skin Imaging Collaboration. ISIC 2019: Skin lesion analysis towards melanoma detection – challenge dataset [Internet]. 2019 [citado 18 set. 2024]. Disponível em: <https://challenge2019.isic-archive.com/>
20. Ultralytics. *Ultralytics YOLO* [repositório GitHub]. Disponível em: <https://github.com/ultralytics/ultralytics>. Acesso em: 10 abril. 2025.
21. Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach Learn Knowl Extr. 2023;5(4):1680–716. doi: <https://doi.org/10.3390/make5040083>
22. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. Deep learning-enabled medical computer vision. NPJ Digit Med. 2021;4(1):5. doi: <https://doi.org/10.1038/s41746-020-00376-2>
23. Cechinel C, Camargo SS. Mineração de dados educacionais: avaliação e interpretação de modelos de classificação. In: Jaques Maillard P, Siqueira S, Pimentel M, Pimentel E, Bittencourt IG, editores. Metodologia de Pesquisa Científica em Informática na Educação: Abordagem Quantitativa. Porto Alegre: Sociedade Brasileira de Computação; 2020. p. 1-. Vol. 2.
24. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016 Jan;69:245-7. doi:



10.1016/j.jclinepi.2015.04.005. Epub 2015 Apr 18. PMID: 25981519; PMCID: PMC5578404.

25. Brasil. Conselho Nacional de Saúde. Resolução nº 466, de 12 de dezembro de 2012. Aprova diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Diário Oficial da União. 2013 jun 13 [citado 2025 ago 27];150(112 Seção 1):59-62. Disponível em: [https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466\\_12\\_12\\_2012.html](https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html)

26. Brasil. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Diário Oficial da União. 2016 maio 24 [citado 2025 ago 27];153(98 Seção 1):44-6. Disponível em: [https://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510\\_07\\_04\\_2016.html](https://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510_07_04_2016.html)