

UNIVERSIDADE FEDERAL FLUMINENSE

ADRIANO LIMA E SOUZA

**CLASSIFICAÇÃO DE LESÕES EM PELE
NEGRA USANDO VGG19 E O CONJUNTO DE
DADOS ISIC 2020**

NITERÓI

2024

ADRIANO LIMA E SOUZA

**CLASSIFICAÇÃO DE LESÕES EM PELE
NEGRA USANDO VGG19 E O CONJUNTO DE
DADOS ISIC 2020**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Computação. Área de concentração: CIÊNCIA DA COMPUTAÇÃO

Orientador:
ESTEBAN WALTER GONZALEZ CLUA

NITERÓI

2024

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

S719c Souza, Adriano Lima
CLASSIFICAÇÃO DE LESÕES EM PELE NEGRA USANDO VGG19 E O
CONJUNTO DE DADOS ISIC 2020 / Adriano Lima Souza. - 2024.
71 f.

Orientador: ESTEBAN WALTER GONZALEZ CLUA.
Dissertação (mestrado)-Universidade Federal Fluminense,
Instituto de Computação, Niterói, 2024.

1. Machine learning. 2. Cancer de pele. 3. Fitz patrick. 4.
Inteligencia artificial. 5. Produção intelectual. I. CLUA,
ESTEBAN WALTER GONZALEZ, orientador. II. Universidade Federal
Fluminense. Instituto de Computação. III. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

Figura 1: Local da ficha catalográfica

ADRIANO LIMA E SOUZA

CLASSIFICAÇÃO DE LESÕES EM PELE NEGRA USANDO VGG19 E O
CONJUNTO DE DADOS ISIC 2020

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Computação. Área de concentração: CIÊNCIA DA COMPUTAÇÃO

Previsão de aprovação em AGOSTO de 2024.

BANCA EXAMINADORA

Prof. ESTEBAN WALTER GONZALEZ CLUA - Orientador, UFF

Prof. FLAVIO LUIZ SEIXAS, UFF

Prof. LUIZ MARCOS GARCIA GONÇALVES, UFRN

Niterói
2024

A Deus, fonte de sabedoria e inspiração constante, À minha querida família, por seu amor incondicional e apoio inabalável, Ao meu dedicado orientador, cuja orientação foi fundamental para este trabalho, Aos estimados professores, pela generosidade em compartilhar conhecimento e experiência.

Agradecimentos

Agradeço a Deus pelas conquistas.

Ao meu orientador, que me mostrou os caminhos a serem seguidos e pela confiança depositada.

À minha família, pelo apoio incondicional e por sempre acreditarem em mim.

Resumo

A classificação de imagens de lesões cutâneas é uma tarefa crucial na detecção precoce do câncer de pele, especialmente o melanoma, que é a forma mais grave. Este artigo propõe a aplicação da arquitetura de rede neural convolucional VGG19 para classificar sete tipos de lesões cutâneas usando o conjunto de dados ISIC 2020. Além disso, o estudo investiga a aplicação da Escala Fitzpatrick para avaliar como diferentes tons de pele podem influenciar a precisão da classificação. Discute-se também como o uso da visão computacional para identificar e classificar a pele de acordo com a Escala Fitzpatrick pode melhorar a precisão do diagnóstico. O objetivo é demonstrar como a extração de características através da VGG19 pode melhorar significativamente a precisão do diagnóstico, beneficiando tanto pacientes quanto médicos. Diferentes métricas de desempenho são aplicadas e discutidas para avaliar a eficácia do modelo.

Keywords: Classificação de Imagens , Lesões de Pele, detecção de melanoma , VGG19, ISIC 2020 dataset, Escala Fitzpatrick , visão computacional, diagnostic accuracy, rede neural convolucional .

Abstract

The classification of skin lesion images is a crucial task in the early detection of skin cancer, particularly melanoma, which is the most severe form. This paper explores the application of the convolutional neural network architecture VGG19 to classify seven types of skin lesions using the ISIC 2020 dataset. Additionally, the study investigates the application of the Fitzpatrick Scale to assess how different skin tones can influence classification accuracy, as well as the use of computer vision to identify skin and classify it according to the Fitzpatrick Scale to improve accuracy and precision. The objective is to demonstrate how feature extraction through VGG19 can significantly improve diagnostic accuracy, benefiting both patients and healthcare professionals. Various performance metrics are applied and discussed to evaluate the model's effectiveness.

Keywords: Image classification, skin lesions, melanoma detection, VGG19, ISIC 2020 dataset, Fitzpatrick Scale, computer vision, diagnostic accuracy, convolutional neural networks.

Lista de Figuras

1	Local da ficha catalográfica	
3.1	FitzPatrick Palette [31]	23
4.1	Volume de cor da pele projetado no plano L^*b^* do espaço Lab (Fonte da imagem: [4]).	28
4.2	Escala de Fitzpatrick	29
5.1	Diagrama do processo	38
6.1	Resultados da extração de características CNN + VGG19 para classes 1 a 6	45
6.2	AUC para classes 1 a 6	45
6.3	Matriz de Confusão para classes 1 a 6	46
6.4	Resultados da extração de características CNN + VGG19 para classes 1 e 2	47
6.5	AUC para classes 1 e 2	47
6.6	Matriz de Confusão para classes 1 e 2	48
6.7	Resultados da extração de características CNN + VGG19 para classes 3 e 4	48
6.8	AUC para classes 3 e 4	49
6.9	Matriz de Confusão para classes 3 e 4	50
6.10	Resultados da extração de características CNN + VGG19 para classes 5 e 6	50
6.11	AUC para classes 5 e 6	51
6.12	Matriz de Confusão para classes 5 e 6	52
6.13	Acurácia para Classe 6	53
6.14	AUC para classes 6	53
6.15	Matriz de Confusão para Classe 6	54
7.1	Comparativo Resnet50 x VGG19	55

Lista de Tabelas

2.1	Resultados de Pesquisa CAPES 2019-2024	15
3.1	Fatores de risco e causas de diferentes cancer de pele no	20
5.1	Distribuição de Imagens por Classe	33
5.2	Distribuição de Imagens por Diagnóstico	35
5.3	Número de Imagens Seleccionadas por Classe e Imagem	36
6.1	Matriz de Confusão, com um total de 108 imagens	43
6.2	Relatório de Classificação para o Modelo	44
6.3	Matriz de Confusão	45
6.4	Relatório de Classificação para Classe 6	53

Sumário

1	Introdução	12
1.0.1	Potencial de Inovação	12
1.0.2	Desafios na Colaboração	13
2	Revisão da Literatura	14
2.0.1	Comparação com Estudos Recentes	14
2.0.2	Inovação e Impacto na Computação	15
2.1	Regularização com Dropout e L2	15
2.1.1	Regularização L2	16
2.1.2	Dropout	16
2.1.3	Comparação com Outros Artigos na Área	16
3	Classificação e Causas dos Tipos de Câncer	18
3.1	Riscos e Causas de Câncer	19
3.2	Câncer de Pele no Brasil	19
3.3	Variantes de Melanoma e Características Clínicas	19
3.4	Redes Neurais Convolucionais e Classificação de Imagens Médicas	22
3.5	Escala de Fitzpatrick e Detecção de Lesões Cutâneas	22
3.6	Abordagens Legais e Desafios na Dermatologia	23
3.7	Aplicações de Redes Neurais em Peles Negras	23
3.8	Normas e Diretrizes de Pesquisa	23
4	Metodologia	25

4.1	Conjunto de Dados ISIC 2020	25
4.1.1	Conjunto de Dados	25
	Melanoma	25
	Nevo	26
	Lentigo NOS	26
	Ceratose liquenoide	26
	Queratose seborreica	26
	Lentigo solar	27
	Mancha café-com-leite	27
4.2	Classificação da Cor da Pele	27
4.3	Pré-processamento de Dados	29
4.4	Aumentação de Dados	29
4.5	Divisão dos Dados	29
4.6	Treinamento do Modelo	30
4.7	Avaliação do Modelo	30
5	Métricas de Desempenho	31
5.1	Acurácia	31
5.2	Precisão	31
5.3	Recall	32
5.4	F1-Score	32
5.5	Considerações sobre o uso das métricas	32
5.5.1	Desequilíbrio de Classes	33
5.5.2	classificando a cor da pele segundo escala FitzPatrick	34
5.5.3	Seleção das Imagens	35
5.5.4	Pré-processamento	36
5.5.5	Aumentação de Dados	36

5.5.6	Divisão dos Dados em Conjuntos de Treinamento e Validação . . .	37
5.5.7	Treinamento do Modelo	38
5.6	Algoritmo	39
6	Resultados	41
6.1	Métricas de Desempenho	41
6.1.1	Matriz de Confusão	43
6.2	Resultados da Avaliação do Algoritmo	44
6.2.1	Todas as Classes (1 a 6)	44
6.2.2	Classes 1 e 2	46
6.2.3	Classes 3 e 4	47
6.2.4	Classes 5 e 6	49
6.2.5	Classe 6	50
6.2.5.1	Análise da Classe 6 (Pele mais escura)	51
7	Comparativo com Resnet50	55
8	Discussão	56
8.1	Desempenho do Modelo	56
8.2	Impacto do Desequilíbrio de Classes	56
8.3	Métricas de Desempenho	56
8.4	Limitações e Considerações Futuras	57
8.5	Ética e Viés nos Sistemas de IA	58
8.5.1	Análise de Viés	59
8.5.2	Mitigação do Viés	59
8.5.3	Tendências Futuras na Pesquisa em Visão Computacional para Di- agnósticos Médicos	59
8.5.3.1	Medicina Personalizada	59

8.5.4	Realidade Aumentada e Virtual	60
8.5.5	Síntese e Impacto Potencial na Prática Clínica e Pesquisa Futura .	60
8.5.5.1	Relevância Clínica	60
8.5.6	Direções Futuras	60
9	Conclusão	61
	REFERENCIAS	63

1 Introdução

O câncer de pele é um dos casos de câncer mais comuns no mundo . Entre os vários tipos, o melanoma é o mais agressivo e potencialmente fatal quando não detectado nos estágios iniciais. A identificação precisa de lesões cutâneas é, portanto, essencial para um tratamento eficaz. Diferentes redes neurais convolucionais (CNNs), como a VGG19, tem mostrado potencial na tarefa de classificação de imagens médicas devido à sua capacidade de extrair características relevantes de imagens complexas [21] .

O termo "lesão melanocítica" refere-se a um grupo de diferentes tipos de lesões cutâneas que surgem devido à proliferação de melanócitos, que podem ser benigna ou maligna. Entre as lesões malignas, o melanoma é o neoplasma cutâneo mais agressivo e frequentemente exibe progressão rápida e disseminação metastática. Com base em suas características clínicas e histológicas, os melanomas podem ser subdivididos em quatro classes principais: melanoma nodular, melanoma lentigo maligno, melanoma acral e melanoma de disseminação superficial. A aplicação da Escala Fitzpatrick é crucial para entender como diferentes tons de pele podem influenciar a detecção e classificação dessas lesões cutâneas.

1.0.1 Potencial de Inovação

O potencial de inovação deste estudo reside na abordagem multifacetada para a classificação de lesões cutâneas. Ao abordar a representatividade étnica e utilizar técnicas avançadas de aprendizado profundo, o modelo desenvolvido pode ser adaptado para uma variedade de contextos clínicos, melhorando a acurácia diagnóstica entre diferentes populações. Isso pode levar a um aumento na equidade nos cuidados de saúde, uma área crítica que tem sido tradicionalmente negligenciada em estudos de visão computacional.

1.0.2 Desafios na Colaboração

A colaboração interdisciplinar é essencial para o sucesso de projetos que envolvem a aplicação de CNNs em diagnósticos médicos, pois integra conhecimentos de áreas como computação, medicina e ética, resultando em soluções mais completas e eficazes. Estudos como o de Turing [40] enfatizam a importância de equipes multidisciplinares no desenvolvimento de sistemas de IA que sejam tecnicamente sólidos e clinicamente relevantes. A relevância desse estudo para a computação está na aplicação prática de técnicas de aprendizado profundo para resolver problemas de saúde pública, como a detecção precoce de melanoma, que pode salvar vidas. A criação de um sistema automatizado robusto auxilia dermatologistas na triagem e diagnóstico de lesões cutâneas, e a combinação de CNNs com técnicas de aumento de dados e análise de variabilidade étnica promove avanços significativos na construção de modelos mais justos e precisos.

Apesar dos benefícios, a colaboração interdisciplinar enfrenta desafios significativos, incluindo diferenças nos termos e metodologias utilizadas por diferentes disciplinas, bem como a necessidade de comunicação eficaz entre membros da equipe. Estudos de Wooliscroft e Howell [41] sugerem que programas de educação e treinamento que promovam a compreensão mútua e a comunicação eficaz podem ajudar a superar esses desafios.

2 Revisão da Literatura

A revisão da literatura sobre a classificação de lesões cutâneas e a detecção precoce de melanoma inclui diversos estudos que investigam diferentes abordagens e tecnologias voltadas para melhorar a precisão diagnóstica e o tratamento. Tais pesquisas analisam desde a aplicação de técnicas convencionais até a implementação de redes neurais avançadas, buscando desenvolver soluções mais eficazes para identificar e tratar lesões potencialmente perigosas, como o melanoma.

O uso de redes neurais convolucionais (CNNs) na classificação de imagens médicas, especialmente para a detecção precoce de melanoma, tem atraído crescente interesse nos últimos anos. Este estudo em particular adota as arquiteturas como EFFICIENTNET, DENSENET, MOBILENET, RESNET50, VGG16, VGG19 e muitas outras para classificar lesões cutâneas, evidenciando como a extração de características melhora significativamente a precisão diagnóstica. A importância dessa pesquisa pode ser contextualizada por sua comparação com outros estudos recentes, reforçando seu impacto na área de computação e visão computacional.

2.0.1 Comparação com Estudos Recentes

O trabalho de Esteva et al. [12] utilizou redes neurais profundas (deep learning) para a classificação de câncer de pele, atingindo níveis de precisão comparáveis aos dos dermatologistas. Os autores treinaram um modelo baseado na arquitetura Inception v3 com um grande conjunto de dados dermatológicos, mostrando que a rede neural poderia identificar melanoma com uma precisão significativamente alta [12]. A abordagem do nosso estudo utilizando VGG19 apresenta resultados similares e complementa essa linha de pesquisa ao incorporar a Escala de Fitzpatrick para avaliar a influência dos tons de pele na acurácia do modelo.

Outro estudo relevante é o de Haenssle et al. [17], que comparou a performance de um modelo baseado em CNN com a de dermatologistas em identificar melanomas

através de imagens dermatoscópicas. O CNN superou a maioria dos dermatologistas na precisão diagnóstica, destacando o potencial das CNNs como ferramentas de suporte no diagnóstico clínico [16]. Nosso estudo contribui adicionalmente ao focar na diversidade étnica dos pacientes, algo não explorado profundamente nesses estudos anteriores, como demonstrado na tabela 2.1.

Pesquisa CAPES 2019-2024	Número de Resultados
melanoma AND deep learning	3658
skin cancer AND deep learning	2618
skin cancer AND machine learning	1134
melanoma AND VGG16	98
black skin AND melanoma AND deep learning	45
black skin AND melanoma AND machine learning	24
black skin AND melanoma AND VGG16	4
black skin AND melanoma AND RESNET	3
black skin AND melanoma AND VGG19	1
skin cancer AND fitzpatrick	945
skin cancer AND fitzpatrick AND machine learning	8
skin cancer AND fitzpatrick AND deep learning	20

Tabela 2.1: Resultados de Pesquisa CAPES 2019-2024

2.0.2 Inovação e Impacto na Computação

Este estudo não apenas valida a eficácia das CNNs na classificação de lesões cutâneas, mas também introduz uma análise detalhada da variabilidade étnica usando a Escala de Fitzpatrick. Este aspecto é inovador pois aborda uma lacuna significativa na representatividade dos conjuntos de dados dermatológicos. A maioria dos modelos de visão computacional são treinados majoritariamente com imagens de indivíduos de pele clara, o que pode levar a vieses diagnósticos.

A metodologia empregada, que inclui aumento de dados específico para diferentes tons de pele, regularização com dropout e L2 (são técnicas utilizadas para prevenir o overfitting, que ocorre quando um modelo se ajusta excessivamente aos dados de treinamento e perde a capacidade de generalizar bem para novos dados), e a avaliação contínua com métricas rigorosas, contribui para o desenvolvimento de um modelo mais robusto e generalizável.

2.1 Regularização com Dropout e L2

No contexto de redes neurais e aprendizado profundo, a regularização com *dropout* e *L2* refere-se a técnicas utilizadas para prevenir o *overfitting*, que ocorre quando um modelo

se ajusta excessivamente aos dados de treinamento, perdendo sua capacidade de generalização.

2.1.1 Regularização L2

A regularização $L2$, também conhecida como *Ridge Regularization*, adiciona um termo de penalização à função de perda do modelo, proporcional ao quadrado dos pesos da rede. O objetivo dessa técnica é incentivar a rede a manter os pesos pequenos, evitando que o modelo se torne excessivamente complexo ou sensível aos dados de treinamento. A função de perda com regularização $L2$ pode ser expressa da seguinte forma:

$$Loss = Loss_{original} + \lambda \sum w^2$$

onde λ é o hiperparâmetro que controla a força da penalidade e w são os pesos da rede neural. Isso ajuda a simplificar o modelo e a torná-lo mais robusto.

2.1.2 Dropout

O *dropout* é uma técnica estocástica de regularização utilizada durante o treinamento de redes neurais, em que algumas unidades (*neurônios*) da rede são "desligadas" aleatoriamente a cada iteração de treinamento, ou seja, suas contribuições são temporariamente ignoradas. Isso força a rede a aprender padrões mais robustos, em vez de depender excessivamente de neurônios específicos. Durante a fase de inferência, todas as unidades são utilizadas, mas seus pesos são ajustados para compensar a ausência de *dropout* durante o treinamento. Essa técnica é amplamente eficaz para redes profundas e densas, ajudando a evitar o *overfitting*.

A inclusão de mais camadas no modelo também potencializa a capacidade de extração de características pela rede neural, permitindo uma detecção mais detalhada e precisa.

2.1.3 Comparação com Outros Artigos na Área

O estudo de Tschandl et al. (2020) investigou o uso de diferentes arquiteturas de CNN para a classificação de lesões de pele, incluindo ResNet e DenseNet, além de Inception v4. Os resultados mostraram que as arquiteturas mais profundas e complexas, como DenseNet, proporcionaram uma melhora marginal na precisão, mas com um custo computacional significativamente maior [39]. Em comparação, nosso uso da VGG19 oferece

um equilíbrio entre complexidade e desempenho, sendo mais eficiente em termos de tempo de treinamento e inferência, enquanto ainda mantém alta precisão.

Além disso, o trabalho de Brinker et al. (2019) focou na transferência de aprendizado, utilizando modelos pré-treinados em grandes conjuntos de dados de imagens gerais antes de fine-tuning com imagens dermatológicas específicas. Esta abordagem mostrou-se eficaz em melhorar a precisão do modelo, especialmente quando os dados específicos do domínio são escassos [6]. Nosso estudo segue uma abordagem similar, utilizando VGG19 pré-treinado, porém introduzindo uma abordagem inovadora ao integrar a análise por fototipos de Fitzpatrick.

3 Classificação e Causas dos Tipos de Câncer

Os cânceres podem ser classificados de acordo com a origem das células malignas e sua localização no corpo. Entre os tipos mais comuns estão o carcinoma, o sarcoma, o linfoma, a leucemia e o melanoma. Cada um desses tipos de câncer possui características e comportamentos distintos, o que influencia diretamente o diagnóstico, o tratamento e o prognóstico dos pacientes.

- **Carcinoma:** O carcinoma é o tipo mais prevalente de câncer e origina-se nas células epiteliais, que revestem a pele e os tecidos internos dos órgãos. Exemplos incluem carcinoma de células basais, carcinoma de células escamosas e adenocarcinoma.
- **Sarcoma:** Esse tipo de câncer surge nas células dos tecidos conjuntivos, como ossos, músculos, cartilagem e gordura. Osteossarcoma (osso) e lipossarcoma (tecido adiposo) são exemplos típicos de sarcomas.
- **Linfoma:** O linfoma afeta as células do sistema linfático, que fazem parte do sistema imunológico. Ele é classificado em dois principais tipos: linfoma de Hodgkin e linfoma não-Hodgkin.
- **Leucemia:** A leucemia é caracterizada pela proliferação anormal de células sanguíneas, especialmente leucócitos. Ela começa na medula óssea e pode se disseminar para o sangue, interferindo na produção normal de células sanguíneas.
- **Melanoma:** O melanoma é um tipo agressivo de câncer de pele que se desenvolve a partir dos melanócitos, células que produzem pigmento na pele. É o tipo mais perigoso de câncer de pele, devido à sua capacidade de metastatizar rapidamente para outras regiões do corpo [37].

3.1 Riscos e Causas de Câncer

Os fatores de risco e as causas dos diferentes tipos de câncer variam amplamente, e cada tipo está associado a diferentes predisposições genéticas, hábitos de vida e exposições ambientais. A seguir, apresenta-se uma tabela que resume alguns dos principais fatores de risco e causas associados aos tipos de câncer mencionados.

3.2 Câncer de Pele no Brasil

De acordo com o artigo de Luz et al. (2023) [26], é apresentada uma análise detalhada das variações no número de casos de melanoma e outras doenças de pele no Brasil. A pesquisa abrange um período de sete anos e examina as diferenças regionais na incidência dessas doenças. Utilizando dados epidemiológicos, o estudo visa identificar tendências e padrões que possam elucidar a distribuição dessas condições em diferentes regiões do país. A introdução destaca a relevância dos cânceres de pele como uma preocupação significativa de saúde pública e a necessidade de um monitoramento contínuo para a formulação de políticas de saúde eficazes.

Os resultados apresentados no artigo mostram variações substanciais entre as regiões brasileiras, sugerindo que fatores ambientais e socioeconômicos desempenham um papel crucial na incidência dessas doenças. A conclusão do estudo enfatiza a importância de considerar essas variáveis ao planejar intervenções de saúde pública e campanhas de conscientização. Além disso, o trabalho sugere a necessidade de mais pesquisas para compreender melhor os fatores subjacentes que contribuem para essas variações, propondo estratégias para melhorar a prevenção e o tratamento do câncer de pele no Brasil. Assim, o artigo de Luz et al. (2023) [26] fornece uma base valiosa para futuras investigações e para o desenvolvimento de políticas de saúde pública mais eficazes no combate às malignidades cutâneas.

3.3 Variantes de Melanoma e Características Clínicas

O melanoma é o neoplasma cutâneo mais agressivo, frequentemente exibindo progressão rápida e disseminação metastática. Cabrera e Recule [8] revisaram as características clínicas e histológicas de apresentações clínicas incomuns de melanoma, enfatizando a importância das descobertas dermatoscópicas. Magro et al. [27] discutiram variantes incomuns de melanoma, destacando a necessidade de reconhecimento e diagnóstico precisos

Tipo de Câncer	Fatores de Risco	Causas
Carcinoma	Idade avançada, exposição ao sol, tabagismo, infecções virais (HPV)	Mutação genética, exposição a carcinógenos, radiação UV, inflamação crônica
Sarcoma	Exposição a produtos químicos, radioterapia, histórico familiar	Alterações genéticas, radiação ionizante, síndromes genéticas hereditárias
Linfoma	Imunossupressão, infecções virais (EBV, HIV), idade avançada	Alterações no DNA dos linfócitos, infecções crônicas, doenças autoimunes
Leucemia	Exposição a radiação, produtos químicos, tabagismo, síndrome de Down	Mutação nas células-tronco hematopoiéticas, fatores genéticos, exposição a benzeno
Melanoma	Exposição ao sol, histórico de queimaduras solares, pele clara, múltiplos nevos	Danos ao DNA causados por radiação UV [20], predisposição genética, exposição a produtos químicos [22]

Tabela 3.1: Fatores de risco e causas de diferentes cancer de pele no

dessas variantes para melhorar os resultados dos pacientes.

A análise das estatísticas globais de câncer de 2018, conforme discutido por Bray et al. [5], oferece dados cruciais sobre a carga global do câncer de pele. O estudo revela que o câncer de pele, excluindo o melanoma, contribuiu significativamente para os números totais de incidência e mortalidade em 2018, destacando a importância contínua da vigilância e intervenção preventiva. Além disso, o câncer de pele melanoma, embora não seja o mais comum entre todos os câncer de pele no Brasil, figura como uma das formas mais letais, sublinhando a necessidade de estratégias eficazes de detecção precoce e tratamento em escala global.

Os dados também destacam a influência de fatores geográficos, econômicos e de estilo de vida na incidência e mortalidade por câncer de pele. A variabilidade geográfica substancial observada no estudo reflete disparidades significativas na distribuição do câncer de pele ao redor do mundo, com áreas de maior exposição solar e desenvolvimento econômico mais baixo frequentemente apresentando maiores taxas de incidência. Essas descobertas sublinham a necessidade de abordagens personalizadas e culturalmente sensíveis para o manejo e prevenção do câncer de pele, visando mitigar sua carga global e melhorar os resultados de saúde pública.

Pessoas negras possuem uma menor incidência de melanoma devido à proteção natural proporcionada pela melanina contra os raios UV prejudiciais, com uma taxa significati-

vamente menor de desenvolvimento da doença em comparação com pessoas brancas não hispânicas. No entanto, quando diagnosticadas com melanoma, pessoas negras enfrentam desafios substanciais na sobrevivência a longo prazo. De acordo com um estudo dos Centros de Controle e Prevenção de Doenças dos Estados Unidos, a taxa de sobrevivência de cinco anos para pacientes negros foi de apenas 66%, enquanto para pacientes brancos não hispânicos foi de 90% durante o período de 2011 a 2015.

O diagnóstico precoce é crucial, mas para pacientes negros, o melanoma muitas vezes se apresenta de forma diferente, como melanoma lentiginoso acral (LMA), localizado em áreas como palmas das mãos, solas dos pés ou abaixo das unhas, onde a exposição solar é mínima. Isso contrasta com a apresentação típica em pele mais clara, onde o melanoma é mais frequentemente encontrado em áreas expostas ao sol. Pesquisas indicam que dermatologistas podem enfrentar desafios na identificação precisa de condições de pele em tons mais escuros, sendo menos propensos a diagnosticar malignidades e mais propensos a recomendar biópsias para condições benignas em comparação com a pele mais clara. Essas discrepâncias destacam a necessidade crítica de aumentar a conscientização e a competência clínica para melhorar os resultados de saúde para pacientes de diferentes etnias e tons de pele.

A ausência de representatividade da diversidade de tons de pele em materiais educativos de medicina e nas ferramentas de diagnóstico tem implicações negativas significativas para a saúde pública, particularmente no que se refere ao câncer de pele. A escassez de imagens de peles escuras em atlas dermatológicos compromete a capacidade de identificar corretamente lesões cutâneas em pessoas negras e pardas, resultando em atrasos no diagnóstico e no início do tratamento. Esses atrasos podem ter consequências graves, como o agravamento das condições de saúde e a redução das chances de tratamento bem-sucedido, especialmente em casos como o melanoma, que exige diagnóstico precoce para melhor prognóstico [19].

Os negros não hispânicos americanos têm uma incidência menor de melanoma do que os brancos não hispânicos americanos. No entanto, eles são frequentemente diagnosticados com melanoma em um estágio mais avançado e têm taxas de sobrevivência mais baixas [9]. Isso provavelmente se deve a uma combinação de fatores, incluindo menor consciência dos fatores de risco do melanoma, menor acesso a cuidados de saúde e diferenças na biologia do melanoma em negros não hispânicos americanos. O tipo mais comum de melanoma em negros não hispânicos americanos é o melanoma lentiginoso acral, que ocorre nas palmas das mãos, plantas dos pés ou leitos ungueais. Este tipo de melanoma é frequentemente

difícil de diagnosticar porque pode ser confundido com um nevo benigno. É necessária uma melhor educação dos profissionais de saúde e do público sobre o melanoma em negros não hispânicos americanos para melhorar o diagnóstico precoce e as taxas de sobrevivência.

A falta de representatividade da diversidade da pele [2] em materiais de ensino médico e ferramentas de diagnóstico impacta negativamente a saúde pública, especialmente no contexto do câncer de pele. A baixa representatividade de tons de pele escura em atlas dermatológicos, por exemplo, dificulta o reconhecimento de lesões em pessoas negras e pardas, levando a atrasos no diagnóstico e tratamento [19], com consequências graves para a saúde e o bem-estar dos pacientes.

Estudos demonstram que a pigmentação da pele pode mascarar os sinais de alerta do câncer de pele [19], como assimetrias, bordas irregulares, variações de cor e crescimento da lesão. A ausência de imagens com tons de pele mais escuros em manuais e ferramentas de diagnóstico dificulta a percepção dessas características em pacientes negros e pardos, por parte dos profissionais de saúde. Essa disparidade racial no acesso a um diagnóstico preciso e oportuno contribui para as taxas mais baixas de sobrevivência do câncer de pele entre a população negra, quando comparadas à população branca [19].

3.4 Redes Neurais Convolucionais e Classificação de Imagens Médicas

As redes neurais convolucionais (CNNs), como a VGG19, têm mostrado grande potencial na tarefa de classificação de imagens médicas devido à sua capacidade de extrair características relevantes de imagens complexas. No trabalho de Simonyan e Zisserman [36] detalha-se a arquitetura da VGG19, que é amplamente utilizada em aplicações de visão computacional para grandes conjuntos de dados de imagens. Esteva et al. [13] demonstraram que a classificação de câncer de pele com redes neurais profundas pode atingir níveis de precisão comparáveis aos dos dermatologistas.

3.5 Escala de Fitzpatrick e Detecção de Lesões Cutâneas

A Escala Fitzpatrick estabeleceu um padrão para classificar o tom de pele (como uma paleta de cores).

Essa definição é crucial para entender como diferentes tons de pele podem influenciar a detecção e classificação de lesões cutâneas. Estudos como o de Goon et al. [15] questionam

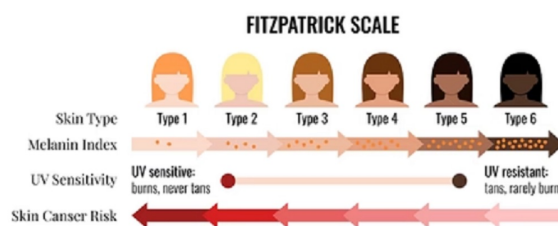


Figura 3.1: FitzPatrick Palette [31]

se a escala oferece uma falsa sensação de segurança para tipos de pele IV-VI, ressaltando a necessidade de modelos mais inclusivos. Kinyanjui et al. [24] investigaram a estimativa do tom de pele e seus efeitos no desempenho da classificação em conjuntos de dados de dermatologia, sublinhando a importância da inclusão de diversas tonalidades de pele nos estudos.

3.6 Abordagens Legais e Desafios na Dermatologia

Troxel [38] e Moshell et al. [30] discutiram aspectos medicolegais relacionados a erros em patologia e reivindicações de responsabilidade profissional contra dermatologistas, respectivamente, destacando a importância de diagnósticos precisos e documentação adequada para evitar litígios. Rayess et al. [34] analisaram litígios por negligência médica relacionados ao melanoma, sugerindo a necessidade de biópsias mais abrangentes para evitar diagnósticos incorretos.

3.7 Aplicações de Redes Neurais em Peles Negras

Barros et al. [3] avaliaram a generalização de modelos baseados em redes neurais profundas para lesões cutâneas em peles negras. Eles descobriram que modelos treinados predominantemente em tons de pele branca não apresentam bom desempenho em peles negras, destacando a necessidade de criar conjuntos de dados diversificados e modelos especializados para garantir diagnósticos inclusivos.

3.8 Normas e Diretrizes de Pesquisa

Xavier [42] oferece diretrizes detalhadas sobre como usar as normas da ABNT para citações, o que é essencial para a padronização e qualidade dos trabalhos acadêmicos. Da

Silva e Menezes [10] fornecem uma metodologia abrangente para a pesquisa e elaboração de dissertações, enquanto Keele [23] e Kitchenham et al. [25] discutem a realização de revisões sistemáticas da literatura em engenharia de software, enfatizando a importância de metodologias rigorosas para garantir a confiabilidade dos resultados.

4 Metodologia

Esta seção descreve os métodos e materiais utilizados para a classificação de lesões cutâneas usando a arquitetura de rede neural convolucional VGG19 com o conjunto de dados ISIC 2020.

4.1 Conjunto de Dados ISIC 2020

O conjunto de dados ISIC 2020, fornecido pela International Skin Imaging Collaboration (ISIC), é amplamente utilizado para a classificação de lesões cutâneas. Ele contém imagens de sete tipos diferentes de lesões cutâneas, incluindo melanoma, nevo, lentigo, ceratose liquenoide, queratose seborreica, lentigo solar e mancha café-com-leite. Este conjunto de dados é conhecido por apresentar desafios significativos, como desequilíbrio de classes e sub-representação de imagens de pele negra.

4.1.1 Conjunto de Dados

Para este estudo, utilizamos o conjunto de dados ISIC 2020, que é composto por imagens de sete tipos diferentes de lesões de pele, incluindo:

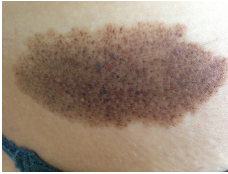
Melanoma

Descrição: Um tipo agressivo de câncer de pele que se desenvolve a partir dos melanócitos, as células que produzem pigmento na pele.



Nevo

Descrição: Também conhecido como "pinta" ou "marca de nascença", é uma lesão pigmentada benigna na pele.



Lentigo NOS

Descrição: Um tipo de mancha escura na pele, muitas vezes causada por exposição solar prolongada.



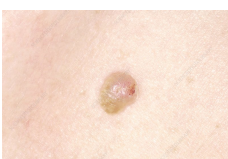
Ceratose liquenoide

Descrição: Uma lesão inflamatória crônica na pele que pode se tornar escamosa e áspera ao longo do tempo.



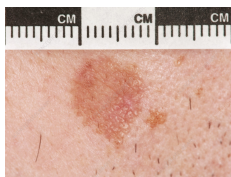
Queratose seborreica

Descrição: Também conhecida como "verruga seborreica", é uma lesão benigna comum que pode parecer uma mancha marrom escura ou preta na pele.



Lentigo solar

Descrição: Uma lesão pigmentada benigna causada pela exposição solar crônica, comum em áreas expostas ao sol.



Mancha café-com-leite

Descrição: Uma mancha na pele que é mais clara que a pele circundante e pode variar em tamanho. Geralmente é benigna.



4.2 Classificação da Cor da Pele

Utilizamos a Escala Fitzpatrick[14] para classificar a cor da pele das imagens. Esta escala categoriza os tipos de pele em seis grupos com base na resposta ao sol e pigmentação. Para determinar o tom de pele das imagens, aplicamos técnicas de visão computacional, incluindo a conversão para o espaço de cores CIELAB e o cálculo do Ângulo de Tipologia Individual (ITA).

$$ITA = \arctan \left(\frac{L - 50}{B} \right) \times \left(\frac{180}{\pi} \right)$$

Os valores calculados de ITA° demonstraram refletir a sensibilidade de um indivíduo à radiação UV e, conseqüentemente, a sensibilidade da pele à exposição ao sol e danos à pele, como alterações de pigmentação, câncer e envelhecimento. De um ponto de vista prático, isso se traduz na determinação do tipo de proteção e/ou tratamentos mais eficazes em tipos específicos de pele.

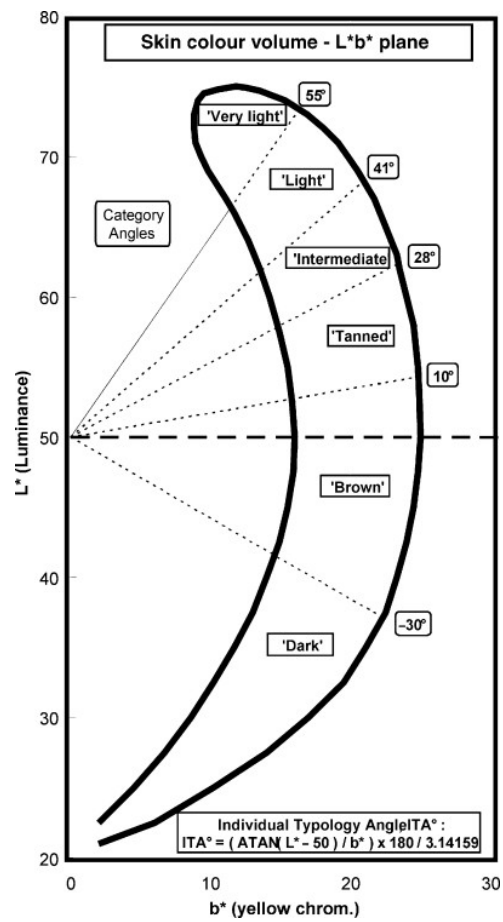


Figura 4.1: Volume de cor da pele projetado no plano L*b* do espaço Lab (Fonte da imagem: [4]).

I		BLONDE		BLUE, GREEN, GREY		WHITE / PALE / FRECKLED Extremely fair skin, always burns, never tans
II		BLONDE/RED		BLUE		WHITE / PALE WITH BEIGE TINT Fair skin, usually burns, sometimes tans
III		LIGHT BROWN		BROWN		WHITE TO LIGHT BROWN White olive skin, sometimes burns, tans mostly uniformly
IV		MEDIUM BROWN		BROWN		LIGHT TO MODERATE BROWN Rarely burns, always tans
V		DARK BROWN		BROWN		MEDIUM TO DARK BROWN Rarely burns, tans more than average
VI		BLACK		BROWN		DARK BROWN TO BLACK Never burns

Figura 4.2: Escala de Fitzpatrick

4.3 Pré-processamento de Dados

As imagens foram redimensionadas para 224x224 pixels e normalizadas. Aplicamos técnicas de aumento de dados para mitigar o desequilíbrio de classes, incluindo rotação, deslocamento, cisalhamento, zoom e flip horizontal.

4.4 Aumentação de Dados

A aumento de dados foi realizada utilizando a classe ImageDataGenerator da biblioteca Keras. As seguintes transformações foram aplicadas:

Rotação de até 20 graus

Deslocamento horizontal e vertical de até 20%

Cisalhamento de até 20%

Zoom de até 20%

Flip horizontal

4.5 Divisão dos Dados

Os dados foram divididos em conjuntos de treinamento e validação usando a função `train_test_split` da biblioteca Scikit-learn, com 80% dos dados para treinamento e 20% para validação. Esta divisão garante uma avaliação objetiva do desempenho do modelo.

4.6 Treinamento do Modelo

Utilizamos um modelo VGG19 pré-treinado, com as camadas finais ajustadas para a classificação das sete categorias de lesões cutâneas. O modelo foi treinado usando a função de perda de entropia cruzada e o otimizador Adam. O treinamento foi realizado em 20 épocas com uma taxa de aprendizado ajustável através de um callback.

4.7 Avaliação do Modelo

O desempenho do modelo foi avaliado utilizando métricas como acurácia, precisão, recall e F1-score. A matriz de confusão foi utilizada para visualizar o desempenho do modelo em classificar corretamente ou incorretamente as imagens de lesões cutâneas.

5 Métricas de Desempenho

No contexto de aprendizado de máquina e classificação, as métricas de desempenho são essenciais para avaliar a eficácia de um modelo. As métricas mais comuns incluem **acurácia**, **precisão**, **recall** e **F1-score**. Cada uma delas oferece uma perspectiva diferente sobre a performance do modelo, sendo úteis em diferentes cenários.

5.1 Acurácia

A **acurácia** é uma das métricas mais simples e amplamente utilizadas. Ela mede a proporção de previsões corretas em relação ao número total de exemplos avaliados. A fórmula para calcular a acurácia é:

$$Acurcia = \frac{TP + TN}{TP + TN + FP + FN}$$

onde:

- TP (True Positives) são os verdadeiros positivos,
- TN (True Negatives) são os verdadeiros negativos,
- FP (False Positives) são os falsos positivos,
- FN (False Negatives) são os falsos negativos.

A acurácia é uma métrica adequada quando as classes estão balanceadas, ou seja, quando há uma quantidade semelhante de exemplos em cada classe.

5.2 Precisão

A **precisão** (*precision*) mede a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo. Ela é particularmente útil quando o custo

de falsos positivos é alto. A fórmula para calcular a precisão é:

$$Preciso = \frac{TP}{TP + FP}$$

Uma alta precisão indica que, dos exemplos classificados como positivos, muitos são realmente positivos.

5.3 Recall

O **recall** (ou *sensitivity*) mede a proporção de verdadeiros positivos em relação ao total de exemplos positivos reais. Ele é importante em cenários onde a identificação correta dos positivos é crucial, como no diagnóstico de doenças. A fórmula para o cálculo do recall é:

$$Recall = \frac{TP}{TP + FN}$$

Um alto recall indica que o modelo é capaz de identificar a maioria dos exemplos positivos, mesmo que cometa alguns erros ao classificar exemplos negativos como positivos.

5.4 F1-Score

O **F1-score** é a média harmônica entre a precisão e o recall. Ele é útil quando há um equilíbrio necessário entre essas duas métricas, especialmente em problemas de classificação com classes desbalanceadas. A fórmula para o cálculo do F1-score é:

$$F1 = 2 \times \frac{Preciso \times Recall}{Preciso + Recall}$$

O F1-score é uma métrica equilibrada que combina tanto a capacidade do modelo de evitar falsos positivos (precisão) quanto a de detectar a maioria dos positivos (recall).

5.5 Considerações sobre o uso das métricas

Dependendo do contexto da aplicação, diferentes métricas podem ser mais relevantes. A acurácia pode ser enganosa quando as classes estão desbalanceadas, enquanto o F1-score oferece uma visão mais equilibrada ao considerar tanto a precisão quanto o recall. A

escolha da métrica deve refletir a natureza do problema e os custos associados a falsos positivos e falsos negativos.

5.5.1 Desequilíbrio de Classes

Um dos desafios significativos ao utilizar o conjunto de dados ISIC 2020 é o desequilíbrio de classes entre as diferentes categorias de lesões de pele. Algumas classes, [5.2](#) podem ter menos exemplos disponíveis em comparação com outras lesões benignas. Este desequilíbrio pode afetar a capacidade do modelo de aprender efetivamente a distinguir entre classes minoritárias, resultando em uma precisão geral inferior para essas categorias menos representadas.

Classe	Número de Imagens
0	32542
1	584
Total Resultado	33126

Tabela 5.1: Distribuição de Imagens por Classe

5.5.2 classificando a cor da pele segundo escala FitzPatrick

Algoritmo 1 Cálculo do Índice de Tipologia Individual (ITA) de uma Imagem

Entrada: Uma imagem no espaço de cores LAB. **Saída:** O valor do ITA para a imagem.

Passo 1: Extração da Luminância (L) Obtenha os valores de luminosidade (L) da imagem, que variam de 0 (preto) a 100 (branco). Ignore os pixels com luminosidade igual a 0, pois são completamente pretos.

Passo 2: Cálculo da Média e Desvio Padrão da Luminância

- Calcule a média dos valores de luminosidade restantes.
- Calcule o desvio padrão, que indica o quanto os valores de luminosidade variam em torno da média.

Passo 3: Filtragem dos Valores de Luminância

- Mantenha apenas os valores de luminosidade que estão dentro de um intervalo de ± 1 desvio padrão da média.

Passo 4: Extração e Filtragem do Componente b (Amarelo-Azul)

- Obtenha os valores do componente b da imagem, que representam a tonalidade amarela-azul.
- Calcule a média e o desvio padrão desses valores.
- Mantenha apenas os valores de b que estão dentro de um intervalo de ± 1 desvio padrão da média.

Passo 5: Cálculo do ITA

- Calcule o ITA usando a relação entre a média ajustada da luminância e a média do componente b:

$$ITA = \arctan \left(\frac{\text{Média da Luminância} - 50}{\text{Média do Componente b}} \right) \times \frac{180}{\pi}$$

Passo 6: Retorno do ITA

- Retorne o valor do ITA, que será usado para classificar a cor da pele da imagem.
-

Além disso, há uma sub-representação notável de indivíduos com tons de pele mais escuros no conjunto de dados. Apenas cerca de 10% das imagens são de indivíduos de pele de fototipo V e VI. Isso é problemático porque os modelos treinados com desequilíbrios semelhantes podem não generalizar bem para populações com diferentes tons de pele. Como resultado, diagnósticos automatizados podem ser menos precisos para esses grupos, exacerbando desigualdades já presentes na prática médica tradicional.

Pesquisas recentes destacam a importância de abordar esses desafios para melhorar a equidade e a precisão dos sistemas de diagnóstico automatizado de câncer de pele. Estratégias para mitigar o desequilíbrio de classes incluem técnicas de aumento de dados específicas para classes minoritárias, ajustes nos pesos das classes durante o treinamento do modelo, e métodos avançados de regularização para evitar que o modelo se concentre apenas nas classes majoritárias.

Diagnóstico	Número de Imagens
atypical melanocytic proliferation	1
cafe-au-lait macule	1
lentigo NOS	44
lichenoid keratosis	37
melanoma	584
nevus	5193
seborrheic keratosis	135
solar lentigo	7
unknown	27124*
Total Resultado	33126

Tabela 5.2: Distribuição de Imagens por Diagnóstico

5.5.3 Seleção das Imagens

As imagens foram selecionadas de acordo com os tons de pele. Essa informação não existia no dataset original, sendo assim foi necessário usar um algoritmo para identificar a cor da pele e classificar as imagens de acordo com a Escala de Fitzpatrick. Inicialmente somente as escalas V e VI foram selecionadas, mas para efeito de comparação, utilizei as classes I e II (Pele clara), III e IV (Pele intermediária) e V e VI (Pele negra) Outro filtro foi utilizado para que as patologias fossem diferentes de "unknown"(desconhecido), mesmo que o indicador estivesse marcado com 0 ou 1 para a presença de melanoma. Desta forma

somente **6.072** foram avaliadas na execução dos testes e validações do algoritmo.

Em nossos teste, avaliamos o algoritmo onde foram filtrados imagens por target (0/1) onde o 1 representa o melanoma, e o 0 representa os outros tipos de diagnósticos, conforme a tabela 5.1 e diagnosticos diferentes de "unknow"(diagnostico desconhecido)

Tabela 5.3: Número de Imagens Selecionadas por Classe e Imagem

Categoria	Classe 1 e 2	Classe 3 e 4	Classe 5 e 6	Classes 1 a 6
Número de Imagens	3684	1094	1292	6072
20% para treinamento	737	219	258	1214

5.5.4 Pré-processamento

As imagens foram redimensionadas para 224x224 pixels, normalizadas e divididas em conjuntos de treinamento e teste. Utilizamos diferentes técnicas de aumento de dados para mitigar o desequilíbrio de classes.

5.5.5 Aumentação de Dados

A aumentoção de dados é uma técnica essencial no treinamento de modelos de aprendizado de máquina, especialmente em problemas onde a quantidade de dados disponíveis é limitada. Ao aplicar transformações aleatórias nas imagens de treinamento, é possível gerar novas amostras a partir das existentes, aumentando assim a variabilidade dos dados e ajudando a prevenir o overfitting.

No presente trabalho, utilizamos a classe `ImageDataGenerator` da biblioteca Keras para realizar a aumentoção de dados. O algoritmo de aumentoção é configurado com uma série de parâmetros que definem as transformações a serem aplicadas às imagens:

- **rotation_range=20**: Rotaciona as imagens aleatoriamente em até 20 graus.
- **width_shift_range=0.2**: Desloca as imagens horizontalmente em até 20% da largura total da imagem.
- **height_shift_range=0.2**: Desloca as imagens verticalmente em até 20% da altura total da imagem.
- **shear_range=0.2**: Aplica uma transformação de cisalhamento às imagens em até 20%.

- **zoom_range=0.2**: Aplica um zoom aleatório nas imagens em até 20%.
- **horizontal_flip=True**: Inverte horizontalmente as imagens.
- **fill_mode='nearest'**: Preenche os pixels vazios gerados pelas transformações com os valores dos pixels mais próximos.

Essas transformações aumentam a robustez do modelo, tornando-o mais capaz de generalizar para novos dados não vistos durante o treinamento. A técnica de aumento de dados é particularmente eficaz em tarefas de classificação de imagens médicas, onde as variações nas imagens de entrada podem ser cruciais para a correta identificação de diferentes classes de lesões cutâneas.

Ao aplicar essas técnicas de aumento de dados, buscamos criar um conjunto de dados de treinamento mais diversificado e representativo, contribuindo para a melhora do desempenho do modelo VGG19 na classificação de lesões cutâneas.

5.5.6 Divisão dos Dados em Conjuntos de Treinamento e Validação

A divisão dos dados em conjuntos de treinamento e validação é um passo fundamental no processo de desenvolvimento de modelos de aprendizado de máquina. Esta etapa garante que o modelo seja treinado em um subconjunto dos dados e avaliado em um subconjunto separado, permitindo uma avaliação objetiva de seu desempenho em dados não vistos.

No presente trabalho, utilizamos a função `train_test_split` da biblioteca Scikit-learn para realizar a divisão dos dados. Esta função divide os dados de entrada `X` e os rótulos correspondentes `y` em conjuntos de treinamento e validação, de acordo com a proporção especificada pelo parâmetro `test_size`. O código utilizado para esta divisão é apresentado a seguir:

A divisão dos dados em conjuntos de treinamento e validação é crucial para garantir que o modelo treinado possa generalizar bem para novos dados. Ao reservar uma parte dos dados para validação, podemos monitorar o desempenho do modelo durante o treinamento e ajustar os hiperparâmetros conforme necessário para evitar o overfitting. Essa abordagem é especialmente importante em tarefas de classificação de imagens médicas, onde a precisão do modelo tem um impacto direto na qualidade do diagnóstico.

5.5.7 Treinamento do Modelo

Nossa solução é baseada em um modelo VGG19[36] pré-treinado, com as camadas finais ajustadas para a classificação das sete categorias. As categorias foram agrupadas de forma binária, para positivo (quando existência de melanoma) ou negativo (quando qualquer uma das 6 categorias), Foi utilizada a função de perda de entropia cruzada e o otimizador empregado foi o Adam. A Figura 5.1 ilustra o processo de treinamento.

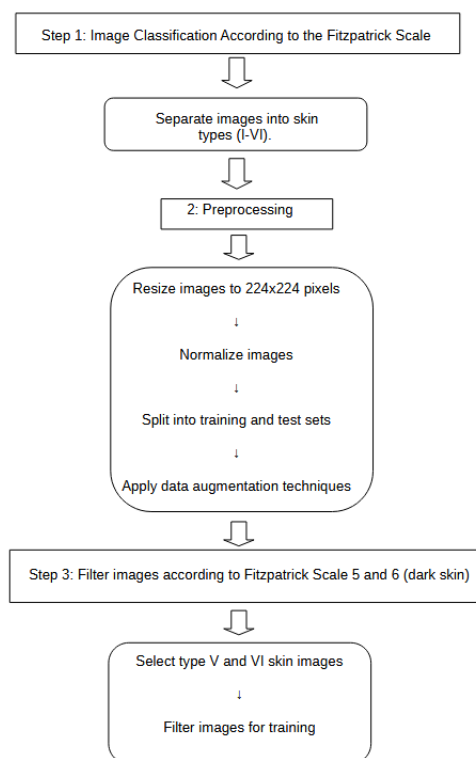


Figura 5.1: Diagrama do processo

5.6 Algoritmo

Importar bibliotecas necessárias:

Importar bibliotecas como pandas, os, numpy, cv2, tensorflow e outras relevantes.

Montar Google Drive:

Montar o Google Drive para acesso aos dados armazenados.

Carregar dados tabulares do ISIC 2020:

Carregar o arquivo CSV contendo os dados de treinamento do ISIC 2020.

Filtrar dados removendo entradas com informações de localização anatômica ou diagnóstico desconhecidas.

Carregar dados de cor da pele:

Carregar arquivo CSV contendo informações sobre a cor da pele.

Remover linhas com valores vazios na coluna de ITA.

Filtrar apenas imagens com cores de pele nos níveis 1 a 6 da escala Fitzpatrick.

Mesclar dados:

Mesclar dados de diagnóstico com dados filtrados por cor da pele.

Filtrar dados para incluir apenas entradas com rótulos de alvo 0 ou 1.

Filtragem e amostragem dos dados:

Definir o número máximo de imagens por tipo de diagnóstico.

Filtrar e amostrar dados para garantir um número equilibrado de imagens por tipo de diagnóstico.

Filtrar e amostrar imagens com rótulo de alvo igual a 1.

Combinar dados filtrados para obter um conjunto de dados final equilibrado.

Remover colunas desnecessárias e garantir que todos os registros tenham valores não nulos nas colunas relevantes.

Função para carregar e pré-processar imagens:

Definir função para carregar e pré-processar imagens usando VGG19.

Redimensionar as imagens, converter para o formato RGB e aplicar pré-processamento específico do VGG19.

Carregar imagens e rótulos:

Definir função para carregar imagens e seus respectivos rótulos.

Carregar imagens do diretório especificado e aplicar pré-processamento.

Dividir dados em treino e validação:

Dividir dados em conjuntos de treino e validação utilizando uma proporção de 80% para treino e 20% para validação.

Carregar modelo VGG19 pré-treinado:

Carregar modelo VGG19 pré-treinado com pesos da ImageNet, excluindo a camada superior.

Congelar camadas convolucionais do VGG19:

Congelar todas as camadas convolucionais do modelo VGG19 para evitar que sejam treinadas.

Adicionar camadas adicionais:

Adicionar camadas adicionais ao modelo VGG19 para a tarefa específica de classificação binária:

Camada Flatten para achatar a saída do VGG19.

Camada Dense com 512 neurônios e ativação ReLU.

Camada Dropout com taxa de 0.5.

Camada Dense com 256 neurônios e ativação ReLU.

Camada Dropout com taxa de 0.5. Camada Dense com 1 neurônio e ativação sigmoide.

Compilar o modelo:

Compilar modelo utilizando otimizador Adam, função de perda de entropia cruzada binária e métricas de acurácia, AUC, precisão e recall.

Callback para reduzir a taxa de aprendizado:

Definir callback para reduzir a taxa de aprendizado quando a perda de validação parar de melhorar.

Treinar o modelo:

Treinar o modelo com dados de treino, validando-o com dados de validação por 20 épocas.

Salvar modelo e histórico:

Salvar modelo treinado e histórico de treinamento no Google Drive. Fazer previsões no conjunto de validação:

Fazer previsões no conjunto de validação utilizando o modelo treinado. Converter previsões em rótulos binários (0 ou 1).

Salvar rótulos e predições:

Salvar rótulos verdadeiros e predições em arquivos CSV no Google Drive. Matriz de confusão e relatório de classificação:

Calcular e imprimir matriz de confusão e relatório de classificação para avaliar desempenho do modelo.

6 Resultados

6.1 Métricas de Desempenho

Métricas de desempenho como acurácia, precisão, recall e pontuação F1 são cruciais na avaliação de modelos usados em diagnósticos médicos. Compreender essas métricas de forma precisa é essencial devido ao impacto direto que falsos positivos e falsos negativos podem ter nos resultados diagnósticos[18].

Falsos positivos ocorrem quando o modelo identifica incorretamente um caso como positivo, indicando a presença de uma condição quando na verdade ela não está presente. Isso pode levar a procedimentos médicos desnecessários, ansiedade do paciente e desperdício de recursos médicos. Por exemplo, no diagnóstico de câncer, um falso positivo pode resultar em biópsias invasivas ou tratamentos agressivos não justificados.

Por outro lado, falsos negativos ocorrem quando o modelo falha em identificar corretamente um caso positivo, indicando incorretamente que a condição está ausente. Isso pode resultar em atrasos no tratamento, permitindo que a doença progrida sem intervenção adequada. No diagnóstico de câncer, um falso negativo pode levar a um diagnóstico tardio, reduzindo significativamente as chances de tratamento eficaz e potencialmente salvador de vidas.

Portanto, entender e minimizar falsos positivos e falsos negativos são cruciais para garantir a precisão dos diagnósticos médicos automatizados, fornecendo resultados confiáveis que apoiam decisões clínicas adequadas e melhoram o cuidado ao paciente.

Métrica	Descrição
Verdadeiro Positivo (VP)	Casos positivos identificados corretamente.
Falso Positivo (FP)	Casos negativos identificados incorretamente.
Verdadeiro Negativo (VN)	Casos negativos identificados corretamente.
Falso Negativo (FN)	Casos positivos identificados incorretamente.

Acurácia:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (6.1)$$

A acurácia é uma medida fundamental de desempenho em modelos de classificação, utilizada para avaliar a proporção de previsões corretas em relação ao total de previsões feitas pelo modelo. Em termos simples, a acurácia mede a precisão geral do modelo em classificar corretamente as amostras.

Precisão:

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (6.2)$$

A precisão é uma métrica crucial para avaliar o desempenho de um modelo de classificação, especialmente em contextos onde o custo de falsos positivos é alto. A precisão mede a proporção de verdadeiros positivos em relação ao total de previsões positivas feitas pelo modelo.

Recall:

$$\text{Recall} = \frac{VP}{VP + FN} \quad (6.3)$$

Recall, também conhecido como sensibilidade ou taxa de verdadeiros positivos, é uma métrica essencial para avaliar o desempenho de modelos de classificação, especialmente em situações onde identificar todos os casos positivos é crucial. Recall mede a proporção de verdadeiros positivos em relação a todos os exemplos que realmente pertencem à classe positiva.

Pontuação F1:

$$\text{Pontuação F1} = 2 \times \left(\frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \right) \quad (6.4)$$

A Pontuação F1 é uma métrica que combina precisão e recall em uma única medida, proporcionando uma visão equilibrada do desempenho de um modelo de classificação. É particularmente útil quando há um desequilíbrio entre as classes de dados, o que é comum em muitos problemas do mundo real.

Essas métricas são cruciais para entender e avaliar a eficácia de modelos de classificação em diversas aplicações.

6.1.1 Matriz de Confusão

O gráfico de matriz de confusão é uma ferramenta poderosa para visualizar rapidamente o número de instâncias classificadas corretamente ou incorretamente pelo modelo.

Abaixo está um exemplo de matriz de confusão resultante de um modelo de classificação: Foram avaliadas 540 imagens, e 20% foram usadas para teste, totalizando 108 imagens.

Para um exemplo de classificação de 108 imagens,

	Classe 0	Classe 1
Classe 0	37 (TN)	9 (FP)
Classe 1	12 (FN)	50 (TP)

Tabela 6.1: Matriz de Confusão, com um total de 108 imagens

- **37 (True Negative - TN):** O modelo classificou corretamente 37 instâncias da classe 0 (lesões benignas).
- **9 (False Positive - FP):** O modelo classificou incorretamente 9 instâncias da classe 0 como sendo da classe 1 (lesões malignas).
- **12 (False Negative - FN):** O modelo classificou incorretamente 12 instâncias da classe 1 como sendo da classe 0.
- **50 (True Positive - TP):** O modelo classificou corretamente 50 instâncias da classe 1 (lesões malignas).

Esses valores ajudam a entender o desempenho do modelo, permitindo identificar não apenas a quantidade de acertos, mas também os tipos de erros cometidos.

6.2 Resultados da Avaliação do Algoritmo

Nesta seção, apresentamos os resultados da avaliação do uso de uma CNN com VGG19 para a extração de características em patches de imagens de dermatoscopia, analisando os dados de acordo com as escalas de Fitzpatrick 1 e 2, 3 e 4, e 5 e 6.

6.2.1 Todas as Classes (1 a 6)

O modelo foi treinado com 6072 imagens, abrangendo todas as classes, para que os resultados pudessem ser avaliados sem separação por cor da pele. A validação foi realizada com 1215 imagens.

- Acurácia de Treinamento: 0.9555
- Acurácia de Validação: 0.9350
- Perda de Treinamento: 0.0852
- Perda de Validação: 0.2143
- AUC de Treinamento: 0.9897
- AUC de Validação: 0.9616
- Precisão de Treinamento: 0.9772
- Precisão de Validação: 0.9162
- Recall de Treinamento: 0.7618
- Recall de Validação: 0.7353

Classe	Precisão	Recall	F1-Score	Suporte
0	0.94	0.98	0.96	977
1	0.92	0.74	0.82	238
Acurácia Média	0.93			
Média Macro	0.93	0.86	0.89	1215
Média Ponderada	0.93	0.93	0.93	1215

Tabela 6.2: Relatório de Classificação para o Modelo

Classe	Número de Imagens
0	961
1	16
2	63
3	175

Tabela 6.3: Matriz de Confusão

A tabela 6.2 mostra o relatório de classificação para o modelo. Já a tabela 6.3 apresenta a matriz de confusão correspondente.

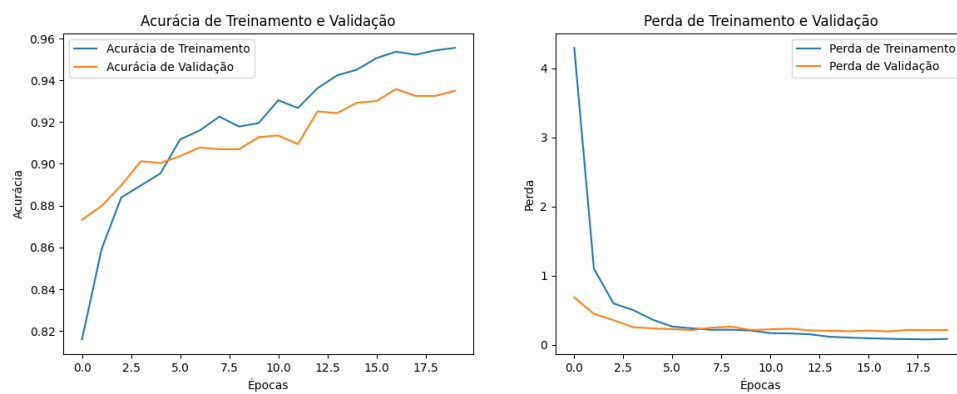


Figura 6.1: Resultados da extração de características CNN + VGG19 para classes 1 a 6

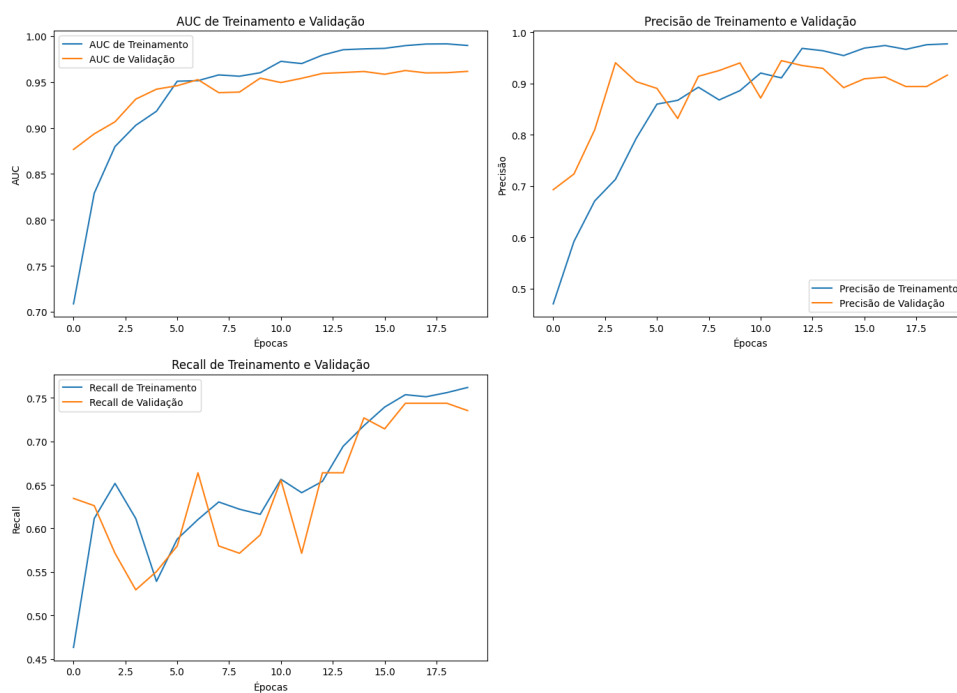


Figura 6.2: AUC para classes 1 a 6

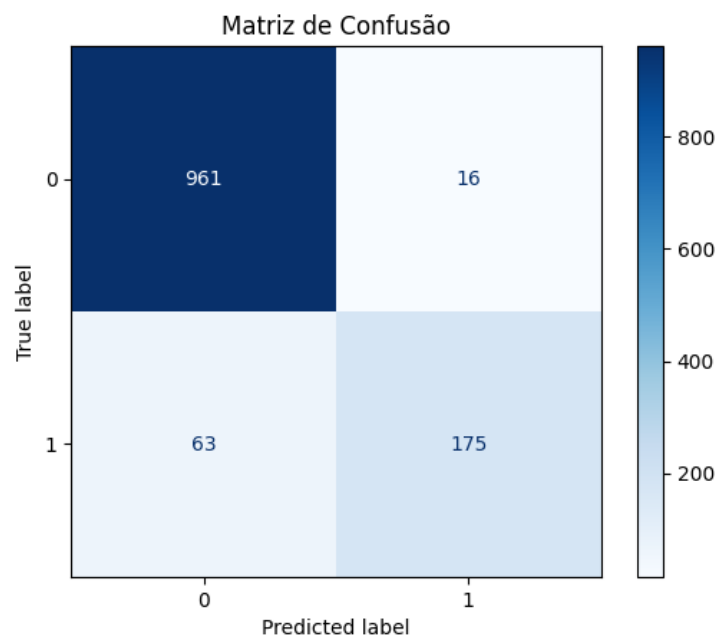


Figura 6.3: Matriz de Confusão para classes 1 a 6

6.2.2 Classes 1 e 2

As classes 1 e 2 representam peles mais claras e que possuem mais amostras no dataset. Foram filtradas 3684 imagens, 20% representa 737 imagens.

- Acurácia de Treinamento: 0.95
- Acurácia de Validação: 0.93
- Perda de Treinamento: 0.08
- Perda de Validação: 0.28
- AUC de Treinamento: 0.98
- AUC de Validação: 0.93
- Precisão de Treinamento: 0.97
- Precisão de Validação: 0.86
- Recall de Treinamento: 0.74
- Recall de Validação: 0.74

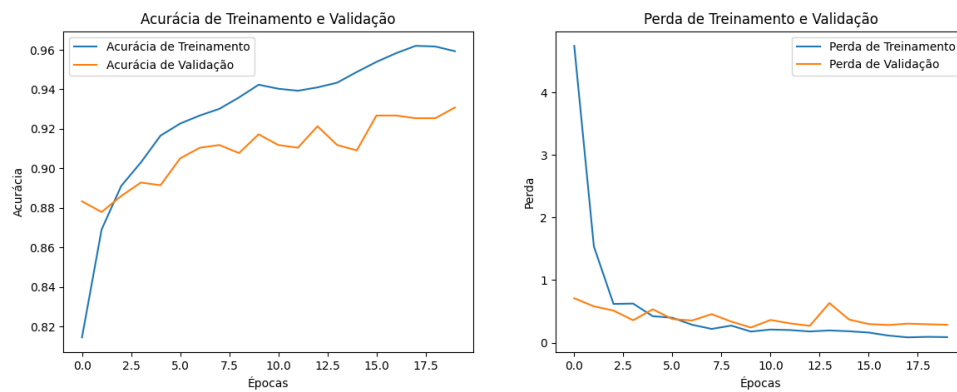


Figura 6.4: Resultados da extração de características CNN + VGG19 para classes 1 e 2

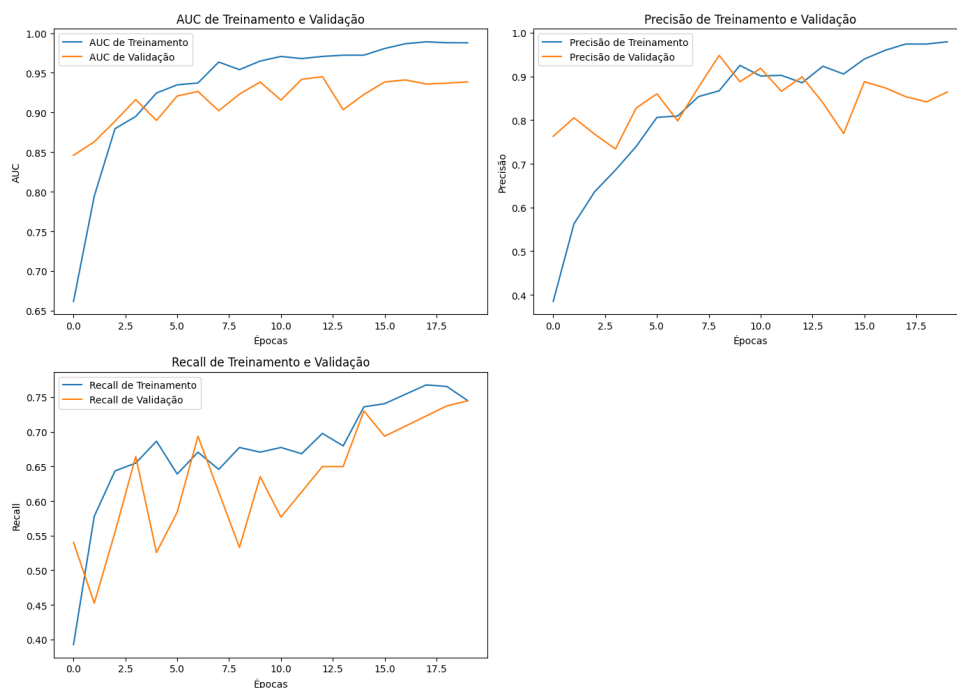


Figura 6.5: AUC para classes 1 e 2

6.2.3 Classes 3 e 4

As classes 3 e 4 representam peles intermediárias amostras no dataset. Foram selecionadas 1095 imagens, sendo 219 representando os 20%

- Acurácia de Treinamento: 0.99
- Acurácia de Validação: 0.96
- Perda de Treinamento: 0.03
- Perda de Validação: 0.40

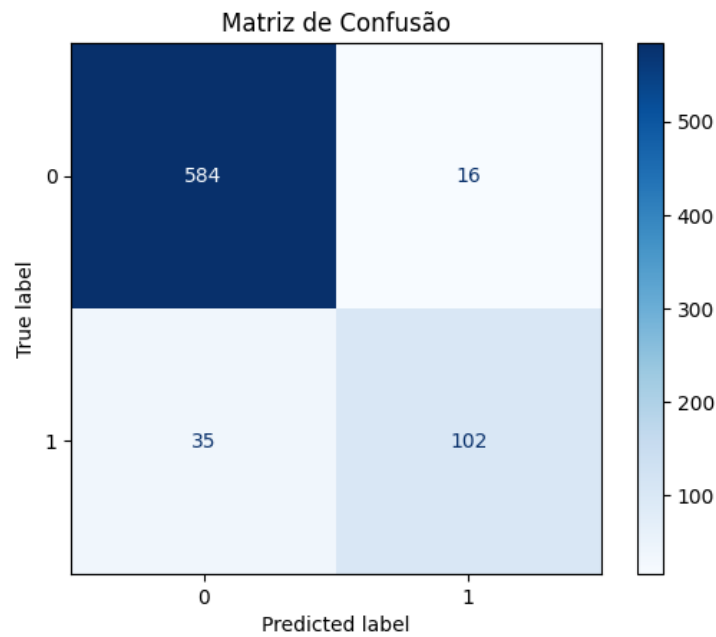


Figura 6.6: Matriz de Confusão para classes 1 e 2

- AUC de Treinamento: 0.99
- AUC de Validação: 0.97
- Precisão de Treinamento: 0.97
- Precisão de Validação: 0.94
- Recall de Treinamento: 0.98
- Recall de Validação: 0.92

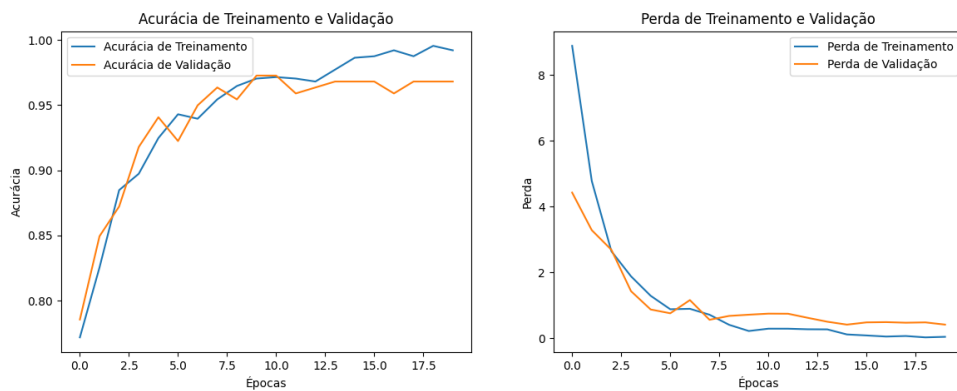


Figura 6.7: Resultados da extração de características CNN + VGG19 para classes 3 e 4

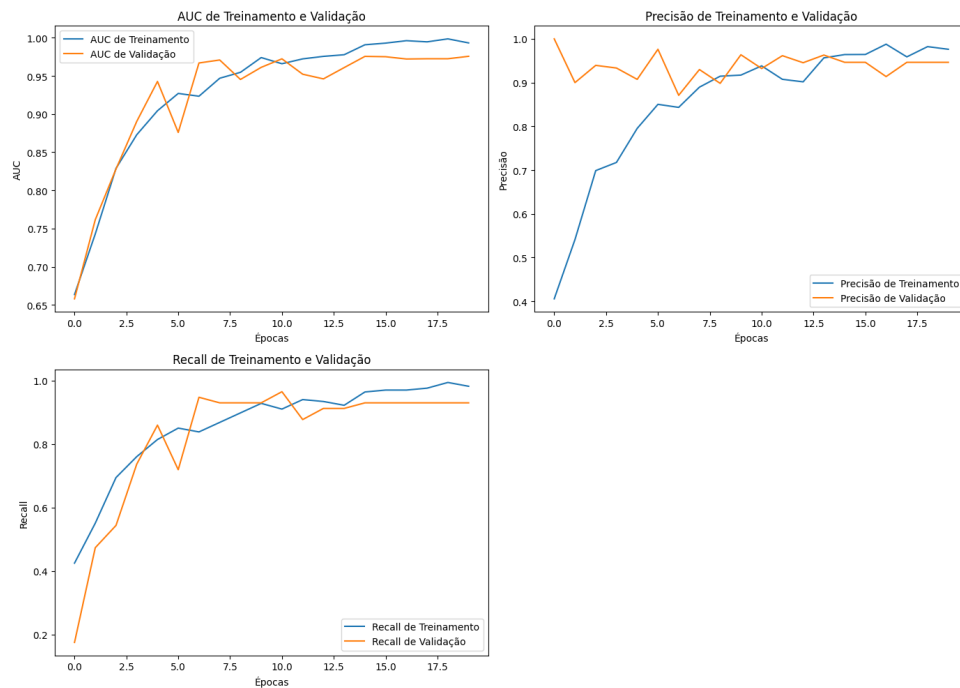


Figura 6.8: AUC para classes 3 e 4

6.2.4 Classes 5 e 6

As classes 5 e 6 representam peles mais escuras e que possuem menos amostras no dataset

Analizadas 1292 imagens das classes 5 e 6 e foram usados 20% (259 imagens)

- Acurácia de Treinamento: 0.99
- Acurácia de Validação: 0.98
- Perda de Treinamento: 0.05
- Perda de Validação: 0.43
- AUC de Treinamento: 0.99
- AUC de Validação: 0.99
- Precisão de Treinamento: 0.99
- Precisão de Validação: 0.92
- Recall de Treinamento: 0.99
- Recall de Validação: 1.0

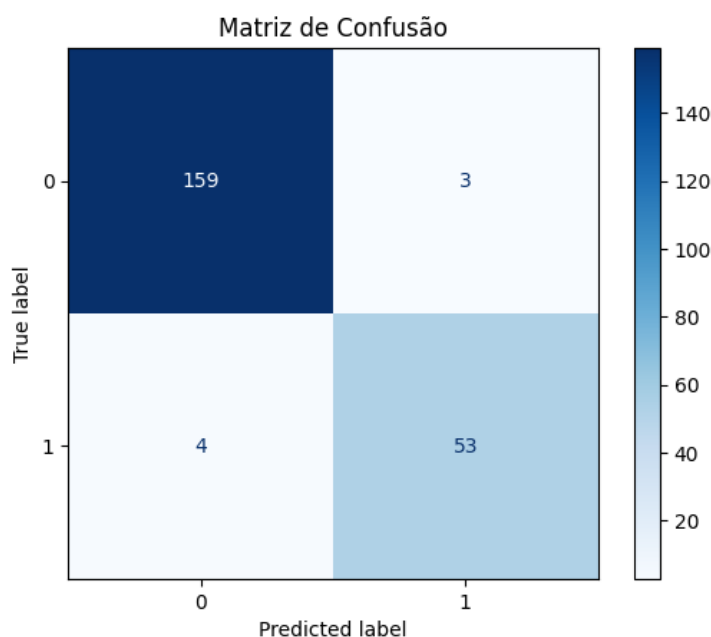


Figura 6.9: Matriz de Confusão para classes 3 e 4

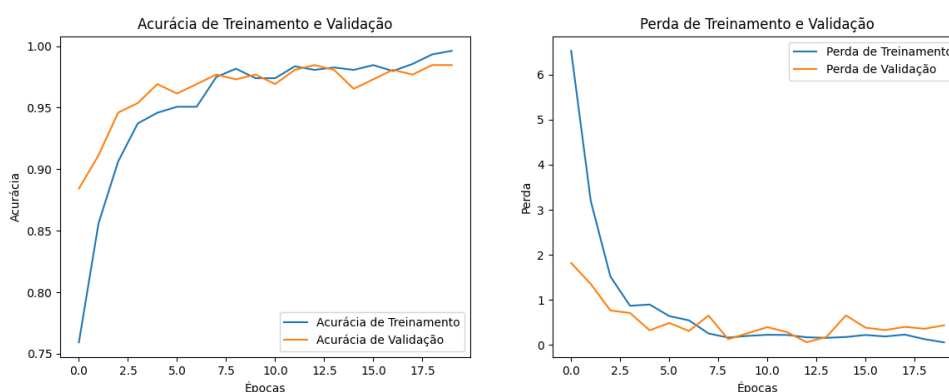


Figura 6.10: Resultados da extração de características CNN + VGG19 para classes 5 e 6

6.2.5 Classe 6

A classe referente ao fototipo VI foi treinada de forma isolada com o objetivo de avaliar a tendência de criação de viés no sentido inverso à análise dos fototipos I e II. Os resultados reforçaram a necessidade de maior variabilidade nos dados e nas imagens de treinamento para se alcançar um modelo menos tendencioso e com maior precisão. Isso evidencia que a diversidade nas amostras utilizadas é essencial para evitar vieses no processo de classificação.

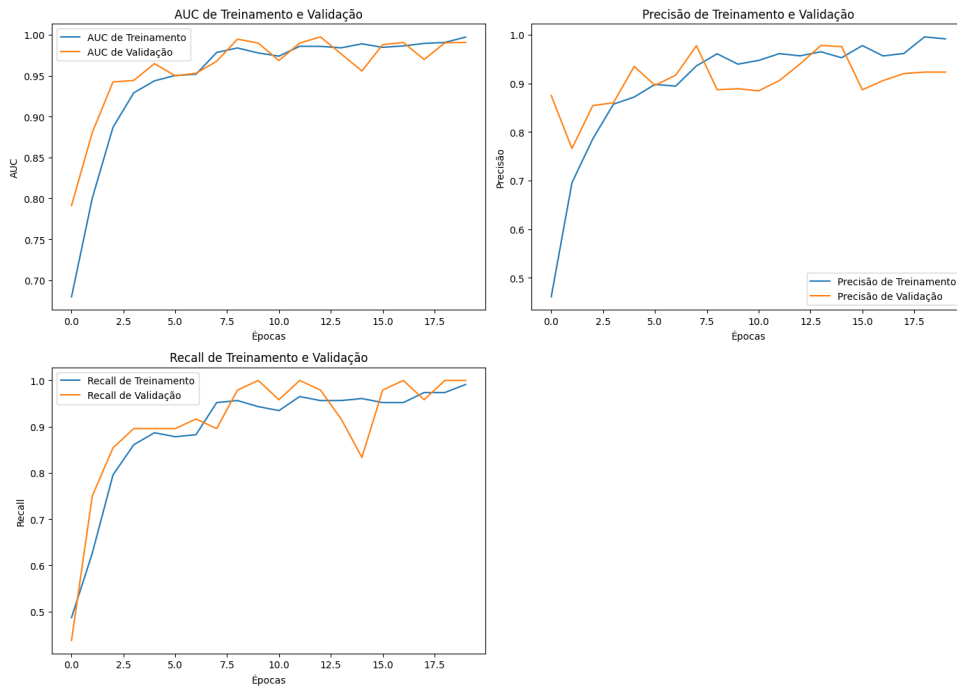


Figura 6.11: AUC para classes 5 e 6

6.2.5.1 Análise da Classe 6 (Pele mais escura)

A avaliação do modelo para a Classe 6, que representa peles mais escuras, mostra um desempenho distinto em comparação com outras classes.

As 960 imagens com filtro para classe 6 avaliadas em 20% totalizando 192 imagens avaliadas.

Durante o treinamento, o modelo alcançou uma acurácia de 99.35% e uma AUC de 0.997, indicando um excelente ajuste aos dados de treinamento. No entanto, no conjunto de validação, a acurácia caiu para 92.71% e a AUC para 0.895, sugerindo um potencial overfitting. A precisão de validação foi de 86.96%, mas o recall foi relativamente baixo, em 83.33%, indicando que o modelo tem dificuldades em identificar corretamente todas as instâncias positivas (melanoma) nesta classe. A matriz de confusão e o relatório de classificação reforçam esses achados, mostrando que, embora a precisão geral seja razoável, a capacidade de detecção de melanomas em peles mais escuras ainda pode ser aprimorada. O F1-Score para ambas as classes é aproximadamente 0.85, refletindo um equilíbrio moderado entre precisão e recall. Esses resultados destacam a necessidade de melhorar a generalização do modelo para peles mais escuras, possivelmente através de técnicas de aumento de dados e ajustes nos hiperparâmetros para abordar melhor as particularidades desta classe.

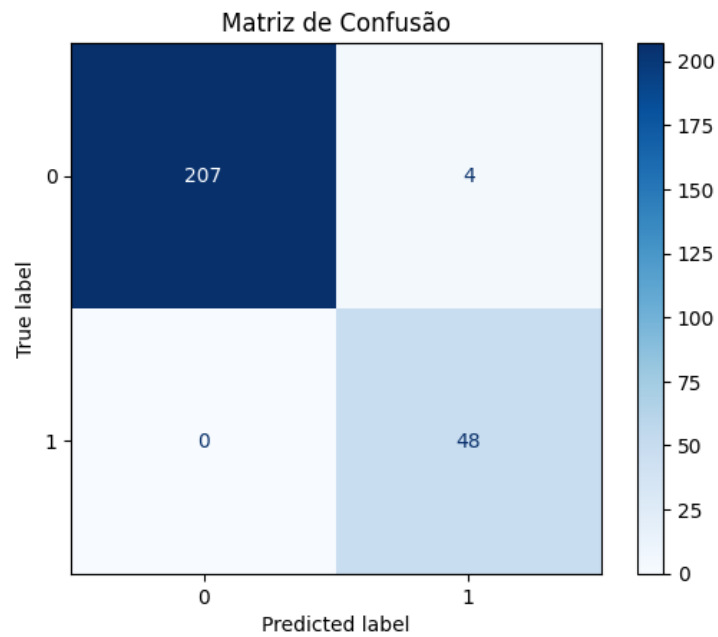


Figura 6.12: Matriz de Confusão para classes 5 e 6

- Acurácia de Treinamento: 0.9935
- Acurácia de Validação: 0.9271
- Perda de Treinamento: 0.0781
- Perda de Validação: 1.4131
- AUC de Treinamento: 0.9974
- AUC de Validação: 0.8954
- Precisão de Treinamento: 0.9735
- Precisão de Validação: 0.8696
- Recall de Treinamento: 1.0
- Recall de Validação: 0.8333

Classe	Precisão	Recall	F1-Score	Suporte
0	0.95	0.96	0.95	144
1	0.87	0.83	0.85	48
Acurácia	0.93			
Média Macro	0.91	0.90	0.90	192
Média Ponderada	0.93	0.93	0.93	192

Tabela 6.4: Relatório de Classificação para Classe 6

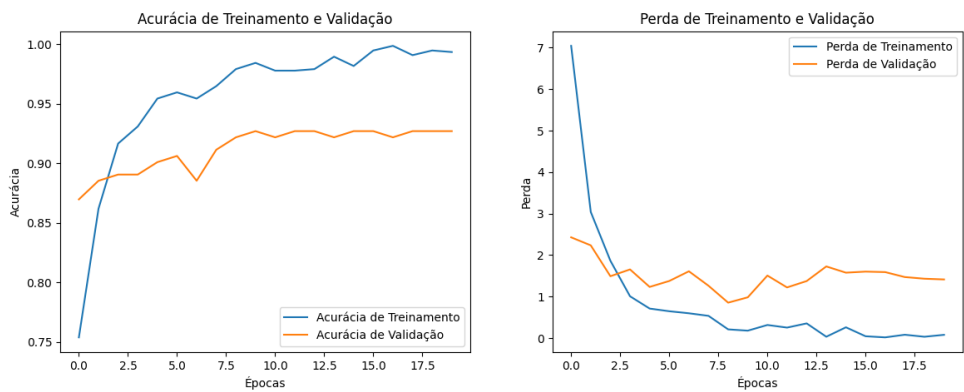


Figura 6.13: Acurácia para Classe 6

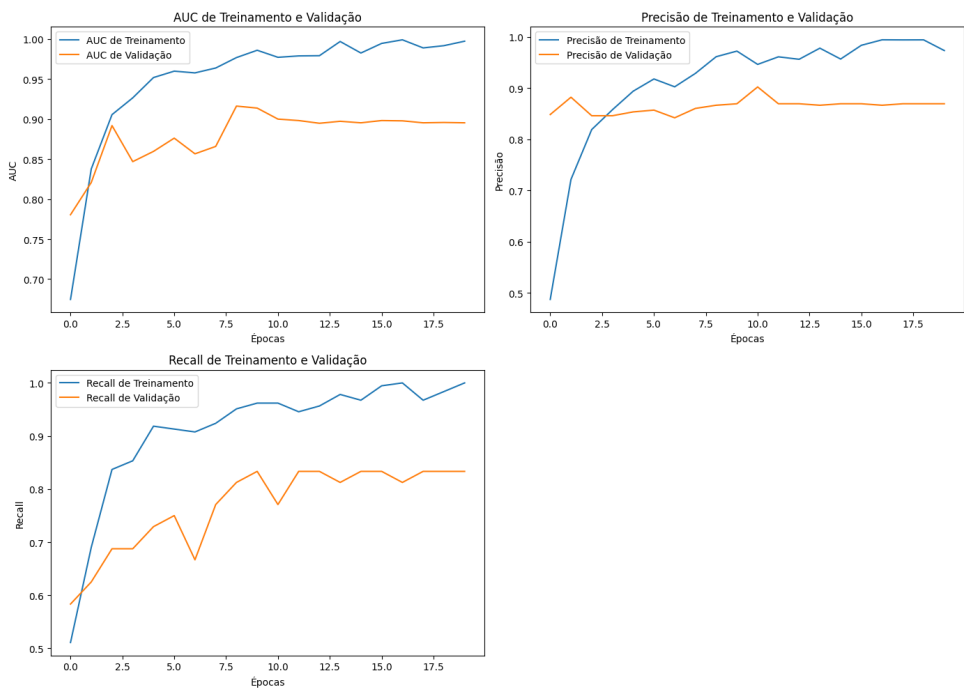


Figura 6.14: AUC para classes 6

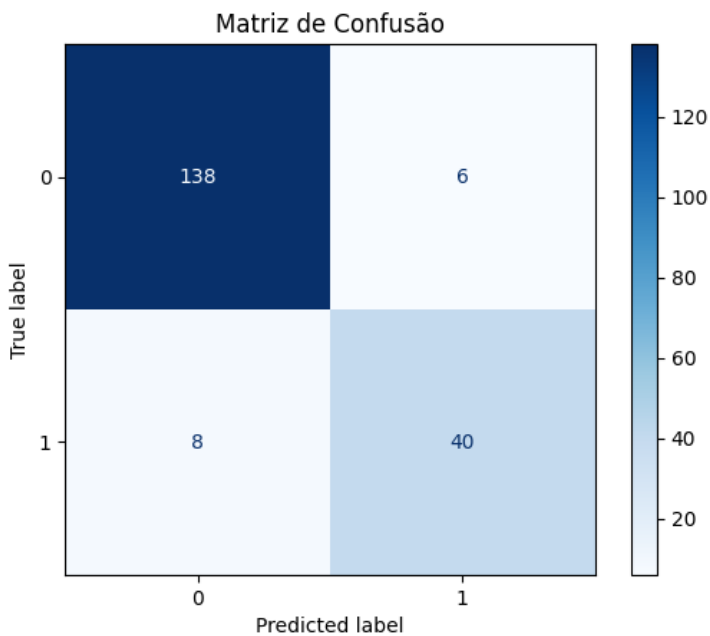


Figura 6.15: Matriz de Confusão para Classe 6

7 Comparativo com Resnet50

Executei os mesmos testes com os mesmos filtros utilizando RESNET50, os resultados seguem na tabela abaixo :

Número de Imagens	Tipos 1-2 364		Tipos 3-4 104		Tipos 5-6 122		Todos (1 a 6) 672	
	ResNet50	VGG19	ResNet50	VGG19	ResNet50	VGG19	ResNet50	VGG19
Acurácia de Treinamento	0.9528	0.9593	0.9783	0.992	0.9884	0.9961	0.9133	0.9555
Acurácia de Validação	0.9254	0.9308	0.9772	0.968	0.9768	0.9846	0.8888	0.9350
Perda de Treinamento	0.0963	0.0879	0.2725	0.0378	0.0699	0.0509	0.159	0.0852
Perda de Validação	0.3648	0.286	0.3043	0.409	0.3071	0.43	0.3757	0.2143
AUC de Treinamento	0.9958	0.9879	0.9891	0.9922	0.998	0.9971	0.969	0.9897
AUC de Validação	0.9439	0.9385	0.9872	0.9756	0.9922	0.9965	0.9235	0.9616
Precisão de Treinamento	0.975	0.9792	0.9458	0.9762	0.9886	0.9913	0.9774	0.9772
Precisão de Validação	0.8796	0.8644	0.9194	0.9464	0.8889	0.9231	0.8859	0.9162
Recall de Treinamento	0.7043	0.7449	0.9401	0.982	0.9609	0.9913	0.513	0.7618
Recall de Validação	0.6834	0.7449	1	0.9089	1	1	0.5546	0.7353

Figura 7.1: Comparativo Resnet50 x VGG19

Ao analisar os resultados apresentados na tabela de comparação entre as arquiteturas VGG19 e ResNet50, é evidente que a VGG19 apresenta uma série de vantagens significativas sobre a ResNet50 em diversos aspectos. A VGG19 demonstra uma superioridade notável em métricas cruciais como a acurácia de validação, onde alcança uma taxa média de 0.9350 para todos os tipos, em comparação com 0.8888 da ResNet50. Este desempenho superior na validação é indicativo de uma capacidade melhor de generalização da VGG19, refletida também em métricas como a precisão de validação, onde a VGG19 atinge uma média de 0.9162, enquanto a ResNet50 apresenta 0.8859. Além disso, a VGG19 mostra-se mais eficaz na redução da perda de validação, com um valor médio de 0.2143, em contraste com 0.3757 da ResNet50, sugerindo uma maior estabilidade no desempenho durante a validação. Esses resultados sugerem que a VGG19, com sua arquitetura mais profunda e sua abordagem simplificada para a construção de redes convolucionais, pode proporcionar um melhor desempenho em tarefas de classificação, especialmente em cenários onde a precisão e a generalização são críticas. A robustez da VGG19 em comparação com a ResNet50 pode ser atribuída à sua capacidade de extrair características mais discriminativas e robustas, resultando em um desempenho global mais consistente e confiável.

8 Discussão

Os resultados obtidos na aplicação do modelo CNN com extração de características VGG19 para a classificação de lesões de pele foram analisados com foco especial na avaliação dos fototipos de Fitzpatrick 1 e 2, 3 e 4, e 5 e 6.

8.1 Desempenho do Modelo

Os resultados do modelo durante o treinamento foram promissores, alcançando alta precisão, recall e AUC no conjunto de treinamento para todas as classes. No entanto, ao avaliar o conjunto de validação, observamos uma queda significativa no desempenho, com uma acurácia de aproximadamente 75-76%, AUC em torno de 0.78-0.85, precisão variando entre 70-80% e recall entre 66-76%. Essa disparidade entre os conjuntos de treinamento e validação sugere a ocorrência de *overfitting*, onde o modelo se ajusta demasiadamente aos dados de treinamento e não generaliza bem para novos dados.

8.2 Impacto do Desequilíbrio de Classes

O desequilíbrio de classes, particularmente notável em fototipos mais escuros (V e VI), apresentou um desafio significativo. Apenas 10% das amostras no conjunto de dados pertenciam a esses fototipos, o que pode levar a um viés no treinamento do modelo e afetar sua capacidade de detectar com precisão lesões de pele em pacientes de tons de pele mais escuros. Estratégias como aumento de dados e ajuste de pesos de classe podem ser necessárias para lidar eficazmente com esse problema.

8.3 Métricas de Desempenho

A análise das métricas de desempenho, como precisão, recall e F1-Score, revelou informações importantes sobre a capacidade do modelo de distinguir entre casos positivos

(melanoma) e negativos (não melanoma). A precisão variou entre 70-80%, indicando a proporção de casos positivos corretamente identificados em relação ao total de identificações positivas feitas pelo modelo, enquanto o recall variou entre 66-76%, destacando a capacidade do modelo de identificar corretamente todos os casos positivos presentes no conjunto de dados.

8.4 Limitações e Considerações Futuras

É crucial reconhecer as limitações deste estudo. A falta de representação adequada de fototipos mais escuros no conjunto de dados pode ter impactado negativamente a capacidade do modelo de generalizar para esses grupos populacionais. Para futuras pesquisas, recomenda-se explorar técnicas avançadas de aumento de dados que sejam específicas para variações de tons de pele, além de considerar o uso de conjuntos de dados mais diversificados e representativos.

Desafios e conjuntos de dados públicos assumem um papel crucial no desenvolvimento de algoritmos confiáveis e eficazes para uso na prática clínica [35]. Tais recursos podem aprimorar a precisão e a generalização dos algoritmos, além de promover a transparência e a colaboração entre pesquisadores. A utilização de conjuntos de dados públicos de alta qualidade garante que os algoritmos sejam treinados em cenários realistas e diversos, o que aumenta sua confiabilidade na prática clínica.

Para melhorar o desempenho do modelo e reduzir o *overfitting*, várias estratégias podem ser consideradas:

- **Realização de Mais Épocas:** Aumentar o número de épocas de treinamento pode permitir que o modelo aprenda melhor as características dos dados, embora seja necessário monitorar para evitar *overfitting*. Técnicas como *EarlyStopping* podem ser úteis para determinar o ponto ideal de parada.
- **Uso de Mais Imagens:** A coleta e o uso de um conjunto de dados maior e mais diversificado podem ajudar a melhorar a capacidade do modelo de generalizar para novos dados. A adição de imagens de fototipos de Fitzpatrick mais escuros pode equilibrar melhor o conjunto de dados e melhorar a precisão para esses grupos.
- **Adicionar Mais Camadas:** A inclusão de mais camadas densas e convolucionais pode ajudar a capturar padrões mais complexos nos dados. No entanto, isso deve

ser feito com cuidado para evitar o aumento excessivo da complexidade do modelo, o que pode levar ao *overfitting*.

- **Técnicas de Regularização:** A aplicação de técnicas de regularização mais robustas, como *dropout* e regularização L2, pode ajudar a mitigar o *overfitting* observado nos resultados do modelo. Essas técnicas adicionam ruído durante o treinamento, o que força o modelo a generalizar melhor.
- **Arquiteturas de Rede Mais Complexas:** Explorar arquiteturas de rede mais avançadas, como *EfficientNet*, *ResNet* ou *Inception*, que têm mostrado excelente desempenho em diversas tarefas de visão computacional. Essas arquiteturas podem oferecer melhor capacidade de extração de características e generalização.
- **Aumento de Dados (Data Augmentation):** Implementar técnicas de aumento de dados pode ajudar a aumentar a variabilidade do conjunto de dados de treinamento, melhorando a robustez do modelo. Técnicas como rotação, translação, zoom, *flip* horizontal e ajuste de brilho podem ser aplicadas.
- **Ajuste de Hiperparâmetros:** Realizar ajustes finos nos hiperparâmetros do modelo, como a taxa de aprendizado, número de neurônios nas camadas densas e parâmetros de regularização, pode ajudar a otimizar o desempenho do modelo.
- **Balanceamento de Classes:** Usar técnicas de balanceamento de classes, como ajuste de pesos de classe ou aumento de dados específico para classes minoritárias, pode melhorar a capacidade do modelo de aprender de forma equitativa para todas as classes.

Essas abordagens oferecem um caminho claro para futuras melhorias no modelo, aumentando sua generalização e desempenho em dados não vistos. Explorar essas direções pode resultar em algoritmos mais robustos e eficazes para a prática clínica.

8.5 Ética e Viés nos Sistemas de IA

A ética e o viés nos sistemas de inteligência artificial (IA) são tópicos críticos na aplicação de CNNs para diagnósticos médicos. Estudos de [32] mostram que algoritmos de IA podem perpetuar ou até amplificar vieses existentes nos dados de treinamento, levando a disparidades nos cuidados de saúde. Esses vieses podem resultar em diagnósticos menos precisos para populações sub-representadas, exacerbando as desigualdades existentes na saúde.

8.5.1 Análise de Viés

A análise de viés é essencial para garantir que os sistemas de IA sejam justos e equitativos. [7] destacam que a falta de diversidade nos dados de treinamento pode levar a algoritmos que funcionam bem para alguns grupos, mas não para outros. Em nosso estudo, a inclusão da Escala de Fitzpatrick e a análise detalhada de diferentes fototipos de pele são passos importantes para identificar e mitigar esses vieses.

8.5.2 Mitigação do Viés

Para mitigar o viés, é importante adotar uma abordagem holística. Isso inclui a coleta de dados diversificados, o uso de técnicas de aumento de dados para simular uma maior variedade de casos, e a validação contínua dos modelos em diferentes subgrupos populacionais. Estudos de [33] sugerem que a transparência nos processos de desenvolvimento e a colaboração com comunidades afetadas são cruciais para criar sistemas de IA mais justos.

8.5.3 Tendências Futuras na Pesquisa em Visão Computacional para Diagnósticos Médicos

A pesquisa em visão computacional para diagnósticos médicos está em constante evolução, com várias tendências emergentes que prometem transformar a prática clínica. Entre essas tendências, destaca-se o uso de arquiteturas de rede neural mais avançadas, como os Transformers em visão computacional, que têm mostrado grande potencial em tarefas de segmentação e classificação de imagens. Estudos de Dosovitskiy et al. (2020) demonstram que essas arquiteturas podem superar as CNNs tradicionais em várias métricas de desempenho [11].

8.5.3.1 Medicina Personalizada

Outra tendência importante é a integração de modelos de visão computacional com dados genômicos e clínicos para promover a medicina personalizada. Estudos de Miotto et al. (2018) sugerem que a combinação de dados de imagem com outras fontes de informação pode levar a diagnósticos mais precisos e tratamentos personalizados, melhorando significativamente os resultados dos pacientes [29].

8.5.4 Realidade Aumentada e Virtual

A aplicação de realidade aumentada (AR) e virtual (VR) em diagnósticos médicos é outra área promissora. Estudos de Maresky et al. (2019) mostram que essas tecnologias podem ser usadas para visualização avançada de dados médicos, permitindo que os médicos vejam imagens de diagnóstico em 3D e em tempo real, o que pode melhorar a precisão e a eficiência do diagnóstico [28].

8.5.5 Síntese e Impacto Potencial na Prática Clínica e Pesquisa Futura

Este estudo demonstrou a eficácia do uso da VGG19 na classificação de lesões cutâneas, destacando a importância da inclusão de variabilidade étnica na avaliação do desempenho do modelo. A integração da Escala de Fitzpatrick permitiu uma análise mais detalhada de como diferentes tons de pele podem influenciar a precisão diagnóstica, oferecendo uma contribuição significativa para a pesquisa em visão computacional e diagnósticos médicos.

8.5.5.1 Relevância Clínica

A relevância clínica deste estudo é evidente na potencial aplicação dos resultados para melhorar o diagnóstico precoce de melanoma, especialmente em populações sub-representadas. A implementação de modelos mais inclusivos pode levar a um aumento na precisão diagnóstica e na equidade nos cuidados de saúde, beneficiando pacientes de diversas origens étnicas.

8.5.6 Direções Futuras

Para pesquisas futuras, é crucial continuar explorando e desenvolvendo técnicas que considerem a diversidade étnica e a variabilidade dos dados. A combinação de CNNs com outras fontes de dados, como genômicos e clínicos, promete avanços significativos na medicina personalizada. Além disso, a aplicação de tecnologias emergentes, como realidade aumentada e virtual, pode abrir novas fronteiras na visualização e interpretação de dados médicos, melhorando ainda mais a prática clínica.

9 Conclusão

Este estudo propõem o uso da VGG19 para classificar lesões de pele, incluindo o melanoma, onde podemos observar que é uma recurso pouco estudado como mostrado na Tabela 2.1. A capacidade do VGG19 de extrair características detalhadas de imagens médicas pode melhorar significativamente a precisão diagnóstica, beneficiando tanto os pacientes quanto os profissionais de saúde.

Outro estudo revela uma forte correlação entre a incidência fatal de melanoma em climas menos ensolarados e a pigmentação da pele, especialmente entre populações com histórico mínimo de migração na Europa. Por outro lado, nos Estados Unidos, há uma associação entre a incidência de melanoma e latitudes mais baixas com índices UV mais altos, afetando tanto indivíduos brancos não hispânicos quanto populações hispânicas e negras.

Embora o câncer de pele seja um dos tipos mais comuns diagnosticados, sua posição não figura entre as principais causas de mortalidade segundo o SEER [1]. Isso ressalta a importância de entender não apenas a incidência da doença, mas também os fatores que influenciam sua prevalência e os desafios enfrentados na prevenção e tratamento.

Esses fatores de risco sugerem mecanismos comuns no desenvolvimento do melanoma entre diferentes grupos étnicos [15]. No entanto, é crucial observar que, de acordo com a Escala de Fitzpatrick, o conjunto de dados ISIC 2020 foi avaliado e constatou-se que imagens dos tons de pele V e VI (marrom e escuro) representam apenas 10% das amostras. Essa sub-representação pode resultar em deficiências na classificação de lesões de pele em indivíduos com tons de pele mais escuros. Para enfrentar esse desafio e melhorar a precisão, técnicas de aprendizado de máquina com categorização de cor da pele podem melhorar significativamente os resultados. Estudos futuros poderiam avaliar resultados separadamente por categoria de cor da pele, explorando como diferentes modelos e técnicas de aprendizado de máquina podem ser adaptados para lidar com a diversidade étnica e melhorar a precisão na classificação de lesões dermatológicas.

A integração de modelos como o VGG19 em sistemas de apoio à decisão clínica representa um avanço promissor para a prática dermatológica, permitindo a detecção precoce e precisa de lesões de pele, potencialmente resultando em melhores resultados clínicos e prognósticos para os pacientes. Além disso, um estudo realizado por Haenssle et al. [17] comparou o desempenho diagnóstico de uma rede neural convolucional (CNN), concluindo que a tecnologia pode fornecer uma abordagem robusta e eficaz para análise de imagens dermatoscópicas, destacando seu potencial para melhorar significativamente a precisão diagnóstica e eficiência.

No entanto, é crucial observar que, de acordo com a Escala de Fitzpatrick, a pele escura não está adequadamente equilibrada no conjunto de dados, potencialmente causando subajuste na classificação. Esse desequilíbrio pode afetar adversamente a eficácia do modelo em identificar corretamente lesões de pele em pacientes de pele mais escura. Para enfrentar esse desafio, estudos futuros devem explorar técnicas robustas de aumento de dados e adaptação do modelo para capturar a diversidade étnica dos pacientes.

A integração de modelos como o VGG19 em sistemas de apoio à decisão clínica representa um avanço promissor para a prática dermatológica, permitindo a detecção precoce e precisa de lesões de pele, o que pode resultar em melhores resultados clínicos e prognósticos para os pacientes.

REFERENCIAS

- [1] National Cancer Institute (NCI). *SEER StatFacts: Common Cancer Facts*. Accessed on July 11, 2024. 2024. URL: <https://seer.cancer.gov/statfacts/>.
- [2] Association of American Medical Colleges. *Why Are So Many Black Patients Dying from Skin Cancer?* <https://www.aamc.org/news/why-are-so-many-black-patients-dying-skin-cancer>. Accessed: 2024-07-11.
- [3] Luana Barros, Levy Chaves e Sandra Avila. “Assessing the Generalizability of Deep Neural Networks-Based Models for Black Skin Lesions”. Em: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Ed. por Verónica Vasconcelos, Inês Domingues e Simão Paredes. Cham: Springer Nature Switzerland, 2024, pp. 1–14. ISBN: 978-3-031-49249-5.
- [4] Sandra Bino et al. “Relationship between skin response to ultraviolet exposure and skin color type”. Em: *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* 19 (jan. de 2007), pp. 606–14. DOI: [10.1111/j.1600-0749.2006.00338.x](https://doi.org/10.1111/j.1600-0749.2006.00338.x).
- [5] Freddie Bray et al. “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. Em: *CA: A Cancer Journal for Clinicians* 68.6 (2018), pp. 394–424. DOI: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492).
- [6] T. J. Brinker et al. “Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task”. Em: *European Journal of Cancer* 113 (2019), pp. 47–54. DOI: [10.1016/j.ejca.2019.05.030](https://doi.org/10.1016/j.ejca.2019.05.030).
- [7] Joy Buolamwini e Timnit Gebru. “Gender shades: Intersectional accuracy disparities in commercial gender classification”. Em: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. PMLR. 2018, pp. 77–91.
- [8] R. Cabrera e F. Recule. “Unusual Clinical Presentations of Malignant Melanoma: A Review of Clinical and Histologic Features with Special Emphasis on Dermatoscopic Findings”. Em: *American Journal of Clinical Dermatology* 19 (nov. de 2018), pp. 15–23. DOI: [10.1007/s40257-018-0373-6](https://doi.org/10.1007/s40257-018-0373-6).

- [9] MaryBeth B Culp e Natasha Buchanan Lunsford. “Melanoma Among Non-Hispanic Black Americans”. Em: *Prev Chronic Dis* 16 (2019). PEER REVIEWED, p. 180640. DOI: [10.5888/pcd16.180640](https://doi.org/10.5888/pcd16.180640).
- [10] Edna Lucia Da Silva e Estera Muszkat Menezes. “Metodologia da pesquisa e elaboração de dissertação”. Em: *UFSC, Florianópolis, 4a. edição* 123 (2005).
- [11] Alexey Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. Em: *arXiv preprint arXiv:2010.11929* (2020).
- [12] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. Em: *Nature* 542.7639 (2017), pp. 115–118. ISSN: 1476-4687. DOI: [10.1038/nature21056](https://doi.org/10.1038/nature21056). URL: <https://doi.org/10.1038/nature21056>.
- [13] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. Em: *Nature* 542 (2017), pp. 115–118. DOI: [10.1038/nature21056](https://doi.org/10.1038/nature21056). URL: <https://doi.org/10.1038/nature21056>.
- [14] T.B. Fitzpatrick. “Ultraviolet-induced pigmentary changes: benefits and hazards”. Em: *Curr Probl Dermatol* 15 (1986), pp. 25–38. DOI: [10.1159/000412090](https://doi.org/10.1159/000412090).
- [15] Philip K. Goon e Susan M. Swetter. “Skin cancer epidemiology in populations of color”. Em: *Dermatologic Clinics* 39.1 (2021), pp. 1–10.
- [16] H. A. Haenssle et al. “Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists”. Em: *Annals of Oncology* 29.8 (2018), pp. 1836–1842. ISSN: 0923-7534. DOI: [10.1093/annonc/mdy166](https://doi.org/10.1093/annonc/mdy166). URL: <https://doi.org/10.1093/annonc/mdy166>.
- [17] HA Haenssle et al. “Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists”. Em: *Ann Oncol* 29.8 (ago. de 2018), pp. 1836–1842. DOI: [10.1093/annonc/mdy166](https://doi.org/10.1093/annonc/mdy166).
- [18] Trevor Hastie, Robert Tibshirani e Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2009.
- [19] Shelley Hu, Raynald Soza-Vento e David F. Parker. “Analysis of skin cancer risk factors in black patients: a retrospective study”. Em: *Dermatology Research and Practice* 2014 (2014), p. 593020. DOI: [10.1155/2014/593020](https://doi.org/10.1155/2014/593020). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4141864/>.

- [20] M. Ichihashi et al. “UV-induced skin damage”. Em: *Toxicology* 189 (2003), pp. 21–39. DOI: [10.1016/S0300-483X\(03\)00150-1](https://doi.org/10.1016/S0300-483X(03)00150-1). URL: [https://doi.org/10.1016/S0300-483X\(03\)00150-1](https://doi.org/10.1016/S0300-483X(03)00150-1).
- [21] Irfan Ali Kandhro et al. “Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification”. Em: *Heliyon* 10.10 (2024), e31488. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2024.e31488>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844024075194>.
- [22] C. Karimkhani et al. “The global burden of melanoma: Results from the Global Burden of Disease Study 2015”. Em: *British Journal of Dermatology* 177 (2017), pp. 134–140. DOI: [10.1111/bjd.15510](https://doi.org/10.1111/bjd.15510). URL: <https://doi.org/10.1111/bjd.15510>.
- [23] Staffs Keele et al. *Guidelines for performing systematic literature reviews in software engineering*. Rel. técn. Citeseer, 2007.
- [24] Newton M. Kinyanjui et al. “Estimating Skin Tone and Effects on Classification Performance in Dermatology Datasets”. Em: *arXiv preprint arXiv:1910.13268* (2019). NeurIPS 2019 Workshop on Fair ML for Health. DOI: [10.48550/arXiv.1910.13268](https://doi.org/10.48550/arXiv.1910.13268). arXiv: [1910.13268 \[cs.CV\]](https://arxiv.org/abs/1910.13268). URL: <https://arxiv.org/abs/1910.13268>.
- [25] Barbara Kitchenham et al. “Systematic literature reviews in software engineering—a systematic literature review”. Em: *Information and software technology* 51.1 (2009), pp. 7–15.
- [26] Renata Garcez da Luz et al. “Analysis of the variation in the number of cases of malignant melanoma and other skin malignancies in all regions of Brazil from 2015 to 2022”. Em: *Research, Society and Development* 12.6 (jun. de 2023), e17712641967. DOI: [10.33448/rsd-v12i6.41967](https://doi.org/10.33448/rsd-v12i6.41967). URL: <https://rsdjournal.org/index.php/rsd/article/view/41967>.
- [27] CM Magro, AN Crowson e MC Mihm. “Unusual variants of malignant melanoma”. Em: *Modern Pathology* 19 (2006), S41–S70.
- [28] Harold S. Maresky et al. “Virtual reality and cardiac anatomy: exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education”. Em: *Clinical Anatomy* 32.2 (2019), pp. 238–243.
- [29] Riccardo Miotto et al. “Deep learning for healthcare: review, opportunities and challenges”. Em: *Briefings in Bioinformatics* 19.6 (2018), pp. 1236–1246.

- [30] AN Moshell, PD Parikh e WJ Oetgen. “Characteristics of medical professional liability claims against dermatologists: data from 2704 closed claims in a voluntary registry”. Em: *Journal of the American Academy of Dermatology* 66.1 (2012), pp. 78–85.
- [31] Novology. *A Guide to Finding the Right Sunscreen for Your Skin Tone*. Accessed: 2024-07-25. 2024. URL: <https://www.novology.com/blogs/pigmentation/a-guide-to-finding-the-right-sunscreen-for-your-skin-tone>.
- [32] Ziad Obermeyer et al. “Dissecting racial bias in an algorithm used to manage the health of populations”. Em: *Science* 366.6464 (2019), pp. 447–453.
- [33] Alvin Rajkomar et al. “Ensuring fairness in machine learning to advance health equity”. Em: *Annals of Internal Medicine* 169.12 (2018), pp. 866–872.
- [34] HM Rayess et al. “A critical analysis of melanoma malpractice litigation: should we biopsy everything?” Em: *Laryngoscope* 127.1 (2017), pp. 134–139.
- [35] Veronica Rotemberg et al. “The Role of Public Challenges and Data Sets Towards Algorithm Development, Trust, and Use in Clinical Practice”. Em: *Seminars in Cutaneous Medicine and Surgery* 38.1 (mar. de 2019), E38–E42.
- [36] Karen Simonyan e Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. Em: *arXiv preprint arXiv:1409.1556* (2014). Submitted on 4 Sep 2014 (v1), last revised 10 Apr 2015 (this version, v6). URL: <https://doi.org/10.48550/arXiv.1409.1556>.
- [37] Strahil Strashilov e Angel Yordanov. “Aetiology and Pathogenesis of Cutaneous Melanoma: Current Concepts and Advances”. Em: *International Journal of Molecular Sciences* 22.6395 (2021). DOI: [10.3390/ijms22126395](https://doi.org/10.3390/ijms22126395). URL: <https://doi.org/10.3390/ijms22126395>.
- [38] DB Troxel. “Medicolegal aspects of error in pathology”. Em: *Archives of Pathology Laboratory Medicine* 130.5 (2006), pp. 617–619.
- [39] P. Tschandl et al. “Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study”. Em: *The Lancet Oncology* 21.1 (2020), pp. 138–149. DOI: [10.1016/S1470-2045\(19\)30761-6](https://doi.org/10.1016/S1470-2045(19)30761-6).
- [40] Alan Turing. “The impact of interdisciplinary collaboration on the development of artificial intelligence”. Em: *AI & Society* 33.4 (2018), pp. 465–479.

-
- [41] John O Woolliscroft e Joel D Howell. “Interdisciplinary collaboration in health care: Implications for teaching and learning”. Em: *Journal of Interprofessional Care* 33.5 (2019), pp. 479–482.
- [42] Andressa Xavier. *Aprenda a usar as Normas da ABNT: Citação (2 de 4)*. Disponível em <https://www.tecmundo.com.br/tutorial/834-aprenda-a-usar-as-normas-da-abnt-citacao-2-de-4-.htm>, acessado em 05/10/2021. 2020.