



UNIVERSIDADE FEDERAL DO MARANHÃO
Programa de Pós-Graduação em Ciência da Computação

Lucas Bezerra Maia

***Aprendizagem Profunda Aplicada ao Diagnóstico de
Melanoma***

São Luís
2019

Lucas Bezerra Maia

Aprendizagem Profunda Aplicada ao Diagnóstico de Melanoma

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação

UFMA

Orientador: Prof. Dr. Geraldo Braz Júnior

Coorientador: Prof. Dr. Anselmo Cardoso de Paiva

São Luís - MA

2019

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Maia, Lucas Bezerra.

Aprendizagem Profunda Aplicada ao Diagnóstico de
Melanoma / Lucas Bezerra Maia. - 2019.

86 f.

Coorientador(a): Anselmo Cardoso de Paiva.

Orientador(a): Geraldo Braz Júnior.

Dissertação (Mestrado) - Programa de Pós-graduação em
Ciência da Computação/ccet, Universidade Federal do
Maranhão, São Luís - MA, 2019.

1. Aprendizagem Profunda. 2. Balanceamento de
Classes. 3. Diagnóstico de Melanoma. 4. Random Forest.
5. Redes Neurais Convolucionais. I. Júnior, Geraldo
Braz. II. Paiva, Anselmo Cardoso de. III. Título.

Lucas Bezerra Maia

Aprendizagem Profunda Aplicada ao Diagnóstico de Melanoma

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Trabalho aprovado. São Luís - MA, 14 de fevereiro de 2019:

Prof. Dr. Geraldo Braz Júnior
Orientador
Universidade Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva
Coorientador
Universidade Federal do Maranhão

Prof. Dr. João Dallyson Sousa de Almeida
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. Antonio Oseas de Carvalho Filho
Examinador Externo
Universidade Federal do Piauí

São Luís - MA
2019

*Aos meus imprescindíveis pais,
Amilton Barra Nova Maia e Francisca Mesquita Bezerra Maia,
que, com muito carinho e apoio, não mediram esforços para que eu chegasse até aqui.*

Agradecimentos

Agradeço ao apoio incondicional dos meus pais, oferecendo-me todas as condições necessárias, estando ao alcance deles ou não. Para eles guardo meu eterno amor e gratidão;

Ao meu orientador e amigo Prof. Dr. Geraldo Braz Junior, de onde veio inspiração e motivação, e sem o qual esse trabalho não seria possível;

Ao meu coorientador e amigo Prof. Dr. Anselmo Cardoso de Paiva, por me depositar confiança fundamental ao sucesso de minha carreira;

À amiga de todas as horas, companheira, psicóloga particular e fiel porto seguro Bárbara Martins de Oliveira Sousa;

Aos meus amigos Nigel da Silva Lima, Alan Carlos, Ítalo Francyles, Antonino Calisto, Jonnison Lima Ferreira e Roberto Matheus Pinheiro Pereira pelos conselhos, pelas lutas, glórias e farras compartilhadas;

Aos meus amigos e companheiros do Núcleo de Telessaúde, pelo apoio durante esse processo, em especial a Humberto Oliveira Serra por sempre acreditar em mim, sendo um combustível importantíssimo para minha motivação;

A todos os meus companheiros de laboratório, por ajudarem com dicas, sugestões e apoio emocional para a construção deste trabalho, em especial a Pedro Thiago Cutrim Santos pelas incontáveis ajudas nas implementações de experimentos;

Ao \LaTeX , ferramenta essencial que nos permite poupar tempo para melhor investimento em tarefas importantes do trabalho;

A todos que participaram de forma direta e indireta para que eu alcançasse mais essa vitória.

"Aprendizagem nunca esgota a mente."

(Leonardo da Vinci)

Resumo

Melanoma é o tipo de câncer mais letal quando comparado às outras neoplasias de pele. Entretanto, quando o diagnóstico é feito em seu estágio inicial, os pacientes têm altas chances de recuperação. Diversas abordagens de detecção e diagnóstico automático de melanoma têm sido exploradas por diferentes autores com intuito de fornecer uma opinião auxiliar aos especialistas. Treinar modelos com as bases de dados existentes tem sido uma tarefa árdua devido ao problema de desbalanceamento entre as classes. Este trabalho tem como objetivo avaliar o desempenho de algoritmos de aprendizagem de máquina combinados com técnicas de treinamento em bases desbalanceadas, levando em consideração a tarefa de diagnóstico de melanoma. As arquiteturas de Redes Neurais Convolucionais VGG16, VGG19, *Inception* e *ResNet* foram utilizadas juntamente com a regra ABCD para extrair padrões de lesões de pele em um conjunto de 200 imagens dermatoscópicas. O classificador *Random Forest* alcançou uma sensibilidade de 92,5% e um índice kappa de 77,15% após o uso de seleção de atributos com *Greedy Stepwise* e balanceamento de dados de treinamento com o *Synthetic Minority Oversampling TEchnique* (SMOTE) e a regra *Edited Nearest Neighbor* (ENN).

Palavras-chave: Diagnóstico de Melanoma, Balanceamento de Classes, Aprendizagem Profunda, Redes Neurais Convolucionais, *Random Forest*, SMOTE, ENN.

Abstract

Melanoma is the most lethal type of cancer when compared to others skin diseases. However, when the diagnosis is made in its initial stage, patients have high rates of recovery. Several approaches to automatic detection and diagnosis of melanoma have been explored by different authors in order to provide an auxiliary opinion to specialists. Training models with the existing data sets have been a difficult task due to the problem of imbalanced data. This work aims to evaluate to the evaluation the performance of machine learning algorithms combined with imbalanced learning technique, regarding the task of melanoma diagnosis. The architectures of Convolutional Neural Networks VGG16, VGG19, Inception, and ResNet were used along with ABCD rule to extract patterns of skin lesions in a set of 200 dermatoscopic images. The Random Forest classifier reached a sensitivity of 92.5 % and a kappa index of 77.15 % after the use of attribute selection with Greedy Stepwise and balancing the training data with Synthetically Minority Oversampling TEchnique (SMOTE) and the Edited Nearest Neighbor (ENN) rule.

Keywords: Melanoma Diagnosis, Imbalanced Learning, Deep Learning, Convolutional Neural Networks, *Random Forest*, SMOTE, ENN.

Lista de ilustrações

Figura 1 – Exame dermatoscópico.	18
Figura 2 – Exemplos de linhas de Langer.	26
Figura 3 – As camadas da pele.	26
Figura 4 – Os extratos componentes da epiderme.	27
Figura 5 – Componentes da derme.	28
Figura 6 – Exemplos de lesões da pele: (a) Efélides ou sardas (b) Melanoses solares ou manchas senis e (c) Nevo.	29
Figura 7 – (a) Carcinoma basocelular. (b) Carcinoma espinocelular.	30
Figura 8 – Imagens dermatoscópicas, apresentando aspectos da Regra ABCD.	31
Figura 9 – Passos fundamentais em processamento de imagens digitais.	32
Figura 10 – Operadores de Sobel usados na detecção de bordas. (a) Horizontal, (b) Vertical.	33
Figura 11 – Exemplo de aplicação do detector de bordas de Sobel. (a) Imagens de uma lente de contato, (b) Gradiente de Sobel.	34
Figura 12 – Neurônios de uma RNA.	36
Figura 13 – Arquitetura simples de uma CNN.	37
Figura 14 – Exemplo de descritores aprendidos através de uma CNN.	37
Figura 15 – Exemplos de convoluções. (a) <i>Stride</i> 1x1, (b) <i>Stride</i> 2x2	38
Figura 16 – Exemplo de <i>pooling</i>	40
Figura 17 – Exemplo de <i>dropout</i>	41
Figura 18 – Visualização dos campos receptivos dos filtros em diferentes camadas para duas arquiteturas de CNNs diferentes.	42
Figura 19 – Arquiteturas VGG.	44
Figura 20 – Um animal alvo que ocupa a maior parte da imagem, depois apenas parte dela e então pouquíssimo espaço.	45
Figura 21 – Módulo <i>Inception</i> Simples.	45
Figura 22 – Módulo <i>Inception</i> para redução de dimensionalidade.	46
Figura 23 – Arquitetura completa de uma rede <i>Inception</i>	46
Figura 24 – Módulo Residual. (a) Fluxo Residual, (b) Fluxo Convencional	47
Figura 25 – Interpretação gráfica de um exemplo sintético com o SMOTE.	49
Figura 26 – Aplicações das variantes do SMOTE. (a) Amostras, (b) <i>Borderline</i> 1, (c) <i>Borderline</i> 2, (d) SVM.	50
Figura 27 – Subamostragens com ENN usando valores diferentes para k	51
Figura 28 – Processo de busca do algoritmo PSO.	54
Figura 29 – Hiperplano de separação para padrões lineares.	56
Figura 30 – Árvore de Decisão para uma porta lógica AND.	59

Figura 31 – Gráfico de entropia.	59
Figura 32 – Exemplo real de Árvore de Decisão.	60
Figura 33 – Fluxograma da metodologia proposta.	62
Figura 34 – Imagem da base PH ² : (a) Dermatoscopia e (b) Máscara binária (b) Região de Interesse Segmentada.	63
Figura 35 – Fluxo de preparação de uma imagem para extração de característica com a arquitetura VGG16.	64
Figura 36 – Fluxograma da extração e seleção de características.	65
Figura 37 – Gráficos de caixa das sensibilidades para o <i>Greedy Stepwise</i>	71
Figura 38 – Mapa de ativações para VGG16 para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha)	72
Figura 39 – Mapa de ativações para VGG19 para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha).	73
Figura 40 – Mapa de ativações para <i>Inception</i> para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha).	73
Figura 41 – Mapa de ativações para ResNet para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha).	74

Lista de tabelas

Tabela 1 – Desempenho das técnicas de diagnóstico de câncer de pele a partir de imagens dermatoscópicas.	24
Tabela 2 – Exemplo de base de dados com instâncias de classes (<i>Attractive</i>) positivas e negativas.	60
Tabela 3 – Número de características por seletor.	65
Tabela 4 – Número de características por descritor de característica para cada seletor.	65
Tabela 5 – Matriz de confusão para duas classes.	67
Tabela 6 – Níveis de precisão de classificação, segundo o índice Kappa.	68
Tabela 7 – Resultados de <i>Random Forest</i> sem uso de SMOTEENN.	69
Tabela 8 – Resultados de <i>Random Forest</i> com uso de SMOTEENN.	69
Tabela 9 – Resultados de SVM sem uso de SMOTEENN.	70
Tabela 10 – Resultados de SVM com uso de SMOTEENN.	70
Tabela 11 – Comparação com trabalhos relacionados.	74
Tabela 12 – Artigos aceitos em relação ao tema do diagnóstico do melanoma com a utilização de imagens dermatoscópicas.	77

Lista de abreviaturas e siglas

ABCD	Assimetria, Bordas, Cor e Diâmetro
AUC	<i>Area Under the Curve</i>
CBC	Carcinoma Basocelular
CEC	Carcinoma Espinocelular
CNNs	<i>Convolutional Neural Networks</i>
DBN	<i>Deep Belief Network</i>
DNNs	<i>Deep Neural Networks</i>
DRNNs	<i>Deep Recurrent Neural Networks</i>
EC	<i>Evolutionary Computation</i>
ENN	<i>Nearest Neighbor Rules</i>
FCRN	<i>Fully Convolutional Residual Network</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
Grad-CAM	<i>Gradient-weighted Class Activation Mapping</i>
HIV	<i>Human Immunodeficiency Virus</i>
HOG	<i>Histogram of Oriented Gradient</i>
kNN	<i>k-Nearest Neighbor</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
LBP	<i>Local Binary Patterns</i>
LICU	<i>Lesion Indexing Calculation Unit</i>
LIN	<i>Lesion Indexing Network</i>
MLP	<i>MultiLayer Perceptron</i>
OMS	Organização Mundial da Saúde

PSO	<i>Particle Swarm Optimization</i>
RBF	<i>Radial Base Function</i>
ReLU	<i>Rectified Linear Unit</i>
RF	<i>Random Forests</i>
RGB	<i>Red, Green, Blue</i>
RNA	Redes Neurais Artificiais
ROC	<i>Receiver Operating Characteristic</i>
ROI	<i>Region Of Interest</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
SMDC	<i>Sydney Melanoma Diagnostic Centre</i>
SMOTE	<i>Synthetic Minority Oversampling TEchnique</i>
SRM	<i>Structural Risk Minimization</i>
SV	<i>Support Vector</i>
SVM	<i>Support Vector Machines</i>
TDE	Transformada de Distância Euclidiana
TN	<i>True Negative</i>
TP	<i>True Positive</i>
t-SNE	<i>t-distributed Stochastic Neighbour Embedding</i>
UMCG	<i>University Medical Center Groningen</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

Sumário

1	INTRODUÇÃO	17
1.1	Objetivos	18
1.1.1	Objetivos Específicos	19
1.1.2	Contribuições	19
1.2	Organização do Trabalho	20
2	TRABALHOS RELACIONADOS	21
3	FUNDAMENTAÇÃO TEÓRICA	25
3.1	A Pele e sua Anatomia	25
3.1.1	Camadas da Pele	26
3.1.2	Tipos de Neoplasias de Pele	28
3.2	Processamento de Imagens Digitais	31
3.2.1	Detecção de Bordas e Descrição Geométrica	33
3.3	Aprendizagem Profunda	35
3.3.1	Redes Neurais Convolucionais	35
3.3.1.1	Transformações Convolucionais	37
3.3.1.2	Subamostragem de dados	39
3.3.1.3	Camada de classificação	40
3.3.2	<i>Deep Features e Transfer Learning</i>	41
3.3.3	Arquiteturas de CNNs Pré-treinadas	43
3.3.3.1	Arquiteturas VGG	43
3.3.3.2	Arquitetura Inception	44
3.3.3.3	Arquitetura ResNet	46
3.4	Balanceamento de Classes	48
3.4.1	Sobre-amostragem de Dados com SMOTE	48
3.4.2	Subamostragem com ENN	50
3.5	Seleção de Características	51
3.5.1	<i>Greedy Stepwise</i>	52
3.5.2	PSO	53
3.5.3	<i>Ranker e RankSearch</i>	54
3.6	Algoritmos de Aprendizagem de Máquina	54
3.6.1	Máquinas de Vetor de Suporte	55
3.6.2	Random Forest	58
4	METODOLOGIA	62

4.1	Aquisição e Preparação de Imagens	62
4.2	Extração e Seleção de Características	63
4.3	Experimentos	66
4.4	Métricas de Desempenho	66
4.4.1	Índice Kappa de Cohen	67
5	RESULTADOS	69
5.1	Comparação com Trabalhos Relacionados	74
6	CONCLUSÃO	76
6.1	Produções Científicas	77
	REFERÊNCIAS	78

1 Introdução

Doenças crônicas já foram consideradas fora do alcance de países com recursos limitados (ANDERSON; ILBAWI; SAGHIR, 2015). A Organização Mundial da Saúde (OMS) focava em doenças como malária, tuberculose e Vírus da Imunodeficiência Humana (HIV - do inglês *Human Immunodeficiency Virus*) que, apesar do progresso significativo, continuam sendo problemas de saúde importantes em países de baixa e média renda. Atualmente, surgiu o reconhecimento de que as doenças não transmissíveis, incluindo o câncer, se tornaram uma preocupação de saúde dominante, pois, com o passar do tempo, os tipos de cânceres característicos de países mais desenvolvidos, como os de mama, pulmão, próstata e pele foram se espalhando para áreas menos favorecidas (SOARES, 2008).

Dentre os tipos de cânceres, o câncer de pele é o mais comum e sua incidência aumenta cada vez mais, por parte, em um reflexo do comportamento humano à exposição excessiva de sol. O melanoma, mesmo sendo considerado o tipo de câncer de pele mais letal, apresenta índices de cura elevados quando diagnosticado em seu estágio inicial. A doença se manifesta em pele normal a partir do surgimento de uma pinta com tonalidade escura e bordas irregulares, ou de uma lesão pigmentada pré-existente, onde o tumor irá evoluir apresentando mudanças de cores e aumento da área da lesão (SOARES, 2008).

A estimativa mundial, em 2018, apresentou que 232.130 novos casos de melanoma seriam diagnosticados em todo o globo. Enquanto que, em termos de mortalidade, foram estimados 55.488 óbitos, sendo os países com população caucasiana¹, os lugares com maiores taxas de incidência. Estima-se, para homens e mulheres, respectivamente, 85.170 e 80.410 novos casos de neoplasia não melanoma no Brasil, cujos valores são correspondentes a um risco estimado de 82,53 casos novos a cada 100 mil homens e 75,84 para cada 100 mil mulheres. O câncer de pele melanoma, mesmo tendo incidência baixa (2.920 e 3.340 novos casos para homens e mulheres, respectivamente), sua letalidade é a mais elevada (INCA, 2018).

As maiores chances de cura do melanoma são quando diagnosticado precocemente, ou seja, quando as células cancerosas ainda estão apenas na superfície da pele. A cirurgia (remoção total da lesão) é o tratamento mais indicado, porém a radioterapia, quimioterapia e imunoterapia também podem ser solicitadas a depender do estágio do câncer. Se não for retirado em seu estágio inicial, pode ocorrer a metástase, que acaba invadindo as camadas sadias da pele, através dos vasos linfáticos e da circulação, disseminando para outras regiões do corpo, o que torna o controle da doença mais difícil.

¹ Que ou quem pertence a uma divisão da espécie humana, originária da Europa, Norte de África e Próximo Oriente, com características físicas e genéticas comuns, nomeadamente pele clara ou morena.

Para diagnosticar o câncer, o paciente deve ser submetido a uma biópsia, que é a única forma de diagnóstico definitivo. Nesse procedimento, toda a lesão, ou pelo menos parte dela, é retirada para ser analisada por um especialista por meio de um microscópio. A análise busca a presença de células cancerosas e, geralmente, deve ser feita por dois ou mais patologistas ([ARNOLD; ODOM; JAMES, 1995](#)).

Além da biópsia, existem alguns exames que servem para detectar e prevenir o câncer de pele. Um deles é a dermatoscopia, técnica não-invasiva, no qual o dermatologista usa um aparelho chamado dermatoscópio, ilustrado na Figura 1, para fotografar a região da lesão, aumentando-a de 10 a 70 vezes. Com uma fonte de luz não-polarizada, uma lente transparente e um líquido que fica entre a luz e a lente, o aparelho permite a avaliação de cores e microestruturas da epiderme, junção dermo-epidérmica e derme papilar que não são visíveis a olho nu. Portanto, esse exame possibilita averiguar regiões que têm chances de ser um tumor maligno, atuando na prevenção do câncer de pele. A precisão do diagnóstico chega a 97% ([SOARES, 2008](#)), evitando cirurgias desnecessárias.

Figura 1 – Exame dermatoscópico.



Fonte: ([KATZ; RABINOVITZ, 2001](#))

Entendendo a importância do diagnóstico do câncer de pele em seu estágio precoce, métodos de processamento digital de imagens estão sendo desenvolvidos com o intuito de melhorar as técnicas de diagnóstico já existentes. A ideia é fornecer um modelo de classificação para diagnóstico de câncer de pele que possa diminuir os casos de diagnósticos incorretos e aumentar as chances do paciente receber um tratamento em tempo hábil.

1.1 Objetivos

O objetivo deste trabalho consiste em propor uma metodologia para diagnóstico de câncer de pele melanoma por meio de um processo supervisionado de aprendizagem de máquina, usando fundamentos de *Deep Learning* em imagens dermatoscópicas e algoritmos de balanceamento de amostras.

1.1.1 Objetivos Específicos

Especificamente, este trabalho busca os seguintes objetivos aplicados ao problema de diagnóstico de melanoma:

- Estudar os conceitos e formas de diagnóstico do melanoma;
- Adaptar e aplicar conceitos de aprendizagem profunda para reconhecimento de padrões em imagens médicas;
- Aplicar arquiteturas de Redes Neurais Convolucionais para extração de características com intuito de auxiliar o reconhecimento de regiões suspeitas de melanoma;
- Estudar a aplicação de técnicas de balanceamento de bases por meio de geração de exemplos sintéticos e diminuição de amostras ruidosas, tais como SMOTE e ENN, para treinamento dos classificadores SVM e *Random Forest* já conhecidos na literatura;
- Avaliar a metodologia proposta por meio de experimentos, usando bases públicas de imagens dermatoscópicas;
- Analisar os resultados obtidos, entendendo as vantagens e limitações da metodologia de classificação.

1.1.2 Contribuições

Destacam-se como principais contribuições:

- Aplicação de balanceamento de classes em uma base de dados de imagens de dermatoscopias, de forma a manter um número baixo de exemplos necessários para o treinamento do modelo de classificação, mesmo utilizando métodos de aprendizagem profunda os quais dependem de uma grande quantidade de amostras;
- Utilização de transferência de aprendizagem para o diagnóstico de melanoma feita por meio da extração de características com arquiteturas de Redes Neurais Convolucionais já treinadas em outro domínio de conhecimento;
- Melhoria da qualidade do diagnóstico médico baseado em imagens no âmbito médico-hospitalar;
- Contribuição para o resgate de casos de câncer, em condições mais precoces e mais aptas ao tratamento curativo, preservando vidas;

- Introdução de uma opinião complementar no que tange o diagnóstico de neoplasias de pele, contribuindo para diminuir os procedimentos desnecessários no caso de erro da detecção/diagnóstico;
- A difusão da tecnologia desta pesquisa pode contribuir para uma redução significativa dos custos diretos ligados à hospitalização e a intervenções cirúrgicas desnecessárias, bem como economias de custos indiretos devido ao afastamento do trabalho de pacientes que se beneficiariam de condutas mais precisas e precoces.

1.2 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 descreve trabalhos relacionados ao tema do diagnóstico melanoma, usando imagens dermatoscópicas.
- O Capítulo 3 trata da fundamentação teórica das técnicas utilizadas. São abordados conceitos referentes ao câncer de pele melanoma, imagens dermatoscópicas, técnicas de pré-processamento de imagens, extração de características, reconhecimento de padrões e aprendizagem profunda;
- O Capítulo 4 apresenta as etapas adotadas que compõem a metodologia proposta para este trabalho. São mostradas as configurações dos parâmetros utilizados para o pré-processamento e a extração das características, bem como os passos tomados para a realização do treinamento e reconhecimento de padrões;
- O Capítulo 5 trata sobre os resultados obtidos e discussões em relação aos experimentos realizados na base de imagens públicas dermatoscópicas; e
- O Capítulo 6 apresenta as considerações finais sobre os resultados e trabalhos futuros e os artigos científicos desenvolvidos.

2 Trabalhos Relacionados

Neste capítulo, são apresentados trabalhos disponíveis na literatura com foco na classificação de melanoma em imagens dermatoscópicas. Cada trabalho coletado apresentou uma maneira diferente de usar Redes Neurais Convolucionais e métodos de aprendizagem de máquina para efetuar o diagnóstico de melanoma, tentando sanar outros problemas inerentes ao uso de aprendizagem profunda como: *overfitting* (quando o classificador se ajusta à base de treinamento e possui baixo poder de generalizar em instâncias desconhecidas para o classificador) (VASCONCELOS; VASCONCELOS, 2017), base limitada (quando é necessário um número demasiado de imagens para ajustar os parâmetros de uma CNN profunda) (SZEGEDY et al., 2015) e base desbalanceada (quando uma base possui mais imagens de uma classe do que da outra, o classificador tende a aprender mais dessa classe, fazendo com que a acurácia não seja uma boa métrica de comparação) (VASCONCELOS; VASCONCELOS, 2017).

Em (ROSADO, 2009), um método supervisionado de classificação de lesões epiteliais é apresentado, baseando-se no sistema de pontuação da regra ABCD. Foi utilizado o esquema de cores em tons de cinza, extraindo o canal azul do RGB por ser o componente que permitiu melhor discriminação da lesão. Para descrever forma e textura, foram utilizados a Transformada de Distância Euclidiana (TDE) e a Entropia. Para a etapa de classificação, foram comparados os classificadores SVM e k-vizinhos mais próximos (kNN) (KELLER; GRAY; GIVENS, 1985).

Alencar (2015) propuseram um modelo de classificação em que as lesões de pele da base PH^2 (MENDONÇA et al., 2013) são segmentadas utilizando limiarização adaptativa de Otsu sobre o canal azul do esquema RGB, baseando-se em (ROSADO, 2009). Para a descrição de contornos, foram utilizadas as funcionalidades *findContours*, *convexHull* e *convexityDefects*, da biblioteca OpenCv (BRADSKI, 2000). O processo de aprendizado de máquina contou com a utilização de uma Rede Neural Artificial *Multilayer Perceptron* (MLP) para classificar as imagens em nevo benigno ou melanoma.

Em (MASOOD; AL-JUMAILY; ANAM, 2015) foi proposto um modelo semi-supervisionado de aprendizagem para detecção de melanoma em imagens dermatoscópicas. No sistema proposto, uma DBN (do inglês - *Deep Belief Network*) e duas Máquinas de Vetor de Suporte Autodeterminadas (SA-SVMs) com kernel RBF (*Radial Base Function*) e Polinomial, respectivamente, foram treinadas em três conjuntos de dados diferentes. Os três conjuntos de imagens foram gerados a partir de dados rotulados e dados não rotulados. A estratégia de *fine-tuning* com uma função de custo exponencial foi adotada no treinamento para maximizar a divisão dos dados rotulados.

Uma técnica que utiliza *Random Forests* (RFs) como algoritmo de reconhecimento de padrões é apresentada em (RASTGO et al., 2016), onde três representações de características são utilizadas: *Scale-Invariant Feature Transform* (SIFT), histogramas de ângulos opostos e intensidade de RGB. O treinamento ocorreu através de validação cruzada com 10 divisões sobre a base de imagens PH².

A arquitetura *Deep Residual Network* (DRN ou ResNet, proposta em (HE et al., 2016) visando o desafio de reconhecimento de padrões da ImageNet (DENG et al., 2009)) foi utilizada por Codella et al. (2016) para a criação de conjuntos de métodos capazes de segmentar lesões cutâneas, além de analisar a área detectada e o tecido adjacente para diagnóstico de melanoma. (YU et al., 2017) integra a arquitetura proposta no trabalho, a *Fully Convolutional Residual Network* (FCRN), para segmentação de lesões de pele e ResNet para classificação, formando uma estrutura de dois estágios para diagnóstico de neoplasia maligna de pele.

Já em (MAJTNER; YILDIRIM-YAYILGAN; HARDEBERG, 2016), uma outra arquitetura de rede neural convolucional profunda, a AlexNet, foi utilizada para extrair características de imagens dermatoscópicas. Esta foi combinada com Máquinas de Vetor de Suporte (SVM) (NAUMOVICH; VAPNIK, 1998) para diagnosticar melanoma. Em (LOPEZ et al., 2017), uma adaptação da arquitetura VGG, proposta em (SIMONYAN; ZISSERMAN, 2014), foi utilizada para estudar os efeitos da inicialização dos pesos da rede de acordo com três paradigmas: (i) treinamento dos pesos ainda não inicializados; (ii) inicialização dos pesos com valores de outras redes já treinadas (conceito de *transfer learning*) e (iii) melhorar o valores de outras redes (*fine-tuning*) com um treinamento nas imagens do problema a ser tratado (imagens dermatoscópicas). Em (MAHBOD; ECKER; ELLINGER, 2017), as arquiteturas pré-treinadas AlexNet e VGG16 foram adotadas para extrair características profundas de imagens dermatoscópicas para classificação de lesões de pele.

Nasr-Esfahani et al. (2016) propuseram uma metodologia usando uma CNN pré-treinada para imagens clínicas de lesões de pele (GIOTIS N. MOLDERS; PETKOV, 2015). As imagens são pré-processadas corrigindo problemas de iluminação. Em seguida uma máscara de segmentação *k-means* de dois grupos é aplicada às imagens que, depois de algumas operações morfológicas, são usadas para extrair a região de interesse das lesões para serem inseridas na arquitetura de *deep learning*. O conjunto de imagens ainda foi utilizado em um processo de *data augmentation*, com intuito de aumentar a base de dados, gerando novas imagens transladadas, redimensionadas e rotacionadas.

Em (SHIN et al., 2016) e (MIRUNALINI et al., 2017), a topologia de rede chamada *Inception*, proposta em (SZEGEDY et al., 2015) para a competição da ImageNet, foi utilizada, juntamente com um treinamento especializado na base de dermatoscopias, para aumentar o desempenho no diagnóstico de neoplasias de pele. A novidade é a

introdução de um módulo que permite convoluções simultâneas e de tamanhos diferentes, aumentando a performance de treinamento da rede, bem como sua classificação. Camadas completamente conectadas são adicionadas nos blocos intermediários da rede para efetuar a retro-propagação do erro de treinamento.

Segundo [Alfed, Khelifi e Bouridane \(2016\)](#), técnicas de extração de características apresentam melhora quando combinadas com *Histogram of Oriented Gradient* (HOG) para aplicações de reconhecimento de melanoma. Em [\(PRIYADARSHINI; RENGINI, 2015\)](#), SVM é utilizado para reconhecimento de características extraídas através da combinação de *Local Binary Patterns* (LBP) e HOG. LBP é uma técnica de análise de textura baseada em padrões locais que, combinada com HOG, aumenta a performance de detecção consideravelmente [\(ALFED; KHELIFI; BOURIDANE, 2016\)](#).

Em [\(BAKHEET, 2017\)](#), características HOG otimizadas são usadas em conjunto com SVM para detecção de melanoma. A região de interesse é segmentada usando limiar adaptativo e operações morfológicas. O experimento foi realizado usando a abordagem *leave-one-out*, que consiste em separar uma imagem para teste, enquanto que as outras são usadas para treinamento. Esse processo se repete para cada imagem, gerando N iterações, sendo N igual ao tamanho da base.

[Han et al. \(2018\)](#) testaram o uso de algoritmos de aprendizagem profunda para classificar imagens clínicas de 12 anomalias de pele (carcinoma basocelular, carcinoma espinocelular, carcinoma intraepitelial, ceratose actínica, ceratose seborreica, melanoma maligno, nevo melanocítico, lentigo, granuloma piogênico, hemangioma, dermatofibroma e verruga). Foi realizado um *fine-tuning* na CNN ResNet-152 por meio de imagens do conjunto de dados da Asan [\(HAN, 2017\)](#), MED-NODE [\(GIOTIS et al., 2015\)](#) e fontes de vários sites dermatológicos (19.398 imagens no total).

Uma estrutura de aprendizagem baseada em CNN que extrai características discriminativas de baixa dimensão foi proposta em [\(SULTANA; MANDAL; PUHAN, 2018\)](#). Sua abordagem minimiza toda a informação de variação dentro da classe e maximiza a informação de variação total da classe. A importância de vários subespaços que surgem na matriz de dispersão dentro da classe seguida de redução de dimensionalidade, usando a informação de variação de classe total, é analisada para detecção de melanoma.

[Safran et al. \(2018\)](#) avaliaram vários algoritmos de aprendizagem de máquina (como MLP, SMV e *Random Forest*) para classificar exames dermatoscópicos. Em [\(ESTEVA et al., 2017\)](#), uma estrutura de CNN com t-SNE (*t-distributed Stochastic Neighbour Embedding* [\(MAATEN; HINTON, 2008\)](#)) foi treinada para classificação geral da lesão cutânea. Então, comparou-se o desempenho de pelo menos 21 dermatologistas testados em três tarefas críticas de diagnóstico: classificação de carcinoma de queratinócitos, classificação de melanoma e classificação de melanoma por dermatoscopia.

A LIN (*Lesion Indexing Network*) foi proposta em (LI; SHEN, 2018) para abordar simultaneamente a segmentação e classificação das lesões. Duas CNNs profundas, FCRNs-88, treinadas com diferentes conjuntos de treinamento, são adotadas para produzir o resultado de segmentação e classificação normal. Uma Unidade de Cálculo de Indexação de Lesão (LICU, do inglês - *Lesion Indexing Calculation Unit*) é proposta para medir a importância de um *pixel* para a decisão da classificação da lesão. O resultado da classificação da lesão é refinado de acordo com o mapa de distância gerado pela LICU.

A Tabela 1 resume o desempenho de trabalhos relacionados, mencionando métodos utilizados, base explorada e resultados alcançados (A - Acurácia, P - Precisão, E - Especificidade e S - Sensibilidade) para o diagnóstico de câncer de pele usando técnicas de aprendizagem profunda em imagens dermatoscópicas.

Tabela 1 – Desempenho das técnicas de diagnóstico de câncer de pele a partir de imagens dermatoscópicas.

Trabalho	Metodologia	Base	A	P	E	S
(ROSADO, 2009)	Cor, TDE + kNN	PH ²	-	-	82,00%	61,00%
(MASOOD; AL-JUMAILY; ANAM, 2015)	DBN + SVM	SMDC	89,00%	90,00%	88,30%	-
(ALENCAR, 2015)	Cor e contornos + MLP	PH ²	86,66%	-	-	-
(CODELLA et al., 2016)	LBP + SC + ResNet	ISIC	83,40%	60,20%	91,12%	53,30%
(KAWAHARA; BENTAIEB; HAMARNEH, 2016)	CNN	Dermofit	81,80%	-	-	-
(RASTGO et al., 2016)	Cor e SIFT + RFs	PH ²	-	-	100%	90,30%
(MAJTNER; YILDIRIM-YAYILGAN; HARDEBERG, 2016)	AlexNet + SVM	ISIC	82,60%	-	89,80%	53,30%
(POMPONI; NEJATI; CHEUNG, 2016)	CNN	DERMIS	93,64%	-	95,18%	92,10%
(NASR-ESFAHANI et al., 2016)	CNN	UMCG	81,00%	75,00%	80,00%	81,00%
(ALFED; KHELIFI; BOURIDANE, 2016)	Cor, SIFT e HOG + SVM e kNN	PH ²	-	-	91,20%	88,10%
(LOPEZ et al., 2017)	VGG16	ISIC	81,33%	79,74%	-	78,66%
(YU et al., 2017)	ResNet	ISIC	94,90%	-	95,70%	91,10%
(BAKHEET, 2017)	Forma (HOG) + SVM	PH ²	97,32%	-	98,21%	87,32%
(VASCONCELOS; VASCONCELOS, 2017)	Inception	ISIC	83,60%	69,90%	86,50%	76,00%
(MIRUNALINI et al., 2017)	Inception	ISIC	82,50%	-	-	-
(ESTEVA et al., 2017)	CNN	Privada	91,00%	-	-	-
(MAHBOD; ECKER; ELLINGER, 2017)	AlexNet, VGG16	ISIC	89,20%	-	84,70%	84,70%
(SAFRAN et al., 2018)	Random Forest	Privada	-	-	83,63%	86,01%
(HAN et al., 2018)	ResNet	Mix	88,00%	80,70%	-	85,50%
(LI; SHEN, 2018)	LIN CNN	ISIC	85,70%	-	96,61%	49,00%
(SULTANA; MANDAL; PUHAN, 2018)	ResNet + Eigen + SVM	PH ²	97,50%	-	100%	85,50%

Nenhuma das propostas apresentadas lidam com o problema de desbalanceamento de classes, onde o número de instâncias de exames dermatoscópicos normais são maiores do que a quantidade de amostras de melanoma. Enquanto algumas abordagens de *deep learning* utilizaram técnicas (como rotações, redimensionamento e translações) para aumentar a base, outras propostas deixam o algoritmo de aprendizagem de máquina ajustar o custo dado para cada classificação. Neste trabalho, é proposta a aplicação de ferramentas de sintetização de dados no domínio de características extraídas das imagens através do *Synthetic Minority Oversampling TEchnique* (SMOTE), bem como redução de instâncias ruidosas, com e a regra *Edited Nearest Neighbor* (ENN), que podem prejudicar o modelo preditivo final.

3 Fundamentação Teórica

Este capítulo apresenta os conceitos explorados para o desenvolvimento do estudo, bem como características da pele, o câncer melanoma, técnicas utilizadas para extração e reconhecimento de padrões, além de explicações sobre *Convolutional Neural Networks* (CNNs) em tarefas de classificação de imagens no intuito de fundamentar a metodologia proposta que tem como objetivo classificar, de forma binária, imagens dermatoscópicas em melanoma e não-melanoma.

3.1 A Pele e sua Anatomia

A pele é o maior órgão do corpo humano e envolve toda a sua superfície, representando aproximadamente 15% do peso, e contém grandes variações ao longo de sua extensão, em partes sendo mais flexíveis, em outras, mais rígidas (LEVER; LEVER, 1991). É considerado o manto que reveste o organismo, ajudando, de forma imprescindível à vida, no isolamento dos componentes orgânicos do meio exterior.

A superfície corpórea possui sulcos e saliências, que se acentuam principalmente nas regiões palmoplantares¹ e pontas dos dedos. A disposição (em redemoinhos, alças e arcos) dessas estruturas são absolutamente únicas e peculiares em cada indivíduo, permitindo sua utilização em datiloscopia² e dermatoglifia³ (SOARES, 2008). Há uma correlação entre dermatóglifos e doenças neuro-cutâneas, mostrando relações entre o desenvolvimento da pele e do sistema nervoso central, que podem ser atingidos por uma mesma perturbação na embriogênese⁴ (LEVER; LEVER, 1991).

A pele apresenta ainda as chamadas **linhas de Langer** (linhas de tensão ou fenda), ilustradas na Figura 2, que são exploradas por cirurgiões com o intuito de efetuarem incisões que deixem cicatrizes menos marcantes (LIBERATO, 1998). Essas linhas são fibras, formadas por colágeno, que definem uma rede de ranhuras rasas de topologia poligonal. Ao envelhecer, as linhas comumente chamadas de rugas demonstram as direções dessas ranhuras.

Na parte inferior da pele, encontram-se espalhados os melanócitos, que produzem a melanina. Com a exposição ao sol, os melanócitos produzem mais pigmentos, resultando no chamado bronzeamento ou escurecimento da superfície epitelial. De certa forma, grupos de melanócitos cercados de tecidos formam uma estrutura não cancerosa e muito comum

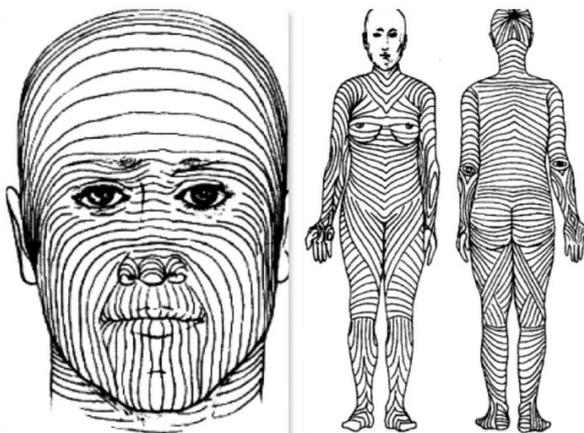
¹ Relativo à palma da mão e à planta do pé.

² Estudo e exame das impressões digitais para identificação de pessoas.

³ Impressão das rugas cutâneas das palmas e dos dedos das mãos e das plantas e dos dedos dos pés.

⁴ Evolução do desenvolvimento dos seres vivos durante o seu estado embrionário.

Figura 2 – Exemplos de linhas de Langer.



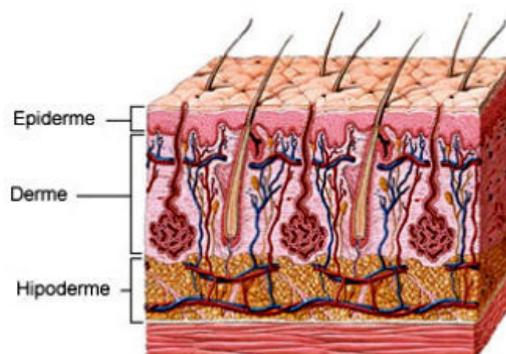
Fonte: (MARX; WALLS; HOCKBERGER, 2013)

denominada *nevo* (SOARES, 2008).

3.1.1 Camadas da Pele

A pele é composta fundamentalmente de três camadas: uma superficial, chamada epiderme, com origem ectodérmica; uma camada intermediária, a derme (ou cório), de origem mesodérmica; e a mais profunda, a hipoderme ou camada subcutânea (LIBERATO, 1998). Uma ilustração dessas camadas podem ser visualizadas na Figura 3.

Figura 3 – As camadas da pele.



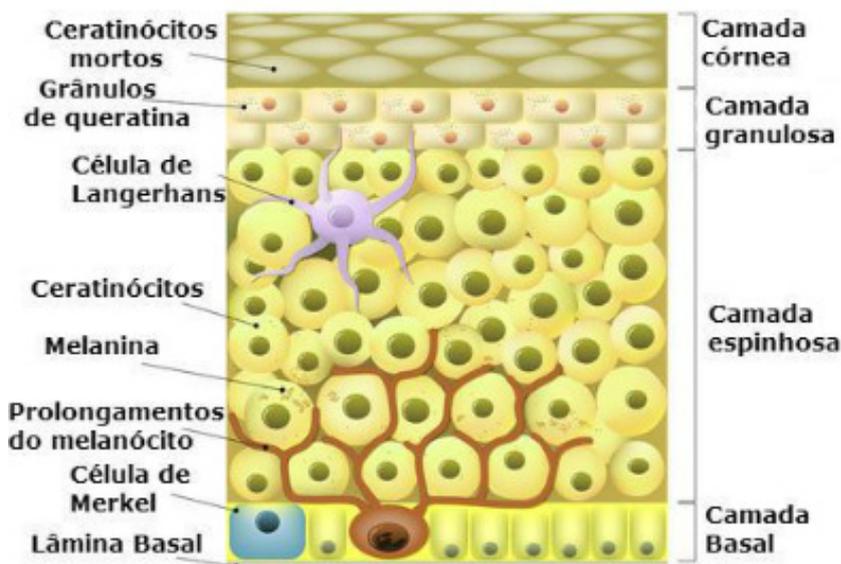
Fonte: (MAGALHÃES, 2018a)

A epiderme é a camada mais superficial da pele, formada por epitélio estratificado queratinizado e não vascularizada. A sua espessura e textura se diferenciam de acordo com a região do corpo, principalmente nas áreas palmoplantares, sendo mais delgada na palma das mãos e mais densa na planta dos pés (DÂNGELO; FATTINI, 1998).

As células que compõem a epiderme são sequencialmente sobrepostas, formando quatro faixas principais, cuja ilustração pode ser apreciada na Figura 4:

- **Estrato córneo:** é uma subcamada fina e superficial, formada por células mortas, denominadas ceratinócitos, sem núcleos e achatadas, as quais apresentam grande quantidade de queratina e estão continuamente descamando;
- **Estrato granuloso:** formado por faixas de células poligonais achatadas e citoplasma acumulado de grânulos (queratino-hialina), os quais darão origem à queratina ao extrato superior;
- **Estrato espinhoso:** formado por subcamadas de células cubóides, pouco achatadas e com núcleo central. Elas apresentam projeções citoplasmáticas com filamentos de queratina, ancorando uma célula às outras, dando resistência ao atrito (LEVER; LEVER, 1991). Apresenta ainda as células de *Langerhans* (têm capacidade de fagocitose e de ativar os linfócitos - mecanismo de defesa);
- **Estrato germinativo ou basal:** camada mais profunda e em contato com a derme. Os queratinócitos (presentes em maior número, cerca de 95%, são responsáveis pela produção de queratina) produzidos são empurrados para as camadas superiores. Apresenta também os melanócitos (responsáveis pela produção de melanina, o pigmento que dá cor a pele) e as células de Merkel (responsáveis pela sensação do tato).

Figura 4 – Os estratos componentes da epiderme.



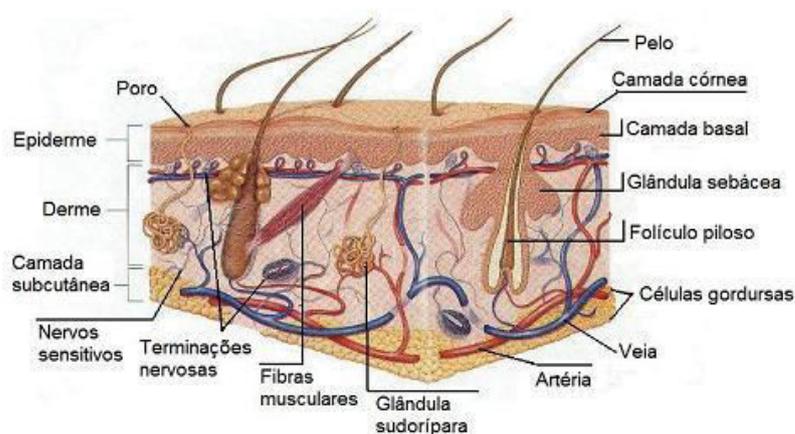
Fonte: (MAGALHÃES, 2018c)

A derme é a camada subsequente à epiderme (Figura 5), cujo limite é formado por saliências chamadas papilas dérmicas (camada papilar), correspondentes a vãos na epiderme. A derme é formada por tecido conjuntivo e apresenta-se mais elástica e firme,

devido à presença de colágeno e elastina. São encontrados, ainda, os vasos sanguíneos e linfáticos, nervos, terminações nervosas, folículos pilosos, glândulas sudoríparas e sebáceas (DÂNGELO; FATTINI, 1998).

Dessa camada, vale ressaltar as glândulas sebáceas e sudoríparas, encontradas praticamente em todo corpo, exceto em lábios e genitais (glande e pequenos lábios) onde se abrem diretamente na superfície. Para as glândulas sebáceas, sua secreção é uma mistura complexa de lipídeos, que deixam a pele oleosa; já as sudoríparas são estruturas tubulosas simples, formando um enovelado imerso na derme, cuja secreção é o suor, um fluido que contém água, sódio, potássio, cloretos, ureia, amônia e ácido úrico.

Figura 5 – Componentes da derme.



Fonte: (MAGALHÃES, 2018b)

A hipoderme, também chamada de tecido subcutâneo, é a camada mais interna. Entretanto, não é considerada parte da pele. Ela é constituída por células adiposas, formando um tecido que armazena gordura subcutânea (panículo adiposo), fibras de colágeno e vasos sanguíneos (LIBERATO, 1998). A quantidade de células adiposas presentes varia de indivíduo para indivíduo e entre as partes do corpo. Essa camada desempenha funções importantes como: isolar o corpo das variações externas do ambiente e fixar a pele aos órgãos e estruturas adjacentes.

3.1.2 Tipos de Neoplasias de Pele

O câncer de pele é o crescimento anormal e descontrolado das células que a compõem e pode surgir a partir de pele normal ou de uma lesão pigmentada pré-existente (SOARES, 2008). Em pele normal, a manifestação pode ocorrer com o aparecimento de uma mancha com bordas irregulares que talvez apresente sintomas como coceira e descamação. A grande maioria das lesões não apresenta sintoma algum (ARNOLD; ODOM; JAMES, 1995). Em

caso de lesão pré-existente, acontece um aumento do tamanho, mudança na pigmentação e na forma da lesão, além de passar a apresentar bordas irregulares.

As lesões nas regiões dos melanócitos podem variar de simples manchas aos chamados nevos atípicos (que podem se tornar o câncer melanoma), tais como: **Efélide** - também conhecidas como sardas (Figura 6(a)), são manchas com distribuição homogênea de pigmento melânico, apresentando bordas tanto regulares, como irregulares; **Melanose solar** - ou mancha senil (Figura 6(b)), tem pigmentos de cores castanhas, de tamanho variado e bordas bem delimitadas, apresentando-se comumente em áreas expostas ao sol, tais como: membros superiores, parte superior das mãos e rosto; **Nevo** - é uma proliferação (benigna) de células (Figura 6(c)), podendo ser congênita ou adquirida (ARNOLD; ODOM; JAMES, 1995).

Figura 6 – Exemplos de lesões da pele: (a) Efélides ou sardas (b) Melanoses solares ou manchas senis e (c) Nevo.



Fonte: (SIMIS; SIMIS, 2006)

Câncer de pele é mais comum em indivíduos com mais de 40 anos, e é relativamente raro em crianças e afrodescendentes. Indivíduos de pele clara, sensível à ação dos raios solares, com doenças cutâneas prévias ou com alterações genéticas são as principais vítimas desse tipo de doença (ARNOLD; ODOM; JAMES, 1995).

Dentre as neoplasias malignas está o Carcinoma Basocelular que é o tipo mais comum de câncer de pele, correspondendo a 60% das neoplasias epiteliais malignas (SIMIS; SIMIS, 2006). Cresce por invasão direta, destruindo o tecido normal da pele, porém muito raramente se evolui para metástase. É caracterizado por um nódulo rosado, enrijecido, apresentando lentamente um crescimento progressivo, levando a uma úlcera, que pode se curar e depois reincidir (ARNOLD; ODOM; JAMES, 1995).

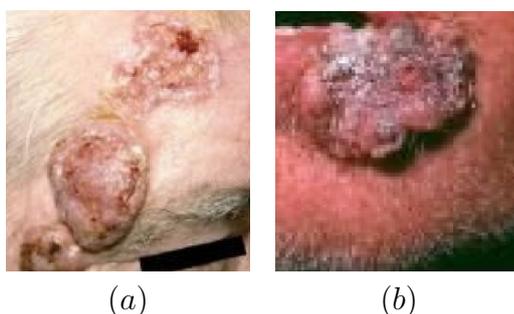
A severidade do carcinoma basocelular (CBC) depende da histologia da região onde se apresenta, podendo ser mais agressivo em locais como: pálpebras, orelhas e nariz, devido à dificuldade de cirurgias (SOARES, 2008). Um exemplo pode ser visto na Figura 7 (a).

Outra proliferação maligna, o Carcinoma Espinocelular, representa cerca de 15% das neoplasias malignas de pele (SIMIS; SIMIS, 2006), originando-se de lesões precursoras

caracterizadas por alterações progressivas e aparência de úlcera que pode se espalhar por meio dos linfonodos e atingir todo o corpo (metástase). Neste caso, há a necessidade de um diagnóstico e tratamento precoce.

O carcinoma espinocelular (CEC) é mais comum no dorso das mãos, lábio inferior, orelhas, face, mucosa bucal e genitália externa. É mais comum em homens, devido a maior exposição a agentes cancerígenos como sol e tabaco, além da predisposição em indivíduos de pele clara (SIMIS; SIMIS, 2006). Um exemplo pode ser visto na Figura 7.

Figura 7 – (a) Carcinoma basocelular. (b) Carcinoma espinocelular.



Fonte: (SIMIS; SIMIS, 2006)

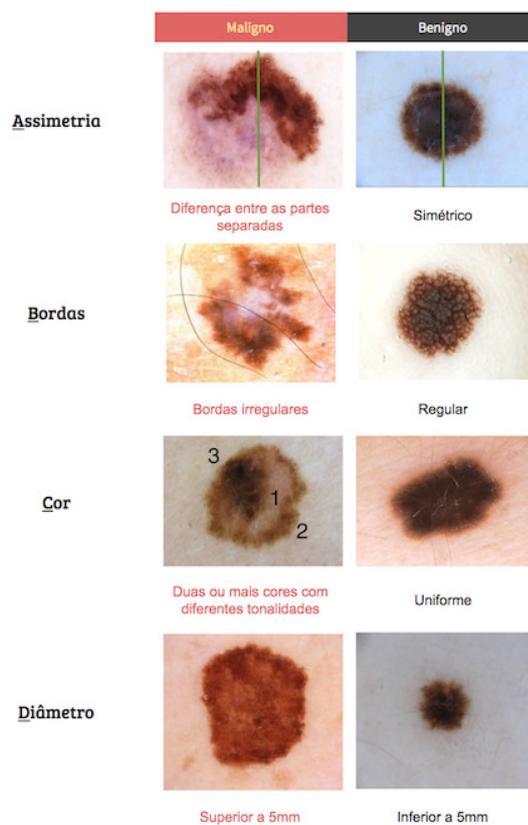
Apesar de ser o menos frequente dentre os tipos de neoplasias de pele, o melanoma é o que possui consequências mais devastadoras. É o mais perigoso devido a grande capacidade de metástase. Entretanto, quando diagnosticado em seu estágio inicial, ao ser feito o tratamento adequado, os índices de cura superam os 90% (ARNOLD; ODOM; JAMES, 1995).

As maiores chances de cura do melanoma são quando diagnosticado precocemente, ou seja, quando as células cancerosas ainda estão apenas na superfície da pele. A cirurgia (remoção total da lesão) é o tratamento mais indicado, porém a radioterapia, quimioterapia e imunoterapia também podem ser solicitadas a depender do estágio do câncer. Se não for retirado em seu estágio inicial, pode ocorrer a metástase, que acaba invadindo as camadas saudáveis da pele, disseminando para outras regiões do corpo, o que torna o controle da doença mais difícil.

Tendo em vista o reconhecimento de sinais e sintomas do câncer de pele melanoma, é necessária a avaliação do próprio corpo, levando em consideração a regra chamada ABCD (SOARES, 2008) (Figura 8).

As iniciais significam: **A**ssimetria - onde, ao realizar um corte mediano na lesão, um lado não é igual ao outro; **B**ordas - as quais geralmente são irregulares no seu contorno, apresentando certas imperfeições; **C**or - essa característica não apresenta uniformidade, ou seja, frequentemente, possui duas ou mais tonalidades, podendo ser escuras, marrons,

Figura 8 – Imagens dermatoscópicas, apresentando aspectos da Regra ABCD.



Fonte: Elaborada pelo autor

avermelhadas ou até azuladas; e, por último, **Diâmetro** - lesões malignas modificam seu tamanho, aumentando a área da lesão que, na maioria dos casos, passam de 5mm (ARNOLD; ODOM; JAMES, 1995).

Os melanomas podem ser notados, quando há uma alteração em um nevo pré-existente, ou se uma nova região escura aparece na pele. Com o avanço da doença, pode-se notar uma leve alteração na textura do nevo, podendo se tornar enrijecida e de alto relevo (SOARES, 2008). Os aspectos do melanoma variam muito de indivíduo para indivíduo. Muitos apresentam características de toda a regra ABCD, porém outros podem sofrer modificações em apenas um ou dois dos itens supracitados. Embora os melanomas possam ter um aspecto diferente, podendo coçar, transpirar ou sangrar, eles geralmente são indolores.

3.2 Processamento de Imagens Digitais

Uma imagem é, geralmente, uma projeção 2D de uma cena 3D capturada por algum sensor (DAWSON-HOWE, 2014). É a representação de um objeto que emite ou recebe luz e que é formada por raios luminosos que passam por uma lente, espelho ou

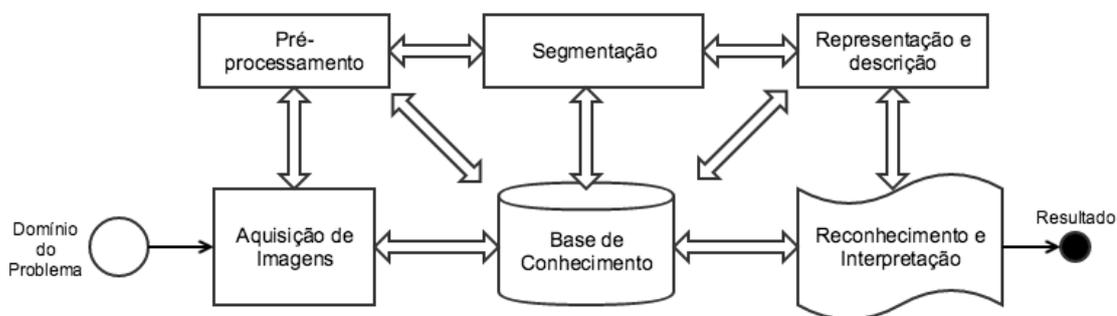
qualquer outro sistema óptico. Consiste em uma função de intensidade de luz de duas dimensões apresentada por $f(i, j)$, onde i e j são as coordenadas espaciais e o valor da função f é proporcional à luminosidade da imagem no ponto (i, j) .

Para representar uma imagem de forma digital, é preciso fazer amostras de $f(i, j)$ na forma de um matriz de dimensão $N \times M$, em que cada elemento é uma quantidade discreta da intensidade luminosa, que é dividida em k intervalos, com valor geralmente igual a 256. Ou seja, para uma imagem monocromática⁵, pode-se ter até 256 tons diferentes de uma única cor para compô-la.

O processamento de imagens pode ser considerado como uma ou mais alterações realizadas sobre os elementos que as compõem, os *pixels*, com intuito de melhorar aspectos visuais de certas estruturas que possibilitem uma determinada interpretação da imagem.

Existem inúmeros algoritmos para processamento de imagens, cada um com finalidades bastante específicas. O conjunto de diferentes algoritmos escolhidos formam uma metodologia, que geralmente acompanha as etapas apresentadas por (GONZALEZ; WOODS, 2006), como pode ser visto na Figura 9.

Figura 9 – Passos fundamentais em processamento de imagens digitais.



Fonte: Adaptado de Gonzalez e Woods (2006)

As etapas envolvidas no processo giram em torno de uma base de conhecimento intrinsecamente ligada ao domínio do problema e, dentre elas, estão: aquisição de imagens, apresentando o problema a ser investigado; pré-processamento, buscando melhorar a visualização e/ou realçar características presentes nas imagens; segmentação, onde a região de interesse é separada das estruturas irrelevantes ao problema; representação e descrição, relacionada à organização das informações; e, finalmente, a etapa de reconhecimento, na qual os padrões encontrados nas informações são interpretados.

Na etapa de representação e descrição, busca-se extrair informações relevantes que melhor rotulam um determinado padrão, ou seja, são aqueles atributos cujos valores servem

⁵ Quando diferentes tons de uma única cor são usados para montar uma imagem.

como assinatura para uma certa categoria dentro do espaço do problema tratado. Esta fase precede a de reconhecimento de padrões, pois seus resultados servem como valores de entrada para a etapa seguinte.

3.2.1 Detecção de Bordas e Descrição Geométrica

A borda é uma característica fundamental de uma imagem, pois é o conjunto de *pixels* onde os tons de cinza alteram gradativamente como em um formato de telhado de casas (MA et al., 2010), além de fazer distinção entre objetos e o fundo e objetos e outros objetos de uma determinada cena.

Uma das ferramentas existentes para detecção de bordas é o operador de Sobel (SOBEL, 1972), que executa uma medição de gradiente bidimensional de uma imagem e enfatiza áreas com alta frequência que correspondem às arestas. Normalmente, ele é usado para localizar a magnitude do gradiente absoluto aproximado em cada ponto em uma imagem em escala de cinza de entrada (GONZALEZ; WOODS, 2006). Ele funciona com derivadas de primeira ordem nos eixos X e Y da imagem. As derivadas são apenas aproximações (porque as imagens não são contínuas). Para aproximá-los, os operadores apresentados na Figura 10 são usados:

Figura 10 – Operadores de Sobel usados na detecção de bordas. (a) Horizontal, (b) Vertical.

-1	0	1	-1	-2	-1
-2	0	2	0	0	0
-1	0	1	1	2	1
(a)			(b)		

Fonte: Elaborada pelo autor

Os operadores podem ser aplicados separadamente para produzir medições em cada eixo, que então podem ser combinados para encontrar a magnitude absoluta do gradiente em cada ponto e a orientação desse gradiente. A magnitude do gradiente é dada aproximadamente por:

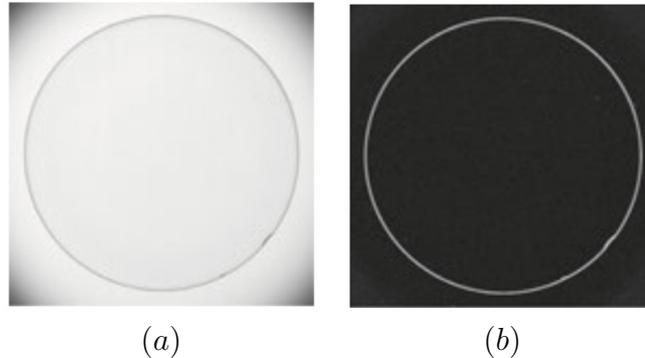
$$|G| = |G_x| + |G_y|, \quad (3.1)$$

O ângulo de orientação da borda que dá origem ao gradiente é dado por:

$$\theta = \arctan \frac{G_y}{G_x}, \quad (3.2)$$

onde G representa a imagem resultante da aplicação dos filtros de Sobel para os eixos x e y . Uma aplicação dos gradientes de Sobel podem ser visualizados na Figura 11.

Figura 11 – Exemplo de aplicação do detector de bordas de Sobel. (a) Imagens de uma lente de contato, (b) Gradiente de Sobel.



Fonte: (GONZALEZ; WOODS, 2006)

Neste trabalho, cujo foco é o diagnóstico de melanoma em imagens dermatoscópicas, explora-se a ideia da Regra ABCD descrita na subseção 3.1.2. As iniciais dessa regra se remetem a quatro características que devem ser analisadas sobre uma suspeita de lesão grave de pele. O câncer melanoma pode apresentar formas **assimétricas**, **bordas** irregulares, diferentes tonalidades de **cores** e tamanho anormal de seu **diâmetro**. Levando em consideração a implementação proposta por Moura et al. (2018), três dessas regras podem ser caracterizadas através de descritores de forma, sendo elas:

Diâmetro:

É a maior distância entre dois pontos ($p_1 = (x_1, y_1)$ e $p_2 = (x_2, y_2)$) pertencentes ao contorno da região de interesse, que pode ser formalizada por:

$$D_{(p_1, p_2)} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3.3)$$

Assimetria:

É calculado a partir do diâmetro, onde a maior distância, em *pixels*, entre a linha que forma a diagonal principal e um ponto pertencente à borda é formulada por:

$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}, \quad (3.4)$$

onde a , b e c são os coeficientes da linha da diagonal e x_0 , y_0 são os pontos de coordenadas da imagem.

Borda:

Os pontos das bordas são ordenados em um vetor unidimensional de forma que seus pontos vizinhos sejam seus mesmos vizinhos no vetor. Percorrendo o vetor sequencialmente

testa-se a diferença nos eixos x e y entre um ponto p e seus vizinhos imediatos ($p - 1$, $p + 1$) para identificar alterações suaves na borda, testa-se também a diferença entre um ponto p e seus vizinhos um pouco mais distantes ($p + d$, $p - d$) para identificar alterações mais drásticas na borda, onde d é o número em pixels relacionado à distância. Conta-se o número de picos e vales tanto suaves quanto drásticos, gerando 4 novas características.

Para a terceira variável da regra ABCD, a cor, conta-se o número de cores encontrados na lesão como uma nova característica. Inicialmente os valores dos *pixels* são divididos entre i intervalos na imagem (tons similares são agrupados). Para cada grupo são contados os números de *pixels* presentes, grupos que possuem um número expressivo de *pixels* são contados, os que possuem um número menor que um limiar são desconsiderados.

3.3 Aprendizagem Profunda

A aprendizagem profunda vem do inglês *Deep Learning*, podendo ser conhecida também como aprendizado hierárquico, ou aprendizado de máquina profundo, e é baseado em uma série de algoritmos que modelam abstrações de dados em alto nível, usando várias camadas de processamento que realizam diversas transformações lineares e não lineares (AHMAD; FARMAN; JAN, 2019). É um ramo do aprendizado de máquina (do inglês *Machine Learning*) que inclui a suposição de que essas camadas fornecem níveis de composição, onde podem ser usadas quantidades e tamanhos de camadas diferentes para que se possa fornecer quantidades variadas de abstração, definindo assim uma arquitetura (BENGIO; COURVILLE; VINCENT, 2013).

Várias arquiteturas de aprendizagem profunda, tais como Redes Neurais Profundas (DNNs - do inglês *Deep Neural Networks*), Redes Neurais Convolucionais (CNNs - do inglês *Convolutional Neural Networks*), Redes de Crenças Profundas (DBN - do inglês *Deep Belief Network*) e Redes Neurais Recorrentes Profundas (DRNNs - do inglês *Deep Recurrent Neural Networks*) têm sido aplicadas em áreas como visão computacional, reconhecimento automático de fala, processamento de linguagem natural, reconhecimento de áudio e bioinformática, onde elas têm se mostrado capazes de produzir resultados do estado-da-arte em várias tarefas (AHMAD; FARMAN; JAN, 2019).

Este trabalho foca em aprendizagem de máquina profunda em imagens médicas, mais especificamente em exames dermatoscópicos. Por tanto, nas seções a seguir, serão abordados os conceitos sobre diversas arquiteturas que foram exploradas no estudo.

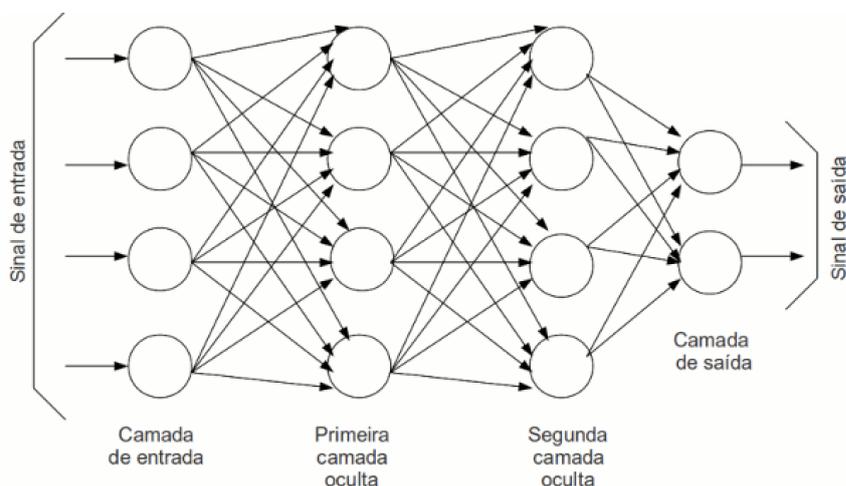
3.3.1 Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNNs - *Convolutional Neural Networks*) têm como objetivo subdividir continuamente imagens com intuito de extrair características presentes nas mesmas, inspirado no córtex visual de animais (HUBEL; WIESEL, 1968). Tais redes

realizam o aprendizado por meio dos métodos de propagação de sinais e retro-propagação de erros (RUMELHART; HINTON; WILLIAMS, 1986).

Sendo a principal característica das CNNs a capacidade de aprender padrões de imagens em vários estágios, a base de seu funcionamento parte dos conceitos de Redes Neurais Artificiais (RNAs), que consiste em diversas células simples de processamento interconectadas, chamadas de neurônios, cada uma produzindo uma série de ativações de valor real. Neurônios de entrada servem de interface com o mundo exterior, pois são ativados através da inserção de qualquer sinal, enquanto que outros neurônios são ativados através de conexões ponderadas de neurônios previamente ativos (SCHMIDHUBER, 2015). As ligações, que podem apresentar diferentes pesos dependendo da rede, entre as células de uma RNA são ilustradas na Figura 12.

Figura 12 – Neurônios de uma RNA.



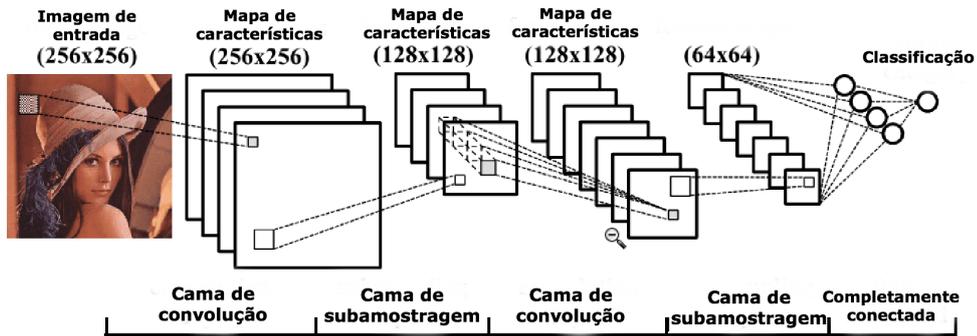
Fonte: Adaptado de Marsland (2015)

O conhecimento de uma rede é representada pelos valores dos pesos nas ligações entre uma célula e outra, que fazem com que a RNA exiba um certo comportamento, dado pelas ativações de um número de células que determinam o padrão que se deseja reconhecer. Se um padrão totalmente diferente for inserido como entrada, outro conjunto de neurônios é ativado, fornecendo assim uma saída diferente.

No caso de uma CNN, existem camadas peculiares que são responsáveis por tarefas diferentes realizadas nos dados de entrada. As camadas são empilhadas de forma que a saída processada por ela serve como entrada para a outra. A Figura 13 ilustra uma arquitetura geral de uma rede neural convolucional.

A imagem de entrada é informada como entrada para CNN, onde sofre um série de transformações, cujos objetivos dependem do estágio no qual ela se encontra. De um modo geral, existem três tipos de camadas em uma rede neural convolucional:

Figura 13 – Arquitetura simples de uma CNN.



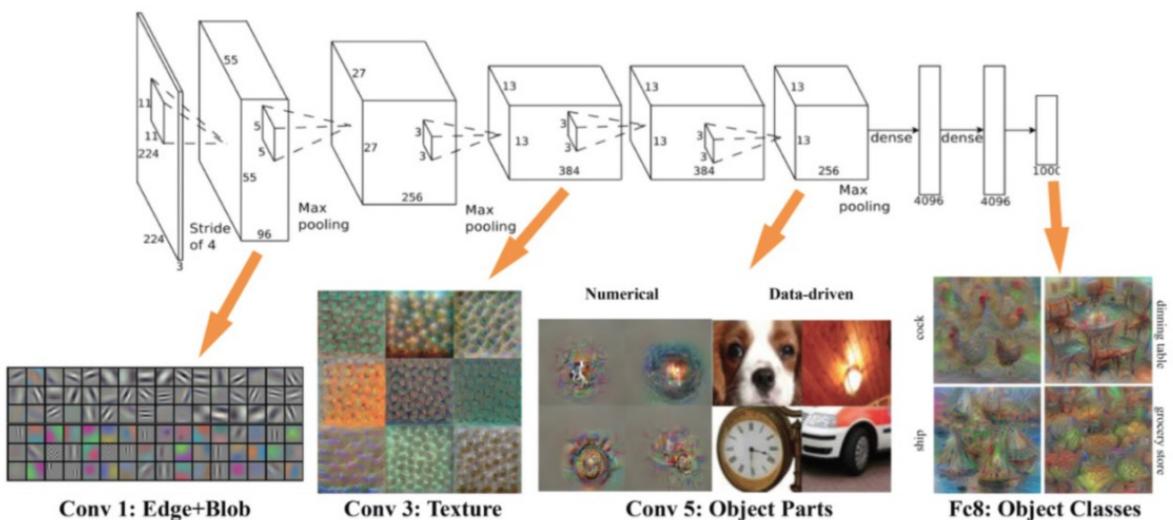
Fonte: Adaptado de [LeCun et al. \(1998a\)](#)

- Camada de Convolução;
- Camada de Subamostragem ou *Pooling*;
- Camada Completamente Conectada.

3.3.1.1 Transformações Convolucionais

Camadas de convolução podem ser entendidas como conjuntos de filtros a serem aprendidos, representados em forma de matrizes de ativação. Cada filtro é aprendido de uma parte diferente da imagem e, então, utilizado como extrator de características, resultando em mapa de ativação representativa daquela característica ([KRIZHEVSKY; SUTSKEVER; HINTON, 2012](#)).

Figura 14 – Exemplo de descritores aprendidos através de uma CNN.



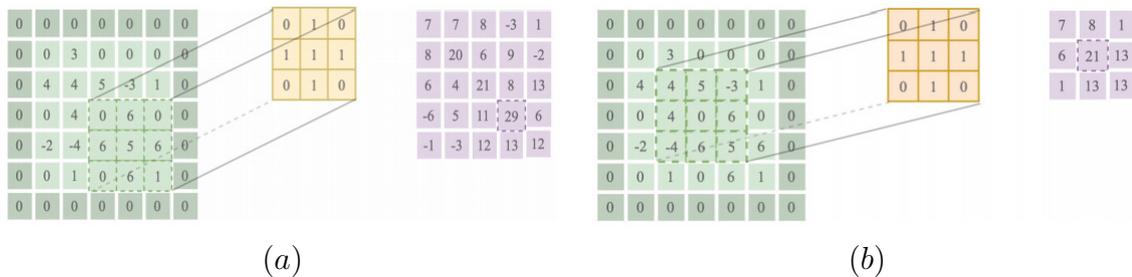
Fonte: ([KRIZHEVSKY; SUTSKEVER; HINTON, 2012](#))

Levando em consideração a Figura 14, a imagem é inserida como entrada do lado esquerdo e sofre transformações de convolução de acordo com que a rede vai ficando

mais profunda. Pode-se notar que as camadas mais rasas da rede tendem a aprender características de mais baixo nível, como filtros de passa baixa e, à medida que convoluções em camadas mais profundas são realizadas, esses filtros evoluem para características mais complexas, como texturas e partes de objetos.

A convolução é uma operação matemática que é realizada por meio do sobreposicionamento entre duas matrizes. Uma matriz menor, denominada *Kernel* ou filtro, desliza sobre outra matriz, considerada entrada da convolução. Para cada valor de células correspondente entre a região do filtro e da entrada, uma operação matemática é realizada, cujo resultado é o valor de saída para a célula localizada no centro do *kernel*. A velocidade com que o filtro passa sobre a imagem de entrada é denominada *stride* ou passo, dado em pixels. A Figura 15 demonstra o funcionamento da operação de convolução.

Figura 15 – Exemplos de convoluções. (a) *Stride* 1x1, (b) *Stride* 2x2



Fonte: Elaborada pelo autor.

A matriz de cor verde representa uma imagem de entrada, a de cor amarela é o *kernel* responsável pela convolução e a de cor lilás representa a saída. Pode-se notar que, ao aumentar o passo da convolução, a imagem de saída é reduzida. A Equação 3.5 demonstra a operação matemática de convolução.

$$Y_k = f(W_k * x), \quad (3.5)$$

onde a imagem de entrada é dada por x ; o *kernel* de convolução relacionado ao k -ésimo filtro da camada é denotado por W_k ; o sinal $*$, neste contexto, significa o operador convolucional 2D usado para calcular o produto interno do modelo de filtro em cada localização da matriz de entrada; e $f(x)$ representa a função de ativação não-linear aplicada ao resultado do produto (LECUN et al., 1998b) e (RAWAT; WANG, 2017).

Uma CNN compartilha pesos. Isso significa que um neurônio é conectado a um volume específico da outra camada. A área desse volume é chamada de campo receptivo e é definida pelo tamanho do filtro da convolução. A profundidade deste volume é sempre igual a profundidade do volume de entrada em cada camada. Sendo assim, para um filtro

de tamanho 5x5 em uma imagem 64x64x3 (RGB), cada neurônio teria 75 (5x5x3) pesos a serem treinados.

As funções de ativação não-lineares permitem a extração de características não-lineares. Atualmente a função ReLu ($f(x)$) está se tornando bastante popular devido à facilidade do cálculo, onde é mantido apenas a parte positiva e transformando a parte negativa para zero, o que aumenta a velocidade do cálculo (LECUN et al., 1998b) e (RAWAT; WANG, 2017).

A função ReLu (do inglês - *Rectified Linear Unit*), geralmente diminui o tempo de treino, pois mantém apenas a parte positiva e transforma a parte negativa em zero, aumentando a velocidade dos cálculos. Na sua forma tradicional, a equação é a mesma do limiar comum, mas com o valor de θ fixo em 0. Normalmente, as camadas de convolução quando implementadas por bibliotecas e *frameworks* possuem uma função linear na sua saída. Formalmente, podemos definir a Equação 3.6:

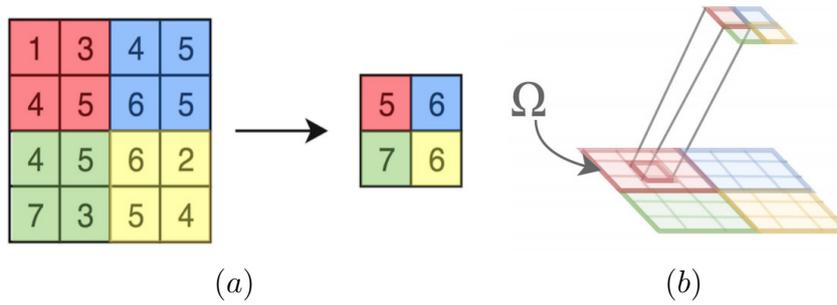
$$f(x) = \max(x, 0) \quad (3.6)$$

3.3.1.2 Subamostragem de dados

Uma camada de convolução pode ser seguida ou não por uma camada denominada *pooling* ou subamostragem, cujo objetivo é representar estatisticamente a camada anterior; pois, além de ser capaz de reduzir a dimensão da imagem, os *pixels* de uma determinada região são resumidos a um único valor, reduzindo ruídos indesejáveis. Usando funções do tipo máximo, mínimo ou média de um conjunto de *pixels*, dentro de um bloco predefinido que desliza sobre a imagem, a quantidade de dados pode ser reduzida significativamente sem grandes perdas de representatividade.

O processo de funcionamento é semelhante ao da camada de convolução, diferenciando-se apenas da operação matemática realizada. Ao invés de um produto interno, realiza-se uma redução de valores por meio de média, valor mínimo ou máximo. Nesse tipo de camada, não existem pesos a serem treinados. A entrada é analisada de acordo com seus valores atuais e resumida para um volume de saída de acordo com o tipo de *pooling* escolhido. Reduzindo a resolução dos mapas de características resultantes da aplicação de convolução torna os descritores invariantes a translação, rotação e escala. Um exemplo de seu funcionamento pode ser visualizado na Figura 16.

No exemplo, pode-se notar a redução da matriz 4x4 para 2x2 através do valor máximo dos números representados pelas cores diferentes. Nas camadas de subamostragem, normalmente, o tamanho do passo é o mesmo tamanho do filtro, de modo que não exista sobreposição em passos consecutivos. A profundidade ou número de canais é sempre a

Figura 16 – Exemplo de *pooling*.

Fonte: Elaborada pelo autor.

mesma do volume inicial. Matematicamente, a Equação 3.7 representa a operação de *pooling* para função Max:

$$Y_{kij} = \max_{(p,q) \in \mathfrak{R}_{ij}} x_{kpq}, \quad (3.7)$$

onde a saída da operação de agrupamento, associada ao k -enésimo mapa de características, é denotada por Y_{kij} , x_{kpq} denota o elemento na localização (p, q) contido pela região de agrupamento \mathfrak{R}_{ij} , que incorpora um campo receptivo ao redor da posição (i, j) , a área que corresponde ao tamanho do *kernel* do *pooling*, ou seja essa definição pode ser vista na Figura 16 (b) (LECUN et al., 1998b) e (RAWAT; WANG, 2017).

3.3.1.3 Camada de classificação

Por último, geralmente, no final das redes neurais convolucionais, são acopladas camadas completamente conectadas para efetuar a classificação das características extraídas pelas camadas anteriores. Neste momento, definem-se as funções de custo a serem utilizadas, que tem como objetivo representar o erro entre o valor esperado e a predição dada pela rede neural. Esse erro é usado no processo de retro-propagação para ajuste de pesos.

As camadas empilhadas anteriormente a essa camada têm como função extrair características significativas da imagem de entrada (LECUN et al., 1998b). No final, essas características são vetorizadas e servem de entrada para uma rede neural artificial completamente conectada, cuja tarefa é categorizar os descritores informados em sua entrada.

Com a classificação dada pela camada completamente conectada a partir de uma imagem informada, necessita-se calcular o erro entre a categorização e o valor verdadeiro do rótulo da entrada. Geralmente, usa-se o operador *softmax* devido à sua simplicidade e interpretação probabilística, que serve para mapear a saída não normalizada da rede para uma distribuição de probabilidade sobre as classes de saída previstas (NASRABADI,

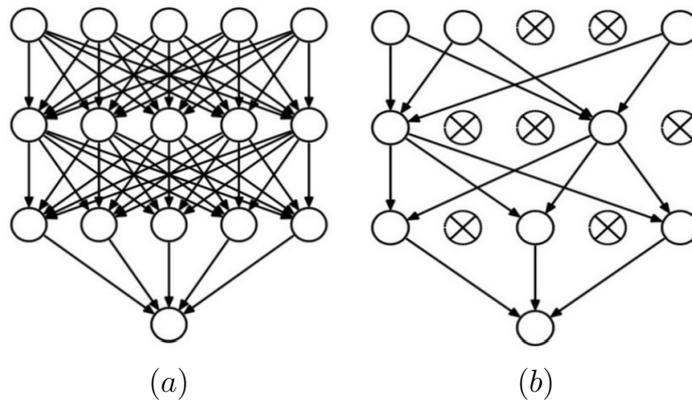
2007). Formalmente, para o i -ésimo vetor de características x_i , que tem classe rotulada por y_i , a função de custo *softmax* (L) pode ser escrita como:

$$L = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right), \quad (3.8)$$

onde o elemento j -ésimo ($j \in [1, K]$, K é o número de classes) do vetor de probabilidades de classe f é representado por f_j e N é o tamanho da base de treinamento. Para o custo, f representa as ativações de uma camada W totalmente conectada; assim, f_{y_i} , pode ser denotado como $f_{y_i} = W_{y_i}^T x_i$ em que W_{y_i} é a y_i -ésima coluna de W .

Com o intuito de prevenir o chamado super-ajuste (*overfitting*), existe uma técnica importante presente em algumas redes neurais que é definida como *dropout* (SRIVASTAVA et al., 2014). Este tipo de técnica é aplicada somente na fase de treino da rede, objetivando prevenir o super-ajuste de neurônios nas redes neurais artificiais. A ideia é simples, porém, eficaz. A cada época de treino ou *batch* treinado, um sorteio sobre os neurônios envolvidos na RNA é feito e alguns são selecionados e excluídos da próxima interação com probabilidade p . A Figura 17 ilustra o funcionamento deste método, na esquerda (a) é ilustrado uma RNA completamente conectada e, na direita (b), a mesma rede após a aplicação do *dropout*.

Figura 17 – Exemplo de *dropout*.



Fonte: Elaborada pelo autor.

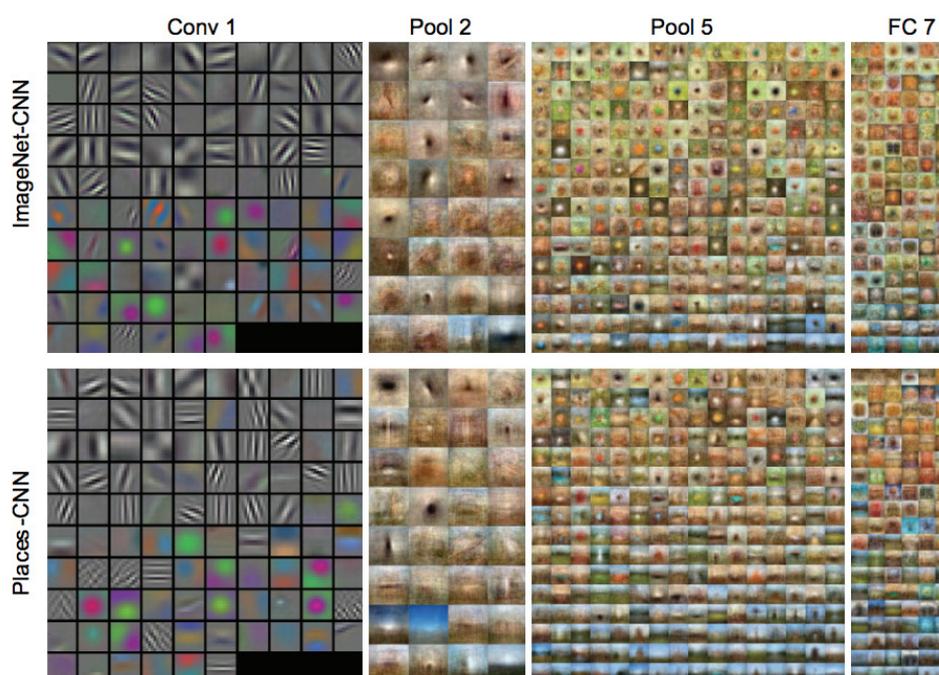
3.3.2 Deep Features e Transfer Learning

Redes neurais, especialmente redes neurais profundas, possuem uma alta capacidade discriminativa. O poder de modelagem não linear faz dessas redes não apenas um bom classificador (MIRUNALINI et al., 2017), mas também superior em tarefas de extração de características (YU et al., 2017). Inspirado em trabalhos como (CODELLA et al., 2016) e (YU et al., 2017), que usam Redes Neurais Convolucionais (CNNs) para engenharia de extração de características no reconhecimento de melanoma, neste trabalho, imagens

dermatoscópicas são processadas pela CNN cujas saídas derivadas de uma camada oculta específica podem ser obtidas. Esses valores derivados da CNN são chamados de *Deep Features*.

Na Figura 18 são demonstradas as características de CNNs pré-treinadas para tarefas de reconhecimento visual. As respostas das camadas de nível mais alto das CNNs provaram ser recursos genéricos eficazes com desempenho de última geração em vários conjuntos de dados de imagens (ZHOU et al., 2014). Pode-se notar que as características passam de simples detectores de contornos e bordas para representação de textura e forma mais complexas de acordo com o avanço da imagem através da rede.

Figura 18 – Visualização dos campos receptivos dos filtros em diferentes camadas para duas arquiteturas de CNNs diferentes.



Fonte: (ZHOU et al., 2014)

Essas características extraídas por meio de arquiteturas de CNNs pré-treinadas podem ser utilizadas não só em uma rede neural artificial simples completamente conectada, mas também para treinar outros classificadores como SVM e *Random Forest*.

CNNs são o estado-da-arte para classificação de imagens (MENEGOLA et al., 2017), mas seu uso para imagens médicas é desafiador, já que esses modelos exigem conjuntos de treinamento muito grandes (de dezenas de milhares até milhões de imagens) (SHIN et al., 2016). Para contornar essa dificuldade, a literatura mais atual emprega o chamado *Transfer Learning*, uma técnica em que um modelo treinado para uma determinada tarefa de origem é parcialmente “reciclado” para uma nova tarefa de destino. *Transfer Learning* varia de simplesmente usar a saída de uma CNN de origem como um vetor de características e treinar um modelo completamente novo (por exemplo, um SVM) para uma determinada

tarefa; até usar uma CNN inicial pré-treinada para inicializar alguns dos pesos de uma CNN de destino e continuar o processo de treinamento normalmente (YOSINSKI et al., 2014).

3.3.3 Arquiteturas de CNNs Pré-treinadas

Grande parte da popularidade atual das redes neurais convolucionais se deve aos seus resultados obtidos pelos autores Krizhevsky, Sutskever e Hinton (2012) na edição da ImageNet (DENG et al., 2009). Os autores utilizaram uma técnica proposta na década de 1990 por LeCun et al. (1998b). Outro fator que influenciou e continua influenciando os estudos sobre redes neurais convolucionais é o aumento do poder computacional que vem acontecendo desde o início do século. Algumas arquiteturas bem conhecidas são discutidas a seguir.

3.3.3.1 Arquiteturas VGG

As arquiteturas VGGs (SIMONYAN; ZISSERMAN, 2014) são um conjunto de redes da mesma família, que variam de acordo com a quantidade de camadas existentes. Duas configurações são amplamente utilizadas, sendo elas a VGG-16 e VGG-19 (Figura 19). A grande contribuição dessas redes na época foi o aumento de profundidade das redes possibilitado pela substituição dos filtros de tamanho grande (11 e 5 na primeira e segunda camada convolucional, respectivamente) por vários filtros de tamanho de núcleo 3x3 um após o outro.

Com um determinado campo receptivo (o tamanho da área efetiva da imagem de entrada na qual a saída depende), um empilhamento múltiplo de *kernels* de tamanho menor é melhor do que aquele com um *kernel* de tamanho maior, porque múltiplas camadas não-lineares aumentam a profundidade da rede que permite com que ela aprenda características mais complexas, com um custo menor (SIMONYAN; ZISSERMAN, 2014).

Por exemplo, três filtros 3x3 empilhados com *stride* (distância de deslize de um bloco em *pixels*) 1 tem um tamanho receptivo de 7, mas o número de parâmetros envolvidos é $3 \times (9c^2)$ em comparação com $49c^2$ parâmetros de *kernels* com um tamanho 7x7. Aqui, presume-se que o número de canais de entrada e saída de camadas é c . Além disso, os *kernels* 3x3 ajudam a reter propriedades de nível mais fino da imagem.

As camadas convolutivas VGG são seguidas por 3 camadas totalmente conectadas. A largura da rede começa em um pequeno valor de 64 e aumenta em um fator de 2 após cada camada de sub-amostragem (*pooling*). O passo é fixado em 1 e o tamanho de filtro em janelas de *pixels* com dimensão 3x3. As camadas de sub-amostragem que totalizam cinco ao todo, o processo é realizado sobre áreas 2x2 com passo do mesmo tamanho, ou seja, sem sobreposição.

Figura 19 – Arquiteturas VGG.

Arquiteturas	
VGG-16	VGG-19
conv3-64	conv3-64
conv3-64	conv3-64
maxpool	
conv3-128	conv3-128
conv3-128	conv3-128
maxpool	
conv3-256	conv3-256
conv3-256	conv3-256
conv3-256	conv3-256
	conv3-256
maxpool	
conv3-512	conv3-512
conv3-512	conv3-512
conv3-512	conv3-512
	conv3-512
maxpool	
conv3-512	conv3-512
conv3-512	conv3-512
conv3-512	conv3-512
	conv3-512
maxpool	
FC-4096	
FC-4096	
FC-1000	
soft-max	

Fonte: Adaptado de: (SIMONYAN; ZISSERMAN, 2014)

3.3.3.2 Arquitetura Inception

Redes muito profundas começaram a ser propensas a *overfitting* no treinamento, além de tornar difícil a propagação de gradiente por toda rede (SZEGEDY et al., 2015). Na tarefa de reconhecimento, partes representativas da classe alvo podem ter uma variação muito grande no seu tamanho em diferentes imagens. Por exemplo, uma imagem de um cachorro pode aparecer conforme mostrado Figura 20. A área ocupada pelo cão é diferente em cada exemplo.

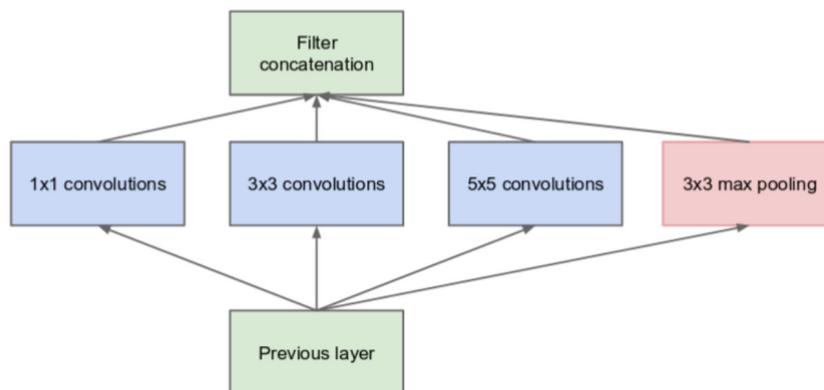
Para sanar problemas como os mencionados anteriormente, foi desenvolvido pela GoogLeNet um módulo denominado *inception* (SZEGEDY et al., 2015) que apresenta filtros com vários tamanhos operando no mesmo nível, tornando a arquitetura mais larga, em vez de mais profunda. A arquitetura, então, usa convoluções de tamanhos diferentes para extrair detalhes em diversas escalas. O módulo *inception* pode ser visto na Figura 21. Na ilustração, pode-se notar que o módulo explora três convoluções de tamanhos diferentes de filtros (1x1, 3x3, 5x5). Além disso, um *max pooling* (usa função de máximo valor) também é executado. As saídas são concatenadas e enviadas para o próximo módulo.

Figura 20 – Um animal alvo que ocupa a maior parte da imagem, depois apenas parte dela e então pouquíssimo espaço.



Fonte: (RAJ, 2018)

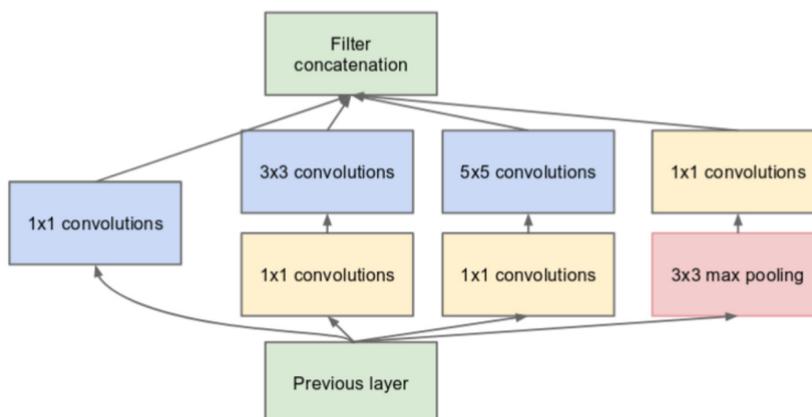
Figura 21 – Módulo *Inception* Simples.



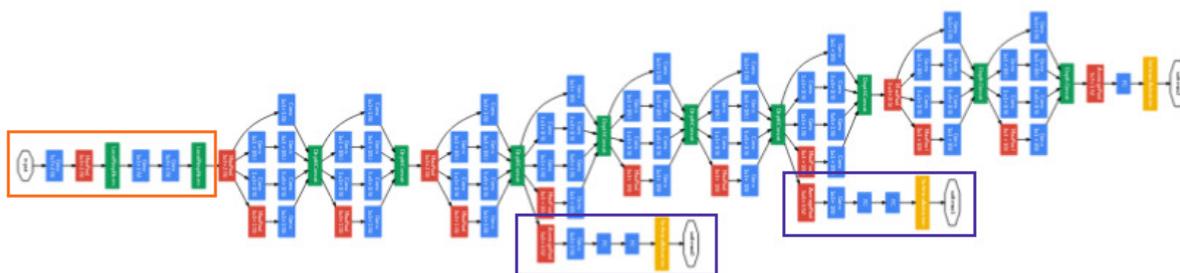
Fonte: (SZEGEDY et al., 2015)

com o intuito de simplificar o custo computacional das redes neurais profundas, Szegedy et al. (2015) limitaram o número de canais de entrada, adicionando uma convolução de *kernel* 1x1 anteriormente às outras convoluções, como ilustra a Figura 22.

Outra mudança que a Inception fez, foi substituir as camadas totalmente conectadas no final por um simples *average pooling* global que mede os valores do canal no mapa de características 2D, após a última camada convolucional. Isso reduz o número total de parâmetros. Os autores introduziram ainda classificadores auxiliares (os retângulos azuis na Figura 23). Esses classificadores intermediários aplicam o *softmax* às saídas de dois módulos *inception* e calculam um custo auxiliar sobre os rótulos dos exemplos. A função de custo total é uma soma ponderada do erro auxiliar e do erro real.

Figura 22 – Módulo *Inception* para redução de dimensionalidade.

Fonte: (SZEGEDY et al., 2015)

Figura 23 – Arquitetura completa de uma rede *Inception*.

Fonte: (SZEGEDY et al., 2015)

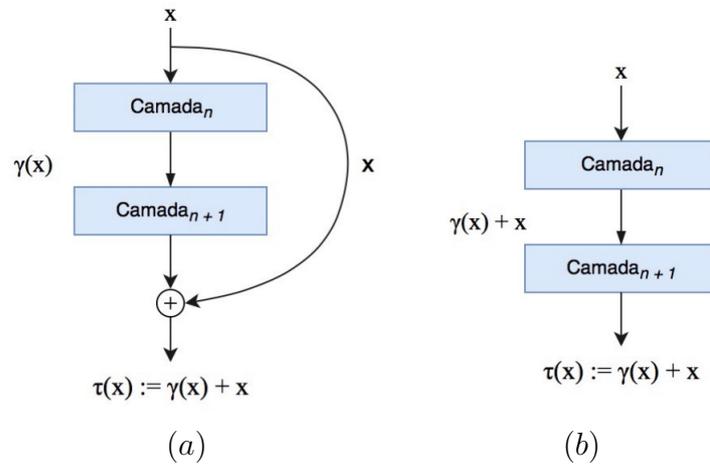
3.3.3.3 Arquitetura ResNet

Ao longo do tempo, as arquiteturas de CNN ficam cada vez mais profundas. Com o aumento da profundidade, deve aumentar a precisão da rede, desde que se tome cuidado com o *over-fitting*. Mas o problema com o aumento da profundidade é que o sinal necessário para atualizar os pesos torna-se muito pequeno nas camadas anteriores. Esse sinal surge do final da rede ao comparar a predição com o valor esperado. Significa, essencialmente, que as camadas anteriores quase não sofrem aprendizado. Isso é chamado de *vanishing gradient* (desaparecimento do gradiente).

O segundo problema com o treinamento de redes mais profundas é realizar a otimização em espaços de parâmetros enormes e, ainda, adicionar ingenuamente camadas que levam a um maior erro de treinamento, problema de degradação. As redes residuais (HE et al., 2016) permitem o treinamento de redes muito profundas construindo-a através

de módulos chamados de modelos residuais como mostrado na Figura 24 (a), dando nome à arquitetura, Redes Neurais Convolucionais Residuais (ResNet).

Figura 24 – Módulo Residual. (a) Fluxo Residual, (b) Fluxo Convencional



Fonte: Adaptado de: (HE et al., 2016)

Imagine uma rede A que produz um erro x . Uma rede B é criada apenas adicionando camadas em A e colocando valores de parâmetros que não causam nada à saída de A, chamando de C. Isso significaria o mesmo erro de treinamento x para a nova rede. Então, enquanto se treina B, o erro de treinamento deveria ser acima do erro de A. E, como isso realmente acontece, a única razão é que aprender o mapa de identidade (fazendo nada à entrada e apenas copiando os valores) com as camadas adicionadas a C não é um problema trivial, que o classificador não consegue atingir.

Para resolver isso, o módulo mostrado na Figura 24 (a) cria um caminho direto entre a entrada e a saída para o módulo, que implica um mapeamento de identidade, enquanto que a camada adicionada C só precisa aprender as características em cima da entrada já disponível. Uma vez que C está aprendendo apenas o residual, todo o módulo é chamado de módulo residual. Na Figura, $\tau(x)$ é o valor que se deseja obter, x a entrada e $\gamma(x)$ o resíduo. Formalmente, podemos definir aprendizado residual como sendo:

$$\tau(x) = \gamma(x, W_{ij}) + x, \quad (3.9)$$

onde W representa os pesos associados às camadas de i até j à serem aprendidos. Experimentos realizados em (HE et al., 2016) mostram que o uso de redes neurais residuais profundas convergem mais rapidamente do que redes neurais planas (convencionais) e exibem erros inferiores durante as etapas de treino e a validação.

3.4 Balanceamento de Classes

Aprender com conjuntos de dados desbalanceados é um grande desafio para muitos dos aplicativos atuais de mineração de dados (PROVOST, 2000). A fase de aprendizado e a predição subsequente de algoritmos de aprendizado de máquina podem ser afetadas pelo problema do conjunto de dados desbalanceados, que corresponde à diferença do número de amostras nas diferentes classes. O desafio se manifesta em duas formas comuns: classe de interesse minoritária e instâncias raras (HE et al., 2008).

Em análise de dados biomédicos, as amostras de informações para diferentes tipos de câncer são normalmente muito limitadas (casos raros) em comparação com casos não cancerosos normais; portanto, a razão entre a classe minoritária e a classe majoritária pode ser significativa, na proporção de 1 para 1.000 ou até mais (PEARSON; GONEY; SHWABER, 2003). Por outro lado, é essencial prever a presença de cânceres ou classificar os diferentes tipos de câncer o mais preciso possível para um tratamento mais precoce e apropriado.

Para lidar com conjuntos de dados desbalanceados, existem tipos de métodos, como os de amostragem (KUBAT; MATWIN et al., 1997), os de movimento de limiares de decisão (MALOOF et al., 2003) (HE; GARCIA, 2008) e os de ajuste de matrizes de custo (MALOOF et al., 2003). O primeiro conjunto de métodos tem como objetivo reduzir o desequilíbrio de dados, removendo instâncias da classe majoritária ou sobre-amostrando (sintetizando) as instâncias de treinamento da classe minoritária ou ambos. O segundo grupo de métodos tenta adaptar os limiares de decisão para impor pré-conceitos à classe das minorias. Da mesma forma, a terceira vertente de métodos melhora o desempenho da previsão ajustando o peso (custo) para cada classe.

3.4.1 Sobre-amostragem de Dados com SMOTE

Para balancear o número de exemplos de diferentes classes, vários algoritmos de sobre-amostragem foram propostos (HUANG et al., 2004) (HE et al., 2008). Neste trabalho, foi abordado o algoritmo *Synthetic Minority Oversampling TEchnique* (SMOTE) (CHAWLA et al., 2002) que é capaz de gerar novas amostras por interpolação das instâncias existentes.

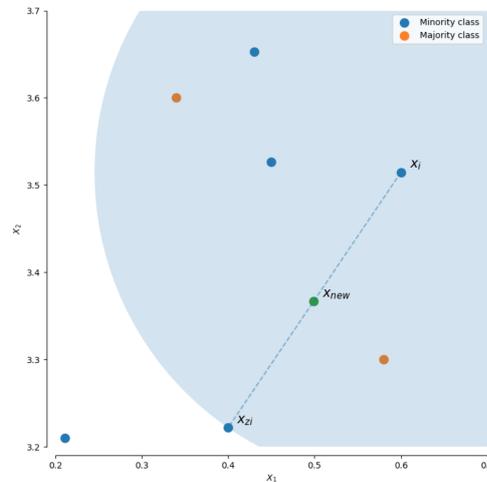
Considerando uma certa vizinhança de tamanho k em relação a um exemplo pivô, uma nova amostra será gerada por meio de seus k -vizinhos mais próximos. Matematicamente, tem-se a Equação 3.10:

$$X_{new} = X_{pivot} - \lambda(X_{ki} - X_{pivot}), \quad (3.10)$$

onde λ é um número aleatório no intervalo $[0, 1]$ e X_{new} é a interpolação que será criada como uma amostra na linha entre X_{pivot} e seu vizinho X_{ki} . De forma gráfica, pode-se

visualizar o funcionamento da criação de um novo exemplo sintético na Figura 25.

Figura 25 – Interpretação gráfica de um exemplo sintético com o SMOTE.



Fonte: (LEMAITRE et al., 2017)

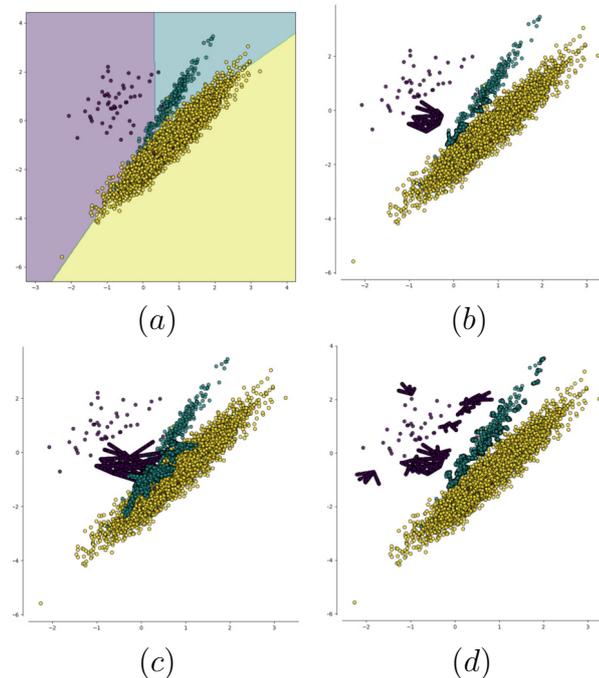
Considere a região azul da imagem como sendo a vizinhança de x_i (x_{pivot}) e x_{zi} o vizinho selecionado para participar da geração de um novo dado. Os pontos azuis representam a classe minoritária, enquanto que os laranjas, a majoritária. As implementações do SMOTE diferem principalmente na maneira como o conjunto de pivôs é selecionado.

Existem três variantes de aplicações do algoritmo SMOTE (LEMAITRE et al., 2017): O regular, *borderline* (que ainda apresenta dois tipos) e SVM. O **regular** não apresenta nenhuma regra especial e selecionará de forma aleatória todos os possíveis x_{pivot} disponíveis. Para o **borderline**, o algoritmo fará uma classificação da vizinhança de cada elemento x_{pivot} . Os vizinhos poderão ser:

- *Ruídos*: vizinhos de classes diferentes da do exemplo x_{pivot} ;
- *Perigosos*: Pelo menos metade dos vizinhos são da mesma classe que x_{pivot} ;
- *Seguros*: Todos os vizinhos são da mesma classes que x_{pivot} .

Depois da classificação, essa versão considera apenas vizinhos perigosos. O tipo *borderline* 1 seleciona x_{zi} (vizinhos) de classes iguais à de x_{pivot} , enquanto que o *borderline* 2 utiliza vizinhos de qualquer classe. Em relação ao do tipo **SVM**, o algoritmo explora um SVM treinado para extrair vetores de suporte para a geração de dados sintéticos. Na Figura 26, são apresentadas os principais variantes do algoritmo SMOTE.

Figura 26 – Aplicações das variantes do SMOTE. (a) Amostras, (b) *Borderline 1*, (c) *Borderline 2*, (d) SVM.

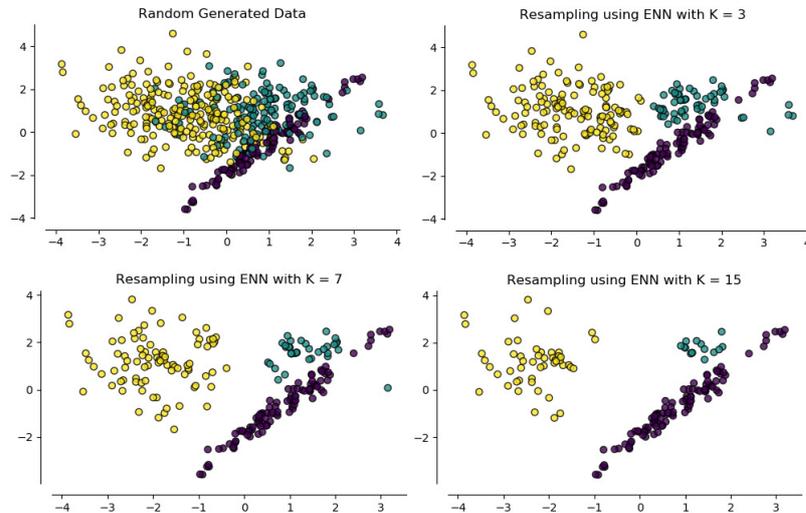


Fonte: (LEMAITRE et al., 2017)

3.4.2 Subamostragem com ENN

Após a geração de dados sintéticos, podem ser geradas possíveis amostras ruidosas (HUANG et al., 2004). Por esse motivo, os algoritmos de subamostragem também se tornaram importantes no processo de balanceamento de bancos de dados. Regras da Vizinhaça mais Próxima (do inglês - *Nearest Neighbor Rules*) são uma coleção de regras simples que pode ter um desempenho satisfatório com apenas algumas amostras pré-classificadas (WILSON, 1972).

Este estudo investiga a regra *Edited Nearest Neighbor* (ENN) (WILSON, 1972), que é usada como estratégia de limpeza de vizinhança capaz de remover qualquer exemplo cujo valor de classe difere da classe de pelo menos dois de seus três vizinhos mais próximos, adotado como critério. No entanto, para aumentar o grau de limpeza de dados e fornecer mais controle ao algoritmo, o ENN pode ser modificado para reduzir um conjunto de dados, apagando amostras que não estão em conformidade com os k -vizinhos mais próximos, por exemplo. Quanto maior o valor de k , mais amostras serão eliminadas do conjunto de dados. Como resultado, uma borda mais simples entre as classes é criada, como mostra a Figura 27.

Figura 27 – Subamostragens com ENN usando valores diferentes para k .

Fonte: (LEMAITRE et al., 2017)

3.5 Seleção de Características

Também chamada de seleção de variáveis ou seleção de atributos, é a escolha automática de características que são mais relevantes para o problema de modelagem preditiva em que se está trabalhando (JOHN; KOHAVI; PFLEGER, 1994).

Os métodos de seleção de variáveis ajudam na criação de um modelo preditivo preciso, exigindo menos dados. Podem ser usados para identificar e remover atributos desnecessários, irrelevantes e redundantes que não contribuem para o desempenho de algoritmo de aprendizagem de máquina. Poucos atributos são preferíveis porque reduzem a complexidade do modelo, e um modelo mais simples é melhor de entender e explicar (GUYON; ELISSEEFF, 2003).

Existem três classes gerais de algoritmos de seleção de características: métodos de filtro, métodos *wrappers* e métodos embutidos (JOHN; KOHAVI; PFLEGER, 1994).

- **Métodos de filtro:** aplicam uma medida estatística para atribuir um valor a cada atributo, que são classificados pela pontuação e selecionados para serem mantidos ou removidos do conjunto de dados. Os métodos geralmente são univariados (permitem a análise de cada variável separadamente), considerando cada característica de forma independente, ou com relação à variável dependente. Alguns exemplos de alguns métodos de filtragem incluem o teste do Qui-Quadrado (PEARSON, 1900), o Ganho de Informação (QUINLAN, 1986) e os resultados dos coeficientes de correlação;
- **Métodos *wrappers*:** consideram a seleção de um conjunto de atributos como

um problema de busca em um espaço de subconjunto de variáveis, em que várias combinações são montadas, avaliadas e comparadas umas às outras. Um modelo preditivo é usado para avaliar uma combinação de características e atribuir uma pontuação com base no desempenho do modelo.

- **Métodos embutidos:** aprendem quais características melhor contribuem para o desempenho de um modelo, enquanto este ainda está sendo criado. Os tipos mais comuns, para esse caso, são os métodos de regularização, cujo enfoque é introduzir restrições adicionais na otimização de um algoritmo de aprendizagem de máquina, que influencia o modelo em direção a uma complexidade menor (menos coeficientes). Exemplos de algoritmos de regularização incluem *Ridge Regression* (HOERL; KENNARD, 1970), LASSO (do inglês - *Least Absolute Shrinkage and Selection Operator*) (TIBSHIRANI, 1996) e *Elastic Net* (ZOU; HASTIE, 2005).

Neste estudo, abordam-se alguns métodos de seleção de características do tipo *wrappers*, que são descritos nas seções a seguir.

3.5.1 Greedy Stepwise

O *Greedy Stepwise* é um algoritmo de paradigma que segue uma heurística de resolução de problemas de modo a fazer a escolha que lhe parece ótima a cada estágio (ótimo local) (GUTIN; YEO; ZVEROVICH, 2002) com a intenção de encontrar um ótimo global. Em muitos problemas, uma estratégia como a descrita geralmente não produz a melhor solução global, mas mesmo assim uma heurística gananciosa pode produzir soluções ótimas localmente que se aproximam de uma solução geral ótima em um período de tempo razoável.

Uma análise discriminante, que utiliza informações das classes associadas a cada padrão para extrair linearmente os atributos mais relevantes, pode ser feita através da computação de uma combinação linear de m variáveis quantitativas que mais separam grupos de amostras em um espaço m -dimensional de forma eficiente, fazendo com que a razão da variância intra-classes e inter-classes seja maximizada (GUTIN; YEO; ZVEROVICH, 2002). O problema se torna, então, definir uma série de atributos que melhor represente esse conjunto de dados. A ideia básica que envolve a análise discriminante é determinar o quão as classes são distintas em relação à média de uma variável, e em seguida usar essa mesma variável adequando-a a um grupo para a nova amostra.

Existem dois tipos de métodos que podem ser utilizados para determinar uma função discriminante: o método simultâneo (direto) e o método *stepwise*. O método direto se caracteriza pela computação da função discriminante e as variáveis independentes são consideradas ao mesmo tempo. Dessa forma, a função discriminante é computada baseada em todo o conjunto de variáveis independentes, sem consideração do poder discriminatório

de cada variável (HAIR; ANDERSON; TATHAM, 2005). O método *stepwise* avalia a relevância de um determinado conjunto de dados ao passo-a-passo, de forma que, a cada iteração, tem-se um novo conjunto de variáveis.

O *Greedy Stepwise* executa uma busca de forma gananciosa de acordo com o método *stepwise*, usando o sentido *forward* ou o *backward*, através do espaço de busca dos subconjuntos de atributos. No caso do *forward*, a busca se inicia com poucos atributos ou de um ponto arbitrário no espaço, e encerra quando todas as inclusões de atributos possíveis resultarem em uma diminuição na avaliação. No caso do sentido *backward*, o conjunto inicial é igual ao vetor inteiro de características e, a cada passo, um atributo é removido do conjunto. Este algoritmo encontra, utilizando uma análise discriminante, as características que melhor representam as classes de imagens dermatoscópicas para o conjunto de características geradas, contendo menos redundâncias que poderiam prejudicar os classificadores durante as próximas etapas (AMBELU; LOCK; GOETHALS, 2010),.

3.5.2 PSO

O PSO (do inglês *Particle Swarm Optimization*) é uma técnica de computação evolutiva (EC para *Evolutionary Computation*) proposta por Kennedy e Eberhart (1995), inspirada no comportamento de enxames, como os bandos de pássaros e os cardumes de peixes (KENNEDY; EBERHART, 1995).

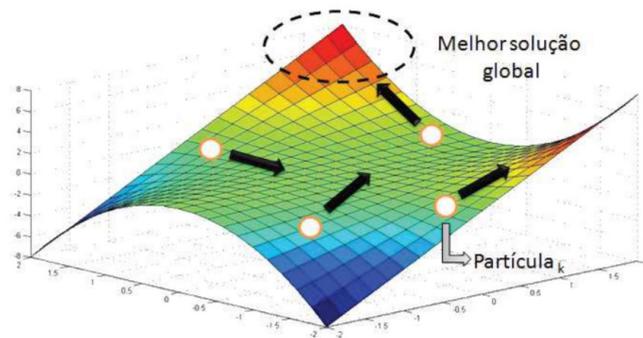
O algoritmo PSO se baseia no princípio de que cada solução pode ser representada como uma partícula em um determinado enxame. Cada partícula possui sua posição no espaço de busca, que é representada por um vetor $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, onde D é a dimensionalidade do espaço de pesquisa. As partículas se deslocam no espaço de busca para investigar as soluções ideais. Portanto, cada partícula tem uma velocidade, que é representada como $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$. Durante o movimento, cada partícula atualiza sua posição e velocidade de acordo com a sua própria experiência e a dos seus vizinhos, baseado na melhor posição global entre as partículas no enxame (DURAISAMY; KAYALVIZHI, 2010). A Figura 28 apresenta o processo de busca do enxame pela melhor solução global.

A melhor posição anterior da partícula é registrada como o melhor p_{best} individual, e a melhor posição obtida pela população inteira até um determinado momento é chamada de g_{best} . Baseado no p_{best} e no g_{best} , o PSO procura as soluções ótimas atualizando a velocidade e a posição de cada partícula de acordo com as seguintes equações:

$$x_i = x_i + v_i \quad (3.11)$$

$$v_i = w.v_i + (C_1.r_1.(p_{best} - x_i)) + (C_2.r_2.(g_{best} - x_i)) \quad (3.12)$$

Figura 28 – Processo de busca do algoritmo PSO.



Fonte: (SILVA et al., 2017)

onde, w é o peso de inércia, que serve para controlar o impacto de velocidades anteriores na atual, C_1 é o parâmetro cognitivo que interfere cada partícula em relação à melhor posição local, C_2 é o parâmetro coletivo que atrai cada partícula para a melhor posição global da população, p_{best} é a melhor posição da partícula i , g_{best} é a melhor posição do enxame e r_1, r_2 são números sorteados entre 0 e 1. A condição de parada do algoritmo é quando um critério predefinido é atendido, podendo ser um bom valor de *fitness* (medida que informa o quão bom uma determinada solução é) ou um número máximo de iterações.

Deve-se entender que nem todos os padrões extraídos por algum método para tarefa de reconhecimento são totalmente relevantes. O algoritmo PSO, cuja implementação foi baseada em (MORAGLIO; CHIO; POLI, 2007), foi utilizado neste trabalho para otimizar e encontrar as melhores características adequadas para classificação.

3.5.3 Ranker e RankSearch

São algoritmos de seleção de atributos que funcionam de forma a ordenar os atributos pelo valor de sua relevância. O *Ranker* difere do *RankSearch* (HALL; HOLMES, 2003) em relação ao método de busca de conjunto de variáveis para serem avaliados. O primeiro não utiliza métodos de busca, pois apenas avalia individualmente a característica, calculando taxa de ganho de informação, ou entropia, para uma determinada variável. O segundo algoritmo usa um avaliador de subconjuntos para classificar todos os atributos. Da lista classificada de atributos, subconjuntos de tamanho crescente são avaliados. O melhor conjunto de atributos é retornado como resposta.

3.6 Algoritmos de Aprendizagem de Máquina

Reconhecimento de padrões iniciou na Engenharia, enquanto que aprendizado de máquina cresceu na área da Ciência da Computação (BISHOP, 2006). Apesar de serem

dois assuntos distintos, podem ser vistos como duas faces da mesma moeda e têm tido crescimento substancial em vários campos nos últimos dez anos. São técnicas com intuito de encontrar padrões para descrever um determinado objeto a partir de características selecionadas na etapa de extração de descritores (BRAZ JUNIOR et al., 2014).

Os algoritmos de aprendizagem de máquina investigam um conjunto de dados atrás de parâmetros que definem classes distintas dispostas de forma visível ou não dentro desse espaço de busca. Esse processo é geralmente dividido em duas fases: uma, chamada de treinamento, na qual um subconjunto dos dados disponíveis são introduzidos a um algoritmo de aprendizado de máquina qualquer que vai tentar separar grupos existentes denominados classes, e outra fase chamada de classificação, onde um novo subconjunto é apresentado ao modelo para que seja determinada a classe de cada indivíduo testado.

O treinamento pode ser feito de diferentes maneiras. Entre elas estão as formas supervisionadas e não-supervisionadas. Em um treinamento supervisionado, o algoritmo de aprendizagem é apresentado a um par (X, Y) , onde X é conjunto de características e Y é o valor de sua respectiva classe. Com os valores dos rótulos previamente conhecidos e das características para cada grupo, a máquina de reconhecimento tenta achar uma associação ou padrão entre uma determinada classe e o conjunto de informações dos indivíduos que pertencem aquele grupo.

Na abordagem não-supervisionada, as classes existentes no conjunto de dados não são conhecidas, sendo, portanto, necessário que o algoritmo de reconhecimento seja capaz de encontrar similaridades entre diferentes indivíduos para que, então, eles possam ser agrupados e as classes definidas por características gerais presentes em cada subgrupo. Essa forma de treinamento é também conhecida simplesmente por agrupamento ou por processo auto-organizável.

Com o modelo devidamente treinado, um novo exemplo, totalmente desconhecido e que não participou da fase de treinamento, é apresentado ao algoritmo que, a partir do conhecimento prévio obtido pela generalização das características usadas, rotula aquele exemplo para uma das classes conhecidas pelo método.

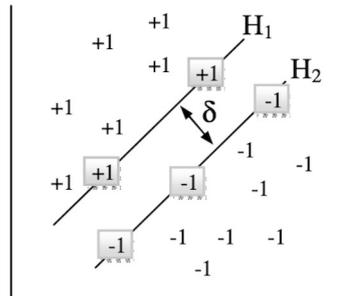
A seguir são apresentados os principais conceitos dos algoritmos de reconhecimento de padrões utilizados neste estudo.

3.6.1 Máquinas de Vetor de Suporte

As Máquinas de Vetor de Suporte, do inglês *Support Vector Machines* (SVM) (NAUMOVICH; VAPNIK, 1998), compõem uma técnica com fundamentação na teoria do aprendizado estatístico, introduzida por Vapnik e colaboradores, que despertou grande interesse da comunidade de aprendizado de máquina (SOARES, 2008) e tem como objetivo maximizar a capacidade de generalização.

Esse classificador realiza a separação de duas classes através de um hiperplano que maximiza uma margem denotada por δ . Esse hiperplano está localizado entre as semi-retas H_1 e H_2 , que separam as classes +1 e -1 respectivamente, atingindo seu ótimo quando a distância para as duas linhas for máxima. Dois pontos da classe +1 definem a reta H_1 e três, da classe -1, definem a reta H_2 . Esses pontos são chamados de vetores de suporte (SV - *Support Vector*) (Figura 29).

Figura 29 – Hiperplano de separação para padrões lineares.



Fonte: (SOARES, 2008)

Ilustra-se, na Figura 29, um exemplo de dados linearmente separáveis. Em caso de padrões não-lineares, as variáveis de entrada são transformadas para um domínio com maior dimensionalidade, cujo objetivo é, então, conseguir separar as classes nesse novo espaço. Porém, encontrar essa função de transformação não é trivial. Essa tarefa acontece com o auxílio de funções não-lineares denominadas *Kernels*, que torna possível a construção de um hiperplano de separação ótimo para o espaço de características sem considerar explicitamente o espaço.

Considerando um conjunto de treinamento (x_i, y_i) , onde $x_i \in R^N$ é o vetor de entrada, $y_i \in \{\pm 1\}$ os rótulos de cada exemplo e N o tamanho da amostra, procura-se estimar a função dada por $f : R^N \rightarrow \{\pm 1\}$ capaz de separar de forma correta as classes dos exemplos informados.

No treinamento a função $f(x) = y_i(w \cdot x_i + b)$ é estimada de forma que w e b satisfaçam a seguinte inequação:

$$y_i(wx_i + b) \geq 1 \quad (3.13)$$

onde w é um vetor normal ao hiperplano de separação e b a distância ou corte da função f em relação à origem. Tem-se agora um problema de otimização dos valores w e b por meio da minimização da Equação 3.14, sujeita à restrição dada pela Equação 3.13 (BISHOP, 2006).

$$\phi(w) = \frac{w^2}{2} \quad (3.14)$$

Como o objetivo é produzir um classificador que apresenta um bom desempenho em relação a amostras excluídas durante o processo de treinamento, ou seja, um classificador com bom grau de generalização, é importante que o SVM seja capaz de minimizar erros de classificação, quando a separação perfeita das classes não for viável. Isso é resolvido, inserindo variáveis de folga de modo que a relação dada pela Equação 3.13 seja alcançada.

Portanto, o novo sistema a ser otimizado é:

$$\phi(w, \xi) = \frac{w^2}{2} + C \sum_{i=1}^N \xi_i \quad (3.15)$$

sujeito a:

$$y_i((w \cdot x_i + b) + \xi_i) \geq 1 \quad (3.16)$$

onde C é um parâmetro de treinamento que estabelece um equilíbrio entre a complexidade do modelo e o erro de treinamento, sendo um valor fornecido pelo próprio usuário.

A função objetivo a ser minimizada é convexa e possui apenas um único mínimo global (LORENA; CARVALHO, 2007), onde pode ser aplicada uma função Lagrangiana, que associa parâmetros denominados de Lagrange α_i (Equação 3.17), englobando as restrições à própria função objetivo.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w \cdot x_i + b) - 1) \quad (3.17)$$

Ao minimizar a função Lagrangiana, implicando em maximizar α_i ao mesmo tempo que minimiza w e b , observa-se ao seguinte ponto de sela, onde:

$$\frac{\partial L}{\partial w} = 0 \quad \text{e} \quad \frac{\partial L}{\partial b} = 0 \quad (3.18)$$

Resolvendo as equações, chega-se às seguintes expressões:

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (3.19)$$

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (3.20)$$

Substituindo as Equações 3.19 e 3.20 em 3.17, tem-se o seguinte problema:

$$\text{Maximizar}_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (3.21)$$

Com restrições:

$$\begin{cases} \alpha_i \geq 0, i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (3.22)$$

Essa é chamada a fórmula dual do problema de otimização que possui restrições mais simples e é definida em função dos dados de treinamento e dos rótulos (LORENA; CARVALHO, 2007).

Os vetores de suporte são pontos em que a restrição dada pela Equação 3.16 é igual à unidade e seus α correspondentes são diferentes de 0 e são utilizados para a definição do separador ótimo.

Os *Kernels*, funções utilizadas para separação não-linear através da transformação de dados em um novo espaço de características, são incluídos na fórmula de otimização da seguinte maneira:

$$\text{Maximizar}_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\phi(x_i) \cdot \phi(x_j)) \quad (3.23)$$

Dentre as funções de Kernel ϕ disponíveis, temos o RBF, do inglês *Radial Base Function*, função de base radial que é bastante explorada em problemas de reconhecimento de padrões e é definida por:

$$R(x_i, x_j) = \epsilon^{-\gamma(x_i - x_j)^2} \quad (3.24)$$

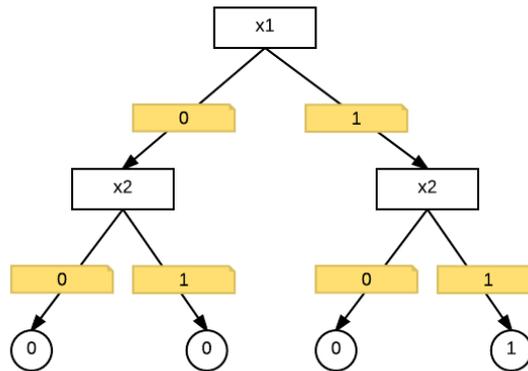
onde γ representa um parâmetro regulador de complexidade, informado a priori pelo usuário.

3.6.2 Random Forest

A árvore, uma das estruturas de dados mais comuns e poderosas em toda a ciência da computação, é usada como base para criação desse algoritmo, onde a classificação é dividida em um conjunto de escolhas sobre cada característica, partindo de sua raiz (*root*) e avançando até as folhas, das quais são tomadas a decisão de classificação (MARSLAND, 2015). As árvores apresentam vantagens muito grandes: além de serem fáceis de entender (no que tange o caminho feito pelo classificador para chegar a uma conclusão), podem ser transformadas em um conjunto de comandos *se-então*, adequadas para uso em um sistema de indução de regras. Na Figura 30, é possível visualizar um exemplo básico de árvore de decisão para uma porta lógica de função AND, cujas entradas de sinais são x1 e x2.

Em termos de busca e otimização, as árvores de decisão exploram heurísticas "gulosas", avaliando as opções disponíveis em um estágio atual de aprendizagem e tomando aquela que aparenta ser ótima naquele ponto (MARSLAND, 2015). O método como uma

Figura 30 – Árvore de Decisão para uma porta lógica AND.



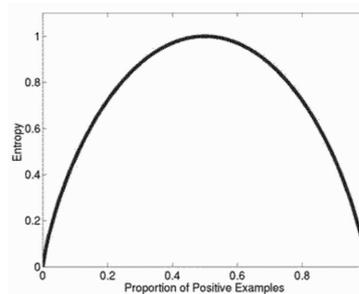
Fonte: Elaborada pelo autor

opção é escolhida em detrimento de outra é chamado de Entropia de Informação, proposto por [Shannon \(1948\)](#). A entropia calcula a quantidade de "impureza" (mistura) em um determinado conjunto de características, definida pela Equação a seguir:

$$H(p) = - \sum_i p_i \log_2 p_i, \quad (3.25)$$

onde H é a entropia para um conjunto de probabilidades p_i , a base 2 representa codificação para duas classes possíveis. O conceito básico é que o valor de H representa quanta informação extra pode-se obter ao escolher uma determinada característica. Um gráfico de entropia pode ser visto na Figura 31.

Figura 31 – Gráfico de entropia.



Fonte: ([MARSLAND, 2015](#))

Suponha-se que há um conjunto de instâncias positivas e negativas em relação a alguma característica dada (sexo, por exemplo). Se todos os exemplos forem positivos ou negativos, não se teria informações extras para saber dessa característica para um determinado exemplo; pois, independentemente do valor, o exemplo será positivo ou negativo, cuja entropia é 0. Entretanto, se a característica separar os exemplos em metade positivos e metade negativos, a entropia calculada será máxima, sendo assim o descritor

teria maior ganho de informação. O ganho é o valor palpável que representa a diferença entre a entropia de todo o conjunto de características e aquela que foi escolhida no momento, cuja Equação é:

$$Gain(S, F) = H(S) - \sum_{f \in valuesOf(F)} \frac{|S_f|}{|S|} H(S_f), \quad (3.26)$$

onde S é o conjunto de características (por exemplo altura e sexo), F é o conjunto de valores possíveis de cada uma (baixo, médio, alto para altura; masculino e feminino para sexo) e o sinal de módulo representa o número de elementos do conjunto.

Considere os exemplos que formam uma pequena base de dados na Tabela 2.

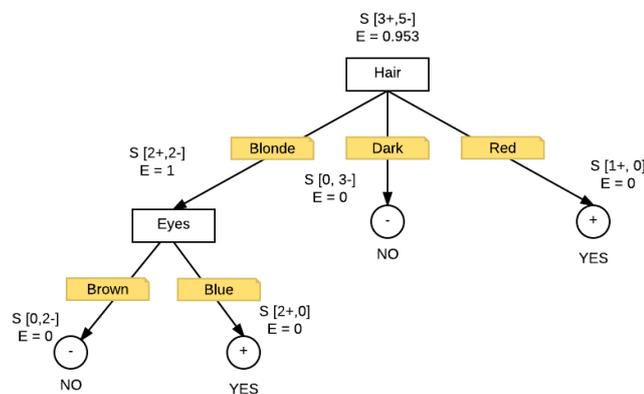
Tabela 2 – Exemplo de base de dados com instâncias de classes (*Attractive*) positivas e negativas.

Height	Hair	Eyes	Attractive?
Small	Blonde	Brown	No
Tall	Dark	Brown	No
Tall	Blonde	Blue	Yes
Tall	Dark	Blue	No
Small	Dark	Blue	No
Tall	Red	Blue	Yes
Tall	Blonde	Brown	No
Small	Blonde	Blue	Yes

Fonte: (MARSLAND, 2015)

Na Figura 32, pode-se ver uma árvore de decisão, apresentando valores de entropia calculada para cada característica (nó) referente aos valores da tabela anterior. Note que, quando o valor de entropia é 0, o valor da classe é dado para a instância restante e o nó se torna folha. Nem todas características são essenciais para a formação de uma boa árvore de decisão, pois o descritor *Height* não foi utilizado.

Figura 32 – Exemplo real de Árvore de Decisão.



Fonte: Elaborada pelo autor

No primeiro nível da árvore, deve-se saber qual característica teria maior ganho de informação. Usando a Equação 3.26, os valores para *Height*, *Hair* e *Eyes* são 0.003, 0.453 e 0.347, respectivamente, sendo escolhida portanto a variável *Hair* como raiz. O processo continua recursivamente até que haja apenas uma classe restante nos dados.

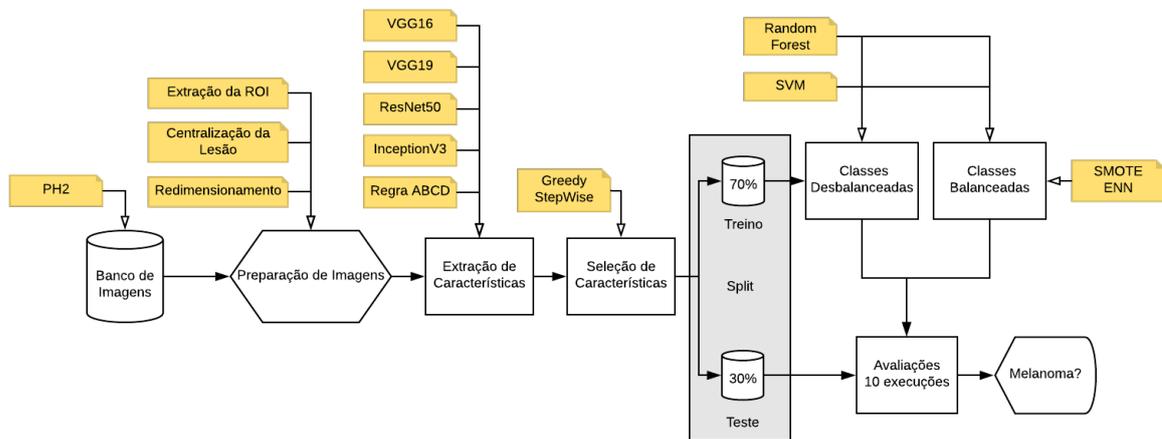
O *Random Forest* (BREIMAN, 2001) é um algoritmo de aprendizado supervisionado que explora diversas árvores de decisão e as mesclam para conseguir maior desempenho e uma precisão mais robusta. A denominação "floresta" (do inglês - *Forest*) trata-se de um conjunto de múltiplas árvores treinadas usando o método de *bagging* (BREIMAN, 1996), cujo intuito é combinar modelos de aprendizado para melhorar um resultado final.

Formalmente, *Random Forest* é um modelo (h) que consiste em uma coleção de classificadores estruturados em árvore $\{h(x, \theta_k), k = 1, \dots\}$, onde os $\{\theta_k\}$ são vetores aleatórios independentes e identicamente distribuídos. Cada árvore vota para a classe mais popular na entrada x (BREIMAN, 2001).

4 Metodologia

Este capítulo apresenta a metodologia proposta para diagnóstico de câncer de pele melanoma em imagens dermatoscópicas. O diagrama do fluxo do conjunto de métodos aplicado neste estudo é sumarizado na Figura 33. O fluxograma mostra o uso de um conjunto de imagens que passa por um passo de preparação, seguido por uma fase de extração de características, as quais serão utilizadas em uma etapa de seleção de características, cuja saída é usada para treinar os classificadores em dois cenários diferentes e, então, serão comparadas as performances dos algoritmos.

Figura 33 – Fluxograma da metodologia proposta.

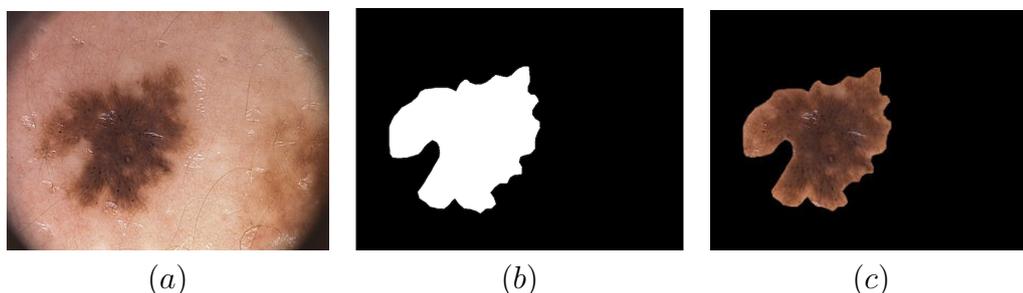


4.1 Aquisição e Preparação de Imagens

Para fins de treinamento e validação, foi usado o conjunto de imagens disponíveis na base PH² (MENDONÇA et al., 2013). Contém 200 exames de dermatoscopia catalogados em três classes diferentes: lesões normais (80 amostras), lesões atípicas (80 amostras) e melanoma (40 amostras). Para este estudo, o conjunto de dados foi dividido em imagens não-melanoma e melanoma, mesclando as classes de lesões normais e atípicas. A base de dados ficou, então, com 160 imagens Não-melanoma e 40 imagens Melanoma.

Juntamente com a base, as imagens vêm com uma máscara binária criada por especialistas, contendo a região da lesão, como mostra a Figura 34 (b). Para preparar as imagens para fase de extração de características, partes que não são relevantes ao problema são removidas. Neste caso, não se deseja extrair informações da pele e apenas da região de interesse (ROI) que é a lesão. Portanto, aplica-se a máscara ao exame original, mantendo-se apenas a ROI.

Figura 34 – Imagem da base PH²: (a) Dermatoscopia e (b) Máscara binária (b) Região de Interesse Segmentada.



Fonte: (MENDONÇA et al., 2013)

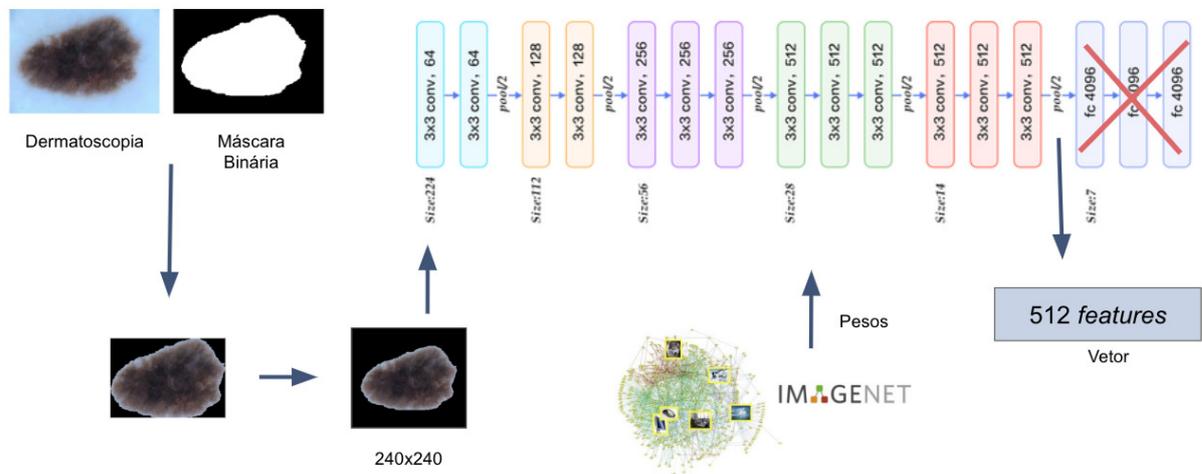
Após essa etapa de extração da ROI, como a base apresenta imagens de diferentes dimensões, aplicou-se, então, uma fase de padronização de tamanhos. Nesta etapa, adicionam-se bordas na vertical ou horizontal a fim de que as alturas e larguras fiquem do mesmo tamanho, sem alterar aspectos da lesão. Isso se faz necessário devido ao fato de que as arquiteturas de Redes Neurais Convolucionais usadas aceitam apenas imagens quadradas.

4.2 Extração e Seleção de Características

Neste trabalho, as arquiteturas de Redes Neurais Convolucionais usadas para extrair características das imagens foram: VGG16 e VGG19 (SIMONYAN; ZISSERMAN, 2014), InceptionV3 (SZEGEDY et al., 2015) e ResNet50 (HE et al., 2016) implementadas pela biblioteca Keras (CHOLLET et al., 2015), carregando pesos treinados no desafio da ImageNet (DENG et al., 2009). Essa técnica é chamada de *Transfer Learning* (MENEGOLA et al., 2017), onde, devido à quantidade insuficiente de dados para criar um novo modelo, o conhecimento de uma grande quantidade de dados conhecidos é transferido para um conjunto de novos dados. Obteve-se as seguintes quantidades de características por extrator: VGG16 (512), VGG19 (512), ResNet50 (2048), InceptionV3 (2048). A Figura 35 apresenta o fluxo de preparação das imagens para serem inseridas nas arquiteturas de CNNs.

Além das características extraídas com as CNNs, foi utilizado um outro descritor, levando em consideração a regra ABCD para análise de lesões de pele. Uma implementação dessa regra, adaptada de Moura et al. (2018), retorna 7 valores, sendo eles: **A**ssimetria (1 valor), **B**ordas irregulares (4 valores), **C**or (1 valor) e **D**iâmetro (1 valor). Para a distância d na contagem de vales e picos da borda o valor 15 foi escolhido. Para a variável de cor, os *pixels* foram divididos em 50 intervalos e o limiar utilizado para desconsiderar determinada cor foi 100. Os valores para tais parâmetros foram escolhidos empiricamente.

Figura 35 – Fluxo de preparação de uma imagem para extração de característica com a arquitetura VGG16.



Todas as características foram concatenadas, totalizando 5.127 descritores para cada imagem.

Logo após, foi realizada uma fase de seleção de atributos, usando os algoritmos chamados *Ranker*, *RankSearch*, *PSO* e *Greedy Stepwise*. O papel dessa etapa é executar uma busca em um espaço de subconjuntos de atributos. O processo termina, quando a adição ou exclusão de quaisquer atributos restantes resulta em uma diminuição no desempenho de um modelo base, no qual uma validação cruzada *k-fold* foi utilizada, onde se divide o conjunto total de dados em *k* subconjuntos mutuamente exclusivos de mesmo tamanho e, a partir disto, um subconjunto é utilizado para teste e os *k-1* restantes são utilizados para treinamento. Este processo é realizado *k* vezes alternando de forma circular o subconjunto de teste. Neste estudo, o valor de *k* utilizado foi 10.

O resultado dos algoritmos de seleção de características, explorados através da ferramenta WEKA (do inglês - *Waikato Environment for Knowledge Analysis*) (HALL et al., 2009), retorna uma lista contendo todos os atributos e o número de conjuntos (dos 10 *folds*, no caso dos algoritmos *RankSearch*, *PSO* e *Greedy Stepwise*) em que aquela determinada característica foi relevante. No caso do algoritmo *Ranker*, um valor entre 0 e 1 é dado para cada atributo de acordo com sua relevância para o modelo. Uma ilustração do que foi descrito pode ser vista na Figura 36.

Um valor de corte, *threshold*, foi determinado para cada algoritmo de seleção. Os atributos que apresentaram relevância inferior ao *threshold* foram eliminados do vetor de característica final. A Tabela 3 apresenta a quantidade de atributos que restaram depois da aplicação de cada seletor separadamente.

O valor de *Threshold* escolhido para *RankSearch*, *PSO* e *Greedy Stepwise* foi 1, representando o número de *folds* mínimo no qual uma determinada característica foi

Figura 36 – Fluxograma da extração e seleção de características.

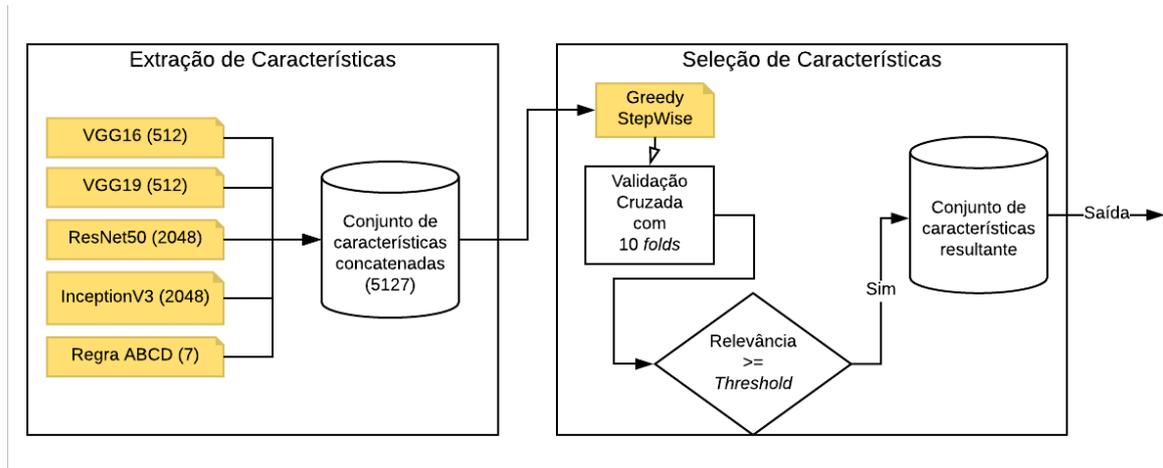


Tabela 3 – Número de características por seletor.

Seletor	Threshold	Nº de Características	Proporção
Ranker	0,1	677	13,20%
RankSearch	1	206	04,01%
PSO	1	916	17,86%
Greedy Stepwise	1	462	09,01%

relevante. Como a abordagem foi de 10-*folds*, neste caso, leva-se em consideração que o atributo seja importante em pelo menos um grupo. Para o *Ranker*, o valor de 0,1 foi escolhido empiricamente.

Para cada descritor de atributos é importante quantificar o número de variáveis que restaram após a aplicação de cada seletor de características. Na Tabela 4, pode-se observar o quantitativo de variáveis restantes por extrator, bem como a porcentagem relativa ao total de atributos de cada cenário.

Tabela 4 – Número de características por descritor de característica para cada seletor.

Seletores	Descritores	Nº de Características	Proporção
Ranker	ResNet	327	0,1597
	VGG16	129	0,2519
	VGG19	120	0,2343
	Inception	94	0,0456
	ABCD	7 (ABCD)	1,0000
RankSearch	ResNet	91	0,0444
	VGG16	38	0,0742
	VGG19	26	0,0508
	Inception	49	0,0240
	ABCD	2 (AC)	0,2857
PSO	ResNet	385	0,1880
	VGG16	112	0,2187
	VGG19	104	0,2031
	Inception	314	0,1533
	ABCD	1 (A)	0,1428
Greedy Stepwise	ResNet	172	0,0840
	VGG16	67	0,1308
	VGG19	42	0,0820
	Inception	179	0,0874
	ABCD	2 (AC)	0,2857

As iniciais ao lado do número de características para o descritor ABCD indicam quais variáveis permaneceram após o algoritmo de seleção de atributos. No caso, para o seletor *Greedy Stepwise*, as variáveis **A**ssimetria e **C**or foram mais relevantes do que **B**ordas e **D**iâmetro.

4.3 Experimentos

O vetor de características resultante da etapa descrita anteriormente foi usado para executar dois tipos de treinamento em dois algoritmos diferentes de aprendizado de máquina: SVM (do inglês - *Support Vectors Machine*) (NAUMOVICH; VAPNIK, 1998), cujo objetivo é usar o princípio de Minimização de Risco Estrutural (SRM - do inglês *Structural Risk Minimization*) para encontrar uma regra de decisão com boa capacidade de generalização, que é sempre única e globalmente ideal, ao contrário dos outros métodos tradicionais de rede estocástica ou neural (CAO; TAY, 2003); e o *Random Forest* (RF) (BREIMAN, 2001), que possui decisão baseada uma votação ponderada de diversas árvores, o que o faz ter bom desempenho mesmo em bases desbalanceadas.

O primeiro tipo de treinamento foi realizado utilizando o conjunto de treinamento desequilibrado. O segundo foi feito com a criação de um pipeline com sobre-amostragem dos dados de treinamento usando o SMOTE e, em seguida, limpando amostras ruidosas com o ENN. Para o propósito deste trabalho, o conjunto de pivôs escolhido para o SMOTE foi o *Support Vectors* (SVs) encontrado por uma Máquina de Vetores de Suporte treinada separadamente com os dados desbalanceados, pois instâncias longe da borda de decisão entre as duas classes não são relevantes para criação do modelo. O valor para k foi configurado como 3 para os algoritmos SMOTE e ENN, pois valores maiores não apresentaram melhora do desempenho dos modelos. Além disso, valor de k muito alto para o ENN tende a eliminar mais instâncias da base, aumentando a probabilidade de excluir exemplos reais que poderiam ser essenciais no treinamento dos classificadores. Depois disso, os exemplos que permaneceram na base foram introduzidos nos classificadores.

Os experimentos deste estudo foram repetidos dez (10) vezes. Em cada experimento a base foi dividida em 70% treino e 30% teste, onde cada classificador teve seus parâmetros automaticamente estimados usando a biblioteca *auto-sklearn* (FEURER et al., 2015) para gerar o melhor modelo de classificação.

4.4 Métricas de Desempenho

Para avaliar os classificadores em relação à sua capacidade de generalização, os critérios de acurácia, precisão, sensibilidade, especificidade e *f-score* foram analisados para cada solução apresentada. As fórmulas foram definidas como:

$$\text{Acurácia:} \quad \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

$$\text{Precisão:} \quad \frac{TP}{TP + FP} \quad (4.2)$$

$$\text{Sensibilidade:} \quad \frac{TP}{TP + FN} \quad (4.3)$$

$$\text{Especificidade:} \quad \frac{TN}{TN + FP} \quad (4.4)$$

$$\text{F-score:} \quad \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4.5)$$

onde TP (*True Positive*) significa os exemplos malignos que foram corretamente classificados como tal, TN (*True Negative*) significa os exemplos benignos corretamente classificados, FP (*False Positive*) e FN (*False Negative*) são respectivamente exemplos benignos classificados incorretamente como malignos e vice-versa. A Tabela 5 contém esses conceitos e é chamada de matriz de confusão.

Tabela 5 – Matriz de confusão para duas classes.

		Predição	
		Não-melanoma	Melanoma
Classes	Não-melanoma	TN	FP
	Melanoma	FN	TP

4.4.1 Índice Kappa de Cohen

O índice Kappa é uma alternativa ao cálculo de taxa de desempenho de classificação, além de ser um método, conhecido há décadas, que compensa acertos que podem ser atribuídos ao acaso (COHEN, 1960). O propósito original do índice se comprometia em medir o grau de concordância ou desacordo entre dois indivíduos observando o mesmo fenômeno. O kappa de Cohen pode ser adaptado para tarefas de classificação e é recomendado que seja empregado porque ele leva em consideração os sucessos aleatórios como um padrão (GARCÍA et al., 2009), da mesma forma que a medida da AUC (BEN-DAVID, 2007).

Segundo Ben-David (2008), existe uma correlação muito forte ainda entre o índice kappa e a curva ROC (do inglês - *Receiver Operating Characteristic*), que descreve uma

escolha de "perde-e-ganha" (*trade-off*) entre TPs e FPs. O valor de kappa pode ser definido como:

$$K = \frac{P_0 - P_c}{1 - P_c}, \quad (4.6)$$

onde P_0 é a probabilidade total de concordância, ou a acurácia (Equação 4.1), e P_c é a probabilidade de concordância que é devido ao acaso, que pode ser traduzida seguinte fórmula:

$$P_c = \frac{TP + FN}{D} \times \frac{TP + FP}{D} + \frac{FP + TN}{D} \times \frac{FN + TN}{D}, \quad (4.7)$$

onde D representa o total de exemplos utilizados no teste, ou seja, a soma de todas as células da matriz de confusão (Tabela 5). Segundo Landis e Koch (1977), existe uma categorização dos níveis de desempenho de classificação pelo índice Kappa que pode ser visualizada na Tabela 6.

Tabela 6 – Níveis de precisão de classificação, segundo o índice Kappa.

Índice Kappa (K)	Qualidade
$K < 0.2$	Ruim
$0.2 \leq K < 0.4$	Razoável
$0.4 \leq K < 0.6$	Bom
$0.6 \leq K < 0.8$	Muito Bom
$K \geq 0.8$	Excelente

5 Resultados

Neste capítulo, são apresentados os resultados e as respectivas análises realizadas durante o desenvolvimento do trabalho, categorizados por classificador, mostrando os diferentes cenários de treinamento. Para se obter um bom modelo de classificação, a capacidade de acertar a porcentagem de melanomas (sensibilidade) deve ser maximizada com a porcentagem correta de exemplos de casos normais (especificidade). A razão para essa escolha está relacionada ao fato de que é melhor detectar o melanoma, gerando ainda mais falsos positivos do que não perceber o melanoma em um paciente de risco.

Os resultados estão apresentados em diferentes tabelas, de acordo com o uso ou não da etapa de balanceamento de classes com o SMOTEENN e por classificador, contendo as médias e desvios padrões, adquiridos por meio das dez execuções realizadas para cada tipo de experimento. Nas Tabelas 7 e 8, são mostrados os resultados do algoritmo *Random Forest* para cada seletor sem e com o uso de balanceamento.

Tabela 7 – Resultados de *Random Forest* sem uso de SMOTEENN.

Seletor	Acurácia	Precisão	Sensibilidade	Especificidade	Fscore	Kappa
Ranker	0,8933 ± 0,0344	0,7970 ± 0,1512	0,6833 ± 0,1511	0,9458 ± 0,0512	0,7163 ± 0,0924	0,6568 ± 0,1084
RankSearch	0,9033 ± 0,0205	0,8319 ± 0,1111	0,6667 ± 0,0556	0,9625 ± 0,0274	0,7379 ± 0,0459	0,6768 ± 0,0585
PSO	0,9083 ± 0,0196	0,8353 ± 0,1124	0,7083 ± 0,1318	0,9583 ± 0,0340	0,7512 ± 0,0694	0,6967 ± 0,0757
Greedy Stepwise	0,9283 ± 0,0261	0,8688 ± 0,1037	0,7833 ± 0,1480	0,9646 ± 0,0326	0,8082 ± 0,0899	0,7656 ± 0,1004

Primeiramente, analisando os resultados, o processo de treinamento dos algoritmos tendem a favorecer a classe com mais dados amostrais, demonstrados pelos valores médios de especificidade de 94,58%, 96,25%, 95,83% e 96,46%, com o classificador RF, para os algoritmos de seleção *Ranker*, *RankSearch*, *PSO* e *Greedy Stepwise* respectivamente, enquanto os percentuais de melanoma corretamente classificados (sensibilidade) foram de 68,33%, 66,67%, 70,83%, 78,33%.

Tabela 8 – Resultados de *Random Forest* com uso de SMOTEENN.

Seletor	Acurácia	Precisão	Sensibilidade	Especificidade	Fscore	Kappa
Ranker	0,8333 ± 0,0567	0,5860 ± 0,1184	0,8250 ± 0,1208	0,8417 ± 0,0730	0,6762 ± 0,0879	0,5753 ± 0,1214
RankSearch	0,9000 ± 0,0351	0,7397 ± 0,1134	0,8167 ± 0,1097	0,9208 ± 0,0469	0,7671 ± 0,0292	0,7045 ± 0,0903
PSO	0,8733 ± 0,0387	0,6700 ± 0,1321	0,8083 ± 0,1043	0,8896 ± 0,0573	0,7215 ± 0,0706	0,6421 ± 0,0929
Greedy Stepwise	0,9200 ± 0,0189	0,7433 ± 0,0484	0,9250 ± 0,0730	0,9188 ± 0,0229	0,8221 ± 0,0413	0,7715 ± 0,0527

Por outro lado, quando técnicas de balanceamento (SMOTE + ENN) foram incluídas antes do treinamento do algoritmo de aprendizado de máquina, um aumento

de sensibilidade pode ser notado, independentemente de qual seletor de característica foi utilizado. Para o classificador RF, teve-se uma melhora de aproximadamente 14 pontos percentuais, usando *Greedy Stepwise* como seletor de atributos. O melhor modelo produzido foi encontrado quando utilizado esta última abordagem. Essa melhora é devido ao fato de o classificador não mais priorizar a classe com mais exemplos já que houve o balanceamento, fazendo com que o peso no erro de classificação seja equilibrado para as duas classes.

Também pode ser notado que o RF, usando SMOTE e ENN, alcançou uma média geral de acurácia parecida com a calculada sem o balanceamento de 92,00%, mantendo uma proporção equilibrada de instâncias corretamente classificadas para as classes normal e de melanoma, 91,88% e 92,50% respectivamente. O coeficiente kappa de Cohen informa que o modelo é muito bom, pois o valor K é maior do que 0,6 e menor que 0,8. Nas Tabelas 9 e 10, são apresentados os resultados do algoritmo SVM.

Tabela 9 – Resultados de SVM sem uso de SMOTEENN.

Seletor	Acurácia	Precisão	Sensibilidade	Especificidade	Fscore	Kappa
Ranker	0,8750 ± 0,0446	0,7321 ± 0,1656	0,6333 ± 0,1255	0,9354 ± 0,0433	0,6696 ± 0,1129	0,5937 ± 0,1390
RankSearch	0,9450 ± 0,0305	0,9300 ± 0,0641	0,7873 ± 0,1255	0,9854 ± 0,0141	0,8466 ± 0,0930	0,8138 ± 0,1101
PSO	0,8967 ± 0,0270	0,8533 ± 0,1223	0,6000 ± 0,1165	0,9798 ± 0,0245	0,6953 ± 0,0870	0,6361 ± 0,1002
Greedy Stepwise	0,9433 ± 0,0326	0,8886 ± 0,0894	0,8250 ± 0,1386	0,9729 ± 0,0198	0,8493 ± 0,0928	0,8148 ± 0,1120

Tabela 10 – Resultados de SVM com uso de SMOTEENN.

Seletor	Acurácia	Precisão	Sensibilidade	Especificidade	Fscore	Kappa
Ranker	0,8550 ± 0,0676	0,6369 ± 0,1307	0,7750 ± 0,0686	0,8750 ± 0,0833	0,6913 ± 0,0916	0,6008 ± 0,1309
RankSearch	0,9267 ± 0,0545	0,8169 ± 0,1281	0,8250 ± 0,1593	0,9521 ± 0,0393	0,8167 ± 0,1325	0,7712 ± 0,1662
PSO	0,8700 ± 0,0450	0,7130 ± 0,1449	0,7083 ± 0,1584	0,9104 ± 0,0798	0,6870 ± 0,0757	0,6082 ± 0,0969
Greedy Stepwise	0,9217 ± 0,0261	0,7998 ± 0,1156	0,8500 ± 0,0766	0,9396 ± 0,0444	0,8157 ± 0,0469	0,7667 ± 0,0631

O mesmo comportamento observado para o algoritmo RF é observado para o SVM. Em média, os valores de sensibilidade foram superiores após aplicar o balanceamento. O melhor modelo SVM, levando em consideração a sensibilidade, foi também utilizando o SMOTEENN, juntamente com o seletor de atributos *Greedy Stepwise*, da mesma forma que o classificador *Random Forest*. Entretanto, o valor da sensibilidade foi menor, 85,00%.

O SVM apresentou acurácias médias mais altas, quando foi treinado com as características resultantes do algoritmo de seleção *RankSearch*. Levando em consideração o número de características muito menor em relação aos outros, pode-se considerar que a maior dimensionalidade interferiu negativamente na formação da fronteira de decisão entre as duas classes. Pode-se ver que o algoritmo PSO, possuindo o maior número de características (mencionado na Tabela 3), teve o pior desempenho entre os seletores na maioria dos testes.

A Figura 37 apresenta os gráficos de caixas das sensibilidades alcançadas pelos modelos RF e SVM, seletor de atributos *Greedy Stepwise*, com uso ou não de balanceamento SMOTEENN, para verificar a distribuição dos resultados.

Figura 37 – Gráficos de caixa das sensibilidades para o *Greedy Stepwise*.



A dispersão, representada pela amplitude dos gráficos, indica que há maior variação de valor, quando não utilizada a ferramenta de sobre-amostragem e sub-amostragem para balanceamento das classes. Com o SMOTEEN, tem-se um valor mínimo de sensibilidade igual ou superior a 75%. Para os modelos treinados balanceadamente, temos que o *Random Forest* obteve mediana superior ao alcançado pelo SVM.

Sobre os extratores de características relacionados à forma e cor da regra ABCD, a variável **Assimetria** demonstrou ser relevante para todos os algoritmos de seleção de atributos, dos quais *Ranker*, *Rank Search* e *Greedy Stepwise* contaram com a variável **Cor**. Os descritores de **Bordas** e **Diâmetro** da lesão permaneceram no conjunto de características apenas para o seletor *Ranker*. Isso pode ter acontecido devido ao fato do *Threshold* (Limiar) de eliminação de atributos usado ter sido 0,1. O valor máximo e mínimo de relevância dada pelo seletor *Ranker* foi de 0,336 e 0, respectivamente. Se o valor máximo for considerado como 100% relevante para o problema, o valor 0,1 representa apenas 29,76% de relevância entre as variáveis existentes. Um valor de limiar que representaria pelo menos 50% de relevância seria 0,168.

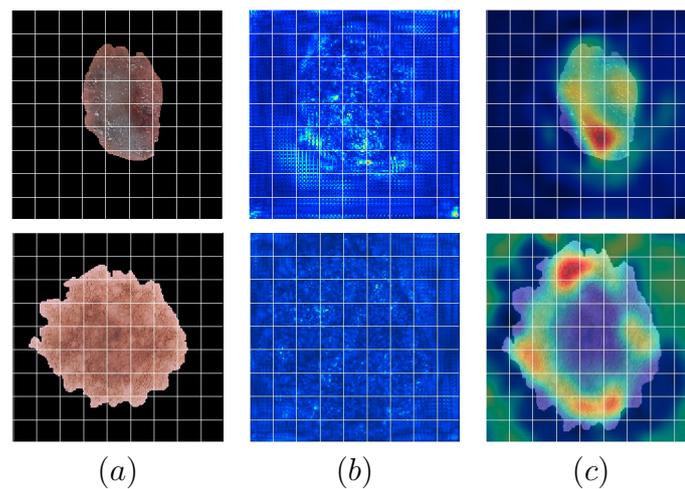
Redes neurais convolucionais tem como comportamento a extração automática de características. Todavia, quando se trata do processo de aprendizado de máquina, é de fundamental importância verificar o porquê de um dado classificador errar em uma determinada instância da amostra, e isso se realiza também observando e analisando como as características foram obtidas.

Essa verificação se faz necessária para identificação de possíveis falhas e melhoria do processo como um todo. Uma técnica para produzir “explicações visuais” para decisões de uma grande classe de modelos baseados em CNN, tornando-os mais transparentes, foi proposta por [Selvaraju et al. \(2017\)](#). O Mapeamento de Ativação de Classe com Gradiente

Ponderado (Grad-CAM, do inglês *Gradient-weighted Class Activation Mapping*) usa os gradientes de qualquer conceito alvo, fluindo para a camada convolucional final para produzir um mapa de localização, destacando as regiões importantes da imagem para efetuar a classificação.

Nas Figuras 38, 39, 40 e 41, são apresentadas as ativações das últimas camadas de convolução das redes VGG16, VGG19, *Inception* e ResNet, respectivamente. Onde (a) representa a região de interesse da lesão, (b) são os pontos da imagem ativados e (c) um mapa de calor, representando quais áreas mais contribuíram para a arquitetura. A primeira linha apresenta um exemplo de melanoma e a segunda, uma imagem não-melanoma. Para realizar as visualizações, levou-se em consideração apenas o desempenho do melhor método de seleção de características que, neste caso, foi o *Greedy Stepwise* para o melhor classificador, *Random Forest*.

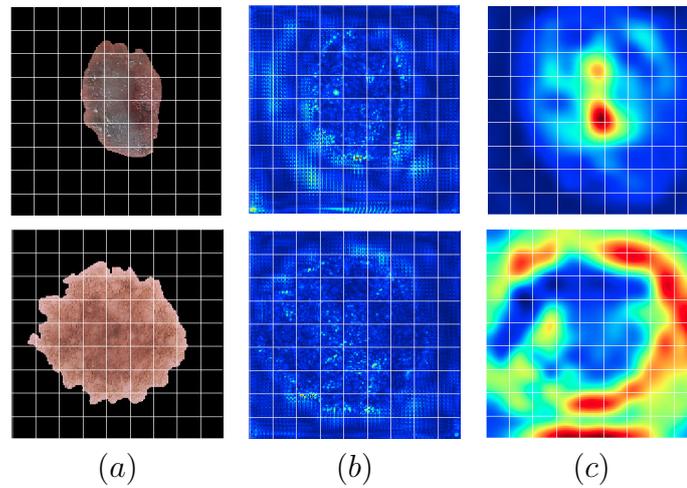
Figura 38 – Mapa de ativações para VGG16 para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha)



Fonte: Elaborada pelo autor

Na Figura 38, pode-se notar que as ativações da arquitetura VGG16 no exemplo de melanoma foram sobre uma região que aparenta ter um agravamento da lesão (pois está mais escura na original, observar letras (b) e (c)), enquanto que para a imagem normal o mais importante foi pegar pigmentos mais claros que ficam perto da borda. O mesmo comportamento, pode ser observado na Figura 41, onde são apresentados mapas para a ResNet.

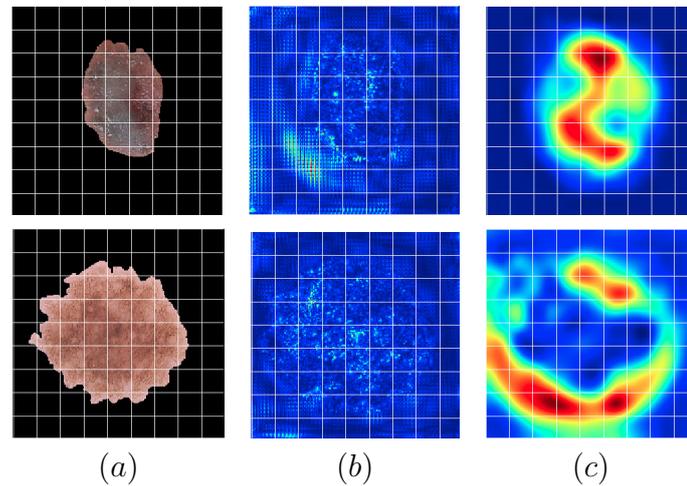
Figura 39 – Mapa de ativações para VGG19 para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha).



Fonte: Elaborada pelo autor

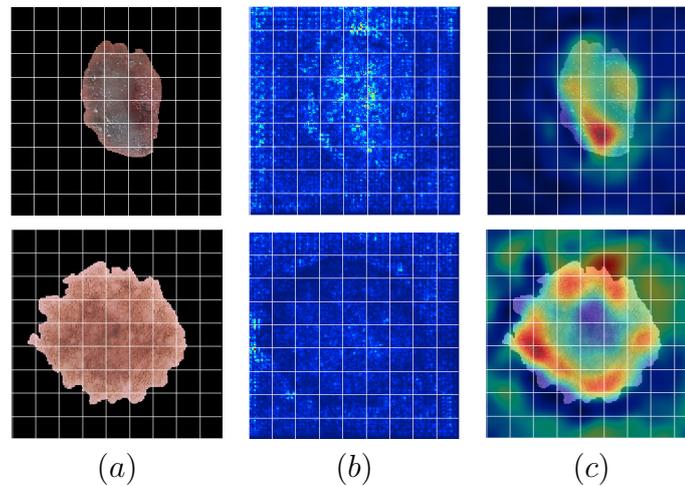
Para as topologias VGG19 (Figura 39) e *Inception* (Figura 40), as regiões mais relevantes para a imagem não-melanoma (segunda linha) foram aquelas ao redor da borda e exteriores a ela. Para as o exemplo de melanoma, as regiões mais escuras da lesão foram consideradas mais relevantes.

Figura 40 – Mapa de ativações para *Inception* para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha).



Fonte: Elaborada pelo autor

Figura 41 – Mapa de ativações para ResNet para um exemplo de melanoma (primeira linha) e para um exemplo de não melanoma (segunda linha).



Fonte: Elaborada pelo autor

Cabe ressaltar que, no processo de seleção de características, nem todas as ativações são consideradas importantes para classificação, sendo portanto eliminadas pelo seletor.

5.1 Comparação com Trabalhos Relacionados

A Tabela 11 apresenta os resultados obtidos em comparação com os alcançados em trabalhos relacionados. Mesmo não utilizando a mesma base de imagens dermatoscópicas, podemos notar resultados promissores, utilizando a combinação entre a *Greedy Stepwise*, SMOTEEN e o classificador *Random Forest*.

Tabela 11 – Comparação com trabalhos relacionados.

Trabalho	Metodologia	Base	Acc	P	E	S
(ROSADO, 2009)	Cor, TDE + kNN	PH ²	-	-	82,00%	61,00%
(MASOOD; AL-JUMAILY; ANAM, 2015)	DBN + SVM	SMDC	89,00%	90,00%	88,30%	-
(ALENCAR, 2015)	Cor e contornos + MLP	PH ²	86,66%	-	-	-
(CODELLA et al., 2016)	LBP + SC + ResNet	ISIC	83,40%	60,20%	91,12%	53,30%
(KAWAHARA; BENTAIEB; HAMARNEH, 2016)	CNN	Dermofit	81,80%	-	-	-
(RASTGO et al., 2016)	Cor e SIFT + RFs	PH ²	-	-	100,0%	90,30%
(MAJTNER; YILDIRIM-YAYILGAN; HARDEBERG, 2016)	AlexNet + SVM	ISIC	82,60%	-	89,80%	53,30%
(POMPONIU; NEJATI; CHEUNG, 2016)	CNN	DERMIS	93,64%	-	95,18%	92,10%
(NASR-ESFAHANI et al., 2016)	CNN	UMCG	81,00%	75,00%	80,00%	81,00%
(ALFED; KHELIFI; BOURIDANE, 2016)	Cor, SIFT e HOG + SVM e kNN	PH ²	-	-	91,20%	88,10%
(LOPEZ et al., 2017)	VGG16	ISIC	81,33%	79,74%	-	78,66%
(YU et al., 2017)	ResNet	ISIC	94,90%	-	95,70%	91,10%
(BAKHEET, 2017)	Forma (HOG) + SVM	PH ²	97,32%	-	98,21%	87,32%
(VASCONCELOS; VASCONCELOS, 2017)	Inception	ISIC	83,60%	69,90%	86,5%	76,00%
(MIRUNALINI et al., 2017)	Inception	ISIC	82,50%	-	-	-
(ESTEVA et al., 2017)	CNN	Privada	91,00%	-	-	-
(MAHBOD; ECKER; ELLINGER, 2017)	AlexNet, VGG16	ISIC	89,20%	-	84,70%	84,70%
(SAFRAN et al., 2018)	<i>Random Forest</i>	Privada	-	-	83,63%	86,01%
(HAN et al., 2018)	ResNet	Mix	88,00%	80,70%	-	85,50%
(LI; SHEN, 2018)	LIN CNN	ISIC	85,70%	-	96,61%	49,00%
(SULTANA; MANDAL; PUHAN, 2018)	ResNet + Eigen + SVM	PH ²	97,50%	-	100,0%	85,50%
Este estudo	SMOTEEN + RF	PH²	92,00%	74,33%	91,88%	92,50%

Muitos trabalhos não apresentam a medida de precisão dos modelos que usaram para seus classificadores. Esta indica uma taxa que informa o quanto o classificador acertou

de melanoma corretamente em relação a todos os exemplos que foram classificados como tal. Ou seja, quanto maior o número de *False Positives*, menor será a precisão. Dos estudos que utilizaram a base PH², nenhum valor de precisão é mostrado, não sendo possível a comparação com os 74,33% alcançados por este trabalho. Em comparação com aqueles que usaram outra base, existem valores mais expressivos como 90% de [Masood, Al-Jumaily e Anam \(2015\)](#).

Em relação à especificidade, taxa dos exemplos não-melanoma corretamente classificados, o maior valor foi 100% alcançados por [Rastgo et al. \(2016\)](#) e [Sultana, Mandal e Puhan \(2018\)](#) que superam os 91,88% presentes neste estudo. Entretanto, a sensibilidade, que informa a taxa dos exemplos melanoma corrente classificados, mais alta entre os trabalhos que usaram a base PH² e os que não utilizaram foi 92,50% alcançados pela aplicação do *pipeline* apresentado neste estudo.

É importante frisar que o custo dado por uma classificação errada de uma pessoa de risco ao dizer que o exame é normal, quando ela na verdade tem neoplasia maligna de pele, é muito maior do que dizer que o indivíduo possui uma doença, mesmo ela sendo normal. A acurácia geral não deixa de ser um valor promissor, 92%, que é superior a muitos outros trabalhos, mesmo mantendo a sensibilidade alta. O que corrobora com o presente estudo, em dizer que o modelo é promissor, é valor do índice Kappa de 0,7715, atribuindo um conceito de “Muito bom” para o modelo preditivo treinado.

6 Conclusão

Este trabalho apresentou uma metodologia para diagnóstico de câncer de pele melanoma, usando características obtidas a partir dos principais modelos de CNNs e regra ABCD. Os conjuntos de atributos foram concatenados e usados em uma fase de seleção de atributos, cujo vetor resultante foi utilizado para treinar os classificadores *Support Vector Machines* e *Random Forest* em vários cenários diferentes. Os testes realizados obtiveram como melhor resultado a sensibilidade de 92,5%, apresentando uma melhora de aproximadamente 14 pontos percentuais após utilização de técnicas de balanceamento de classes.

Como contribuição, o pipeline aplicado ao processo de treinamento efetuando uma balanceamento das classes existentes na base de dados com processos simples de geração de exemplos sintéticos (SMOTE) e diminuição de amostras ruidosas (ENN) fez com que os classificadores aumentassem seus respectivos desempenhos para a tarefa de diagnóstico de câncer de pele melanoma, no que tange o objetivo de aumentar a sensibilidade do classificador. Com isso, o modelo preditivo busca manter a acurácia geral, enquanto que aumenta a classificação dos exemplos de melanoma, em detrimento da classe não-melanoma. As consequências de classificar erroneamente um exemplo de caso de risco podem ser catastróficas, levando até uma possível morte do paciente, comparadas com o custo de dizer que uma determinada pessoa apresenta diagnóstico positivo, quando ela, na verdade, não possui.

Como trabalhos futuros, a metodologia proposta pode ser melhorada, por meio de:

- Utilização de outras bases públicas de imagens dermatoscópicas, como ISIC e HAM10000;
- Implementação de segmentação automática da região de interesse da lesão;
- Usar variações de técnicas de sobre e subamostragem, de forma que seja possível estimar os melhores parâmetros dos algoritmos a serem explorados;
- Testar outros classificadores, como: Regressão Linear, *Bagging - Bootstrap aggregating* (implementação com intuito de melhorar a estabilidade e a precisão dos algoritmos de aprendizado de máquina usados em classificações estatísticas e regressão, reduzindo variância e ajudando a evitar *overfitting*);
- Usar técnicas de aprimoramento de qualidade de imagem e redução de ruído que não foram exploradas neste estudo, como: Equalização de Histograma e Filtro de Mediana;

6.1 Produções Científicas

A Tabela 12 lista os artigos aceitos, a partir dos resultados da metodologia proposta, em relação ao tema de diagnóstico de melanoma com a utilização de imagens dermatoscópicas, no ano de 2018, com o qualis avaliado na área da ciência da computação. Os trabalhos foram produzidos como autor principal.

Tabela 12 – Artigos aceitos em relação ao tema do diagnóstico do melanoma com a utilização de imagens dermatoscópicas.

Título do artigo	Congresso	Qualis
Evaluation of Melanoma Diagnosis using Deep Features	25º IWSSIP	B1
Evaluation of Melanoma Diagnosis using Imbalanced Learning	18º SBCAS	B4

Referências

- AHMAD, J.; FARMAN, H.; JAN, Z. Deep learning methods and applications. In: *Deep Learning: Convergence to Big Data Analytics*. USA: Now Publishers, 2019. p. 31–42. Citado na página 35.
- ALENCAR, F. E. S. Desenvolvimento de um sistema para classificação automática de imagens dermatoscópicas para dispositivos móveis. *Universidade do Estado do Rio Grande do Norte, Biblioteca Central Orlando Teixeira, BR-RN*, 2015. Citado 3 vezes nas páginas 21, 24 e 74.
- ALFED, N.; KHELIFI, F.; BOURIDANE, A. Improving a bag of words approach for skin cancer detection in dermoscopic images. In: IEEE. *Control, Decision and Information Technologies (CoDIT), 2016 International Conference on*. Saint Julian's, Malta, 2016. p. 024–027. Citado 3 vezes nas páginas 23, 24 e 74.
- AMBELU, A.; LOCK, K.; GOETHALS, P. Comparison of modelling techniques to predict macroinvertebrate community composition in rivers of ethiopia. *Ecological Informatics*, v. 5, n. 2, p. 147 – 152, 2010. ISSN 1574-9541. Citado na página 53.
- ANDERSON, B. O.; ILBAWI, A. M.; SAGHIR, N. S. E. Breast cancer in low and middle income countries (Imic s): A shifting tide in global health. *The breast journal*, Wiley Online Library, v. 21, n. 1, p. 111–118, 2015. Citado na página 17.
- ARNOLD, H. L.; ODOM, R. B.; JAMES, W. D. A pele: estrutura básica e função. doenças básicas da pele de andrews: Dermatologia clínica. *Dermatologia clínica*, Editora Manole Ltda., São Paulo/SP, 1995. Citado 5 vezes nas páginas 18, 28, 29, 30 e 31.
- BAKHEET, S. An svm framework for malignant melanoma detection based on optimized hog features. *Computation*, Multidisciplinary Digital Publishing Institute, v. 5, n. 1, p. 4, 2017. Citado 3 vezes nas páginas 23, 24 e 74.
- BEN-DAVID, A. A lot of randomness is hiding in accuracy. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 20, n. 7, p. 875–885, 2007. Citado na página 67.
- BEN-DAVID, A. About the relationship between roc curves and cohen's kappa. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 21, n. 6, p. 874–882, 2008. Citado na página 67.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 35, n. 8, p. 1798–1828, 2013. Citado na página 35.
- BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN 0387310738. Citado 2 vezes nas páginas 54 e 56.
- BRADSKI, G. Opencv library. *Dr. Dobb's Journal of Software Tools*, 2000. Disponível em: <<https://opencv.org/>>. Citado na página 21.

- BRAZ JUNIOR, G.; FONSECA NETO, J. V. da; SILVA, A. C.; PAIVA, A. C. *Detecção de Regiões de Massas em Mamografias usando índices de Diversidade, Geoestatística e Geometria Côncava*. Tese (Doutorado) — Universidade Federal do Maranhão, 2014. Citado na página 55.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996. Citado na página 61.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado 2 vezes nas páginas 61 e 66.
- CAO, L.-J.; TAY, F. E. H. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks*, IEEE, v. 14, n. 6, p. 1506–1518, 2003. Citado na página 66.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, v. 16, p. 321–357, 2002. Citado na página 48.
- CHOLLET, F. et al. *Keras*. 2015. Disponível em: <<https://keras.io>>. Citado na página 63.
- CODELLA, N.; NGUYEN, Q.-B.; PANKANTI, S.; GUTMAN, D.; HELBA, B.; HALPERN, A.; SMITH, J. R. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, Munich, v. 61, 2016. Citado 4 vezes nas páginas 22, 24, 41 e 74.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. Citado na página 67.
- DÂNGELO, J.; FATTINI, C. *Anatomia humana sistêmica e segmentar: para o estudante de medicina*. [S.l.]: Atheneu, 1998. (Biblioteca biomédica). ISBN 9788573790733. Citado 2 vezes nas páginas 26 e 28.
- DAWSON-HOWE, K. *A Practical Introduction to Computer Vision with OpenCV*. 1st. ed. New Delhi, India: Wiley Publishing, 2014. ISBN 1118848454, 9781118848456. Citado na página 31.
- DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *CVPR09*. Miami, FL, USA: IEEE, 2009. p. 248–255. Citado 3 vezes nas páginas 22, 43 e 63.
- DURASAMY, S. P.; KAYALVIZHI, R. A new multilevel thresholding method using swarm intelligence algorithm for image segmentation. *Journal of Intelligent Learning Systems and Applications*, Scientific Research Publishing, v. 2, n. 03, p. 126, 2010. Citado na página 53.
- ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, Nature Publishing Group, v. 542, n. 7639, p. 115, 2017. Citado 3 vezes nas páginas 23, 24 e 74.

FEURER, M.; KLEIN, A.; EGGENSPERGER, K.; SPRINGENBERG, J.; BLUM, M.; HUTTER, F. Efficient and robust automated machine learning. In: CORTES, C.; LAWRENCE, N. D.; LEE, D. D.; SUGIYAMA, M.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems 28*. Montreal, Canada: Curran Associates, Inc., 2015. p. 2962–2970. Citado na página 66.

GARCÍA, S.; FERNÁNDEZ, A.; LUENGO, J.; HERRERA, F. A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. *Soft Computing*, Springer, v. 13, n. 10, p. 959, 2009. Citado na página 67.

GIOTIS, I.; MOLDERS, N.; LAND, S.; BIEHL, M.; JONKMAN, M. F.; PETKOV, N. Med-node. *Expert Syst. Appl.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 42, n. 19, p. 6578–6585, nov. 2015. ISSN 0957-4174. Citado na página 23.

GIOTIS N. MOLDERS, S. L. M. B. M. J. I.; PETKOV, N. Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images. In: UMCG. *Expert Systems with Applications*. Louisiana, 2015. p. 6578–6585. Citado na página 22.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X. Citado 3 vezes nas páginas 32, 33 e 34.

GUTIN, G.; YEO, A.; ZVEROVICH, A. Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the tsp. *Discrete Applied Mathematics*, Elsevier, v. 117, n. 1-3, p. 81–86, 2002. Citado na página 52.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of machine learning research*, v. 3, n. Mar, p. 1157–1182, 2003. Citado na página 51.

HAIR, J.; ANDERSON, R.; TATHAM, R. *Análise Multivariada de Dados*. Campina Grande, PB: Bookman, 2005. ISBN 9788536304823. Citado na página 53.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM, v. 11, n. 1, p. 10–18, 2009. Citado na página 64.

HALL, M. A.; HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering*, IEEE, v. 15, n. 6, p. 1437–1447, 2003. Citado na página 54.

HAN, S. S. *Asan and Hallym Dataset (Thumbnails)*. figshare, 2017. Disponível em: <https://figshare.com/articles/Asan_and_Hallym_Dataset_Thumbnails_/5406136/1>. Citado na página 23.

HAN, S. S.; KIM, M. S.; LIM, W.; PARK, G. H.; PARK, I.; CHANG, S. E. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, Elsevier, 2018. Citado 3 vezes nas páginas 23, 24 e 74.

HE, H.; BAI, Y.; GARCIA, E. A.; LI, S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*. *IEEE International Joint Conference on*. Hong Kong, China: IEEE, 2008. p. 1322–1328. Citado na página 48.

- HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, Ieee, n. 9, p. 1263–1284, 2008. Citado na página 48.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, USA: IEEE, 2016. p. 770–778. Citado 4 vezes nas páginas 22, 46, 47 e 63.
- HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Taylor & Francis Group, v. 12, n. 1, p. 55–67, 1970. Citado na página 52.
- HUANG, K.; YANG, H.; KING, I.; LYU, M. R. Learning classifiers from imbalanced data based on biased minimax probability machine. In: IEEE. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Washington, DC, USA, 2004. v. 2, p. II–II. Citado 2 vezes nas páginas 48 e 50.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, Wiley Online Library, v. 195, n. 1, p. 215–243, 1968. Citado na página 35.
- INCA. Estimativas 2018: Incidência de câncer no brasil. <http://www.inca.gov.br/estimativa/2018/>. 2018. Citado na página 17.
- JOHN, G. H.; KOHAVI, R.; PFLEGER, K. Irrelevant features and the subset selection problem. In: *Machine Learning Proceedings 1994*. New Brunswick, Canada: Elsevier, 1994. p. 121–129. Citado na página 51.
- KATZ, B.; RABINOVITZ, H. S. Introduction to dermoscopy. *Dermatologic clinics*, Elsevier, v. 19, n. 2, p. 221–258, 2001. Citado na página 18.
- KAWAHARA, J.; BENTAIEB, A.; HAMARNEH, G. Deep features to classify skin lesions. In: IEEE. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. Prague, Czech Republic, 2016. p. 1397–1400. Citado 2 vezes nas páginas 24 e 74.
- KELLER, J. M.; GRAY, M. R.; GIVENS, J. A. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics*, IEEE, n. 4, p. 580–585, 1985. Citado na página 21.
- KENNEDY, J.; EBERHART, R. Particle swarm optimization. *IEEE International Conference on Neural Networks*, IEEE, Boston, MA, v. 4, n. 6, p. 1942–1948, 1995. Citado na página 53.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. USA: Curran Associates Inc., 2012. (NIPS'12), p. 1097–1105. Citado 2 vezes nas páginas 37 e 43.
- KUBAT, M.; MATWIN, S. et al. Addressing the curse of imbalanced training sets: one-sided selection. In: *Icml*. Nashville, USA: Morgan Kaufmann, 1997. v. 97, p. 179–186. Citado na página 48.

- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, Wiley, International Biometric Society, v. 33, n. 1, p. 159–174, 1977. ISSN 0006341X, 15410420. Citado na página 68.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 37.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, Nov 1998. ISSN 0018-9219. Citado 4 vezes nas páginas 38, 39, 40 e 43.
- LEMAITRE, G.; NOGUEIRA, F.; OLIVEIRA, D.; ARIDAS, C. *Over-sampling*. 2017. Disponível em: <https://imbalanced-learn.readthedocs.io/en/stable/over_sampling.html>. Citado 3 vezes nas páginas 49, 50 e 51.
- LEVER, W. F.; LEVER, G. S. *Histologia da Pele*. São Paulo/SP, Brasil: Editora Mande Ltda, 1991. Citado 2 vezes nas páginas 25 e 27.
- LI, Y.; SHEN, L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 2, p. 556, 2018. Citado 2 vezes nas páginas 24 e 74.
- LIBERATO, J. A. D. D. *Tratado de anatomia aplicada*. São Paulo/SP, Brasil: Editora Póluss Editorial, 1998. Citado 3 vezes nas páginas 25, 26 e 28.
- LOPEZ, A. R.; NIETO, X. Giro-i; BURDICK, J.; MARQUES, O. Skin lesion classification from dermoscopic images using deep learning techniques. In: IEEE. *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*. Innsbruck, Austria, 2017. p. 49–54. Citado 3 vezes nas páginas 22, 24 e 74.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado 2 vezes nas páginas 57 e 58.
- MA, C.; GAO, W.; YANG, L.; LIU, Z. An improved sobel algorithm based on median filter. In: IEEE. *Mechanical and Electronics Engineering (ICMEE), 2010 2nd International Conference on*. Kyoto, Japan, 2010. v. 1, p. V1–88. Citado na página 33.
- MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. Nov, p. 2579–2605, 2008. Citado na página 23.
- MAGALHÃES, L. *Camadas da pele*. 2018. Disponível em: <<https://www.todamateria.com.br/camadas-da-pele>>. Citado na página 26.
- MAGALHÃES, L. *Derme*. 2018. Disponível em: <<https://www.todamateria.com.br/derme>>. Citado na página 28.
- MAGALHÃES, L. *Epiderme*. 2018. Disponível em: <<https://www.todamateria.com.br/epiderme>>. Citado na página 27.
- MAHBOD, A.; ECKER, R.; ELLINGER, I. Skin lesion classification using hybrid deep neural networks. *arXiv preprint arXiv:1702.08434*, 2017. Citado 3 vezes nas páginas 22, 24 e 74.

- MAJTNER, T.; YILDIRIM-YAYILGAN, S.; HARDEBERG, J. Y. Combining deep learning and hand-crafted features for skin lesion classification. In: *2016 6th International Conference on Image Processing Theory Tools and Applications (IPTA)*. Oulu, Finland: IEEE, 2016. p. 1–6. Citado 3 vezes nas páginas [22](#), [24](#) e [74](#).
- MALOOF, M. A.; LANGLEY, P.; BINFORD, T. O.; NEVATIA, R.; SAGE, S. Improved rooftop detection in aerial images with machine learning. *Machine Learning*, Springer, v. 53, n. 1-2, p. 157–191, 2003. Citado na página [48](#).
- MARSLAND, S. *Machine learning: an algorithmic perspective*. New York, USA: CRC press, 2015. Citado 4 vezes nas páginas [36](#), [58](#), [59](#) e [60](#).
- MARX, J.; WALLS, R.; HOCKBERGER, R. *Rosen's Emergency Medicine-Concepts and Clinical Practice E-Book*. Philadelphia, USA: Elsevier Health Sciences, 2013. Citado na página [26](#).
- MASOOD, A.; AL-JUMAILY, A.; ANAM, K. Self-supervised learning model for skin cancer diagnosis. In: IEEE. *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. Montpellier, France, 2015. p. 1012–1015. Citado 4 vezes nas páginas [21](#), [24](#), [74](#) e [75](#).
- MENDONÇA, T.; FERREIRA, P. M.; MARQUES, J. S.; MARÇAL, A. R. S.; ROZEIRA, J. Ph2 - a dermoscopic image database for research and benchmarking. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, v. 2013, p. 5437–40, 2013. Citado 3 vezes nas páginas [21](#), [62](#) e [63](#).
- MENEGOLA, A.; FORNACIALI, M.; PIRES, R.; BITTENCOURT, F. V.; AVILA, S.; VALLE, E. Knowledge transfer for melanoma screening with deep learning. In: IEEE. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. Melbourne, VIC, Australia, 2017. p. 297–300. Citado 2 vezes nas páginas [42](#) e [63](#).
- MIRUNALINI, P.; CHANDRABOSE, A.; GOKUL, V.; JAISAKTHI, S. Deep learning for skin lesion classification. *Computer Vision and Pattern Recognition (cs.CV)*, 2017. Citado 4 vezes nas páginas [22](#), [24](#), [41](#) e [74](#).
- MORAGLIO, A.; CHIO, C. D.; POLI, R. Geometric particle swarm optimisation. In: SPRINGER. *European conference on genetic programming*. Valencia, Spain, 2007. p. 125–136. Citado na página [54](#).
- MOURA, N.; VERAS, R.; AIRES, K.; MACHADO, V.; SILVA, R.; ARAÚJO, F.; CLARO, M. Abcd rule and pre-trained cnns for melanoma diagnosis. *Multimedia Tools and Applications*, Springer, p. 1–20, 2018. Citado 2 vezes nas páginas [34](#) e [63](#).
- NASR-ESFAHANI, E.; SAMAVI, S.; KARIMI, N.; SOROUSHMEHR, S. M. R.; JAFARI, M. H.; WARD, K.; NAJARIAN, K. Melanoma detection by analysis of clinical images using convolutional neural network. In: IEEE. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL, USA, 2016. p. 1373–1376. Citado 3 vezes nas páginas [22](#), [24](#) e [74](#).
- NASRABADI, N. M. Pattern recognition and machine learning. *Journal of electronic imaging*, International Society for Optics and Photonics, v. 16, n. 4, p. 049901, 2007. Citado na página [41](#).

- NAUMOVICH, V.; VAPNIK. *Statistical learning theory*. New York: Wiley New York, 1998. v. 1. Citado 3 vezes nas páginas 22, 55 e 66.
- PEARSON, K. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Taylor & Francis, v. 50, n. 302, p. 157–175, 1900. Citado na página 51.
- PEARSON, R.; GONEY, G.; SHWABER, J. Imbalanced clustering for microarray time-series. In: *2003 Workshop on Learning from Imbalanced Data Sets (II)*. Washington, DC, USA: Proceedings of the ICML, 2003. v. 3. Citado na página 48.
- POMPONI, V.; NEJATI, H.; CHEUNG, N.-M. Deepmole: Deep neural networks for skin mole lesion classification. In: *IEEE. Image Processing (ICIP), 2016 IEEE International Conference on*. Oulu, Finland, 2016. p. 2623–2627. Citado 2 vezes nas páginas 24 e 74.
- PRIYADARSHINI, D.; RENGINI, D. Automatic melanoma detection using local binary pattern and support vector machine. In: . India: IJIRCCE, 2015. Citado na página 23.
- PROVOST, F. Machine learning from imbalanced data sets 101. In: *Proceedings of the AAAI'2000 workshop on imbalanced data sets*. Texas, USA: AAAI, 2000. p. 1–3. Citado na página 48.
- QUINLAN, J. R. Induction of decision trees. *Machine learning*, Springer, v. 1, n. 1, p. 81–106, 1986. Citado na página 51.
- RAJ, B. *A Simple Guide to the Versions of the Inception Network*. 2018. Disponível em: <<https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202>>. Citado na página 45.
- RASTGO, M.; LEMAITRE, G.; MOREL, O.; MASSICH, J.; GARCIA, R.; MERIAUDEAU, F.; MARZANI, F.; SIDIBÉ, D. Classification of melanoma lesions using sparse coded features and random forests. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *SPIE Medical Imaging*. California, 2016. p. 97850C–97850C. Citado 4 vezes nas páginas 22, 24, 74 e 75.
- RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, v. 29, n. 9, p. 2352–2449, 2017. PMID: 28599112. Citado 3 vezes nas páginas 38, 39 e 40.
- ROSADO, L. F. C. M. G. Sistema automático para diagnóstico de lesões cutâneas baseado em imagens dermoscópicas. *Universidade do Estado do Rio Grande do Norte, Biblioteca Central Orlando Teixeira, BR-RN*, 2009. Citado 3 vezes nas páginas 21, 24 e 74.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533, 1986. Citado na página 36.
- SAFRAN, T.; VIEZEL-MATHIEU, A.; CORBAN, J.; KANEVSKY, A.; THIBAudeau, S.; KANEVSKY, J. Machine learning and melanoma: The future of screening. *Journal of the American Academy of Dermatology*, Elsevier, v. 78, n. 3, p. 620–621, 2018. Citado 3 vezes nas páginas 23, 24 e 74.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural networks*, Elsevier, v. 61, p. 85–117, 2015. Citado na página 36.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE, 2017. p. 618–626. Citado na página 71.

SHANNON, C. E. A mathematical theory of communication. *Bell system technical journal*, Wiley Online Library, v. 27, n. 3, p. 379–423, 1948. Citado na página 59.

SHIN, H.-C.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1285–1298, 2016. Citado 2 vezes nas páginas 22 e 42.

SILVA, G. L. F. d. et al. Diagnóstico de nódulos pulmonares em imagens de tomografia computadorizada usando redes neurais convolucionais evolutivas. Universidade Federal do Maranhão, 2017. Citado na página 54.

SIMIS, T.; SIMIS, D. R. C. Doenças da pele relacionadas à radiação solar. *Revista da Faculdade de Ciências Médicas de Sorocaba*, v. 8, n. 1, p. 1–8, 2006. Citado 2 vezes nas páginas 29 e 30.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition (cs.CV)*, 2014. Citado 4 vezes nas páginas 22, 43, 44 e 63.

SOARES, H. B. *Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte*. Tese (Doutorado) — Universidade Federal do Rio Grande do Norte, 2008. Citado 10 vezes nas páginas 17, 18, 25, 26, 28, 29, 30, 31, 55 e 56.

SOBEL, I. *Camera models and machine perception*. Haifa, Israel, 1972. Citado na página 33.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014. Citado na página 41.

SULTANA, N. N.; MANDAL, B.; PUHAN, N. Deep residual network with regularised fisher framework for detection of melanoma. *IET Computer Vision*, IET, v. 12, n. 8, p. 1096–1104, 2018. Citado 4 vezes nas páginas 23, 24, 74 e 75.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Boston, USA: IEE, 2015. p. 1–9. Citado 6 vezes nas páginas 21, 22, 44, 45, 46 e 63.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 267–288, 1996. Citado na página 52.

VASCONCELOS, C. N.; VASCONCELOS, B. N. Experiments using deep learning for dermoscopy image analysis. *Pattern Recognition Letters*, Elsevier, 2017. Citado 3 vezes nas páginas 21, 24 e 74.

WILSON, D. L. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, IEEE, v. 3, p. 408–421, 1972. Citado na página 50.

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? In: *Advances in neural information processing systems*. Washinton, DC, USA: NIPS, 2014. p. 3320–3328. Citado na página 43.

YU, L.; CHEN, H.; DOU, Q.; QIN, J.; HENG, P.-A. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging*, IEEE, v. 36, n. 4, p. 994–1004, 2017. Citado 4 vezes nas páginas 22, 24, 41 e 74.

ZHOU, B.; LAPEDRIZA, A.; XIAO, J.; TORRALBA, A.; OLIVA, A. Learning deep features for scene recognition using places database. In: *Advances in neural information processing systems*. Montreal, Canada: Curran Associates, Inc., 2014. p. 487–495. Citado na página 42.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 67, n. 2, p. 301–320, 2005. Citado na página 52.