



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Comparação entre técnicas de extração de características para detecção de melanoma em imagens dermatoscópicas

Lorrane Lorena Oliveira Ribeiro

Monografia apresentada como requisito parcial
para conclusão do Curso de Computação — Licenciatura

Orientadora
Prof.a Dr.a Roberta Barbosa Oliveira

Brasília
2023



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Comparação entre técnicas de extração de características para detecção de melanoma em imagens dermatoscópicas

Lorrane Lorena Oliveira Ribeiro

Monografia apresentada como requisito parcial
para conclusão do Curso de Computação — Licenciatura

Prof.a Dr.a Roberta Barbosa Oliveira (Orientadora)
CIC/UnB

Prof. Dr. Jan Mendonça Correa Prof. Dr. Marcus Vinicius Lamar
CIC/UNB CIC/UNB

Prof. Dr. Jorge Henrique Cabral Fernandes
Coordenador do Curso de Computação — Licenciatura

Brasília, 10 de Fevereiro de 2023

Dedicatória

Dedico este trabalho a minha família e aos meus amigos, que muito me ajudaram em todos os momentos, principalmente me motivando a sempre continuar, dedico também a todos os professores que me ajudaram a trilhar o caminho até aqui, cada um teve um pedaço na minha jornada.

Agradecimentos

Eu agradeço, a minha família pelo apoio prestado nesses últimos anos que estive na UnB, agradeço principalmente aos meus avós que me ajudaram na jornada e a minha irmã Larissa que me ajudou a não desistir de escrever minha monografia. também agradeço aos meus amigos que sempre estiveram do meu lado. Agradeço Profa Dra Roberta Barbosa Oliveira, que foi uma excelente orientadora, que me ajudou e que muito me ensinou e estava sempre me motivando.

Resumo

O diagnóstico do melanoma é uma tarefa importante para identificar a presença da doença em seu estágio inicial, pois tem mais chances de cura quando detectada precocemente. O diagnóstico clínico é geralmente realizado por um dermatologista, que pode solicitar uma biópsia para confirmar a presença da doença. No entanto, a biópsia é um procedimento invasivo. Os avanços tecnológicos têm trazido novos recursos para ajudar no diagnóstico do melanoma. O uso de modelos baseados em aprendizado de máquina tem sido um desses recursos, como por exemplo os modelos tradicionais que usam diversas características extraídas a partir da área da lesão, tais como cor e textura. Além disso, também são utilizados modelos híbridos que combinam aprendizado profundo e aprendizado de máquina tradicional para extrair características de maneira automática e classificar as lesões, respectivamente. Esses modelos têm ajudado a melhorar a eficiência do diagnóstico automático do melanoma, tornando-o mais fácil e rápido como forma de auxiliar os médicos. O presente trabalho tem como objetivo a comparação entre técnicas de extração de características com métodos tradicionais e a extração de características de modo automático, com imagens dermatoscópicas para detecção de melanoma. São considerados o classificador Floresta Aleatória e a arquitetura de Rede Neural Convolutiva VGG16, bem como um modelo híbrido desenvolvido a partir da combinação de ambas as técnicas. Nessa mesma linha é proposto o uso de um pré-processamento para o aumento da classe de melanomas para resolver o problema de desbalanceamento. Para os experimentos, a base do ISIC 2017 contendo 2750 imagens é utilizada. O modelo híbrido teve maior assertividade na detecção de melanomas, no qual obteve uma acurácia de 0.685, precisão de 0.789, revocação de 0.685 e *F1-Score* de 0.725 considerando os dados não balanceados.

Palavras-chave: redes neurais convolucionais, floresta aleatória, modelo híbrido, melanoma, imagem dermatoscópica

Abstract

The diagnosis of melanoma is an important task for identifying the presence of the disease in its early stage, as it has more chances of cure when detected early. Clinical diagnosis is usually performed by a dermatologist, who may request a biopsy to confirm the presence of the disease. However, biopsy is an invasive procedure. Technological advances have brought new resources to help diagnose melanoma. The use of machine learning-based models has been one of these resources, such as traditional models that use various features extracted from the lesion area, such as color and texture. In addition, hybrid models that combine deep learning and traditional machine learning are also used to automatically extract features and classify lesions, respectively. These models have helped improve the efficiency of automatic melanoma diagnosis, making it easier and faster to assist doctors. Here, the main objective is to compare traditional techniques for feature extraction with automatic feature extraction based on deep learning from dermoscopic images for melanoma detection. The Random Forest classifier and the VGG16 Convolutional Neural Network architecture are considered, as well as a hybrid model developed from the combination of both techniques. Furthermore, the use of pre-processing for increasing the melanoma class is proposed to solve the imbalance problem. For the experiments, the ISIC 2017 database containing 2750 images is used. The hybrid model had greater results in the detection of melanomas, with an accuracy of 0.685, precision of 0.789, Recall of 0.685, and F1-Score of 0.725, considering the unbalanced data.

Keywords: convolutional neural networks, random forest, hybrid model, melanoma, dermoscopic image

Sumário

1	Introdução	1
1.1	Objetivos gerais	2
1.2	Objetivos específicos	2
1.3	Estruturação do trabalho	2
2	Fundamentação Teórica	4
2.1	Visão Computacional	4
2.1.1	Categoria de Dados	5
2.1.2	Espaço de cor	5
2.1.3	Textura	8
2.2	Aprendizado de Máquina	8
2.2.1	Tipos de Aprendizado	9
2.2.2	Aspectos Básicos dos Algoritmos de Classificação	9
2.3	Floresta Aleatória	10
2.4	Redes Neurais Artificiais	10
2.4.1	Funcionamento das Redes Neurais Artificiais	11
2.4.2	Arquitetura	12
2.5	Redes Neurais Convolucionais	12
2.5.1	Camada de Convolução	13
2.5.2	Camada de Subamostragem ou <i>Pooling</i>	13
2.5.3	Camada Completamente Conectada	14
2.5.4	Arquiteturas VGG	14
2.5.5	<i>Transferência de Conhecimento</i>	15
2.6	Balanceamento de Classes	16
2.7	Métricas de Desempenho	17
3	Revisão de Literatura	20
4	Metodologia	24
4.1	Base de Dados	26

4.2	Pré-processamento das Imagens	27
4.3	Extração de características	28
4.4	Aplicação de <i>Transfer Learning</i> e Análise de Hiperparâmetros	29
4.5	Experimentos com extração de características manual	30
4.6	Experimentos com extração de características automática	33
5	Resultados	36
5.1	Ambiente Experimental	36
5.2	Experimento I: Floresta Aleatória	36
5.3	Experimento II: Floresta Aleatória + Aumento de Dados	37
5.4	Experimento III: Floresta Aleatória + Aumento de Dados + otimização dos hiperparâmetros	38
5.5	Experimento IV: VGG16 Sem Aumento dos Dados	39
5.6	Experimento V: VGG16 com Aumento dos Dados	39
5.7	Experimento VI: VGG16 + Floresta Aleatória	41
5.8	Experimento VII: VGG16 + Floresta Aleatória + Aumento dos dados . . .	42
6	Discussão dos Resultados	44
7	Conclusão e Trabalhos Futuros	46
	Referências	47

Lista de Figuras

2.1	Camadas do espaço de cor RGB - <i>Red, Green e Blue</i>	6
2.2	Camadas do espaço de cor HSV - Matiz, Saturação e Valor	7
2.3	Diagrama cromático de cor CIELUV	7
2.4	Espaço de cor CIE L^*a^*b	8
2.5	Floresta Aleatória	10
2.6	Camadas de uma Rede Neural Artificial	11
2.7	Camadas de uma Rede Neural Convolutiva	13
2.8	Extração utilizando CNN e Classificação através de Floresta Aleatória	15
4.1	Método Usando Somente Floresta Aleatória	25
4.2	Classificação Com Modelo Híbrido CNN + Floresta Aleatória	25
4.3	Exemplo de imagem da base de dados: (b) <i>Ground Truth</i> da segmentação (imagem binarizada) de (a) uma lesão não melanoma. Fonte: ISIC 2017 [1].	27
4.4	Exemplos de Imagens criadas após o pré-processamento.	28
4.5	Canais de cores aplicados nas figuras de lesões de pele	29
4.6	Método Sem Aumento de Dados	31
4.7	Etapas da Floresta Aleatória com Aumento de Dados	32
4.8	Métodos com Aumento de Dados mais Hiperparâmetros	32
4.9	VGG16 Sem Aumento de Dados	33
4.10	VGG com aumento de dados	34
4.11	VGG16 + Floresta Aleatória	34
4.12	VGG com Floresta Aleatória com a base balanceada	35
5.1	Matrizes de Confusão do Experimento III	38
5.2	Resultados do modelo VGG16 sem aumento de dados	40
5.3	Resultados do modelo VGG16 com aumento de dados	40
5.4	Matriz de Confusão do Experimento VI	41
5.5	Matriz de Confusão do Experimento VII	43

Lista de Tabelas

3.1	Trabalhos Relacionados	23
4.1	Tabela a a quantidade de imagens utilizadas para treinamento	26
4.2	Tabela com os descritores e características utilizadas nos experimentos . .	30
5.1	Resultados da Floresta Aleatória sem aumento dos dados	37
5.2	Resultados da Floresta Aleatória com aumento dos dados	37
5.3	Tabela com Resultados da Floresta Aleatória com aumento dos dados. . .	38
5.4	Tabela com Resultados da CNN	41
5.5	Tabela com Resultados da CNN	42
6.1	Tabela com Resultados da CNN	45
6.2	Tabela com Resultados da CNN	45

Lista de Abreviaturas e Siglas

ABCD Assimetria, Borda, Cor e Diâmetro.

BCC *Basal Cell Carcinoma.*

BNN *Back-propagation Neural Network.*

CIE *Commission Internationale de l'Eclairage.*

CNN *Convolutional Neural Network.*

CSV *Comma Separated Values.*

ENN *Edited Nearest Neighbor.*

FCL *Fully Connected Layer.*

FN Falso Negativo.

FP Falso Positivo.

HSB *Hue, Saturation e Brightness.*

HSV *Hue, Saturation e Value.*

IA Inteligência Artificial.

IEC *Intra-Epidermal Carcinoma.*

INCA Instituto Nacional de Câncer.

ISIC *International Skin Imaging Collaboration.*

KNN *K-Nearest Neighbors.*

LAB L representa *perceptual lightness* e AB representa as quatro únicas cores visíveis ao olho humano: vermelho, verde, azul e amarelo.

LMM Lentigo Maligno Melanoma.

LUV L representa *brightness* e UV representa *chroma*.

MLP *Multilayer Perceptron*.

NM *Nodular Melanoma*.

OPF *Optimum-path Forest*.

ReLU *Rectified Linear Unit*.

RF *Random Forest*.

RGB *Red, Green e Blue*.

RNA Rede Neural Artificial.

SCC *Squamous Cell Carcinoma*.

SSM *Superficial Spreading Melanoma*.

SVM *Support Vector Machines*.

VN Verdadeiro Negativo.

VP Verdadeiro Positivo.

Capítulo 1

Introdução

O câncer é resultado de uma mutação genética, que se inicia através de uma expansão anormal e incontida das células, estas podem invadir tecidos adjacentes ou órgãos e sem tratamento pode levar a doenças letais. Em sua fase mais agressiva ocorre um processo chamado de metástase que corresponde à formação de tumores que se espalham por diversas regiões do corpo [2].

Existem diferentes tipos de câncer de pele dentre eles estão os carcinomas, sarcomas, linfomas e mielomas. Os sarcomas surgem nas células de tecido conjuntivo, como músculos, tendões, vasos sanguíneos, ossos, entre outros. Já os carcinomas e têm origem nas células epiteliais, as quais revestem a pele e a maioria dos órgãos. E os linfomas e mielomas, surgem nas células do sistema sanguíneo.

O câncer de pele causado pelo desencadeamento de células anormais e seu aumento, é muito comum que sua concepção se dê por conta de exposição excessiva aos raios UV do sol, mas outros fatores que podem ser genéticos ou tipo de pele podem influenciar. De acordo com dados do Instituto Nacional de Câncer (INCA) [3], o câncer de pele é um dos mais comuns no Brasil, ele representa mais de 30% dos tumores malignos. O câncer de pele acomete mais na população de pele clara e de fotótipo queimado e não bronzeado é mais propensa ao câncer de pele, enquanto que hispânicos, asiáticos e negros tendem a serem menos afetados por esse tipo de câncer. Existem diferentes tipos de câncer de pele, mas os três tipos principais são o câncer de pele não melanoma, o carcinoma epidermoide e o carcinoma basocelular [4].

O uso da Inteligência Artificial, especificamente do Aprendizado Profundo, tem sido proposto como uma solução eficiente para o problema de detecção precoce de melanomas [5][6]. Dentre os diversos modelos propostos, uma abordagem híbrida tem sido considerada uma alternativa interessante, consistindo na combinação de dois modelos diferentes, tal como ilustrado no estudo realizado por Tanna et al. [7]. Esta combinação de modelos permite uma abordagem mais completa e precisa no reconhecimento de melanomas,

contribuindo para a efetividade do sistema de detecção.

1.1 Objetivos gerais

O objetivo deste estudo é comparar o desempenho de diferentes modelos de classificação aplicados ao problema de detecção de melanoma em imagens dermatoscópicas. Além disso, é comparado um modelo clássico que utiliza características extraídas manualmente com um modelo que possui extração automática de características através de uma *Convolutional Neural Network* (CNN).

1.2 Objetivos específicos

Para alcançar os objetivos descritos anteriormente os seguintes objetivos específicos são definidos:

- Verificar o desempenho da detecção de melanoma considerando diferentes descritores de características e o classificador floresta aleatória;
- Verificar o desempenho do modelo VGG16 para a extração de característica e detecção de melanoma;
- Avaliar o comportamento dos modelos com o uso de aumento de dados e sem aumento de dados;
- Avaliar modelos híbridos considerando VGG16 e Floresta Aleatória para a extração de características e detecção de melanoma, respectivamente.

1.3 Estruturação do trabalho

A estrutura deste trabalho é formada pelos seguintes capítulos:

- O Capítulo 2 fornece uma base sólida de conhecimento teórico para que seja possível uma introdução breve aos conceitos que serão abordados ao longo do presente trabalho. Este referencial teórico é fundamental para o entendimento das discussões e análises que serão realizadas no decorrer deste trabalho, e proporcionará uma base sólida para que se possa avaliar e compreender os resultados apresentados no final;
- O Capítulo 3 apresenta uma ampla revisão da literatura, destacando os principais autores e artigos científicos que foram essenciais para a construção da base teórica

e foram uma fonte de inspiração para a proposta deste trabalho. Nele, são apresentadas as contribuições mais relevantes para o tema em questão e seus principais conceitos e teorias. Esta revisão ajudará a embasar a proposição apresentada no presente trabalho;

- O Capítulo 4 fornece uma descrição detalhada da metodologia utilizada nos experimentos propostos no presente trabalho. Nesta seção, são apresentados os passos críticos que foram seguidos para garantir a confiabilidade e a validade dos resultados obtidos. Além disso, é descrita a base de dados que foi utilizada para alimentar os experimentos, incluindo informações sobre sua coleta, tratamento e organização. Por fim, é explicado como o processamento dos experimentos foi realizado, incluindo as técnicas e métodos utilizados para analisar e interpretar os dados obtidos;
- O Capítulo 5 apresenta uma análise detalhada dos resultados obtidos a partir dos experimentos propostos no capítulo anterior. A principal finalidade desta seção é encontrar a melhor combinação de resultados para a classificação de melanomas, considerando as diferentes abordagens e técnicas utilizadas nos experimentos;
- O Capítulo 6 propõe uma reflexão aprofundada sobre os resultados obtidos ao longo do presente trabalho, a fim de compreender melhor as implicações e contribuições desses achados para a área de estudo em questão;
- O Capítulo 7 apresenta conclusões e sugestões para trabalhos futuros, visando aprimorar e expandir o conhecimento e estudos sobre melanomas, bem como a identificação de novos desafios e oportunidades.

Capítulo 2

Fundamentação Teórica

O capítulo que se segue tem como objetivo demonstrar as técnicas utilizadas para a realização do trabalho, além de expor os conceitos para auxiliar o entendimento das mesmas. Para uma melhor compreensão faz-se uma apresentação geral do tema de forma expositiva com auxílio de uma base de dados em que há uma análise para a fundamentação aprimorada dos resultados da pesquisa realizada. Dessarte o conteúdo presente trata-se de um aprofundamento dos conceitos iniciais que são de extrema importância para uma melhor compreensão da proposta do projeto.

2.1 Visão Computacional

A visão computacional é uma área da inteligência artificial que visa compreender informações visuais, podendo ser imagens ou vídeos. É resultado da combinação de teorias, algoritmos e técnicas de processamento de imagens, identificação de padrões e aprendizado de máquina, com o objetivo de imitar a percepção humana do mundo.

Desta forma, o processamento de imagens é uma etapa importante na visão computacional, que se concentra em tratar e interpretar imagens digitais. É um conjunto de técnicas que possibilita a manuseio, análise e compreensão das informações obtidas, com a finalidade de extrair informações úteis e significativas, as etapas do processamento da imagem são [8]:

- Aquisição de imagem: Os dados foram obtidos através de um desafio do *International Skin Imaging Collaboration* (ISIC);
- Pré-processamento: Limpeza e preparação da imagem para o processamento subsequente;
- Segmentação: Divisão da imagem em regiões homogêneas, separando objetos e fundo.

- Caracterização: Extração de informações relevantes como formas, cores, texturas, entre outras;
- Interpretação: Compreensão do contexto da imagem e tomadas de decisão com base nas informações processadas.

2.1.1 Categoria de Dados

Os dados são uma das fontes mais importantes para a tomada de decisões e a solução de problemas em muitas áreas. Por esse motivo, é importante entender como eles podem ser classificados.

Os dados estruturados são aqueles que possuem uma forma clara e organizada de apresentação. Eles têm campos definidos com formatos específicos e são geralmente encontrados em bancos de dados com relações projetadas. Isso significa que esses dados são facilmente interpretáveis e utilizáveis, o que os torna uma escolha comum para análises e decisões em muitos setores.

Por outro lado, os dados não estruturados não possuem um padrão ou estrutura comum, como imagens, anotações e áudios. Eles precisam ser organizados antes da análise e, por vezes, podem ser mais desafiadores de se trabalhar, mas ainda são uma fonte importante de informação.

Os dados semiestruturados são aqueles que possuem uma estrutura irregular ou limitada e não possuem formatos definidos para os campos. Um exemplo disso é o arquivo *Comma Separated Values* (CSV), que integra fontes diversas. Embora esses dados não sejam tão organizados quanto os dados estruturados, ainda são úteis em muitos casos, pois combinam informações de diferentes fontes em um único lugar [9].

Neste trabalho são utilizadas imagens associada a arquivos CSV com rótulos, como fonte de informação e treino, validação e teste do modelo.

2.1.2 Espaço de cor

A partir de uma imagem é possível extrair atributos, dentre eles podem conter diversas cores. Existem várias características de cores das lesões de pele que os especialistas usam no diagnóstico. A cor geral da lesão de pele pode ser avaliada em relação ao resto da pele circundante. Algumas lesões podem ter uma cor uniforme, enquanto outras podem ter variações de cor. A homogeneidade refere-se à uniformidade da cor da lesão de pele. Uma lesão homogênea terá uma cor uniforme em toda a sua extensão, enquanto uma lesão não homogênea terá áreas com diferentes tons ou padrões de cores. A variação de cor refere-se a diferenças de cor dentro da lesão de pele. As áreas de variação de cor podem ser indicativas de uma lesão maligna. A distribuição de cor se refere a como a cor é

distribuída dentro da lesão de pele. Algumas lesões podem ter uma distribuição simétrica da cor, enquanto outras podem ter uma distribuição assimétrica. A cor da borda da lesão de pele pode ser avaliada em relação ao resto da lesão e da pele circundante [10].

No caso de uma imagem RGB, será separada pelos canais de cores *Red*, *Green* e *Blue* conforme visto na Figura 2.1, sendo as três cores primárias de acordo com a teoria das cores, assim sendo possível captar somente o vermelho ou somente o verde ou somente o azul, mas também existem outros canais de cores que são o *Hue*, *Saturation* e *Brightness* (HSB) ou *Hue*, *Saturation* e *Value* (HSV), LUV (L representa *brightness* e UV representa *chroma*), LAB (L representa *perceptual lightness* e AB representa as quatro únicas cores visíveis ao olho humano: vermelho, verde, azul e amarelo) e níveis de cinza [11]. Uma imagem é formada por um conjunto de *pixels*, que por sua vez costuma ser baseada no modelo de RGB. A partir disso, também é possível converter o modelo RGB em modelos como HSV ou LUV [12].

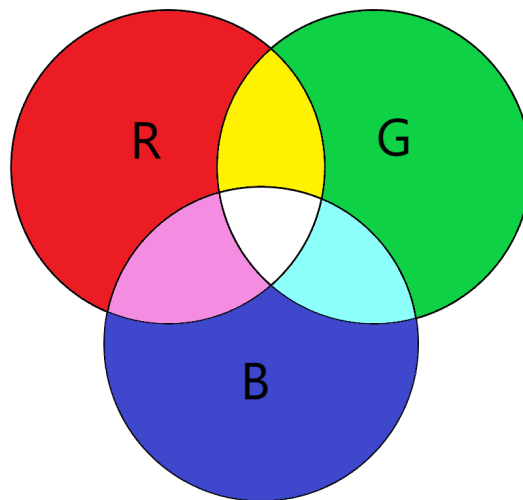


Figura 2.1: Camadas do espaço de cor RGB - *Red*, *Green* e *Blue*
Fonte: Autoria própria.

Hue, *Saturation* e *Value* (HSV) também conhecido como HSB, é abreviação de um modelo que tem como base a matiz (*hue*), a matiz é a tonalidade dominante da área, que estende-se desde o vermelho até o violeta, já a saturação é a pureza da cor, uma imagem mais pura possui mais tons de cinza, enquanto o valor se refere a uma dimensão de luminosidade da cor, na Figura 2.2 é visto esses espaços. O espaço de cor foi desenvolvido por Smith [13] com o intuito de ajudar na manipulação de cores de forma mais instintiva.

O modelo CIELUV, conforme Figura 2.3, é um modelo de cor, utilizado para mostrar um sistema de coordenadas tridimensionais, é baseado em um polígono que abrange todas as cores e leva em consideração a falta de percepção dos olhos humanos, desenvolvido pela *Commission Internationale de l'Eclairage* CIE [14].

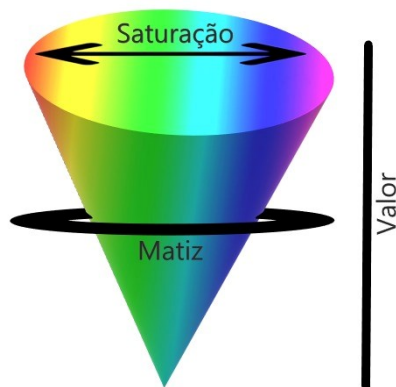


Figura 2.2: Camadas do espaço de cor HSV - Matiz, Saturação e Valor
 Fonte: Autoria própria.

Diagrama cromático do espaço de cores

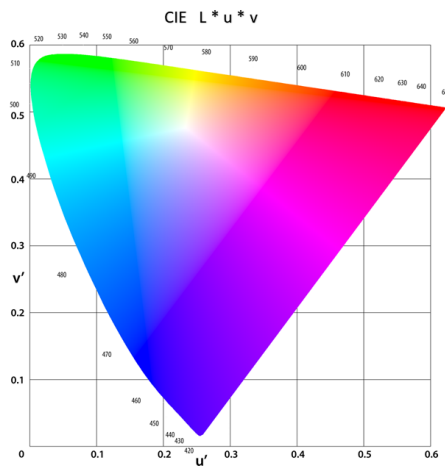


Figura 2.3: Diagrama cromático de cor CIELUV
 Fonte: Adaptado de Hisour [15].

Outro modelo também desenvolvido pela CIE é o LAB ou CIELAB possui duas coordenadas, onde a coordenada A varia entre vermelho e verde e a coordenada B varia entre amarelo e azul. Além disso, a luminosidade é representada pela coordenada L . Na literatura, é comum usar a colometria para fins diagnósticos, como o diagnóstico de doenças, como por exemplo no artigo de Arifin et al. [16].

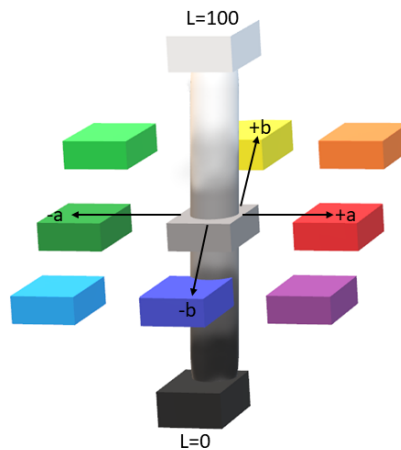


Figura 2.4: Espaço de cor CIE L^*a^*b
 Fonte: Autoria própria.

2.1.3 Textura

Textura é uma propriedade visual e tátil de uma superfície que pode ser percebida por meio do toque ou da visão. Na área de processamento de imagens, a textura é uma característica importante que descreve as variações nas intensidades dos *pixels* em uma imagem. Essas variações podem ocorrer em diferentes escalas e direções, criando padrões de textura que podem ser utilizados para identificar objetos ou regiões em uma imagem [17].

Existem diferentes tipos de textura, incluindo texturas regulares, irregular e quase regular [18]. A textura pode ser descrita qualitativamente, com termos como rugoso, liso, granulado ou áspero, ou quantitativamente, com medidas estatísticas que descrevem a distribuição e a variação de intensidades de *pixel* em uma determinada área da imagem.

A análise de textura é amplamente utilizada em várias áreas, incluindo medicina, análise de alimentos, geologia, análise de materiais e processamento de imagens [19]. Alguns dos métodos mais comuns para análise de textura em imagens médicas incluem a matriz de cocorrência de níveis de cinza [20] e análise da dimensão fractal [21].

2.2 Aprendizado de Máquina

Com o mundo cada vez mais conectado e com mais informação, a cada instante cria-se uma demanda por entender o volume de dados gerados. Tendo em vista a complexidade dos problemas já existentes e futuros, viabiliza uma busca por ferramentas com maior independência capazes de solucionar problemas através de experiências passadas denominada Aprendizado de Máquina [22].

Aprendizado de Máquina do inglês *Machine Learning*, é uma área da Inteligência Artificial (IA) que utiliza técnicas e algoritmos computacionais na construção de sistemas capazes de analisar dados para se obter o conhecimento de forma automática, na construção de modelos analíticos [23]. Deste modo, busca aprender com dados, além de identificar padrões e até mesmo tomar decisões com a menor quantidade de interferência humana. O aprendizado de máquina está em crescente evolução e ascensão, tendo em vista o seu potencial, tem se tornado cada vez mais utilizado na tomada de decisões [24].

2.2.1 Tipos de Aprendizado

Existem diferentes tipos de sistemas nos quais o aprendizado de máquina pode ser classificado, as quatro principais categorias consistem em aprendizado supervisionado, aprendizado não supervisionado, semi-supervisionado e aprendizado por reforço [25].

No modelo supervisionado, que é o utilizado no desenvolvimento deste trabalho, os dados foram apresentados ao algoritmo com os dados de entrada e com os resultados que o acompanha aquele conjunto, para incentivar e possibilitar o aprendizado. Já no aprendizado não supervisionado, os dados fornecidos para o treinamento do modelo não são rotulados, ou seja, o modelo tenta aprender de maneira autônoma sem a ajuda de um supervisor. Aprendizado semi-supervisionado é uma forma de possibilitar a máquina a aprender de forma autônoma, isto é, primeiro ensina-se com condições normais, logo após ensina-se casos que fogem do padrão, neste método é possível mexer com dados rotulados ou não-rotulados. Por fim, na aprendizagem por reforço, o algoritmo interage repetidamente com um ambiente dinâmico com o objetivo específico de chegar à solução mais otimizada para o problema por meio de repetidos erros e tentativas.

2.2.2 Aspectos Básicos dos Algoritmos de Classificação

A classificação é uma tarefa amplamente usada no aprendizado de máquina, no qual possui o objetivo de prever a categoria ou classe de novos dados baseada nas amostras de treinamento rotulados anteriormente. Na etapa de classificação existem dois possíveis métodos, o primeiro consiste no método tradicional no qual se aplica aprendizagem de máquina e extrai de forma manual as características da imagem como por exemplo: cor, textura e borda, já no método não tradicional aplica uma extração automática, no qual usa de uma rede neural convolucional. Os métodos de classificação, podem ser usados de diversas formas, dentre elas é possível classificar se um e-mail é *spam* ou não, se uma imagem trata-se de um animal, mas também é possível assimilar padrões, como, Assimetria, Borda, Cor e Diâmetro (ABCD), desta forma identificando possíveis doenças de pele [26].

2.3 Floresta Aleatória

Floresta Aleatória, do inglês *Random Forest*, é um método de aprendizado de máquina [27]. Ele funciona criando várias árvores de decisão, por isso o termo floresta, em seguida, utiliza uma votação entre as árvores para prever uma classificação ou valor. Cada uma das árvores é montada a partir de uma amostra do conjunto de dados, com um número aleatório de características escolhidas para tomadas de decisão. A votação é realizada agregando e analisando as previsões de todas as árvores.

Floresta Aleatória é um modelo bastante disseminado na área de inteligência artificial para tomada de decisões baseadas em previsões, porque é um modelo de pouco risco além de ser capaz de lidar com problemas de classificação e regressão [28].

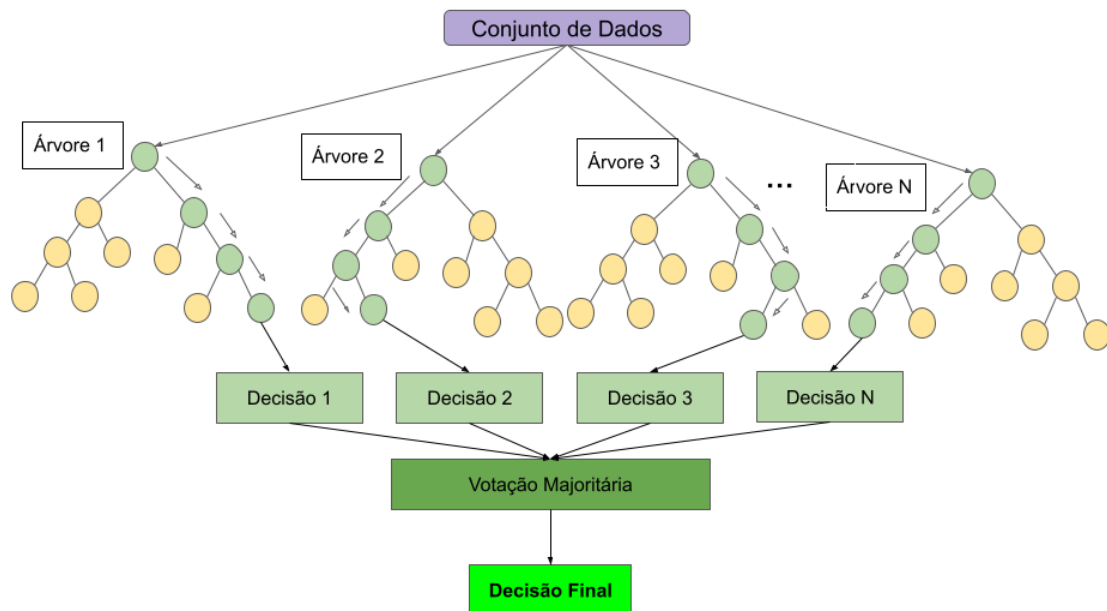


Figura 2.5: Floresta Aleatória
Fonte: Autoria própria.

2.4 Redes Neurais Artificiais

A Aprendizagem Profunda, também conhecida como *Deep Learning*, foi apresentada por Frank Rosenblatt [29]. O algoritmo utiliza redes neurais e permite o uso de imagens digitais. Desde então, houve muitas aplicações, incluindo na área da saúde para detecção de imagens clínicas. Algoritmos de Aprendizagem Profundo já foram comprovados a superar especialistas humanos em alguns problemas de diagnóstico de imagens médicas [30]. Desta maneira as RNA são um modelo de aprendizagem de máquina inspirados em no

cérebro humano [31], pois usam camadas de neurônios que são modelados matematicamente, essas são versáteis e capazes de classificar imagens, reconhecimento de fala e assim aprendendo sozinha e configurando próprios parâmetros de acordo com que se aprende, na Figura 2.6 é possível ver algumas dessas camadas. As redes neurais são conjuntos de neurônios dispostos em camadas, ligados por uma sinapses ou elos de conexão [32]. Como Alpaydin [33] expõe, “as redes até o momento atuam de maneira restrita, mas é notável os resultados conforme novas redes são treinadas e com maior número de dados”. Além disso, quanto maior o número de camadas na rede neural, mais neurônios há, e conseqüentemente, mais características podem ser extraídas e analisadas para melhorar a precisão da rede.

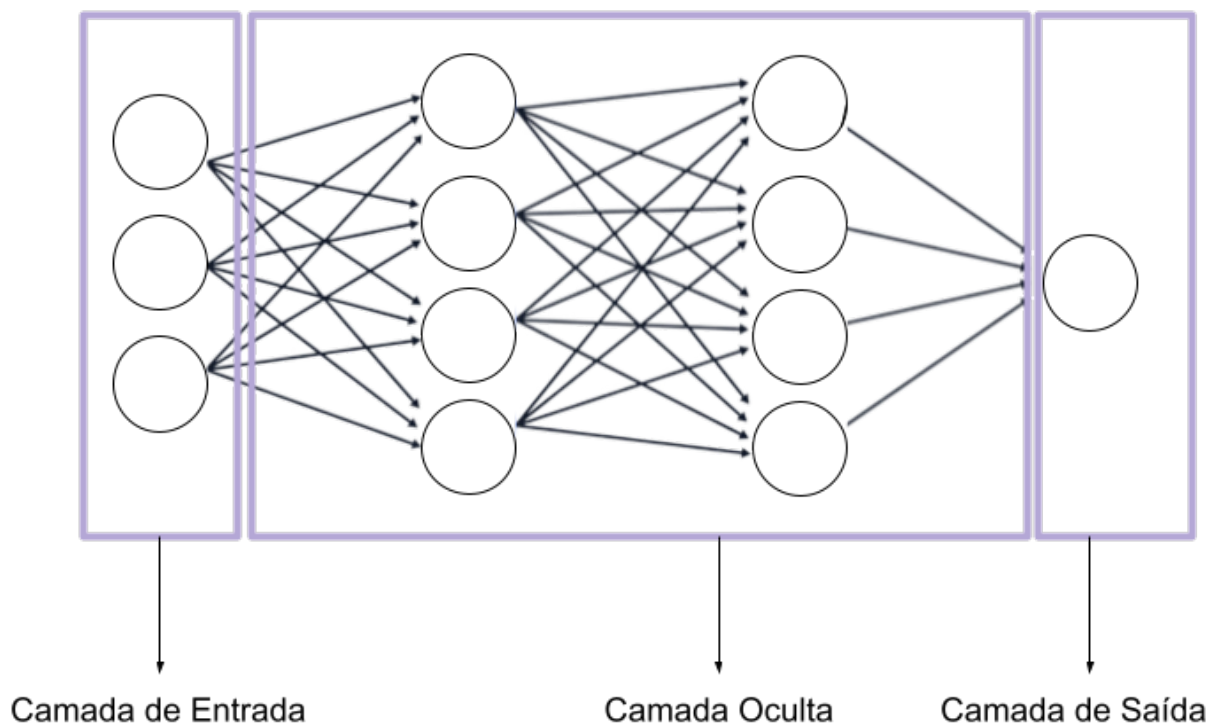


Figura 2.6: Camadas de uma Rede Neural Artificial
Fonte: Autoria própria.

2.4.1 Funcionamento das Redes Neurais Artificiais

A arquitetura de uma Rede Neural Artificial (RNA) é composta por camadas interconectadas de neurônios, cada uma delas pode conter vários neurônios e pode ser organizada de diferentes formas de acordo com o tipo de RNA e o problema a ser resolvido. A camada de entrada é a primeira camada da rede e é crucial para o funcionamento da RNA, pois é nesta camada que os dados de entrada são recebidos e processados. Cada neurônio

nesta camada é responsável por representar uma característica importante dos dados de entrada, sejam eles *pixels* de uma imagem ou palavras em um texto.

2.4.2 Arquitetura

Uma das arquiteturas mais comuns é a rede neural *feedforward*, também conhecida como *Multilayer Perceptron* (MLP), na qual as camadas estão organizadas de forma linear, sem ciclos. As camadas são conectadas entre si de forma que os dados fluem em uma direção única, da entrada para a saída. Outra arquitetura comum é a rede neural recorrente, na qual as camadas são conectadas de forma que os dados podem fluir de volta para camadas anteriores. Isso permite que as informações sejam mantidas ao longo do tempo, o que é útil para problemas de séries temporais.

2.5 Redes Neurais Convolucionais

A Rede Neural Convolucional, do inglês *Convolutional Neural Network* (CNN), é especialmente eficiente para trabalhar com imagens e outros tipos de dados específicos. CNNs são fundamentadas na ideia de que uma característica importante de uma imagem pode ser encontrada em qualquer lugar da imagem, e que essa característica pode ser encontrada usando um conjunto de filtros que procuram por padrões específicos [34].

As CNNs podem conceber modelos mais robustos que são capazes de extrair mais informações concretas na distinção de imagens em classes, por indicar precisamente características espaciais relevantes.

CNNs são compostas por uma quantidade numerosa de camadas, cada uma das quais efetua operações matemáticas específicas, como mostrado na Figura 2.7 . As primeiras camadas de uma CNN geralmente consistem na aplicação de diversos filtros a imagem de entrada para extrair características. Esses filtros são preparados de detectar padrões específicos, como bordas, texturas e formas. Essas características são então passadas para as camadas seguintes, que são combinadas com a finalidade de produzir características mais complexas [35].

As CNNs são extensamente aplicadas em diferentes tarefas de processamento de imagem, como reconhecimento de rostos, classificação de imagens, detecção de objetos e segmentação de imagens. Além disso, as CNNs têm sido usadas para processar vídeos, áudios e outras séries temporais.

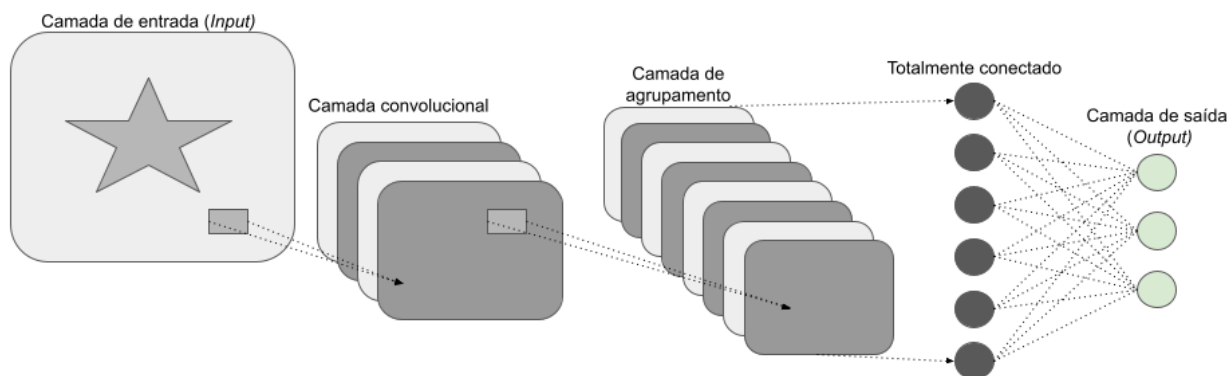


Figura 2.7: Camadas de uma Rede Neural Convolutiva
 Fonte: Autoria própria.

2.5.1 Camada de Convolução

A Camada de Convolução é uma das principais componentes das CNN, é destinada à extração de características pertinentes das imagens de entrada. A camada de Convolução trabalha empregando os filtros nas imagens, no qual cada filtro é responsável por reconhecer uma característica específica, como bordas ou texturas. Os filtros são aplicados nas imagens, sendo cada filtro responsável por reconhecer uma característica específica, como bordas, texturas, etc. Esses filtros são matrizes de pesos que são ajustadas durante o processo de treinamento da CNN. O resultado da aplicação dos filtros na imagem é uma nova imagem conhecida como mapa de ativação, onde cada *pixel* representa a ativação do filtro naquele ponto da imagem.

2.5.2 Camada de Subamostragem ou *Pooling*

A Camada de Subamostragem, também conhecida como camada *Pooling* [34] tem o principal foco em diminuir a dimensionalidade dos dados de entrada, dessa maneira mantendo uma maior quantidade de características importantes. Isso é obtido por meio da combinação de diversos *pixels* em uma única unidade, normalmente utilizando operações matemáticas básicas como média ou máximo. Essas operações são aplicadas em janelas deslizantes, conhecidas como *kernels*, que percorrem a imagem de entrada.

Essa camada é importante para a CNN, pois auxilia a evitar o sobreajuste e melhorar o desempenho no qual ocorre o treinamento e o teste da rede. Desta maneira, auxilia em que a rede obtenha uma maior competência de generalização, isso porque ela auxilia na identificação de características mais abstratas e mais genéricas na imagem.

Uma outra vantagem dessa camada é sua capacidade de preservar a invariância de translação. Isso significa que a rede neural pode identificar objetos mesmo que eles apareçam em diferentes posições na imagem de entrada. Além disso, como a camada de subamostragem agrega informações de vários *pixels* em uma única unidade, ela ajuda a identificar características mais robustas na imagem. Como resultado, essa camada pode melhorar significativamente a capacidade da rede de reconhecer objetos em diferentes contextos visuais.

2.5.3 Camada Completamente Conectada

A camada totalmente conectada, do inglês *Fully Connected Layer* (FCL), é uma camada fundamental na estrutura de uma CNN. Cada neurônio na camada FCL está conectado a todos os neurônios da camada anterior, dando origem ao seu nome totalmente conectada [36]. A principal função da FCL é transformar as características detectadas nas camadas anteriores em uma representação final que pode ser usada para fazer previsões. Isso é realizado através do uso de pesos e bias, que são aprendidos durante o treinamento da CNN. Os pesos são usados para combinar as entradas da camada anterior para calcular a saída do neurônio.

2.5.4 Arquiteturas VGG

A arquitetura VGG, desenvolvida pela equipe do grupo *Visual Geometry* da Universidade de Oxford [37], é amplamente reconhecida como uma das mais eficientes para o reconhecimento de imagens. Introduzida em 2014, essa arquitetura consiste em uma sequência de camadas de convolução, *pooling* e camadas totalmente conectadas, que em conjunto oferecem um desempenho interessante em tarefas de classificação de imagens. Devido à sua capacidade de melhorar significativamente o reconhecimento de objetos em imagens, a arquitetura VGG tem sido amplamente adotada em aplicações de visão computacional.

A arquitetura VGG é caracterizada por sua estrutura de camadas de convolução intercaladas com camadas de *pooling* e finalizadas com camadas totalmente conectadas. O que a diferencia de outras arquiteturas é o uso de uma grande quantidade de camadas de convolução de baixa dimensionalidade (3×3), em vez de camadas de alta dimensionalidade (como 5×5 ou 7×7). Além disso, a VGG inclui várias camadas de *pooling* entre as

camadas de convolução para reduzir a dimensionalidade dos dados e aumentar a eficiência do modelo.

2.5.5 *Transferência de Conhecimento*

A transferência de conhecimento, do inglês *Transfer Learning*, é uma técnica que consiste na possibilidade de aperfeiçoar um modelo já treinado para executar outra demanda parecida. Alguns modelos partilham de características semelhantes ao lidar com tarefas parecidas. Isto é, ao invés de precisar treinar um modelo do início, a transferência de conhecimento pode ser usada a partir de informações que ela aprendeu no primeiro treinamento, isso torna a tarefa mais rápida e com melhor desempenho.

A aprendizagem por transferência permite aproveitar o conhecimento adquirido em um problema específico para melhorar a performance em outros problemas relacionados. Ela permite que os domínios, tarefas e distribuições utilizados durante o treinamento e teste sejam diferentes, economizando tempo na rotulação de dados. Isso é possível porque existem dados semelhantes que podem ajudar no aprendizado. Um exemplo disso é o *ImageNet*, onde um modelo treinado para reconhecer objetos em imagens, como animais e veículos, pode ser transferido para detectar objetos em vídeos, pois muitas das características importantes para reconhecer objetos em imagens são relevantes para outras tarefas de processamento de imagem, conforme é possível visualizar na Figura 2.8.

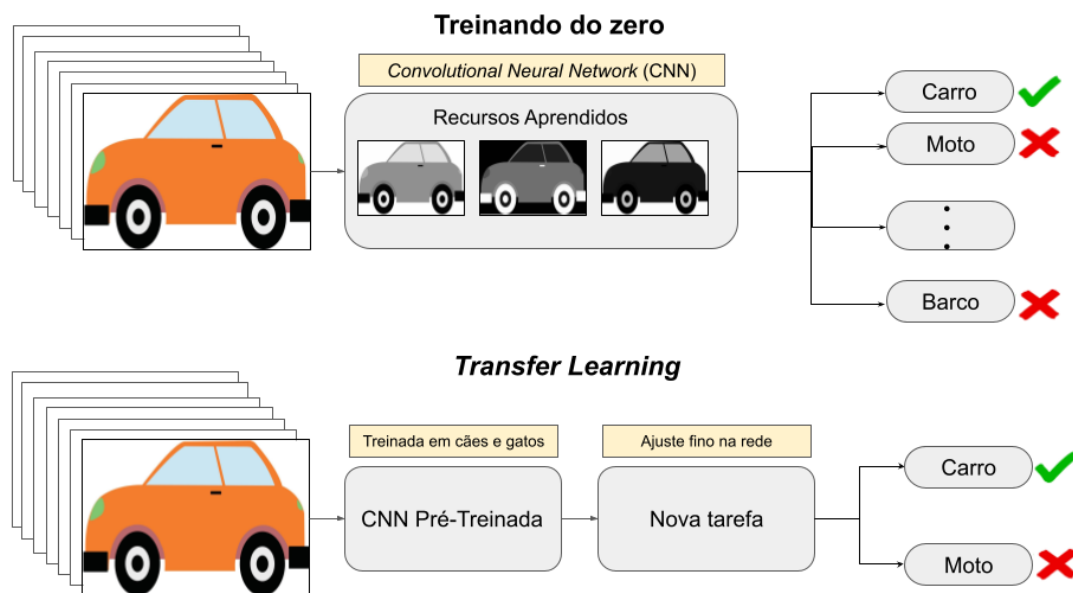


Figura 2.8: Extração utilizando CNN e Classificação através de Floresta Aleatória
Fonte: Autoria própria.

A necessidade de agilizar o treinamento de grandes modelos de aprendizado profundo e suprir a escassez de dados de treinamento resultou na utilização da técnica de transferência de conhecimento, que permite aproveitar o conhecimento adquirido por modelos anteriormente treinados com grandes volumes de dados [38].

Em cenários em que a classe minoritária possui uma quantidade muito inferior de amostras em relação à classe majoritária, a transferência de conhecimento pode ser uma estratégia eficaz para enfrentar esse desafio [39]. Essa técnica permite o aproveitamento de informações provenientes de outros treinamentos, mesmo que não sejam relacionados ao problema específico em questão. O aprendizado por transferência é uma abordagem mais eficiente e acessível, o que o torna particularmente adequado para grandes conjuntos de dados, como o *ImageNet*, em modelos supervisionados e não supervisionados. Com isso, é possível melhorar a acurácia do modelo e lidar com problemas de desbalanceamento de classes, ampliando as possibilidades de aplicação em diferentes áreas da visão computacional.

Na literatura, tem sido uma opção interessante, pois a transferência de conhecimento tem gerado bons resultados e é uma alternativa muito vantajosa, pois dispensa o tempo que seria gasto treinando uma rede neural robusta do zero para um domínio específico [38] [40].

De acordo com Morid et al. [41], é possível transferir o conhecimento obtido a partir de dados não necessariamente relacionados à área médica para resolver problemas específicos nessa área. Mais especificamente, os parâmetros de modelos de redes convolucionais treinados em dados do *ImageNet*, que consiste em imagens naturais (como as usadas nas arquiteturas AlexNet[42], VGGNet[37] e ResNet[43]), podem ser utilizados em um modelo de redes convolucionais destinado a solucionar um problema de imagem médica. Essa abordagem de transferência de conhecimento tem se mostrado promissora para melhorar o desempenho de modelos em tarefas de imagem médica, permitindo que dados de outras áreas sejam aproveitados para resolver problemas específicos da área médica.

Isto é, o *ImageNet* consiste em um projeto com foco em rotular e classificar imagens em um conglomerado de frases e palavras, no projeto existem mais de 14 milhões de imagens nas quais podem ser reaproveitadas. Podendo ser utilizada com as redes VGG16, VGG19 e ResNet50.

2.6 Balanceamento de Classes

Além dessas técnicas, a aprendizagem por transferência também pode ser uma solução eficaz para o desequilíbrio de classes. Ao aproveitar conhecimento adquirido em um problema específico, a transferência permite a utilização de dados adicionais que podem com-

plementar e equilibrar a distribuição das classes no treinamento. Dessa forma, o modelo pode aprender a lidar com as classes minoritárias de maneira mais eficiente, melhorando sua performance na tarefa de classificação. É importante lembrar que a escolha da técnica deve ser baseada nas características do conjunto de dados e no objetivo específico da tarefa. Em resumo, a aprendizagem por transferência pode ser uma abordagem complementar e valiosa para lidar com o desequilíbrio de classes em tarefas de classificação.

Aumento de dados

A técnica de aumento de dados, do inglês *Data Augmentation*, é uma técnica utilizada para aumentar o número de exemplos de treinamento em um conjunto de dados de imagem. Isso é feito através da aplicação de transformações aleatórias nas imagens existentes, como rotação, escala e deslocamento, que é uma técnica que pode ser utilizada para auxiliar no balanceamento das classes [44].

Desta maneira, o aumento de dados é uma técnica que foi proposta e estudada por vários autores ao longo dos anos. A ideia de utilizar transformações aleatórias para aumentar o número de exemplos de treinamento em um conjunto de dados. Os autores Krizhevsky et al. [35] mostraram que o uso de aumento de dados de imagens juntamente com uma arquitetura de CNN ajudou a melhorar significativamente a precisão de reconhecimento de imagens no grande banco de dados *ImageNet*.

2.7 Métricas de Desempenho

Considerando que a natureza dos testes de aprendizado de máquina pode resultar em erros e acertos, é importante lembrar que nem todas as tentativas serão perfeitas ou terão 100% de precisão. Embora o objetivo seja obter resultados o mais próximos possível da realidade, ainda pode haver variações nos resultados. A avaliação dos modelos de aprendizado de máquina é feita através de métricas de desempenho, que são usadas para medir a eficácia dos modelos de classificação.

A escolha das métricas de avaliação tem um impacto significativo no desempenho dos resultados da classificação. Por isso, é relevante que essa escolha seja feita de forma criteriosa, já que existem métricas específicas indicadas para diferentes objetivos e tipos de modelos testados. A escolha deve verificar e ponderar fatores ligados ao tipo de dado utilizado e sua distribuição, nesse caso é importante validar as implicações direta das métricas utilizadas ao modelo.

Métricas de desempenho podem ter diferentes níveis de dificuldade em sua aplicação. Nesse caso, a mais simples é a chamada Acurácia, cujo o principal objetivo é analisar o percentual de acertos em detrimento a quantidade total de dados utilizados. Para as

métricas é observado a quantidade de acertos e a quantidade de erros, dessa forma há as seguintes maneiras de se classificar os resultados obtidos: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN). Desta maneira, o verdadeiro positivo e negativo indica que o teste aponta de forma precisa e que o resultado está certo. Ou seja, quando a imagem é analisada e se obtém indícios claros e precisos de que o indivíduo possui câncer de pele, será um verdadeiro positivo, através das imagens foi diagnóstico o melanoma, e da mesma maneira quando a análise não localiza esses indícios, se qualifica como verdadeiro negativo, pois o indivíduo não possui câncer de pele.

O falso positivo e o falso negativo representam resultados incorretos que podem ser apresentados por testes de diagnósticos. Um falso positivo ocorre quando o teste indica uma condição positiva, mas na realidade o paciente não a possui. Por exemplo, uma pessoa pode receber um diagnóstico de melanoma com base na análise de imagens, mas exames clínicos posteriores revelam que ela não tem câncer de pele. Já um falso negativo acontece quando o teste indica uma condição negativa, mas na verdade o paciente a possui. Por exemplo, um indivíduo pode receber um resultado negativo para melanoma, mas exames clínicos posteriores confirmam a presença da doença, a partir desses resultados é possível entender as métricas que ajudam na avaliação de um modelo.

Dentre as métricas, acurácia, precisão, revocação e F1-Score são comumente usadas na avaliação de classificadores [45].

- *Acurácia*: É a métrica que diz a respeito das previsões corretas em relação ao total de previsões, é calculado da seguinte maneira:

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (2.1)$$

- *Precisão*: É a proporção de verdadeiros positivos corretamente identificados em relação ao total de verdadeiros positivos e falsos positivos, é calculada a partir de:

$$Precisão = \frac{VP}{VP + FP} \quad (2.2)$$

- *Revocação*: É a proporção de verdadeiros positivos identificados em relação ao número total de verdadeiros positivos e falsos negativos, é calculada conforme:

$$Revocação = \frac{VP}{VP + FN} \quad (2.3)$$

- *F1-Score*: É a média harmônica entre precisão e revocação.

$$F1 - Score = 2 \times \frac{Precis\tilde{a}o \times Revoca\tilde{c}a\tilde{o}}{Precis\tilde{a}o + Revoca\tilde{c}a\tilde{o}} \quad (2.4)$$

Capítulo 3

Revisão de Literatura

Este capítulo apresenta as principais fontes de inspiração para a construção deste trabalho. Aqui, serão abordados temas relacionados à aplicação de métodos tradicionais de aprendizado de máquina e métodos de aprendizado profundo, com foco na classificação do câncer de pele. O objetivo principal é sintetizar as principais contribuições presentes em diversos estudos. Alguns trabalhos consideram essa classificação como binária, ou seja, determinado elemento é classificado apenas como presente ou ausente. Outras abordagens, por outro lado, são definidas como multiclasse, onde existem várias categorias para o problema

A *International Skin Imaging Collaboration* (ISIC) é uma plataforma colaborativa que reúne imagens voltadas para o tema de detecção e classificação de câncer de pele, usando técnicas avançadas de aprendizado de máquina. O objetivo é fomentar o conhecimento na área por meio da participação ativa da comunidade científica, oferecendo dados para pesquisas e estimulando a participação de pessoas externas ao meio acadêmico.

O PH2 [46] contempla um propósito semelhante, no qual é um banco de dados de imagens dermatoscópicas, que foi obtido através do Serviço de Dermatologia do Hospital Pedro Hispano, Matosinhos, Portugal, em que possui 200 imagens para classificação de melanoma. Essas bases de imagens são algumas das principais utilizadas na literatura quando é falado em categorização de imagens de câncer de pele.

Alguns trabalhos na literatura compartilham o mesmo campo de estudo, como o de Thomas et al. [47], que teve como objetivo caracterizar e classificar o tecido em 12 classes dermatológicas significativas, incluindo glândulas, inflamação, folículos capilares, hipoderme, derme reticular, derme papilar, epiderme, queratina, fundo, *Basal Cell Carcinoma* (BCC), *Squamous Cell Carcinoma* (SCC), e *Intra-Epidermal Carcinoma* (IEC). Neste estudo, as imagens de BCC (140), SCC (600) e IEC (90), bem como outras imagens de pele que podem ser machucados, dentre outros, foram usadas para treinamento. As lâminas de câncer de pele foram fotografadas usando um microscópio DP27 Olympus e o

MyLab Pathology forneceu acesso a elas. Para análise, um classificador CNN com 5 épocas e o otimizador Adam foram usados, enquanto o Random Forest com o parâmetro padrão de 100 árvores, K-Nearest Neighbors com parâmetros padrão e Perceptron de camada única com 64 neurônios ocultos (ativações *Rectified Linear Unit* (ReLU)) implementado em Keras foram usados para comparação. O classificador CNN alcançou um precisão geral da classificação nas classes saudáveis, BCC, SCC e IEC de 93.6%. Quando é visto os resultados de classificação para um câncer e não câncer, a CNN e a Floresta Aleatória tiveram um desempenho parecido, onde houve uma precisão geral de 97.9%. No caso do modelo KNN e o Perceptron de Camada Única, houve a performance onde os classificadores pontuaram 80.9% e 85.1% para o problema de quatro classes, já na classificação binária, 85.1% e 89.4% respectivamente. Nesse trabalho, há uma amostra bem pequena e por isso houve uma limitação, tendo em vista isso, é notável que a variação de classe não é bem aplicada e por isso houveram precisões diferentes conforme a densidade de *pixels* [47].

Em uma abordagem feita por Winkler et al. [48] usa o método CNN para diagnóstico de melanomas possui o foco em apurar qual seria o desempenho de uma CNN no mercado europeu em diferentes localizações e subtipos de melanoma. A base é formada por um conjunto de 30 imagens de melanomas e 100 não melanomas. Portanto, pode ser observado que a base de imagens contempla um conjunto levemente desbalanceado, assim a principal ideia é a classificações em seis conjuntos das imagens dermatoscópicas (melanomas de disseminação superficial e nevos maculares (SSM), melanomas de lentigo maligno e lentigos solares faciais/ceratoses seborreicas/nevos (LMM), melanomas nodulares e nevos papilomatosos/dérmicos/azuis (NM), melanomas mucosos e melanoses mucosas/máculas/nevos (Mucosa), melanomas acrolentiginoso e nevos, melanomas subungueais e nevos subungueais/lentigos/pigmentação de tipo étnico. No teste é possível observar que a CNN mostrou boas métricas para classes de SSM, NM e LMM, nas quais foram avaliadas pelas especificidade, sensibilidade, curva ROC-AUC, com respectivamente cerca de 93.3%, 65% e 0.926, já na classe AM com 83.3%, 91.0%, 0.928, enquanto na classe mucosa apresentou 93.3%, 38.0%, 0.754, na classe AM-unha 53.3%, 68.0%, 0.621. Um dos principais pontos desse estudo é que as imagens foram retiradas de locais distintos e escolhidas de maneira aleatória, isto é, não pode garantir que valerá para toda a população.

Outro método abordado é a de Oliveira et al. [49] onde é proposto a combinação de características, com base nas propriedades da forma, variação de cor e análise de textura, a ser extraída usando vários métodos de extração de características. Assim avaliar as características mais relevantes para o cálculo de lesões de pele para o diagnóstico com base na regra ABCD, incluindo propriedades de forma, variação de cor e análise de textura usando vários métodos diferentes, no qual é utilizado o repositório ISIC (2016) e foi

aplicado em conjunto de 1104 imagens dermatoscópicas por meio de um procedimento de validação cruzada. O diferencial está principalmente no critério de estruturas que é definido pela presença de redes de pigmento, vascularização, estruturas de regressão, estrias e até pontos/glóbulos. Os melhores resultados foram obtidos através do classificador Floresta Aleatória com ótimos resultados. Seis diferentes categorias de classificadores foram adotadas, a saber KNN, redes *Bayes*, árvore de decisão C4.5, *Multilayer Perceptron* (MLP), *Support Vector Machines* (SVM) e *Optimum-path Forest* (OPF). O classificador OPF obteve os melhores resultados de classificação usando algoritmos de seleção de características para o sistema de diagnóstico computacional de lesões cutâneas. O modelo obtido foi medido através de acurácia, sensibilidade e especificidade, para o conjunto completo obteve 92.3%, 87.5%, 97.1%, quando foram escolhidas 50 características, 91.6%, 87% e 96.2%, neste caso os resultados foram bem parecidos. Estudo não utilizou todas as imagens do conjunto de dados original.

Outro autor que também cita o método ABCD é Thanh et al. [50], no qual propõe processamento automático de imagens para detecção de câncer de pele do tipo melanoma, mas a principal técnica é a segmentação, utilizando imagens da coleção ISIC, onde é feito o pré-processamento de imagens de lesões de pele por tipo de curvatura principal adaptativa, lesões de pele segmentadas pela normalização de cor e características de extração pela regra ABCD, com imagens 2000 dermatoscópicas, além de mais 150 imagens de validação e para o teste 600 imagens, utilizadas em conjunto com um arquivo em formato CSV no qual possui a variável resposta. Também foram usadas variações de sistemas de cores como tons de cinza, HSV e LAB, outra característica extraída é homogeneidade usando a matriz de co-ocorrência de nível de cinza. Os resultados adquiridos na detecção do câncer de pele do melanoma indicam que o método proposto tem alta precisão e, no geral, um bom desempenho: para a etapa de segmentação, precisão, Dice, pontuações de Jaccard são 96.6%, 93.9% e 88.7%, respectivamente; e para o estágio de detecção de melanoma, a precisão é de até 100% para um subconjunto selecionado do conjunto de dados ISIC, mostrando um alto nível de precisão e com o uso dos métodos para normalização de cor, além de usar a técnica de retirada de pelos que podem ser encontrados nas imagens, torna o modelo bastante eficaz.

Seguindo a linha de extração de características com a técnica ABCD, há a abordagem proposta por Majumder [51], que consiste em extrair características geométrica distintas de imagens dermatoscópica para diferenciar e classificar o melanoma em benigno ou maligno, a base deste artigo é do Hospital Pedro Hispano (PH2) [46], onde é utilizada a rede neural *Back-propagation Neural Network* (BNN) com 200 imagens para treinamento, dentre elas sendo 160 melanomas benignos e 40 melanomas malignos. No estudo contém características como perímetro, área da lesão, método ABCD, a média para as cores RGB,

e a diferença entre os diâmetros máximo e mínimo da elipse que melhor se ajusta à lesão, com isso obtiveram uma precisão de 98%, sensibilidade de 95%, especificidade 98.8%, isto é, o aumento da característica aprimorou os resultados.

O trabalho de Bansal et al. [52] tem como objetivo extrair as características usando *Convolutional Neural Network* (CNN) com aprendizado por transferência e um classificador composto por *Random Forest* (RF), *K-Nearest Neighbors* (KNN) e *AdaBoost* [53] para detectar melanoma através das características extraídas, utiliza também a base de dados ISIC 2017 com o conjunto de treinamento possuindo 2000 imagens divididas entre 374 melanomas, 254 verruga seborreica e 1372 benignas. Pelo fato da coleção de dados ser desbalanceada foi utilizada uma técnica para aumento dos dados, sendo gerada imagens a partir da rotação das mesmas.

Uma abordagem muito interessante é apresentada por Maia [54], que possui o objetivo analisar e classificar melanomas observando e medindo quais algoritmos de aprendizagem de máquina combinados dispõe do melhor resultado. Levando em consideração um desbalanceamento, utiliza-se 200 imagens dermatoscópica divididas entre 80 lesões normais, 80 lesões atípicas e 40 melanomas, isto é, 160 não melanomas e 40 melanomas, com a base de dados PH2. Utilizando o método ABCD para diagnóstico de lesão de pele afim de comparar características extraídas por cada CNN treina-se VGG-16, VGG-19, ResNet50, ResNet. Logo após os aspectos derivados são passados para dois classificadores Máquina de Vetores de Suporte e RF, que é ajustado para SMOTE [55] e *Edited Nearest Neighbor* (ENN) [56] para obter um resultado melhor, no qual passa por um seletor, o classificador Floresta Aleatória obteve uma sensibilidade de 92.5% e um índice kappa de 77.15% após o balanceamento, nesta linha observa-se que é utilizado métodos tradicionais para extração e rede neural para classificação. A quantidade de amostra para treinamento foi consideravelmente pequena. Na Tabela 3.1, é possível notar os melhores resultados apresentados nessa seção, a partir disso é possível compreender as técnicas aplicadas em cada problema e os resultados.

Tabela 3.1: Trabalhos Relacionados

Autor	Extração de características	Técnica	Resultado
Thomas <i>et al.</i> [47]	Automática	CNN e RF	Precisão: 97.9%
Winkler <i>et al.</i> [48]	Automática	CNN	Precisão: 93.3%
Oliveira <i>et al.</i> [49]	Manual	OPF	Acurácia: 92.3%
Majumder <i>et al.</i> [51]	Automática	BNN	Precisão: 98%
Bansal <i>et al.</i> [52]	Automática	CNN, RF e KNN	Precisão: 81.5%
Maia <i>et al.</i> [54]	Automática	CNN e RF	Precisão: 92.5%

Capítulo 4

Metodologia

O propósito deste trabalho consiste em comparar o desempenho de dois modelos de aprendizado de máquina na tarefa de extração de características de melanomas: um modelo baseado em floresta aleatória e um modelo híbrido. Nos experimentos, é realizada a extração automática de características das imagens por meio de uma CNN combinada com um modelo de aprendizado de máquina (floresta aleatória) para a classificação de melanomas e somente a extração manual com características de cor e textura.

Diversos experimentos foram realizados com o objetivo de avaliar qual modelo apresenta o melhor desempenho para a tarefa de classificação, desta maneira foram testados sete diferentes tipos de abordagens envolvendo modelos de aprendizado de máquina, com o objetivo de encontrar uma solução eficiente para a detecção precoce de melanomas.

Os principais pilares deste trabalho se baseiam em avaliar como os modelos desempenhariam o papel com características extraídas previamente em relação aos modelos que fazem essa extração de maneira automática, com o uso de aumento de dados e sem aumento de dados, com modelos separados e também unificados. Desta forma, ao final desses testes, será verificado como as métricas de desempenho se comportam entre cada modelo, identificando se existe semelhança, se um modelo apresenta desempenho superior ou inferior em relação ao outro.

A Figura 4.1 descreve-se o processo experimental utilizado para a extração manual das bases de processamento e fases do método proposto. Para aumentar os dados, utilizou-se a técnica de rotação, seguida da extração de características por meio de uma floresta aleatória e da classificação. No segundo caso, conforme representado na Figura 4.2, empregou-se um modelo híbrido que combinou a extração de características por meio de uma CNN, utilizando a arquitetura VGG16, com a classificação por meio de uma floresta aleatória, aplicados no diagnóstico de melanomas. Para a segunda comparação, utilizou-se a técnica de transferência de conhecimento a partir do conjunto de dados *ImageNet*, combinada com a extração automática de características por meio de uma CNN. Essa abordagem tem sido

amplamente adotada na literatura devido aos bons resultados que tem demonstrado [57].

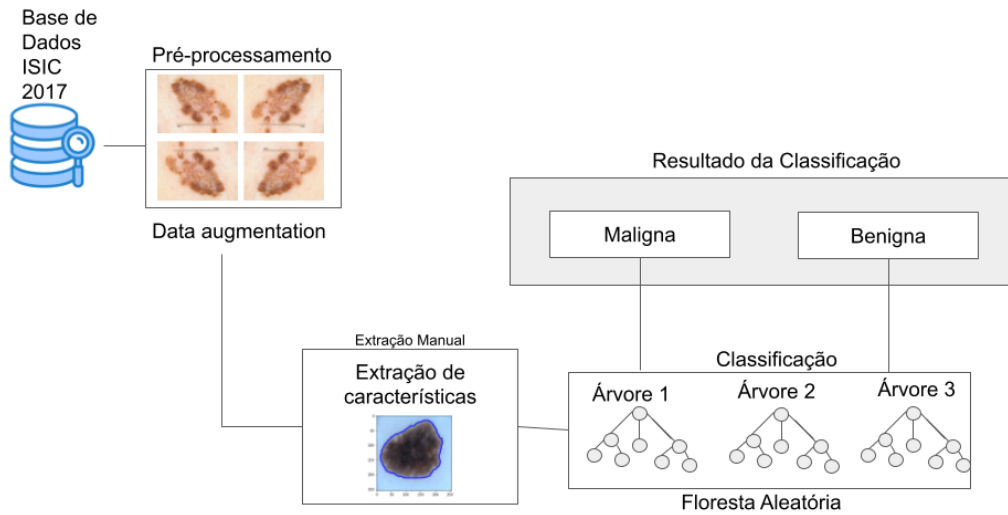


Figura 4.1: Método Usando Somente Floresta Aleatória
Fonte: Autoria própria.

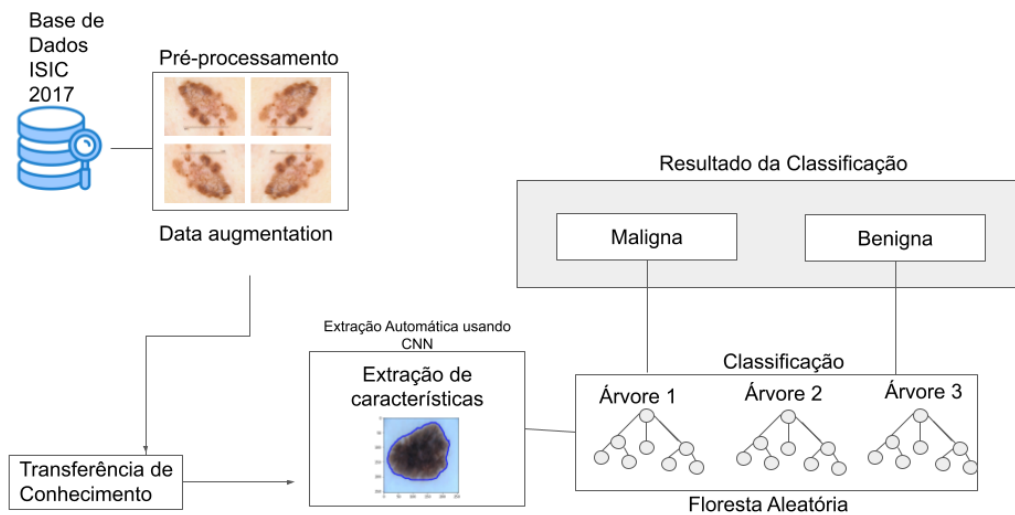


Figura 4.2: Classificação Com Modelo Híbrido CNN + Floresta Aleatória
Fonte: Autoria própria.

4.1 Base de Dados

A base de dados consumida é a do desafio *International Skin Imaging Collaboration* (ISIC) 2017¹, no qual o desafio fornece 2750 imagens, sendo elas divididas em treino (2000 imagens), teste (600 imagens), validação (150 imagens).

O ISIC foi desenvolvido por estudiosos e empresários para viabilizar o uso de imagens da pele com melanoma, tendo em vista que quando identificado no início o problema pode ser prontamente resolvido. O desafio ISIC visa ajudar a reduzir a mortalidade por melanoma através da instrução visual tanto da população quanto da comunidade médica.

O desafio ISIC de 2017 foi uma competição de aprendizado de máquina com o objetivo de classificar lesões de pele em diferentes categorias, incluindo melanoma, nevus e ceratose seborreica. Dentre o conjunto de imagens de treinamento possuem 374 com melanomas, e as demais são consideradas benignas (1626) veja na Tabela 4.1.

Todas as imagens estão com formato .JPEG, o que facilitou o uso, uma vez que não é preciso fazer conversão no formato da imagem. Também foram fornecidos pela base de dados ISIC o *Ground Truth* de todas as imagens, termo usado para se referir a verdadeira classificação ou rótulo de uma amostra de dados.

Tabela 4.1: Tabela a a quantidade de imagens utilizadas para treinamento

Classes	Base ISIC 2017	Ground Truth
Lesão Benigna	1626	1626
Melanomas	374	374
Total	2000	2000

Ground Truth da segmentação é usado como parâmetro para mensurar a eficácia de um modelo de aprendizado de máquina. Por exemplo, no desafio ISIC 2017 mencionado anteriormente, as imagens de lesões de pele foram previamente rotuladas por especialistas em dermatologia. O objetivo do desafio era desenvolver um modelo de aprendizado de máquina capaz de prever esses rótulos com alta precisão, para qual é possível observar na imagem da Figura 4.3.

Os dados conhecidos como *Ground Truth* da segmentação são normalmente coletados por profissionais humanos com experiência em diagnósticos de doenças de pele, como dermatologistas ou outros profissionais da área da saúde. No entanto, é possível obter esses dados também por meio de instrumentos de medição mais precisos, como sensores eletrônicos, equipamentos de medição avançados e outros tipos de tecnologias.

¹<https://challenge.isic-archive.com/data/#2017>

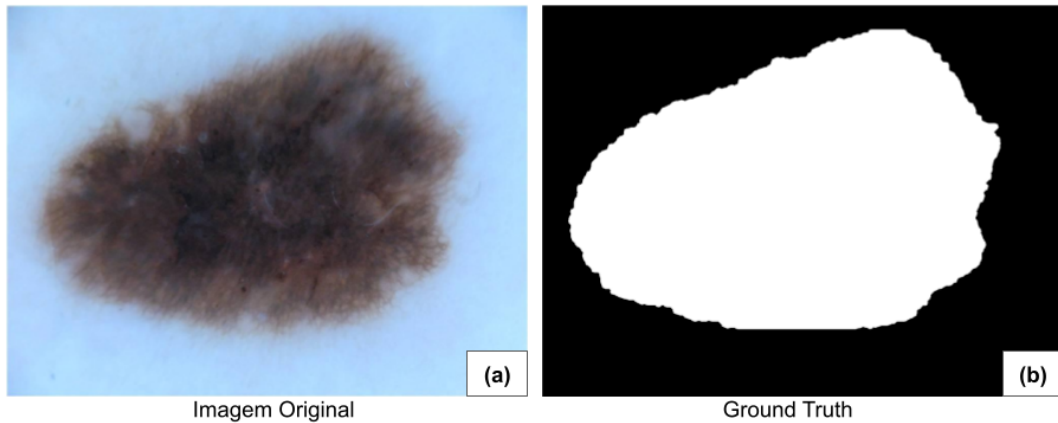


Figura 4.3: Exemplo de imagem da base de dados: (b) *Ground Truth* da segmentação (imagem binarizada) de (a) uma lesão não melanoma. Fonte: ISIC 2017 [1].

4.2 Pré-processamento das Imagens

As imagens da base de treinamento fornecidas pela ISIC 2017 apresentam um desbalanceamento significativo, com apenas 18,7% da base sendo composta por imagens de melanoma e 81,3% por imagens benignas. Para lidar com esse problema, uma estratégia adotada foi aumentar a base de dados por meio da geração de novas imagens. Foram criadas 1496 imagens adicionais de melanoma e 346 imagens adicionais de lesões benignas, simplesmente rotacionando as imagens originais em 40 graus. Com essa adição, a base de dados final agora contém 1870 imagens de melanoma e 1972 imagens de lesões benignas, totalizando 3842 imagens, foram criadas mais imagens a partir da classe de não melanoma para equiparar as classes. Os exemplos dessas imagens estão ilustrados na Figura 4.4.

A base de treino original foi ampliada através da geração de quatro imagens dermatoscópicas adicionais. Além das novas imagens geradas, as imagens originais também são utilizadas na base de treino. Essa ampliação resultou em um aumento de 1842 imagens na base de treino, além das imagens já existentes e ajudou a equilibrar a classe de melanoma, garantindo uma distribuição mais justa de exemplos para o treinamento do modelo de inteligência artificial.

A base de validação e teste apresenta desequilíbrio de dados, com a base de validação contendo 30 casos de melanoma para 120 casos de lesões benignas, representando apenas 20% da classe minoritária em comparação com 80% da classe majoritária. Já na base de teste, há 117 casos de melanoma e 483 casos de lesões benignas, o que significa 19.5% da classe minoritária e 81.5% da classe majoritária. Portanto, é importante ter em mente que essa distribuição desigual de dados pode impactar a eficiência e precisão dos modelos de classificação.

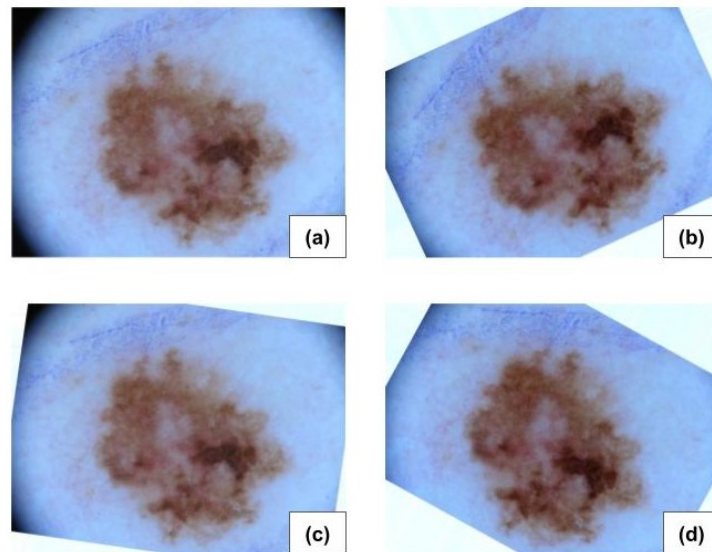


Figura 4.4: Exemplos de Imagens criadas após o pré-processamento.
Fonte: Autoria própria.

4.3 Extração de características

A extração de características é uma parte fundamental do processo de diagnóstico do melanoma. Esta técnica compreende a utilização de descritores de cor e textura para identificar as propriedades específicas da lesão. Entre os espaços de cores utilizados, estão o RGB, CIELUV (LUV), HSV, CIELAB (LAB), que são separados em 3 canais de cores, cada um com sua própria função específica, bem como a escala de cinza. A Figura 4.5 mostra os canais de cores para exemplificação.

Para extrair a cor da lesão, as imagens que correspondem aos *Ground Truths* da segmentação são utilizadas para delimitar a área da lesão na imagem. Em seguida, são extraídas medidas separadas para cada canal de cor, incluindo a média, o mínimo, o máximo e o desvio padrão da quantidade de cada canal de cor presente nas imagens de treinamento, validação e teste, baseada em Oliveira et al. [49] que totalizam 39 características. Além disso, são extraídas características de textura, baseadas em dimensão

fractal [58] e a matriz de coocorrência (Contraste, Entropia, Variância, Homogeneidade) [59], A dimensão fractal utilizada é a *Box-counting* que é uma das medidas mais comuns de dimensão fractal. Ela mede a complexidade da textura pela contagem do número de caixas necessárias para cobrir a imagem.

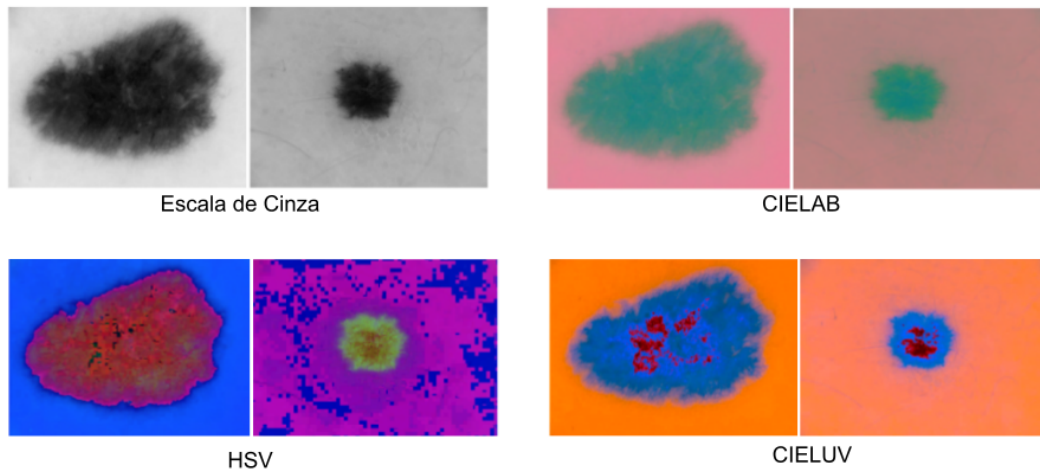


Figura 4.5: Canais de cores aplicados nas figuras de lesões de pele
Fonte: Autoria própria.

4.4 Aplicação de *Transfer Learning* e Análise de Hiperparâmetros

Conforme mencionado nos capítulos anteriores, é possível entender que a transferência de conhecimento ajuda a melhorar os resultados dos modelos de forma considerável, neste trabalho foram usadas as seguintes técnicas CNN, VGG16 por apresentar métricas altas de precisão, é incluído o parâmetro pré-treinada a partir da base de *ImageNet* para aprimoramento, os métodos são extraídos da biblioteca *keras*² do *tensorflow*³. Conta atualmente com 14 197 122 imagens e inúmeros rótulos para cada uma delas, esses rótulos também são conhecidos por conjunto de sinônimos, isso facilita o uso dos dados.

A base de dados ISIC 2017 é relativamente pequena e a transferência de conhecimento é uma abordagem útil para lidar com a falta de dados, a transferência de conhecimento é particularmente útil para o treinamento da rede VGG16.

²<https://keras.io/about/>

³<https://www.tensorflow.org/resources/libraries-extensions?hl=pt-br>

Os experimentos foram significativamente melhorados através do uso de hiperparâmetros, que permitiram tomar decisões informadas sobre quais seriam os melhores valores para os parâmetros utilizados no processo de aprendizado. Para alcançar esse objetivo, foram testadas uma ampla variedade de combinações de valores padronizados, visando identificar a melhor configuração para o processo de aprendizado.

Para modelos que usam a Floresta Aleatória foram utilizados o *GridSearchCV* que pertence a biblioteca *sklearn*, já no modelo onde possui a VGG16 sem utilização de outro método em conjunto é utilizado o método do otimizador Adam com um *learning rate* de 0.00001 combinado com o método *Binary Cross Entropy* e 20 épocas para treinamento.

4.5 Experimentos com extração de características manual

Com a finalidade de melhorar os resultados, os experimentos abordados envolveram a extração manual de dados, que resultou na criação de um arquivo detalhado com inúmeras linhas de informação. Cada linha deste arquivo representa um valor distinto para uma determinada característica, incluindo: (1) cor, como média, mínimo, máximo e desvio padrão para cada canal de cor, (2) textura, como a análise de coocorrência de níveis de cinza e a análise da dimensão fractal), e (3) a combinação desses dois aspectos mencionados anteriormente, conforme mostrado na Tabela 4.2. Este processo buscou identificar a melhor combinação de características para obter resultados ainda mais precisos.

Tabela 4.2: Tabela com os descritores e características utilizadas nos experimentos

Descritores	Características Extraídas
Para cada canal RGB	cor, média, mínimo, máximo, desvio padrão
Para cada canal LUV	cor, média, mínimo, máximo, desvio padrão
Para cada canal LAB	cor, média, mínimo, máximo, desvio padrão
Para cada canal HSV	cor, média, mínimo, máximo, desvio padrão
Escala de Cinza	cor, média, mínimo, máximo, desvio padrão
Dimensão Fractal	<i>Box-counting</i>
Matriz de Coocorrência	contraste, entropia, variância, homogeneidade

Experimento I: Floresta Aleatória

O experimento um tem como objetivo principal realizar uma comparação detalhada do comportamento da técnica de Floresta Aleatória, descrita na Seção 2.3. Para isso, serão avaliados diferentes cenários de uso da técnica, incluindo o uso da floresta aleatória sozinha, juntamente com todas as características reunidas, bem como com cada característica

avaliada separadamente. Além disso, serão testados os métodos utilizando otimizadores de métricas, com o objetivo de otimizar e melhorar os resultados obtidos, na Figura 4.6 é demonstrado como foi a etapa de treinamento.

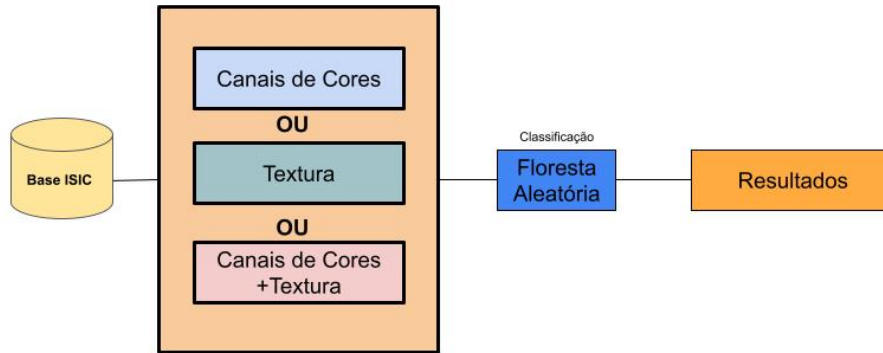


Figura 4.6: Método Sem Aumento de Dados
Fonte: Autoria própria.

Experimento II: Floresta Aleatória + Aumento de Dados

No segundo experimento, serão realizados os testes de comportamento a partir do aumento de dados, visando a comparação com os resultados obtidos sem o aumento de dados. Estes testes serão realizados de acordo com o Experimento I descrito na Seção 4.5, conforme ilustrado na Figura 4.7.

Experimento III: Floresta Aleatória + Aumento de Dados + Hiperparâmetros

No Experimento III será executado os comportamentos a partir do aumento de dados para comparação com os resultados sem o aumento de dados, executando os mesmos testes do Experimento 1 da Seção 4.5, como visto na Figura 4.7. Com o aumento de bases e o uso de hiperparâmetros compõe a tentativa de um aprimoramento dos resultados em relação ao teste anterior conforme Figura 4.8.

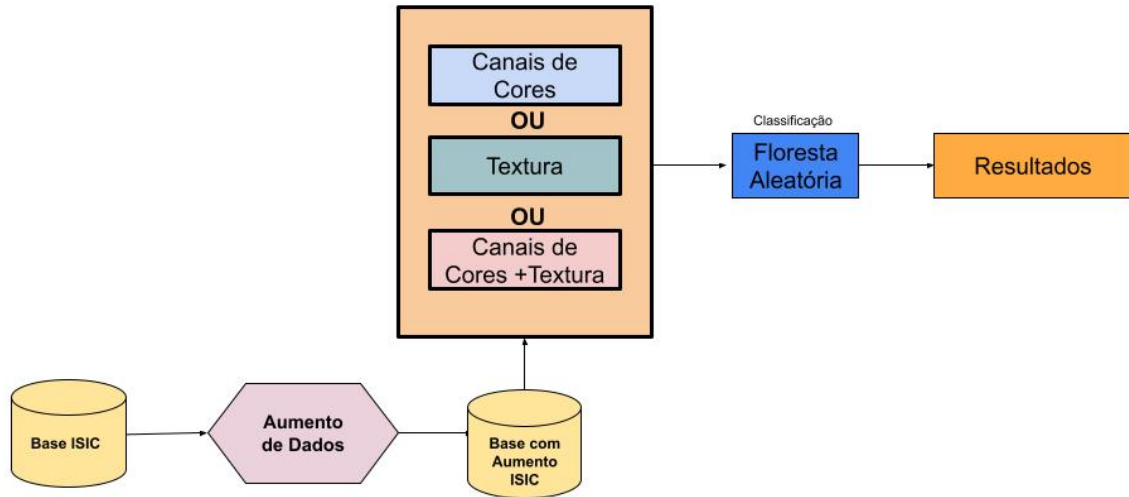


Figura 4.7: Etapas da Floresta Aleatória com Aumento de Dados
 Fonte: Autoria própria.

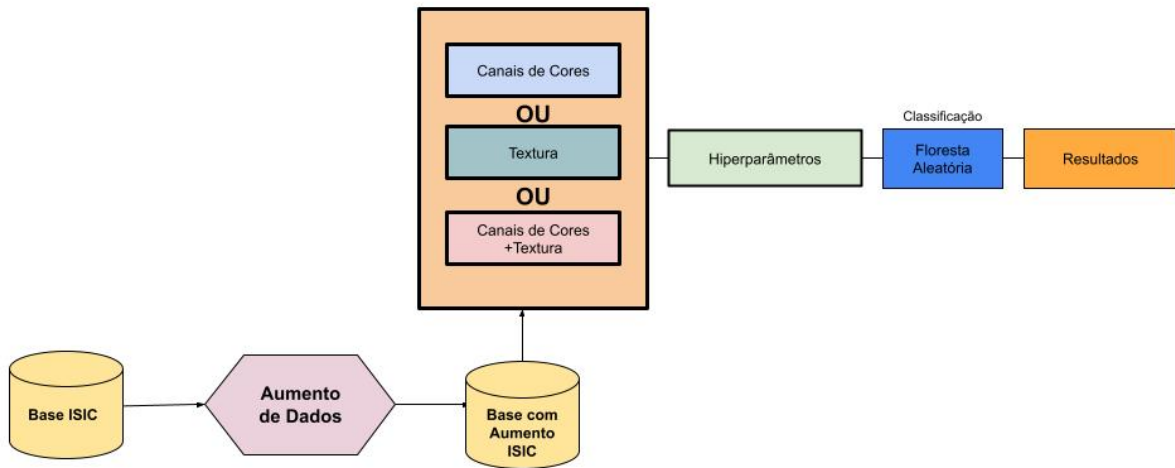


Figura 4.8: Métodos com Aumento de Dados mais Hiperparâmetros
 Fonte: Autoria própria.

4.6 Experimentos com extração de características automática

Nesta Seção, o objetivo é avaliar a eficiência da extração de características automáticas de uma forma abrangente. Para isso, é feito o uso da rede VGG16 como base e, em seguida, combiná-la com técnicas de aumento de dados e utilizar a Floresta Aleatória como classificador. A combinação dessas técnicas permitirá avaliar o quão bem a extração de características é realizada e como isso afeta o desempenho do classificador. É importante destacar que a avaliação da eficiência da extração de características é fundamental para o sucesso do processo de classificação.

Experimento IV: VGG16 Sem Aumento dos Dados

No experimento quatro, foi considerada a VGG16 em todas as etapas do processo, desde a classificação de características até a extração, sendo passada somente os conjuntos de treinamento e associada a *ImageNet*. Foi incluído o otimizador *Adam*, para um resultado mais performático, como forma de teste e entendimento do comportamento da VGG16, nesse caso são usados os dados desbalanceados conforme a Figura 4.9.

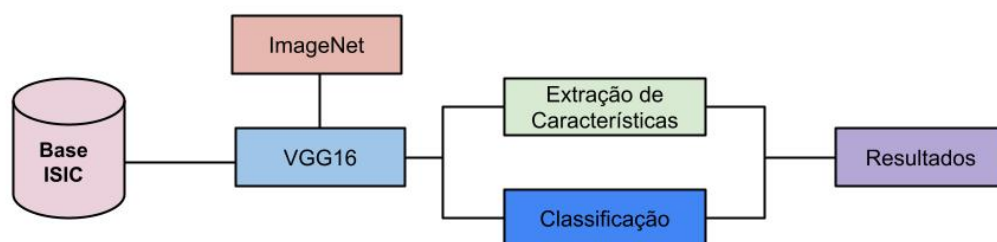


Figura 4.9: VGG16 Sem Aumento de Dados
Fonte: Autoria própria.

Experimento V: VGG16 com Aumento dos Dados

No experimento cinco, foi considerada a VGG16 em todas as etapas do processo Figura 4.10, sendo passada somente os conjuntos de treinamento e associada a *ImageNet*. Foi incluído otimizador *Adam*, para um resultado mais performático. Como forma de teste e entendimento do comportamento da VGG16 usou-se a base com aumento de dados sem a criação de novas imagens.

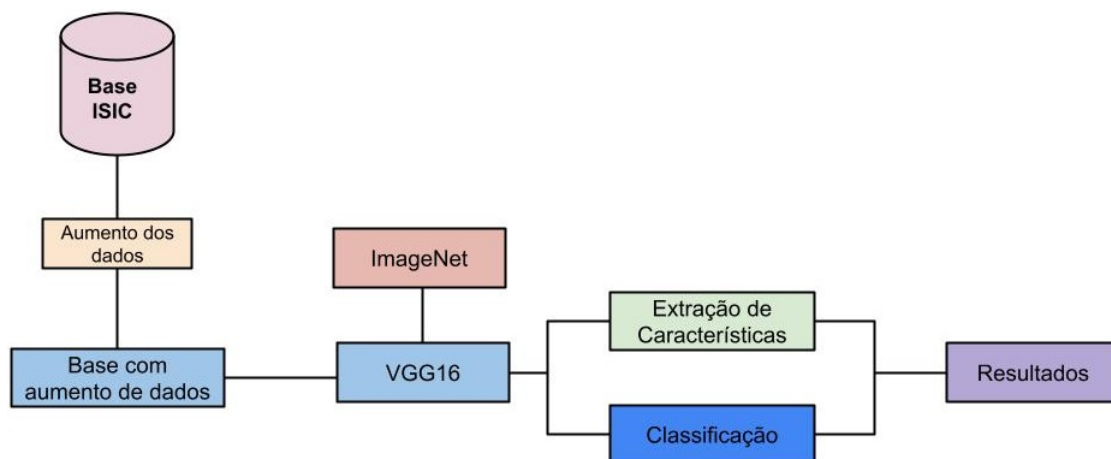


Figura 4.10: VGG com aumento de dados
Fonte: Autoria própria.

Experimento VI: VGG16 + Floresta Aleatória

Como o modelo híbrido pode melhorar as métricas, no experimento seis foi utilizado um modelo híbrido onde a CNN possui o papel de extrair as características, e a Floresta Aleatória possui o papel de classificar com a finalidade de otimizar as métricas anteriores e trazer resultados Figura 4.11. Dessa maneira os testes feitos foram baseados em usar a base balanceada e desbalanceada, além de usar hiperparâmetros. Foram testados inúmeros cenários na qual a Floresta Aleatória poderia ter o melhor resultado.

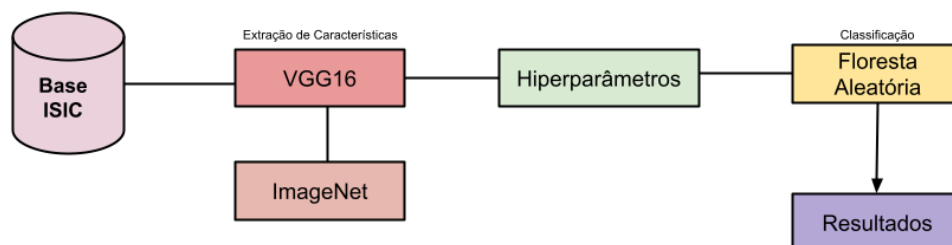


Figura 4.11: VGG16 + Floresta Aleatória
Fonte: Autoria própria.

Experimento VII: VGG16 + Floresta Aleatória + Aumento dos dados

O experimento sete tem como o bjetivo aprimorar o uso da VGG16 ao treiná-la com um conjunto maior de imagens, como descrito na Seção 2.6. Além disso, o experimento visa aprender com as características das imagens do conjunto minoritário para equilibrar a base de dados pré-processada e, assim, melhorar os resultados das análises, como na Figura 4.12. Acredita-se que esse experimento seja o mais promissor em termos de métricas, já que utiliza um modelo híbrido.

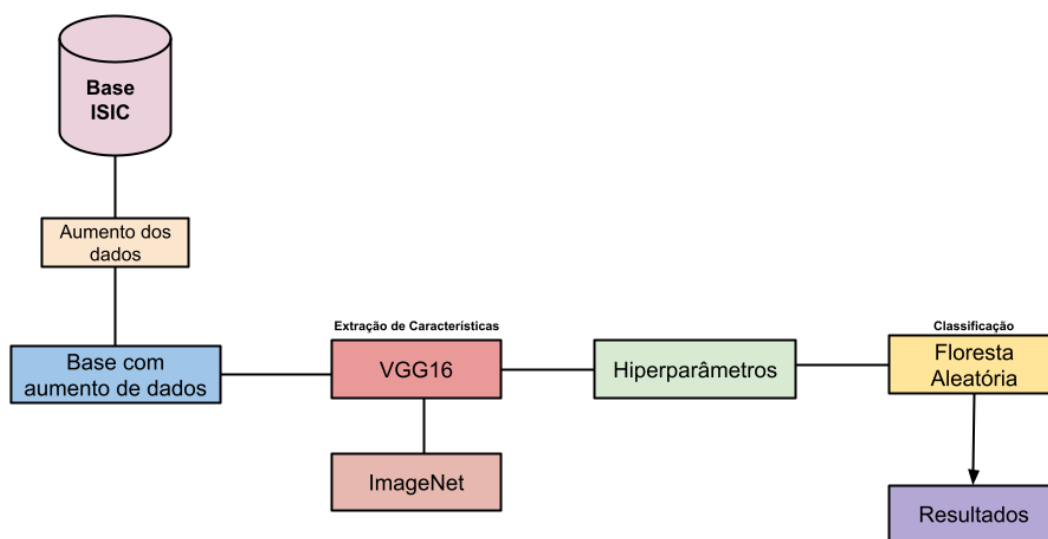


Figura 4.12: VGG com Floresta Aleatória com a base balanceada
Fonte: Autoria própria.

Capítulo 5

Resultados

Este capítulo visa demonstrar os resultados dos experimentos abordados com apresentação das métricas obtidas e encontrar a melhor combinação de resultados para classificação de melanomas, métricas como acurácia, F1-score, precisão, revocação são analisadas, abordados na Seção 2.7. A distribuição de imagens utilizadas para treinamento, validação e teste foram as mesmas disponibilizadas pelo ISIC, conforme mencionado na Seção 4.1.

5.1 Ambiente Experimental

Os experimentos foram executados em uma máquina com processador i5-11400H 2.70GHz com 16 Gb de *RAM* e disco rígido de 500 Gb, utilizando o sistema operacional Windows 10. A linguagem de programação utilizada foi *Python* na versão 3.9.7, com as bibliotecas *os* [60], *NumPy* [61], *Pandas* [62], *Scikit-learn* [63], *Math* [64], *TensorFlow* [65], *Keras*, *Matplotlib* [66].

5.2 Experimento I: Floresta Aleatória

Como mencionado anteriormente na Seção 4.5, este experimento contém o objetivo de comparar os resultados da classificação das imagens dermatoscópicas a partir dos descritores de cor e textura separadamente e a combinação das mesmas, utilizando Floresta Aleatória para treinamento da base e compreender se a Floresta Aleatória sozinha performa bem. Na arquitetura da Floresta Aleatória são aplicados os métodos padrões da biblioteca *sklearn*¹, no qual o parâmetro padrão da biblioteca é 100 árvores. Como apresentado na Tabela 5.1 os resultados dos experimentos, as métricas de acurácia, precisão revocação F1-Score estão bem próximas.

¹<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

Tabela 5.1: Resultados da Floresta Aleatória sem aumento dos dados

Descritores	Acurácia	Precisão	Revocação	F1-Score
Canal de cor	0.795	0.711	0.795	0.726
Textura	0.606	0.654	0.606	0.628
Canal de Cor + Textura	0.753	0.714	0.753	0.730

Os resultados na Tabela 5.1 indicam que a classificação com características baseadas na cor apresentou a melhor acurácia, atingindo 0.795. Já os demais experimentos apresentaram valores de 0.611 e 0.753. Quando analisado a métrica de precisão, a classificação com características baseadas na cor se destacou, com uma precisão de 0.654. As demais classificações com características baseadas na cor experimentos apresentaram valores de precisão de 0.711 e 0.714. Já no caso do F1-Score e Revocação apontaram que o experimento que combina canais de cores e texturas obteve melhor o desempenho, com resultado de 0.730 e 0.795 respectivamente.

Embora tenham sido observados resultados semelhantes, conforme mencionado na Seção 2.7, é importante considerar que a métrica que classifica falsos negativos é a revocação, enquanto o *F1-Score* é a média harmônica entre revocação e precisão. Dessa forma, a situação em que o paciente não sabe que possui câncer é considerada de alta complexidade. Neste experimento, o descritor de canal de cor demonstrou um desempenho superior, o que pode ser atribuído à sua capacidade de capturar informações relevantes das imagens, mesmo em situações de maior dificuldade.

5.3 Experimento II: Floresta Aleatória + Aumento de Dados

Neste experimento foram aplicadas os mesmos métodos do Experimento I da Seção 4.5, entretanto, foram incluídas 1892 imagens além das disponibilizadas pelo ISIC. Com isso é possível comparar o desempenho de uma Floresta Aleatória com classes mais balanceada.

Tabela 5.2: Resultados da Floresta Aleatória com aumento dos dados

Descritores	Acurácia	Precisão	Revocação	F1-Score
Canal de cor	0.751	0.709	0.752	0.726
Textura	0.703	0.747	0.700	0.721
Canal de Cor + Textura	0.706	0.652	0.706	0.628

Os resultados na Tabela 5.2 indicam que o experimento com canais de cores apresentou a melhor acurácia, atingindo 0.751. Já os demais experimentos apresentaram valores de 0.703 e 0.706, sendo que 1.0 é o maior valor possível. Quando analisado a métrica de

precisão, o experimento que utiliza textura se destacou, com uma precisão de 0.747. Os demais experimentos apresentaram valores de 0.709 e 0.652.

A métrica de revocação mostrou que o experimento com canais de cores apresentou o melhor desempenho, com resultado total de 0.752 e no caso do F1-Score apontou que o experimento canais de cores obteve a melhor desempenho, com resultado de 0.726. No geral, o descritor com somente o uso do canal de cor se saiu melhor, possuindo uma acurácia, revocação e *F1-Score* mais alto.

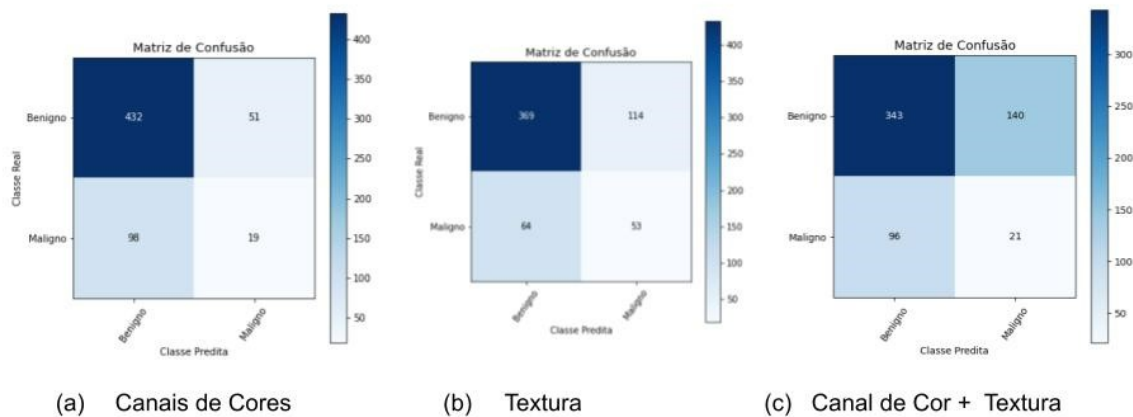


Figura 5.1: Matrizes de Confusão do Experimento III
Fonte: Autoria própria.

5.4 Experimento III: Floresta Aleatória + Aumento de Dados + otimização dos hiperparâmetros

Neste experimento, diferentemente dos anteriores, não utiliza-se os itens padrão da biblioteca. Em vez disso, é realizada uma busca pelos melhores parâmetros através do *GridSearch*², propondo diversas combinações e valores antes do treinamento.

Tabela 5.3: Tabela com Resultados da Floresta Aleatória com aumento dos dados.

Descritores	Acurácia	Precisão	Revocação	F1-Score
Canal de cor	0.723	0.737	0.723	0.729
Textura	0.763	0.689	0.763	0.717
Canal de Cor + Textura	0.703	0.747	0.700	0.721

Os resultados mostram que o experimento usando a textura teve o melhor desempenho, com uma acurácia de 0.763. Já os experimentos usando apenas canais de cor e

²https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

combinando canais de cor e textura apresentaram resultados mais baixos, com acurácias de 0.723 e 0.703, respectivamente. No quesito precisão, os canais de cor tiveram o melhor resultado, com 0.747. Quando se considera a revocação, o experimento com textura teve um resultado ligeiramente melhor, superando o experimento com canais de cor em 0.011. Por fim, em relação ao *F1-Score*, os canais de cor tiveram o melhor desempenho.

No caso apresentado, o experimento com textura apresentou a melhor acurácia e revocação, enquanto o experimento com canais de cor teve o melhor desempenho em precisão e *F1-Score*, desta maneira seguindo a lógica da Seção 4.5 no experimento 1, para este caso a textura ficou com uma revocação e a acurácia melhor, por isso é possível entender que o falso negativo para o caso de melanoma, teria maior peso, por isso para o experimento três a textura tem o melhor resultado.

5.5 Experimento IV: VGG16 Sem Aumento dos Dados

O objetivo do experimento IV visa entender o comportamento de uma CNN VGG16 atuando como modelo único, isto é, desempenhando os papéis de extrair as características das figuras e classificá-las. O modelo VGG16 foi treinado com 20 épocas e com taxa de aprendizado de 0.00001, nas quais foram aplicadas 200 passos de treino e 50 para validação, dos quais foram testados diversas vezes para para entendimento de qual seria a melhor combinação de época e taxa de aprendizado.

Os valores apresentados na Figura 5.2 mostram que o modelo tem uma acurácia de 0.939 durante o treinamento e de 0.815 na validação.

5.6 Experimento V: VGG16 com Aumento dos Dados

O Experimento V possui os mesmos parâmetros apresentados pelo Experimento IV da Seção 4.5, entretanto é acrescido com as imagens do *data augmentation* e com 384 etapas para treinamento e 96 etapas para teste, na qual foram medidas através de perda e acurácia na validação todas apresentaram resultados semelhantes, a perda acontece de maneira estável.

As métricas apresentadas na Figura 5.3 mostram o desempenho de um modelo de classificação equilibrado. A acurácia treinada varia de 0.509 a 0.511, enquanto a acurácia de validação é de 0.527. A perda no treinamento varia de 0.072 a 0.078 e a perda de validação é de 0.071.

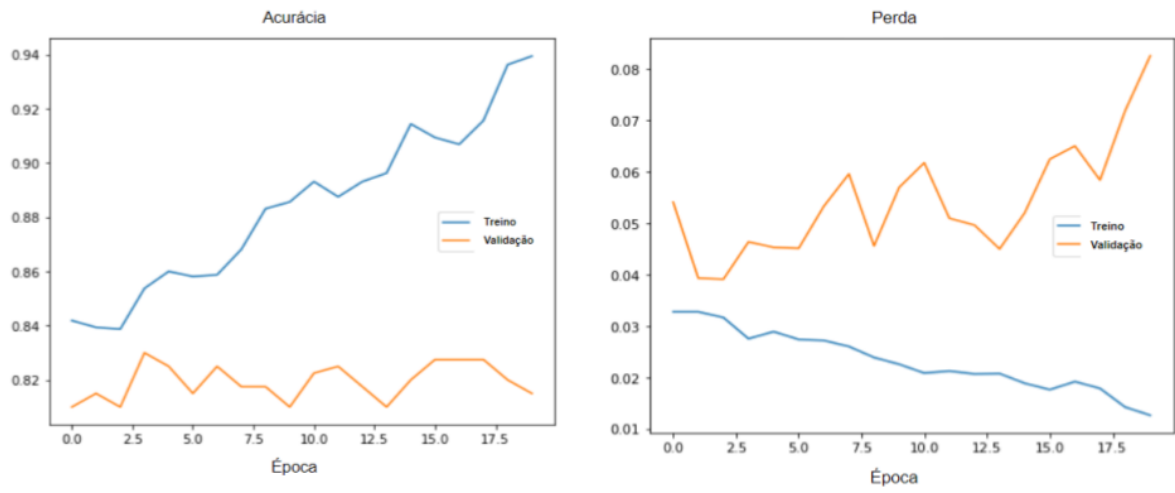


Figura 5.2: Resultados do modelo VGG16 sem aumento de dados
 Fonte: Autoria própria.

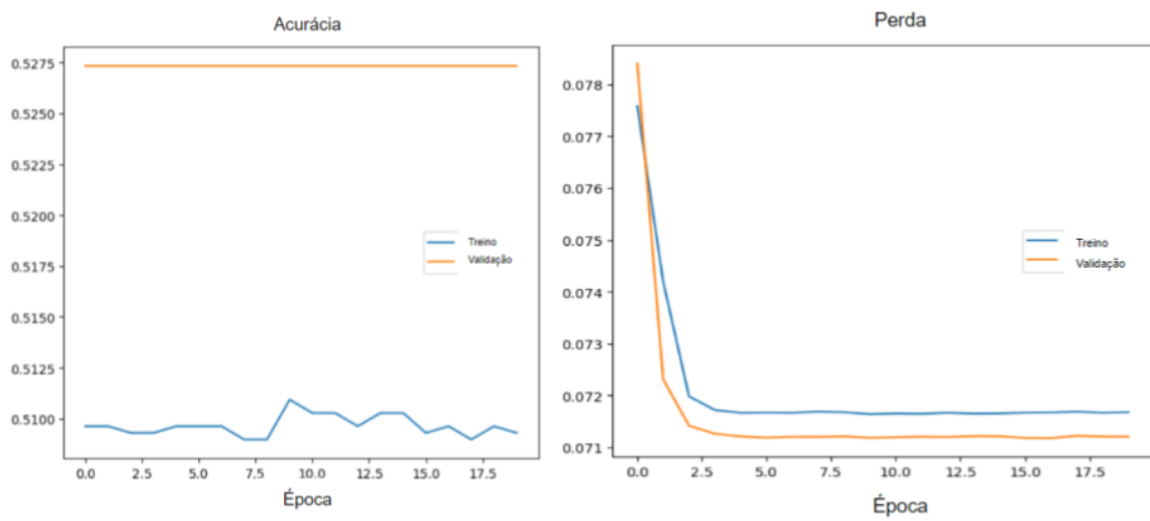


Figura 5.3: Resultados do modelo VGG16 com aumento de dados
 Fonte: Autoria própria.

5.7 Experimento VI: VGG16 + Floresta Aleatória

O Experimento VI é baseado em uma arquitetura híbrida, em que é proposto uma mistura das arquiteturas anteriores sendo VGG16 para extração das características e Floresta Aleatória para classificação, com a finalidade de melhorar as métricas e a assertividade na detecção de presença de melanoma, esse método se mostra bem interessante, principalmente ao usar desbalanceado teve um desempenho interessante, na qual acertou de 79 lesões malignas e 318 benignas.

Tabela 5.4: Tabela com Resultados da CNN

Descritores	Acurácia	Precisão	Revocação	F1-Score
Dados Desbalanceados	0.685	0.789	0.685	0.725

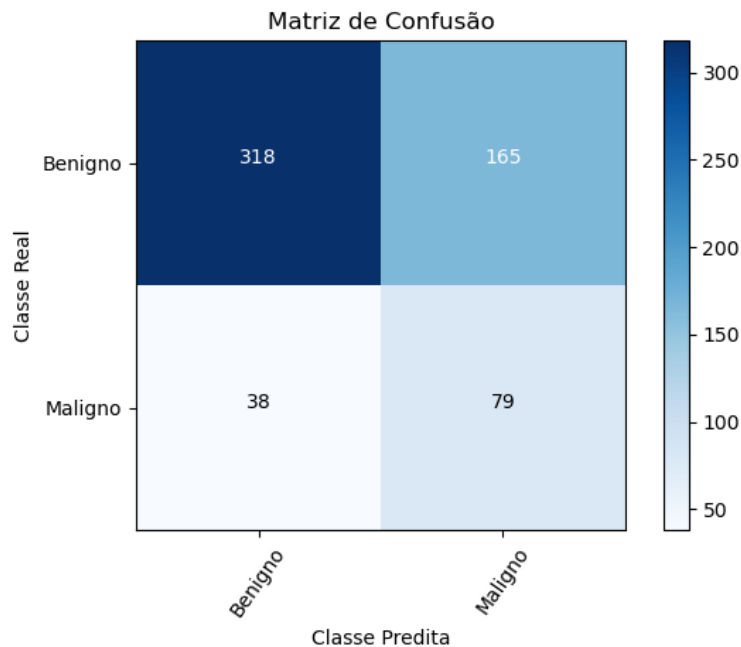


Figura 5.4: Matriz de Confusão do Experimento VI

Fonte: Autoria própria.

As métricas indicadas pela Tabela 5.4 mostram que o modelo tem uma acurácia moderada. No entanto, a precisão do modelo é relativamente alta 0.789, o que significa que ele é bom em identificar verdadeiros positivos, conforme explicado na Seção 2.7. A revocação mostrou-se baixa, 0.685, o que significa que ele não é muito bom em encontrar todos os positivos. O *F1-Score* mostrou-se baixa também, logo é (0.725) um modelo que não é tão bom quanto se esperaria. Em resumo, essas métricas indicam que o modelo apresenta desempenho limitado, particularmente em relação à revocação, devido ao desequilíbrio de classes.

5.8 Experimento VII: VGG16 + Floresta Aleatória + Aumento dos dados

O Experimento VII utiliza a mesma arquitetura do Experimento VI da Seção 4.6, ou seja, a VGG16 + Floresta Aleatória. No entanto, há uma mudança importante na base de dados, que agora está mais equilibrada e calibrada, com uma quantidade semelhante de imagens para ambas as classes (melanoma e lesões benignas). Essa mudança tem como objetivo aumentar a precisão, acurácia, revocação e *F1-Score* do modelo, buscando uma maior probabilidade de acertos.

Tabela 5.5: Tabela com Resultados da CNN

Descritores	Acurácia	Precisão	Revocação	F1-Score
Dados Balanceados	0.751	0.709	0.751	0.726

Essas métricas indicadas pela Tabela 5.5 demonstra que o modelo tem uma boa acurácia (0.751), o que significa que ele acerta 75,1% das vezes. A precisão do modelo é de 0.709, o que significa que ele é capaz de identificar verdadeiros positivos com bastante precisão. A revocação é também boa (0.751), o que significa que o modelo é capaz de encontrar a maioria dos positivos. O *F1-Score* (0.726) é uma métrica que combina precisão e revocação e indica que o modelo tem um desempenho aceitável. Em resumo, essas métricas indicam que o modelo tem um desempenho satisfatório, particularmente porque o desequilíbrio de classes foi corrigido, conforme demonstrado na Figura 5.5

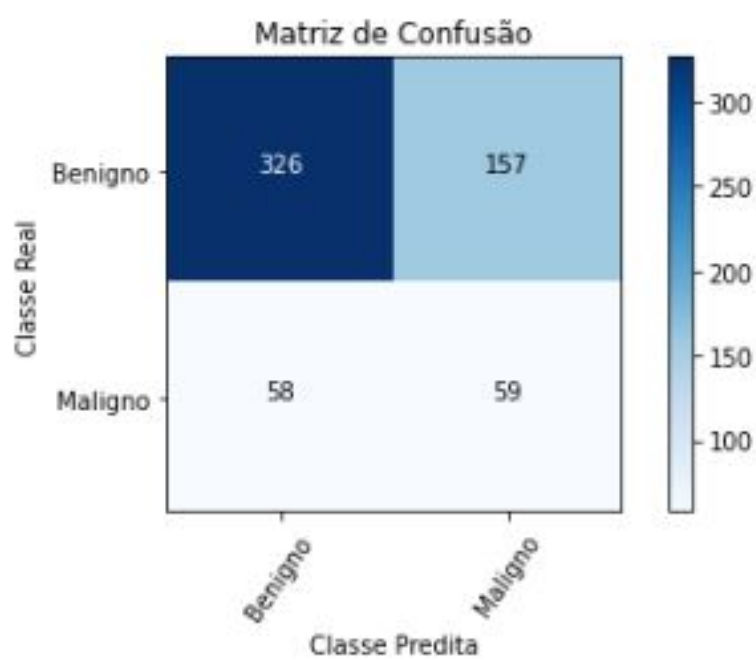


Figura 5.5: Matriz de Confusão do Experimento VII
Fonte: Autoria própria.

Capítulo 6

Discussão dos Resultados

Mediante aos resultados obtidos na Seção 5, o Experimento I, II e III, houve uso dos descritores de cor e textura separadamente e a combinação desses descritores, utilizando Floresta Aleatória para treinamento, validação e teste da base. Resultados da Floresta Aleatória sem aumento de dados, onde a classificação com características baseadas na cor apresentou a melhor acurácia, com um valor de 0.795, e a melhor precisão, com um valor de 0.654.

Em geral, os resultados indicam que a classificação com características baseadas na cor se destaca em termos de acurácia e precisão, entretanto no caso de um melanoma considera-se que o fato de uma pessoa não saber acaba sendo pior do que ela achar que tem e na verdade não possui um melanoma, isto porque em casos iniciais são removidos com cirurgia, já em casos de falsos positivos, onde a pessoa não sabe que tem e ela tem, o melanoma pode avançar, desta maneira uma métrica que é levada em consideração é a revocação e acurácia, na qual no experimento I é a melhor.

O Experimento II com canal de cor apresentou a melhor acurácia, com um valor de 0.751, e a melhor métrica de revocação, com um valor de 0.752. Já o experimento com textura apresentou a melhor precisão, com um valor de 0.747. Além disso, o experimento com canal também apresentou o melhor desempenho na métrica F1-Score, com um resultado de 0.726.

Já no Experimento III, o descritor de textura apresentou o melhor desempenho, com uma acurácia de 0.763. No entanto, quando se considera a precisão, o experimento com canais de cor teve o melhor resultado, com 0.747. Além disso, o experimento com textura teve uma melhor revocação, enquanto o experimento com canais de cor teve o melhor desempenho em relação ao F1-Score. Dessa forma, os melhores resultados foram considerados os que possuíam uma maior revocação, conforme a Tabela 6.1

Tabela 6.1: Tabela com Resultados da CNN

Experimento	Acurácia	Precisão	Revocação	F1-Score
Experimento I - Canal de Cor	0.795	0.711	0.795	0.726
Experimento II - Canal de Cor	0.751	0.709	0.752	0.726
Experimento III - Textura	0.763	0.689	0.763	0.717

Os experimentos IV e V utilizaram a arquitetura VGG16 para classificar as imagens. Surpreendentemente, a classificação a partir dos dados sem aumento obteve um resultado superior em comparação com os dados aumentados. Mais especificamente, a VGG16 sem aumento alcançou uma acurácia de 0.939, enquanto a VGG16 com dados aumentados obteve uma acurácia de apenas 0.509.

Nos casos dos experimentos VI e VII, temos o uso dos dados com e sem aumento, partindo de uma VGG16 para extração das características e a Floresta Aleatória para classificação, conforme é possível observar na Tabela 6.2.

Tabela 6.2: Tabela com Resultados da CNN

Experimento	Acurácia	Precisão	Revocação	F1-Score
Sem Aumento	0.685	0.789	0.685	0.725
Com Aumento	0.751	0.709	0.751	0.726

No Experimento VII a acurácia, revocação e *F1-Score* maior que a do Experimento VI. Entretanto o Experimento VII acerta mais a classe benigna do que a maligna de fato. Nesse caso possuir um modelo desequilibrado trouxe uma maior assertividade na classe com lesões malignas. Embora todos os modelos apresentassem desempenho semelhante em geral, cada um se destacou em uma métrica específica, demonstrando uma desempenho superior à dos demais.

Capítulo 7

Conclusão e Trabalhos Futuros

Neste trabalho, foram apresentadas técnicas para detecção de melanoma utilizando modelos como a VGG16 e a Floresta Aleatória, com o objetivo de aprimorar os resultados obtidos. A combinação dos modelos, com e sem aumento de imagens, apresentou resultados semelhantes, com acurácia de 0.685, precisão de 0.789, revocação de 0.685 e *F1-Score* de 0.725 para o conjunto de dados desbalanceado, e acurácia de 0.751, precisão de 0.709, revocação de 0.751 e *F1-Score* de 0.726 para o conjunto de dados balanceado. Apesar dos valores de desempenho próximos, o modelo sem o aumento de dados apresentou maior assertividade e menos falsos negativos.

Apesar dos resultados promissores apresentados, é possível incluir diversas melhorias, tais quais a inclusão de outras técnicas para criar o aumento de dados, como por exemplo aumentar a invariância à rotação e reflexões, mudando os ângulos, combinados com inversões horizontais e verticais [67].

Usar outros métodos de combinação com CNNs para avaliação de possíveis melhoras [68], além disso também o experimento de otimização dos hiperparâmetros nos testes do modelo híbrido também poderia ser incluído na tratativa de otimização do desempenho.

Referências

- [1] Codella, Noel CF, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler *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)*. Em *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, páginas 168–172. IEEE, 2018. ix, 27
- [2] Schmidt, Carolina: *O Que E Cancer?: Um Livro Para Crianças*. Carolina Witchmichen Penteadó Schmidt, 2016. 1
- [3] *Câncer de pele melanoma*. <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/pele-melanoma>. Accessed: 2023-2-4. 1
- [4] *Skin cancer information*. <https://www.skincancer.org/skin-cancer-information/>, julho 2018. Accessed: 2023-2-4. 1
- [5] Codella, Noel CF, Q B Nguyen, Sharath Pankanti, David A Gutman, Brian Helba, Allan C Halpern e John R Smith: *Deep learning ensembles for melanoma recognition in dermoscopy images*. IBM Journal of Research and Development, 61(4/5):5–1, 2017. 1
- [6] Adegun, Adekanmi A e Serestina Viriri: *Deep learning-based system for automatic melanoma detection*. IEEE Access, 8:7160–7172, 2019. 1
- [7] Tanna, Riya e Toshita Sharma: *Binary classification of melanoma skin cancer using svm and cnn*. Em *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, páginas 1–4. IEEE, 2021. 1
- [8] Gonzalez, Rafael C e Richard E Woods: *Image processing*. Digital image processing, 2(1), 2007. 4
- [9] Warren, James e Nathan Marz: *Big Data: Principles and best practices of scalable realtime data systems*. Simon and Schuster, 2015. 5
- [10] Moura, Nayara H de, Romuere R V e Silva, Flávio HD Araújo *et al.*: *Extração de atributos e classificação de lesões de pele utilizando combinação de componentes de modelos de cores*. Revista de Sistemas e Computação-RSC, 5(1), 2015. 6
- [11] Kavitha, JC e A Suruliandi: *Texture and color feature extraction for classification of melanoma using svm*. Em *2016 International conference on computing technologies and intelligent data engineering (ICCTIDE'16)*, páginas 1–6. IEEE, 2016. 6

- [12] Giorgianni, Edward J e Thomas E Madden: *Digital color management: Encoding solutions*, volume 13. John Wiley & Sons, 2009. 6
- [13] Smith, Alvy Ray: *Color gamut transform pairs*. ACM Siggraph Computer Graphics, 12(3):12–19, 1978. 6
- [14] Ohta, Noboru: *Correspondence between cielab and cieluv color differences*. Color Research & Application, 2(4):178–182, 1977. 6
- [15] *CIELUV – HiSoUR arte cultura exposição*. <https://www.hisour.com/pt/cieluv-24851/>. Accessed: 2023-2-5. 7
- [16] Arifin, M Shamsul, M Golam Kibria, Adnan Firoze, M Ashraful Amini e Hong Yan: *Dermatological disease diagnosis using color-skin images*. Em *2012 international conference on machine learning and cybernetics*, volume 5, páginas 1675–1680. IEEE, 2012. 7
- [17] Haralick, Robert M, Karthikeyan Shanmugam e Its' Hak Dinstein: *Textural features for image classification*. IEEE Transactions on systems, man, and cybernetics, (6):610–621, 1973. 8
- [18] Liu, Yanxi e Wen Chieh Lin: *Deformable texture: the irregular-regular-irregular cycle*. Carnegie Mellon University, the Robotics Institute, 2003. 8
- [19] Yuan, Xiaojing, Zhenyu Yang, George Zouridakis e Nizar Mullani: *Svm-based texture classification and application to early melanoma detection*. Em *2006 international conference of the IEEE engineering in medicine and biology society*, páginas 4775–4778. IEEE, 2006. 8
- [20] DALVI, RODOLFO DE F, PAULO G DE MORAES, JORGE LA SAMATELO e PATRICK M CIARELLI: *Análise de metodologias para detecção de câncer em mamografias digitais*. 8
- [21] George, Loay E e Esraa Z Mohammed: *Cancer tissues recognition system using box counting method and artificial neural network*. Em *2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, páginas 5–9. IEEE, 2011. 8
- [22] Coelho, André LV, Everlândio Fernandes e Katti Faceli: *Multi-objective design of hierarchical consensus functions for clustering ensembles via genetic programming*. Decision Support Systems, 51(4):794–809, 2011. 8
- [23] Hastie, Trevor, Robert Tibshirani, Jerome H Friedman e Jerome H Friedman: *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009. 9
- [24] Jayatilake, Senerath Mudalige Don Alexis Chinthaka e Gamage Upeksha Ganegoda: *Involvement of machine learning tools in healthcare decision making*. Journal of healthcare engineering, 2021, 2021. 9
- [25] Géron, Aurélien: *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow*. Alta Books, 2019. 9

- [26] Elakya.: *Melanoma skin cancer detection*. *ijitee*, 9(1):4073–4076, 2019. 9
- [27] Breiman, Leo: *Random forests*. *Machine learning*, 45:5–32, 2001. 10
- [28] Liaw, Andy, Matthew Wiener *et al.*: *Classification and regression by randomforest*. *R news*, 2(3):18–22, 2002. 10
- [29] Van Der Malsburg, Christoph: *Frank rosenblatt: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Em *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory, October 1–4, 1984*, páginas 245–248. Springer, 1986. 10
- [30] Liu, Xin, Tsuyoshi Murata, Kyoung Sook Kim, Chatchawan Kotarasu e Chenyi Zhuang: *A general view for network embedding as matrix factorization*. Em *Proceedings of the Twelfth ACM international conference on web search and data mining*, páginas 375–383, 2019. 10
- [31] Fausett, Laurene V: *Fundamentals of neural networks: architectures, algorithms and applications*. Pearson Education India, 2006. 11
- [32] Haykin, Simon: *Redes neurais: princípios e prática*. Bookman Editora, 2001. 11
- [33] Alpaydin, Ethem: *Machine learning: the new AI*. MIT press, 2016. 11
- [34] LeCun, Yann, Léon Bottou, Yoshua Bengio e Patrick Haffner: *Gradient-based learning applied to document recognition*. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 12, 13
- [35] Krizhevsky, Alex, Ilya Sutskever e Geoffrey E Hinton: *Imagenet classification with deep convolutional neural networks*. *Communications of the ACM*, 60(6):84–90, 2017. 12, 17
- [36] Bishop, Christopher M *et al.*: *Neural networks for pattern recognition*. Oxford university press, 1995. 14
- [37] Simonyan, Karen e Andrew Zisserman: *Very deep convolutional networks for large-scale image recognition*. *arXiv preprint arXiv:1409.1556*, 2014. 14, 16
- [38] Menegola, Afonso, Michel Fornaciali, Ramon Pires, Sandra Avila e Eduardo Valle: *Towards automated melanoma screening: Exploring transfer learning schemes*. *arXiv preprint arXiv:1609.01228*, 2016. 16
- [39] Johnson, Justin M e Taghi M Khoshgoftaar: *Survey on deep learning with class imbalance*. *Journal of Big Data*, 6(1):1–54, 2019. 16
- [40] Shin, Hoo Chang, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura e Ronald M Summers: *Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning*. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016. 16

- [41] Morid, Mohammad Amin, Alireza Borjali e Guilherme Del Fiol: *A scoping review of transfer learning research on medical image analysis using imagenet*. *Computers in biology and medicine*, 128:104115, 2021. 16
- [42] Krizhevsky, Alex, Ilya Sutskever e Geoffrey E Hinton: *Imagenet classification with deep convolutional neural networks*. Em Pereira, F., C.J. Burges, L. Bottou e K.Q. Weinberger (editores): *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. 16
- [43] He, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun: *Deep residual learning for image recognition*. Em *Proceedings of the IEEE conference on computer vision and pattern recognition*, páginas 770–778, 2016. 16
- [44] Shorten, Connor e Taghi M Khoshgoftaar: *A survey on image data augmentation for deep learning*. *Journal of big data*, 6(1):1–48, 2019. 17
- [45] Domingos, Pedro: *A few useful things to know about machine learning*. *Communications of the ACM*, 55(10):78–87, 2012. 18
- [46] Mendonça, Teresa, Pedro M Ferreira, Jorge S Marques, André RS Marcal e Jorge Rozeira: *Ph 2-a dermoscopic image database for research and benchmarking*. Em *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, páginas 5437–5440. IEEE, 2013. 20, 22
- [47] Thomas, Simon M, James G Lefevre, Glenn Baxter e Nicholas A Hamilton: *Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer*. *Medical Image Analysis*, 68:101915, 2021. 20, 21, 23
- [48] Winkler, Julia K, Katharina Sies, Christine Fink, Ferdinand Toberer, Alexander Enk, Teresa Deinlein, Rainer Hofmann-Wellenhof, Luc Thomas, Aimilios Lallas, Andreas Blum *et al.*: *Melanoma recognition by a deep learning convolutional neural network—performance in different melanoma subtypes and localisations*. *European Journal of Cancer*, 127:21–29, 2020. 21, 23
- [49] Oliveira, Roberta B, Aledir S Pereira e João Manuel RS Tavares: *Computational diagnosis of skin lesions from dermoscopic images using combined features*. *Neural Computing and Applications*, 31(10):6091–6111, 2019. 21, 23, 28
- [50] Thanh, Dang NH, VB Prasath, Le Minh Hieu e Nguyen Ngoc Hien: *Melanoma skin cancer detection method based on adaptive principal curvature, colour normalisation and feature extraction with the abcd rule*. *Journal of Digital Imaging*, 33(3):574–585, 2020. 22
- [51] Majumder, Sharmin e Muhammad Ahsan Ullah: *Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer*. Em *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, páginas 185–188, 2018. 22, 23

- [52] Bansal, Priti, Sumit Kumar, Ritesh Srivastava e Saksham Agarwal: *Using transfer learning and hierarchical classifier to diagnose melanoma from dermoscopic images*. International Journal of Healthcare Information Systems and Informatics (IJHISI), 16(2):73–86, 2021. 23
- [53] Freund, Yoav e Robert E Schapire: *A decision-theoretic generalization of on-line learning and an application to boosting*. Journal of computer and system sciences, 55(1):119–139, 1997. 23
- [54] Maia, Lucas Bezerra *et al.*: *Aprendizagem profunda aplicada ao diagnóstico de melanoma*. 2019. 23
- [55] Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall e W Philip Kegelmeyer: *Smote: synthetic minority over-sampling technique*. Journal of artificial intelligence research, 16:321–357, 2002. 23
- [56] Wilson, Dennis L.: *Asymptotic properties of nearest neighbor rules using edited data*. IEEE Trans. Syst. Man Cybern., 2:408–421, 1972. 23
- [57] Arsa, Dewa Made Sri e Anak Agung Ngurah Hary Susila: *Vgg16 in batik classification based on random forest*. Em *2019 International Conference on Information Management and Technology (ICIMTech)*, volume 1, páginas 295–299, 2019. 25
- [58] Liebovitch, Larry S e Tibor Toth: *A fast algorithm to determine fractal dimensions by box counting*. physics Letters A, 141(8-9):386–390, 1989. 29
- [59] Ribeiro, Campus Darcy: *Fusão hierárquica de medidas de textura e níveis de descontinuidade para extração de estradas em imagens de satélite de alta resolução*. 29
- [60] Python Software Foundation: *The Python Standard Library - Operating System Interface*. Python Software Foundation, 2021. <https://docs.python.org/3/library/os.html>. 36
- [61] Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke e Travis E. Oliphant: *Array programming with NumPy*. Nature, 585(7825):357–362, setembro 2020. <https://doi.org/10.1038/s41586-020-2649-2>. 36
- [62] McKinney, Wes *et al.*: *Data structures for statistical computing in python*. Em *Proceedings of the 9th Python in Science Conference*, volume 445, páginas 51–56. Austin, TX, 2010. 36
- [63] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot e E. Duchesnay: *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12:2825–2830, 2011. 36

- [64] Python Software Foundation: *The Python Standard Library - Math Module*. Python Software Foundation, 2021. <https://docs.python.org/3/library/math.html>. 36
- [65] Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu e Xiaoqiang Zheng: *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. <https://www.tensorflow.org/>, Software available from tensorflow.org. 36
- [66] Hunter, J. D.: *Matplotlib: A 2d graphics environment*. Computing in Science & Engineering, 9(3):90–95, 2007. 36
- [67] Vasconcelos, Cristina Nader e Bárbara Nader Vasconcelos: *Convolutional neural network committees for melanoma classification with classical and expert knowledge based image transforms data augmentation*. arXiv preprint arXiv:1702.07025, 2017. 46
- [68] Daghrrir, Jinen, Lotfi Tlig, Moez Bouchouicha e Mounir Sayadi: *Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach*. Em *2020 5th international conference on advanced technologies for signal and image processing (ATSIP)*, páginas 1–5. IEEE, 2020. 46