



**UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO**



José Carlos Monte Silva Júnior

Técnicas preditivas para auxílio no diagnóstico de melanomas via imagens

Recife

2024

José Carlos Monte Silva Júnior

Técnicas preditivas para auxílio no diagnóstico de melanomas via imagens

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Prof. Dr. Rodrigo G. F. Soares

Recife

2024

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Ana Catarina Macêdo – CRB-4 1781

S586t Silva Júnior, José Carlos Monte.
Técnicas preditivas para auxílio no diagnóstico de melanomas via imagens / José Carlos Monte Silva Júnior. - Recife, 2024.
77 f.; il.

Orientador(a): Rodrigo Gabriel Ferreira Soares.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal Rural de Pernambuco, Bacharelado em Sistemas da Informação, Recife, BR-PE, 2024.

Inclui referências e apêndice(s).

1. Pele - Câncer. 2. Aprendizado do computador. 3. Diagnóstico por imagem. 4. Redes neurais (Computação) 5. Inteligência artificial - Aplicações médicas. I. Soares, Rodrigo Gabriel Ferreira, orient. II. Título

CDD 004

José Carlos Monte Silva Júnior

Técnicas preditivas para auxílio no diagnóstico de melanomas via imagens

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 02 de Outubro de 2024.

BANCA EXAMINADORA

Rodrigo Gabriel Ferreira Soares (Orientador)
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Cícero Garrozi
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

À Deus, minha luz e fortaleza, ajudador presente em todos os meus caminhos.

Agradecimentos

Agradeço primeiramente a Deus e ao meu Senhor Jesus Cristo, o qual, através da presença constante do Seu Espírito Santo, tem me encorajado, guiado e sustentado ao longo desta trajetória. Dou-Te graças Senhor, porque, através da Tua imensa bondade e misericórdia, tens demonstrado o Teu agir na minha vida.

Agradeço à minha amada esposa Larissa, por todo o seu apoio e incentivo, por estar sempre ao meu lado, acreditando que eu seria capaz. Não poderia me esquecer da nossa filha Luna Ester, a alegria de nossas vidas e dona de uma energia incrivelmente cativante.

Sou imensamente grato aos meus pais, José Carlos e Rosimere, que sempre me apoiaram e investiram em mim, proporcionando-me uma educação sem igual.

Agradeço ao meu sogro Ronaldo e à minha sogra Evina, pelo suporte e por sempre me incentivarem a buscar conhecimento.

Gostaria de agradecer aos meus professores da graduação, especialmente ao meu orientador, o professor Rodrigo Soares, por todo o suporte e tempo dedicados.

Por fim, gostaria de expressar os meus sinceros agradecimentos a todos aqueles que se fizeram presentes ao longo deste percurso, inclusive os que não chegaram a ser citados. Saibam que o seu apoio e incentivo foram importantes para que este sonho pudesse ser concretizado.

*“Consagre ao Senhor tudo o que você faz, e os seus planos serão bem-sucedidos.”
(Provérbios 16.3)*

Resumo

O câncer de pele é o tipo mais comum de câncer no mundo, dividido em dois tipos principais: melanoma e não melanoma. Embora mais raro, o melanoma é o mais letal devido ao seu potencial de causar metástase. Métodos não invasivos, como a dermatoscopia e a regra ABCDE, têm sido utilizados para evitar procedimentos cirúrgicos desnecessários e têm ajudado na identificação de lesões, contribuindo para diagnósticos mais rápidos. Com o avanço da tecnologia, a Inteligência Artificial (IA) ganhou destaque, mostrando-se uma solução promissora para a análise de dados médicos, especialmente com o uso de Redes Neurais Convolucionais (RNCs), que podem reconhecer padrões em imagens dermatoscópicas e ajudar na classificação de lesões como melanoma ou não melanoma de forma automatizada. Este projeto propõe um comitê de classificadores baseado em Redes Neurais Convolucionais para classificar imagens dermatoscópicas como melanoma ou não melanoma, comparando seu desempenho com arquiteturas validadas, como AlexNet e VGG-16, utilizando técnicas de Transfer Learning. As análises de Precisão, Revocação e Pontuação F1 mostraram que o comitê de Redes Neurais Convolucionais superou os modelos que utilizam técnicas de Transfer Learning, com a AlexNet apresentando desempenho superior à VGG-16. O comitê de Redes Neurais Convolucionais demonstrou uma maior capacidade de generalização, mostrando-se promissor ao capturar as características relevantes das imagens, revelando potencial para aplicações médicas, embora ainda precise ser refinado para atingir padrões clínicos.

Palavras-chave: Câncer de Pele, Aprendizado de Máquina, Técnicas de Comitês, Redes Neurais Convolucionais, Aprendizado por Transferência.

Abstract

Skin cancer is the most common type of cancer worldwide, divided into two main types: melanoma and non-melanoma. Although rarer, melanoma is the most lethal due to its potential to cause metastasis. Non-invasive methods, such as dermoscopy and the ABCDE rule, have been used to avoid unnecessary surgical procedures and have helped in the identification of lesions, contributing to faster diagnoses. With advances in technology, Artificial Intelligence (AI) has gained prominence, proving to be a promising solution for medical data analysis, especially with the use of Convolutional Neural Networks (CNNs), which can recognize patterns in dermoscopic images and help classify lesions as melanoma or non-melanoma in an automated manner. This project proposes an ensemble of classifiers based on Convolutional Neural Networks to classify dermoscopic images as melanoma or non-melanoma, comparing its performance with validated architectures, such as AlexNet and VGG-16, using Transfer Learning techniques. The analyses of Precision, Recall, and F1 Score showed that the ensemble of Convolutional Neural Networks outperformed the models using Transfer Learning techniques, with AlexNet showing better performance than VGG-16. The ensemble of Convolutional Neural Networks demonstrated a greater generalization capability, proving to be promising in capturing relevant features from the images, revealing potential for medical applications, although it still needs refinement to meet clinical standards.

Keywords: Skin Cancer, Machine Learning, Ensemble Techniques, Convolutional Neural Networks, Transfer Learning.

Lista de ilustrações

Figura 1 – Melanomas apresentando características da regra ABCDE	17
Figura 2 – As coordenadas x e y de uma imagem RGB de dimensão 32×28 <i>pixels</i>	20
Figura 3 – Representação de uma imagem digital RGB com dimensão de 4×4 <i>pixels</i>	21
Figura 4 – Um processo de treinamento típico	22
Figura 5 – Esquematização de um Neurônio biológico	24
Figura 6 – Representação de um Neurônio artificial	25
Figura 7 – Imagem de uma Rede Neural Perceptron Multicamada (MLP)	26
Figura 8 – Representação da arquitetura da LeNet-5	28
Figura 9 – Representação de uma imagem em escala de cinza de 28×28 <i>pixels</i> com um campo receptivo local de 5×5 conectado a um neurônio oculto na primeira camada oculta	29
Figura 10 – Mostra o campo receptivo apresentado na Figura 9 deslizando nos <i>pixels</i> da imagem de entrada com um <i>stride</i> de tamanho 1	30
Figura 11 – Operação de convolução, onde o filtro de detecção de bordas, um <i>kernel</i> 3×3 é aplicado na entrada.	30
Figura 12 – Antes (esquerda) e depois (direita) de uma imagem que foi submetida a um filtro de detecção de bordas	31
Figura 13 – Exemplo de operações de <i>Max Pooling</i> e <i>Average Pooling</i>	31
Figura 14 – Representação da arquitetura da AlexNet	33
Figura 15 – Representação da arquitetura da VGG-16	34
Figura 16 – Relação Performance x Quantidade de dados em algoritmos tradicionais de AM e RNA	35
Figura 17 – Representação visual de um comitê de classificadores do tipo <i>Bagging</i>	36
Figura 18 – Etapas do <i>Fine-Tuning</i>	38
Figura 19 – Representação da Validação Cruzada	39
Figura 20 – Exemplo de transformações aplicadas com o uso do <i>Data Augmentation</i>	41
Figura 21 – Matriz de Confusão.	41
Figura 22 – Arquitetura da SkinLesionClassifier	47
Figura 23 – Fluxo de execução do método <i>fit</i>	50
Figura 24 – Fluxo de execução do método <i>predict</i>	50
Figura 25 – Processo de <i>Fine-tuning</i> das redes VGG-16 e AlexNet	52
Figura 26 – Amostras da base de dados do ISIC <i>Challenge</i> 2019	53
Figura 27 – Amostras de imagens antes e após aplicação das transformações	58
Figura 28 – Curva de Aprendizado dos modelos	62

Figura 29 – Curva de Aprendizado de cada estimador do comitê de RNCs	62
Figura 30 – Matriz de confusão do Comitê de RNCs	63
Figura 31 – Matriz de confusão da AlexNet com o uso do <i>Transfer Learning</i> e <i>Fine-tuning</i>	63
Figura 32 – Matriz de confusão da VGG-16 com o uso do <i>Transfer Learning</i> e <i>Fine-tuning</i>	64
Figura 33 – Comparação gráfica das métricas obtidas pelas abordagens	65
Figura 34 – Comparação gráfica das métricas obtidas pelas abordagens para a classe Não Melanoma.	66
Figura 35 – Comparação gráfica das métricas obtidas pelas abordagens para a classe Melanoma.	67
Figura 36 – Curva PR-AUC dos modelos	67

Lista de tabelas

Tabela 1 – Distribuição das imagens do ISIC <i>Challenge</i> 2019 de acordo com as suas classes.	54
Tabela 2 – Nova distribuição das imagens após a subamostragem de acordo com as suas classes.	55
Tabela 3 – Distribuição final das imagens de acordo com as novas classes. . .	55
Tabela 4 – Distribuição de hiperparâmetros para as Camadas convolucionais. .	59
Tabela 5 – Distribuição de hiperparâmetros para as Camadas totalmente conectadas e regularização.	59
Tabela 6 – Distribuição de hiperparâmetros para o treinamento.	59
Tabela 7 – Precisão, Revocação e Pontuação F1 das abordagens.	65
Tabela 8 – Precisão, Revocação e Pontuação F1 das abordagens para a classe Não Melanoma.	66
Tabela 9 – Precisão, Revocação e Pontuação F1 das abordagens para a classe Melanoma.	66
Tabela 10 – Média e Desvio Padrão obtidos dos resultados dos estimadores do comitê.	68
Tabela 11 – Hiperparâmetros selecionados para as Camadas convolucionais da <i>SkinLesionClassifier</i>	69
Tabela 12 – Hiperparâmetros selecionados para as Camadas totalmente conectadas e regularização da <i>SkinLesionClassifier</i>	69
Tabela 13 – Hiperparâmetros selecionados para o treinamento da <i>SkinLesionClassifier</i>	69

Lista de abreviaturas e siglas

ACS	<i>American Cancer Society</i>
AM	Aprendizado de Máquina
AP	Aprendizado Profundo
AUC	Área Sob a Curva
CPU	Unidade de Processamento Central
DA	<i>Data Augmentation</i>
GPU	Unidade de Processamento Gráfico
IA	Inteligência Artificial
INCA	Instituto Nacional de Câncer
ISIC	<i>International Skin Imaging Collaboration</i>
FN	Falso Negativo
FP	Falso Positivo
MLP	Perceptron Multicamadas
ReLU	Unidade Linear Retificada
RGB	<i>Red-Green-Blue</i>
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
ROC	Característica de Operação do Receptor
TL	<i>Transfer Learning</i>
VGG	Grupo de Geometria Visual
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

Sumário

1	INTRODUÇÃO	15
1.1	Objetivos	18
1.1.1	Objetivos Específicos	18
1.2	Organização do Trabalho	19
2	REFERENCIAL TEÓRICO	20
2.1	Fundamentos da Imagem Digital	20
2.2	Aprendizado de Máquina	21
2.2.1	Redes Neurais Artificiais	23
2.2.2	Redes Neurais Convolucionais	27
2.2.3	Aprendizado Profundo	34
2.3	Comitê de Classificadores	35
2.4	Transfer Learning	37
2.4.1	Fine-Tuning	37
2.5	Métodos de Amostragem e Seleção de Hiperparâmetros	38
2.5.1	Divisão do conjunto de dados	38
2.5.2	Validação Cruzada	39
2.5.3	Busca Aleatória	40
2.6	Data Augmentation	40
2.7	Métodos de Avaliação de Desempenho	41
2.7.1	Precisão, Revocação e Pontuação F1	42
3	TRABALHOS RELACIONADOS	44
4	TÉCNICAS PREDITIVAS PARA AUXÍLIO NO DIAGNÓSTICO DE MELANOMAS VIA IMAGENS	47
4.1	Comitê de Redes Neurais Convolucionais	47
4.2	Redes Neurais Convolucionais e Transfer Learning	51
5	EXPERIMENTOS	53
5.1	Conjunto de Dados	53
5.2	Ambiente experimental	56
5.3	Pré-processamento das imagens e técnicas de Data Augmentation	56
5.4	Seleção de modelos	58
5.5	Análise e Resultados	61
6	CONCLUSÕES	71

REFERÊNCIAS	74
APÊNDICE A - Código Fonte Disponível no GitHub	78

1 Introdução

O câncer de pele é considerado o tipo de câncer mais frequente no mundo, sendo também o mais constante no Brasil, de acordo com o Instituto Nacional de Câncer (INCA) (INCA, 2022b). A doença manifesta-se em sua grande maioria em pessoas adultas, por volta dos 40 anos de idade, e com maior frequência em pessoas que possuem uma pele clara, geralmente devido à exposição contínua aos raios solares os quais são submetidos ao decorrer de suas vidas. Casos preexistentes envolvendo doenças cutâneas do próprio indivíduo ou no histórico dos membros do grupo familiar, distúrbios genéticos como o albinismo ou tratamentos envolvendo medicamentos contra doenças autoimunes ou inflamatórias, os imunossupressores, também são fatores de risco no desenvolvimento da doença e considerados agentes determinantes no número dos casos registrados (INCA, 2020).

Correspondendo a cerca de 27% dos casos de todos os tumores malignos que foram registrados no país, o câncer de pele pode ser classificado em dois tipos distintos, o câncer de pele não melanoma e o câncer de pele melanoma. O câncer de pele não melanoma é o tipo de câncer mais comum no Brasil, e apesar da possibilidade de ocasionar mutilações como sequelas, apresenta um grande índice de recuperação quando diagnosticado de forma precoce. Segundo estatísticas do INCA (2020), a estimativa é que ocorram cerca de 177 mil novos casos de câncer do tipo não melanoma por ano no Brasil, enquanto a estimativa para novos casos de melanoma é de 8,4 mil casos ao ano.

O melanoma é um tipo de câncer incomum quando comparado aos demais cânceres de pele e que surge quando os melanócitos, que são células especializadas na produção de melanina, a proteína que garante a coloração da pele, começam a crescer descontroladamente (ACS, 2019b). O INCA (2022a) aponta que apenas 4% do número total de cânceres de pele registrados no país correspondem ao melanoma. Mesmo com baixa incidência entre os diagnósticos de câncer de pele, o melanoma ainda pode ser considerado o mais letal entre todos os tipos de neoplasias malignas na pele, pois possui grandes chances de causar a metástase, que é quando o tumor maligno passa a infectar outros órgãos, conseqüentemente, podendo levar o paciente a óbito.

Apresentando sintomas, os pacientes devem ser encaminhados a um dermatologista, um médico especializado em problemas relacionados à pele, o qual irá submeter os pacientes a alguns métodos e técnicas para saber se o problema dermatológico pode se tratar de algum tumor maligno ou qualquer outro tipo de problema cutâneo. O

médico irá levar em conta no momento da avaliação os fatores de riscos, sintomas e histórico do paciente com a doença. Caso o médico julgue alta a possibilidade do melanoma, pode ser considerada a realização de uma biópsia de pele, um exame moroso e invasivo, onde a área afetada pela lesão dermatológica é colhida, deixando sempre uma cicatriz nos pacientes. Essa área colhida é levada a um laboratório, onde será avaliada de forma microscópica (ACS, 2019a).

Os tratamentos não invasivos surgiram de uma demanda para encontrar novas técnicas capazes de prover um diagnóstico preciso, sem a necessidade de realizar cirurgias desnecessárias ou incisões consideráveis no corpo dos pacientes, como a biópsia. A dermatoscopia é uma técnica não invasiva que utiliza um instrumento chamado dermatoscópio e permite ao médico olhar as lesões de pele com mais clareza, além de fornecer a possibilidade da realização de capturas e armazenamento de imagens digitais das lesões (WELZEL; SCHUH, 2017).

Segundo HOFF (2013), a identificação precoce do melanoma e o tratamento adequado são essenciais para o sucesso do tratamento do paciente. A regra ABCDE é um método usual que vem se mostrando eficaz na avaliação das lesões melanocíticas quanto a sua classificação, definindo cinco características que auxiliam a determinar uma lesão como melanoma ou não, as quais são:

- **A é para assimetria:** Quando repartida ao meio, a lesão tem forma assimétrica, já a não melanoma é oposta, sendo simétrica;
- **B é para bordas:** O melanoma possui bordas recortadas, irregulares ou borradas e a não melanoma possui bordas lisas;
- **C é para coloração:** A coloração do melanoma é heterogênea podendo ser em cores diferentes (vermelho, branco, preto, marrom, tons cinzas e azuis), a não melanoma possui geralmente apenas tons de marrom, sendo claro ou escuro;
- **D é para diâmetro:** Pode-se considerar melanomas quando estão maiores de 0,6 cm, embora ainda seja possível encontrar melanomas menores.
- **E é para evolução:** Deve-se analisar se há modificações no tamanho ou forma da lesão ou algum crescimento gradativamente.

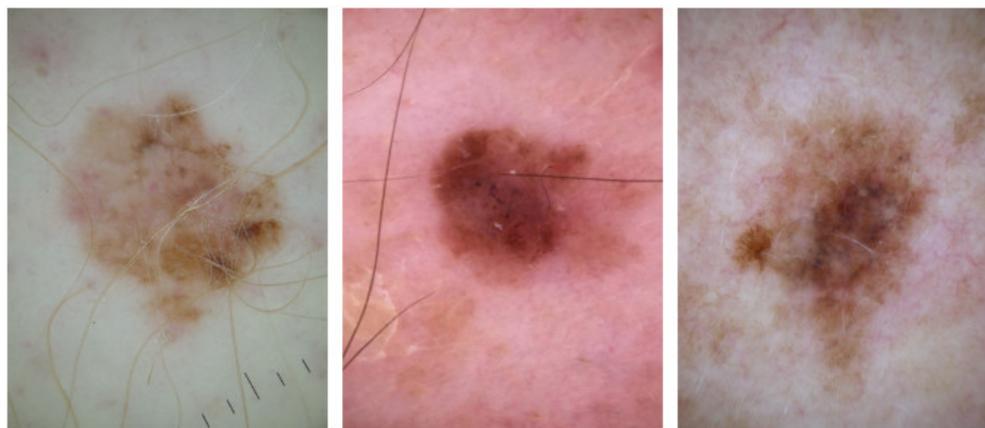


Figura 1 – Melanomas apresentando características da regra ABCDE
Fonte: (ISIC, 2019b)

Apresentar dois ou mais critérios da regra ABCDE já é o suficiente para ser considerado um sinal de alerta e suficiente para busca de ajuda médica especializada. Na Figura 1 é possível visualizar três exemplos de melanomas que apresentam um ou mais critérios presentes na regra ABCDE. Sendo utilizado como um autoexame ou como complemento à dermatoscopia, o método tem se mostrado eficaz no auxílio da tomada de decisão, tanto dos pacientes na procura de ajuda, quanto para médicos, no momento da classificação da lesão (HADIMI; PENG; COCKBURN, 2010).

Porém, determinar o prognóstico de uma lesão dermatológica como benigna ou melanoma, ainda é uma tarefa bastante difícil. Geralmente, médicos com pouca experiência possuem dificuldades na execução do método, o que faz com que seja necessário a prática constante, até que o médico consiga utilizar a técnica com eficácia (CHAPPUIS et al., 2016). Mesmo entre os mais experientes, se faz necessário em alguns casos, uma segunda avaliação realizada por outro médico especializado. Entretanto, alguns estudos mostram que a concordância dos diagnósticos apresentados entre os especialistas pode ser considerada variável, o que pode ocasionar em um diagnóstico tardio do tumor maligno, acarretando em implicações nas chances de recuperação de um paciente (CORONA et al., 1996; FARMER; GONIN; HANNA, 1996).

Conforme demonstrado por ELLIOTT et al. (2017), alguns dados mostram que para os médicos não especialistas, cerca de 22,3 lesões dermatoscópicas benignas chegam a ser removidas para cada melanoma que é diagnosticado, e mesmo entre os médicos especialistas e mais experientes, o número das lesões benignas que são removidas é aproximado de 4,5 para cada diagnóstico de melanoma.

Em razão da necessidade de novos métodos e técnicas que possuam a capacidade de complementar ou realizar por completo um diagnóstico das lesões melano-cíticas de maneira não invasiva, o uso da Inteligência Artificial (IA), vem ganhando espaço e notoriedade, se tornando uma importante auxiliadora da medicina moderna

(CUI et al., 2019). Isso se deve ao fato da IA possuir a capacidade de aprender com a experiência adquirida através da análise dos exemplos de entrada, o que a possibilita prever ou classificar dados com precisão e de forma automatizada. Dessa maneira, a IA consegue analisar aspectos e características que a primeiro momento seriam imperceptíveis aos médicos, fazendo com que o uso dessa tecnologia possa contribuir no diagnóstico precoce não somente do melanoma, como também no de outras neoplasias malignas, auxiliando os médicos na tomada de decisão (GIAVINA-BIANCHI; CORDIOLI; MACHADO, 2021).

A atuação da IA no diagnóstico de melanomas ainda é de fato algo recente, porém, vem sendo amplamente utilizada por empresas privadas e instituições acadêmicas pelo globo, onde alguns estudos já se mostram bastante promissores, conseguindo obter um nível de precisão nos diagnósticos próximo aos obtidos por médicos especialistas (HAENSSLE et al., 2018; TSCHANDL et al., 2019).

Impulsionadas pelo crescente avanço da IA, novas tecnologias passaram a ser criadas e exploradas, como o caso do Aprendizado de Máquinas (AM) e das Redes Neurais Convolucionais (RNC). Uma RNC se trata de um algoritmo de Aprendizado Profundo (AP) que vem sendo utilizado de forma bem sucedida na análise e processamento de imagens digitais, e é capaz de reconhecer padrões presentes nas imagens com uma curva de aprendizagem. Enquanto que através da aplicação da regra ABCDE, os médicos especializados conseguem determinar características visuais nas lesões que auxiliam na identificação dos melanomas, uma RNC irá analisar as imagens digitais das lesões dermatológicas obtidas por meio da dermatoscopia, para encontrar características nas imagens de casos confirmados de melanomas e aprender com essas características para realizar previsões ou classificações de forma precoce e não invasiva (ANWAR et al., 2018; ESTEVA et al., 2017).

1.1 Objetivos

Devido ao avanço da tecnologia e o impacto da utilização da Inteligência Artificial na medicina e no diagnóstico de doenças dermatoscópicas, este trabalho tem como objetivo, analisar técnicas de Aprendizado de Máquina para reconhecimento de imagens, propondo um modelo de comitê (*ensemble*) de Redes Neurais Convolucionais, capaz de classificá-las como melanoma ou não melanoma, e comparar o seu desempenho com o de modelos já treinados e utilizados largamente, através de técnicas de *Transfer Learning* (Aprendizado por Transferência).

1.1.1 Objetivos Específicos

- Identificar e estudar técnicas de pré-processamento de imagens;

- Estudar e executar técnicas de Aprendizado de Máquina capazes de extrair informações das imagens digitais;
- Estudar e aplicar técnicas de *Transfer Learning*, com a intenção de utilizar o conhecimento prévio dos modelos escolhidos, e analisar o impacto do seu uso no consumo de recursos computacionais e tempo;
- Realizar análise experimental dos modelos para a comparação dos seus desempenhos preditivos.

1.2 Organização do Trabalho

O trabalho está estruturado em seis capítulos. O capítulo 2 retrata o referencial teórico necessário para abordar os conceitos utilizados no desenvolvimento e entendimento do trabalho: fundamentos da imagem digital, técnicas de Aprendizado de Máquina, técnicas de *Transfer Learning* e técnicas de avaliação de desempenho de modelos. No capítulo 3, são apresentados os trabalhos relacionados baseados em técnicas de AM, que utilizam modelos baseados em técnicas de TL, para a classificação de doenças dermatoscópicas através de imagens. No capítulo 4, é apresentada e detalhada a Metodologia do modelo proposto e dos modelos utilizados com a técnica de TL. O capítulo 5 aborda de forma detalhada a base de dados utilizada e os experimentos realizados. Ainda no mesmo capítulo, são realizadas análises e discussões acerca dos resultados obtidos. Finalmente no capítulo 6, são apresentadas as conclusões finais.

2 Referencial Teórico

No presente capítulo, serão abordados temas relevantes para os experimentos e avaliação dos resultados obtidos, como os Fundamentos da Imagem Digital, Aprendizado de Máquina, Redes Neurais Artificiais e *Transfer Learning*.

2.1 Fundamentos da Imagem Digital

Uma imagem digital pode ser descrita como uma matriz digital finita, onde cada índice de suas linhas e colunas indicam um ponto na imagem. Sendo considerados a menor unidade de uma imagem digital, estes pontos são conhecidos como *pixels*, e cada *pixel* contém coordenadas espaciais (x, y) e um determinado nível de brilho. Podemos então, nos referir à uma imagem digital monocromática como uma função bidimensional $f(x, y)$, onde x e y correspondem às coordenadas espaciais e o valor da função f em um dado *pixel* é relativo ao seu brilho (GONZALEZ; WOODS, 2008). Tratando-se de imagens coloridas, cada *pixel* passa a apresentar diferentes canais e para cada canal, se faz necessário uma função $f(x, y)$ (MARQUES FILHO; VIEIRA NETO, 1999).

A origem de uma imagem digital se encontra no *pixel* no canto superior esquerdo da imagem e é especificada com a notação $(0, 0)$, onde o primeiro 0 representa a coordenada x e o segundo 0 a coordenada y . O valor da coordenada x começa em 0 na origem e aumenta à medida que vai se deslocando para a direita, enquanto o valor da coordenada y começa em 0 na origem e vai aumentando à medida que vai se deslocando para baixo (Figura 2), diferente do formato que vemos em um Sistema Cartesiano Ortogonal na Geometria Analítica (SWEIGART, 2015).

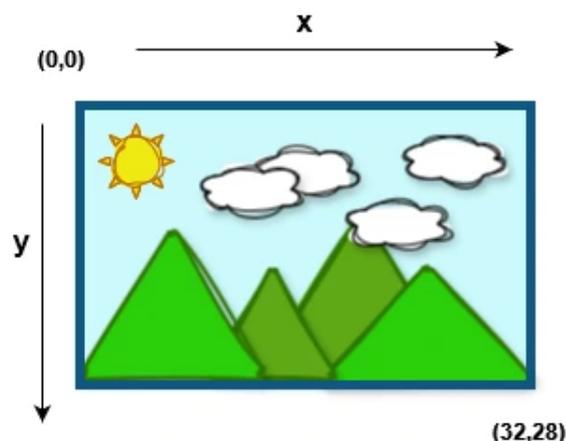


Figura 2 – As coordenadas x e y de uma imagem RGB de dimensão 32×28 *pixels*
Fonte: Autor (2024)

Comumente, os programas de computador representam as cores presentes nos *pixels* no formato RGB (*Red-Green-Blue*), o qual é um formato padronizado de cores que identifica os valores de vermelho (*Red*), verde (*Green*) e azul (*Blue*) de uma determinada cor (Figura 3). Quando falamos de imagens coloridas do formato RGB, temos que cada *pixel* possui 3 canais, com cada canal possuindo um valor inteiro que varia entre 0 e 255, onde 0 é a ausência total de brilho (cor preta) e o valor 255 é o valor máximo de brilho (cor branca) que um canal pode apresentar. As imagens também possuem um valor *alpha*, que no caso do padrão RGB é definido como RGBA. O valor *alpha* refere-se à opacidade da imagem, onde esse valor também é um inteiro que varia entre 0 e 255. O valor 0 significa que a imagem é transparente, enquanto o valor máximo de 255, informa que temos uma imagem totalmente opaca (SWEIGART, 2015).

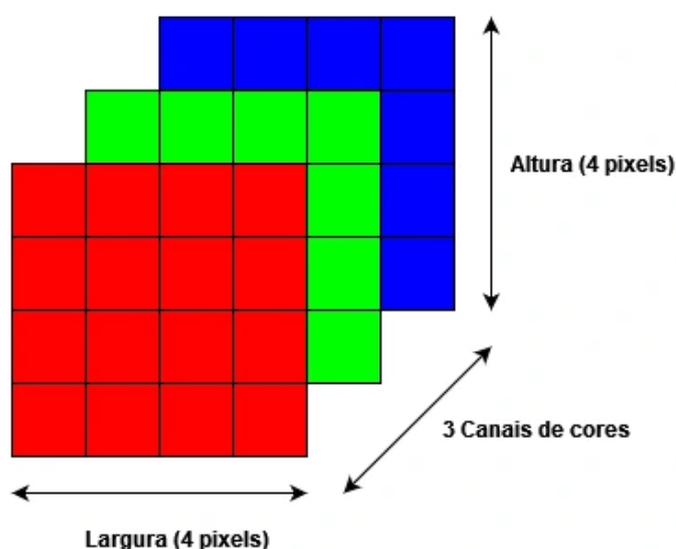


Figura 3 – Representação de uma imagem digital RGB com dimensão de 4×4 *pixels*
Fonte: Autor (2024)

Devido ao impulso nas diversas técnicas de processamento de imagens, tornou-se possível analisar, extrair e processar uma imensa quantidade de informações das imagens de forma automática, rápida e precisa, de tal maneira que seria impensável ao ser humano realizar ao olho nu, permitindo a aplicabilidade dessas técnicas em inúmeras áreas (MARQUES FILHO; VIEIRA NETO, 1999).

2.2 Aprendizado de Máquina

A Inteligência Artificial (IA) é um campo em rápido crescimento, com diversas aplicações práticas, incluindo automação, processamento de linguagem e imagem, diagnósticos médicos e pesquisa científica. Surgindo como uma solução para tarefas

intuitivas, permitindo que computadores aprendam hierarquias de conceitos e construam representações complexas a partir de conceitos mais simples (GOODFELLOW; BENGIO; COURVILLE, 2016).

O Aprendizado de Máquina (AM) pode ser considerado como o estudo de técnicas que permitem a melhoria em uma tarefa ao longo do tempo através da experiência. Essa abordagem envolve a definição de um programa flexível com parâmetros variáveis, os quais são ajustados através de um processo de treinamento (Figura 4), usando conjuntos de dados para melhorar o desempenho do programa em uma medida específica de desempenho (ZHANG et al., 2023).

Um algoritmo de AM é capaz de aprender com os dados para executar uma tarefa específica, e são compostos através da combinação de vários componentes, como algoritmos de otimização, função de custo, modelo e conjunto de dados. As tarefas de AM envolvem o processamento de exemplos, onde cada exemplo consiste em conjuntos de características quantitativamente medidas de objetos ou eventos. Os algoritmos de AM geralmente se dividem em Aprendizado Supervisionado e Aprendizado Não Supervisionado (GOODFELLOW; BENGIO; COURVILLE, 2016).

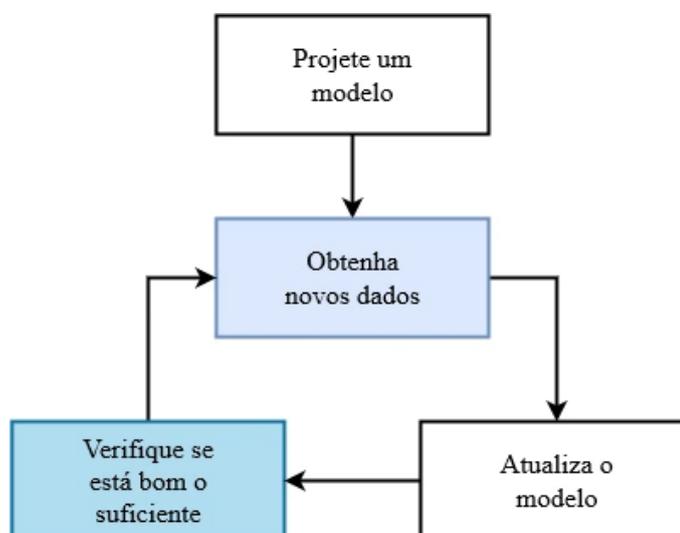


Figura 4 – Um processo de treinamento típico
Fonte: (ZHANG et al., 2023)

O objetivo do Aprendizado Supervisionado é prever um rótulo desconhecido com base em entradas nunca antes vistas. O modelo é treinado com exemplos rotulados, onde cada entrada consiste em características e um rótulo correspondente (ZHANG et al., 2023). No aprendizado supervisionado observamos exemplos de um vetor aleatório x e um valor associado ou vetor y , onde o propósito do algoritmo, é prever y a partir de x , estimando $p(y|x)$ (GOODFELLOW; BENGIO; COURVILLE, 2016).

Com um agente externo, comumente chamado de supervisor, indicando ao programa as respostas desejadas para os dados de entradas, é possível usar o erro en-

contrado, que é a diferença entre o resultado esperado e o valor obtido pelo algoritmo, para ajustar os parâmetros do modelo, de forma que, executando esse processo repetidas vezes, seja possível minimizar o erro, até que o modelo atinja um nível satisfatório. Ao atingir esse ponto, pode-se dizer que o modelo adquiriu conhecimento (FURTADO, 2019).

Segundo NG (2018), hoje existem muitas formas de AM, mas a maior parte do valor prático vem do aprendizado supervisionado, como os problemas de regressão e classificação e os algoritmos de Regressão Linear, Regressão Logística e as Redes Neurais Artificiais.

Os algoritmos utilizados nos experimentos deste trabalho se tratam de algoritmos de Aprendizado Supervisionado, tendo em vista que o objetivo deste trabalho é, através do uso do AM, analisar e categorizar imagens de lesões dermatoscópicas como melanoma ou não, indicando que se trata de um problema de classificação.

2.2.1 Redes Neurais Artificiais

Sendo extremamente poderoso e complexo, o cérebro humano é capaz de processar rapidamente uma quantidade incrível de informação através de suas unidades principais conhecidas como neurônios (Figura 5). Responsável pela transmissão de informações, o neurônio consiste em três partes principais: o corpo celular, os dendritos e o axônio (Data Science Academy, 2021). Os dendritos recebem estímulos, aumentando a área de superfície para integrar impulsos de terminais axônicos. O axônio conduz impulsos para transmitir informações a outras células. A transmissão do impulso nervoso entre neurônios depende das sinapses, locais de contato entre axônios e dendritos ou corpo celular de outros neurônios. Cada neurônio responde a estímulos do ambiente por meio de alterações na diferença de potencial elétrico através de sua membrana celular. Essas mudanças de potencial são propagadas para outros neurônios, músculos ou glândulas, formando um impulso nervoso (FURTADO, 2019).

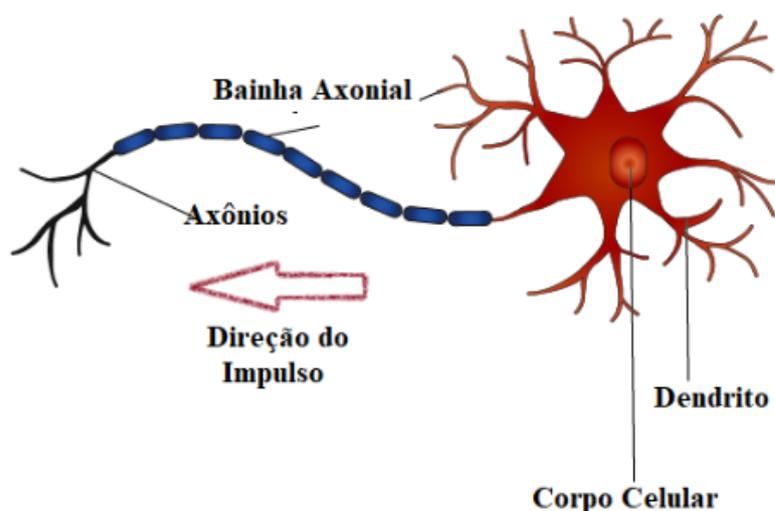


Figura 5 – Esquemática de um Neurônio biológico
Fonte: (FURTADO, 2019)

Motivados pela capacidade de aprendizagem por experiência do cérebro humano, pesquisadores tentaram simular o seu funcionamento, desenvolvendo sistemas inteligentes capazes de realizar tarefas complexas como classificação, reconhecimento de padrões, processamento de imagens, entre outras atividades. Em 1943, Warren McCulloch e Walter Pitts, tentaram descrever em um artigo como deveria ser o funcionamento de um neurônio. Em seguida, criaram um modelo computacional baseado em matemática e algoritmos denominados lógica de limiar. Inspirado pelos trabalhos desenvolvidos anteriormente, Frank Rosenblatt, desenvolveu na década de 50, o Perceptron, um algoritmo para o reconhecimento de padrões, que calcula uma soma ponderada das entradas, subtrai um limite e passa um dos dois valores possíveis como resultado (Data Science Academy, 2021).

Considerado um tipo de neurônio artificial (Figura 6), o Perceptron proposto por Rosenblatt, recebe um vetor de entradas $x_1, x_2, x_3, \dots, x_n$ e produz uma única saída binária. Ele atribui pesos $w_1, w_2, w_3, \dots, w_n$, às entradas para calcular a saída, onde a soma ponderada determina se a saída é 0 ou 1, com base em um valor limite. Os pesos e o limite são números reais e parâmetros do neurônio, permitindo que ele tome decisões com base em diferentes tipos de evidência. Semelhantes aos Perceptrons, os neurônios sigmóides também são um tipo de neurônio artificial, onde suas entradas podem assumir quaisquer valores entre 0 e 1. Também possuem pesos para cada entrada, e um viés geral. São modificados de modo que pequenas mudanças em seus pesos e vieses causam apenas uma pequena mudança em suas saídas, resultando em valores diferentes de 0 ou 1. Esse é um fato crucial que permite uma rede de neurônios sigmóides aprender (NIELSEN, 2015).

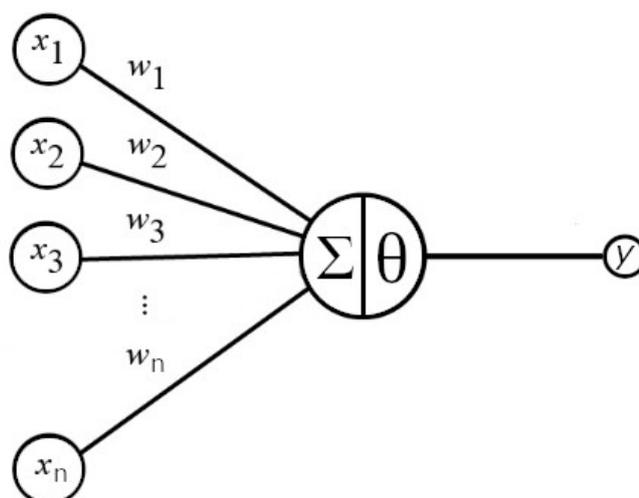


Figura 6 – Representação de um Neurônio artificial
Fonte: Autor (2024)

O surgimento de modelos de neurônios artificiais, possibilitou o desenvolvimento de um sistema composto de vários neurônios artificiais interconectados, que posteriormente ficou conhecido como Rede Neural Artificial (RNA) (Data Science Academy, 2021). Um RNA é composta por várias unidades processadoras, os neurônios, que são organizados em camadas e são amplamente interligados com as unidades de camadas consecutivas através das sinapses (conexões ponderadas por pesos) (BEZERRA, 2016). Todo o processamento de uma RNA ocorre distributivamente entre seus neurônios, cada um realizando seu processamento de maneira isolada e paralela, e embora individualmente realizem tarefas simples, sua associação permite resolver problemas altamente complexos (FURTADO, 2019).

Durante o treinamento, os pesos da RNA normalmente são ajustados através do algoritmo de Retropropagação (*Backpropagation*) (Algoritmo 1), que consiste basicamente em duas etapas: o passo para frente (*forward pass*) e o passo para trás (*backward pass*). O passo para frente tem como objetivo propagar os dados de entrada através da rede, aplicando operações de multiplicação entre os vetores (dados, pesos e vieses) e funções de ativação, até a saída da rede. A partir dos dados obtidos da camada de saída, o passo para trás é o responsável por calcular através de uma função de perda (também conhecida como função de custo) o gradiente, o qual é utilizado para atualizar os pesos da rede de forma recursiva através da regra da cadeia. Essas duas etapas fazem do *backpropagation* um dos algoritmos mais importantes e essenciais para as RNAs (Data Science Academy, 2021).

Algorithm 1 Backpropagation

- 1: **Entrada:** Dados de entrada X , rótulos Y , taxa de aprendizado η , número de épocas E
- 2: **Inicializar:** Pesos W e bias b aleatoriamente
- 3: **Para** época = 1 **até** E **faça**
- 4: **Para** cada exemplo x_i em X **faça**
- 5: **Forward Pass (Feedforward):**
- 6: Calcular a saída da rede $a = f(W \cdot x_i + b)$
- 7: **Cálculo do Erro:**
- 8: $erro = Y_i - a$
- 9: **Backward Pass (Backpropagation):**
- 10: $d_erro = erro \cdot f'(z)$, onde $z = W \cdot x_i + b$
- 11: **Atualização dos Pesos:**
- 12: $W = W + \eta \cdot d_erro \cdot x_i^T$
- 13: $b = b + \eta \cdot d_erro$
- 14: **Fim Para**
- 15: **Fim Para**

Onde:

- x_i^T denota a transposta do vetor x_i .

A primeira camada em uma RNA é chamada de Camada de entrada (*Input layer*) e os neurônios pertencentes a esta camada são chamados de neurônios de entrada. Enquanto a camada mais à direita, a última, é chamada de Camada de saída (*Output layer*) e contém os neurônios de saída. Todas as camadas intermediárias entre a Camada de entrada e a Camada de saída são referidas como Camadas ocultas (*Hidden layers*) (NIELSEN, 2015).

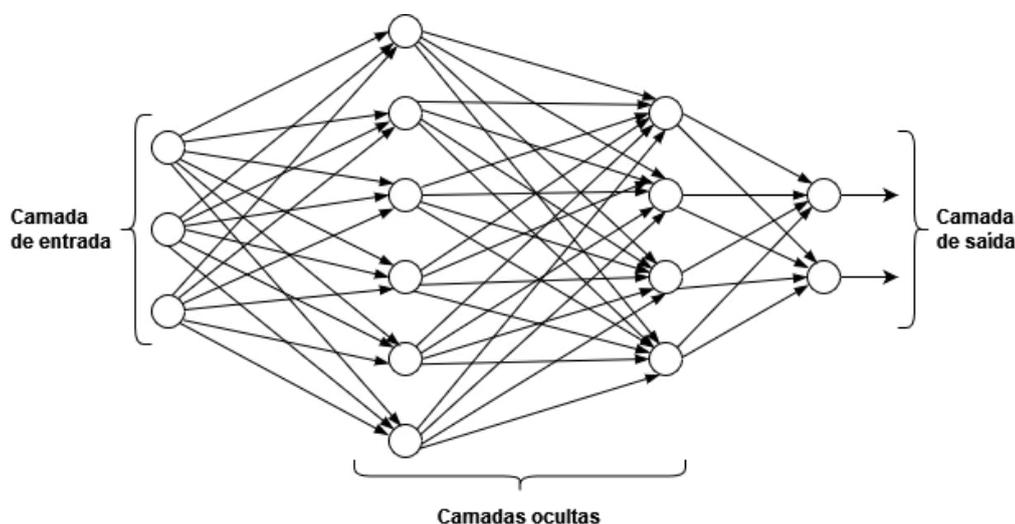


Figura 7 – Imagem de uma Rede Neural Perceptron Multicamada (MLP)
Fonte: Autor (2024)

A arquitetura de uma rede diz respeito ao modo pelo qual suas unidades de processamento estão conectadas. Essa arquitetura influencia diretamente nos problemas que uma RNA pode resolver e na forma de realizar seu treinamento (BEZERRA, 2016). Apesar de em sua grande maioria serem compostas essencialmente de neurônios sigmóides, RNAs com múltiplas camadas geralmente são chamadas de Perceptrons Multicamadas (MLPs) (Figura 7) ou de Redes Neurais Feedforward, como também são conhecidas. Nessas RNAs, a saída obtida de uma camada é usada como entrada para a próxima camada. Sem *loops* na rede, a informação é sempre alimentada para a frente, nunca retroalimentada (NIELSEN, 2015).

Usando os exemplos para inferir automaticamente regras para reconhecer padrões, ao aumentar o número de exemplos de treino, uma RNA pode aprender mais sobre os padrões e, assim, melhorar a sua precisão (NIELSEN, 2015). Com um conjunto de dados suficientemente grande, processadores rápidos e um algoritmo suficientemente sofisticado, as RNAs podem realizar tarefas que até então só podiam ser realizadas apenas por seres humanos. A evolução natural das Redes Neurais, são as Redes Neurais Profundas (Data Science Academy, 2021).

2.2.2 Redes Neurais Convolucionais

Algumas redes como as Redes Neurais com Camadas Totalmente Conectadas (*Fully Connected Networks*), não são ideais para a classificação de imagens, pois não consideram a estrutura espacial das imagens e tendem a tratar todos os *pixels* da mesma maneira, independentemente de sua proximidade ou distância na imagem (NIELSEN, 2015). Para superar essas limitações, as Redes Neurais Convolucionais (RNC), do inglês *Convolutional Neural Networks*, adicionam esparsidade às conexões entre neurônios de camadas próximas e usam o compartilhamento de parâmetros, o que resulta em um menor número de parâmetros, tornando as RNCs mais resistentes a variações nos dados, como translação e distorção (BOCHIE et al., 2020).

Em 1979, Kunihiko Fukushima desenvolveu a Neocognitron, uma RNA que aprendeu a reconhecer padrões visuais e que viria a ser considerada como a primeira Rede Neural Convolutiva. Ela foi estruturada com uma arquitetura composta por multicamadas agrupadas e com convoluções de forma hierárquica (Data Science Academy, 2021). Porém, somente em 1998, através do artigo “*Gradient-based learning applied to document recognition*”, em livre tradução “Aprendizagem baseada em gradiente aplicada ao reconhecimento de documentos”, elaborado por Yann LeCun, Léon Bottou, Yoshua Bengio e Patrick Haffner, que o termo Redes Convolutivas foi estabelecido. Em seu artigo, LeCun e os demais desenvolveram uma rede nomeada de LeNet-5 (Figura 8), que era capaz de reconhecer dígitos manuscritos com precisão. Utilizando o algoritmo *backpropagation* em uma Rede Neural *Feedforward* com muitas camadas

ocultas e mapas de unidades replicadas em cada camada, a LeNet-5 permitiu o reconhecimento de vários caracteres ao mesmo tempo, mesmo que sobrepostos (LECUN et al., 1998).

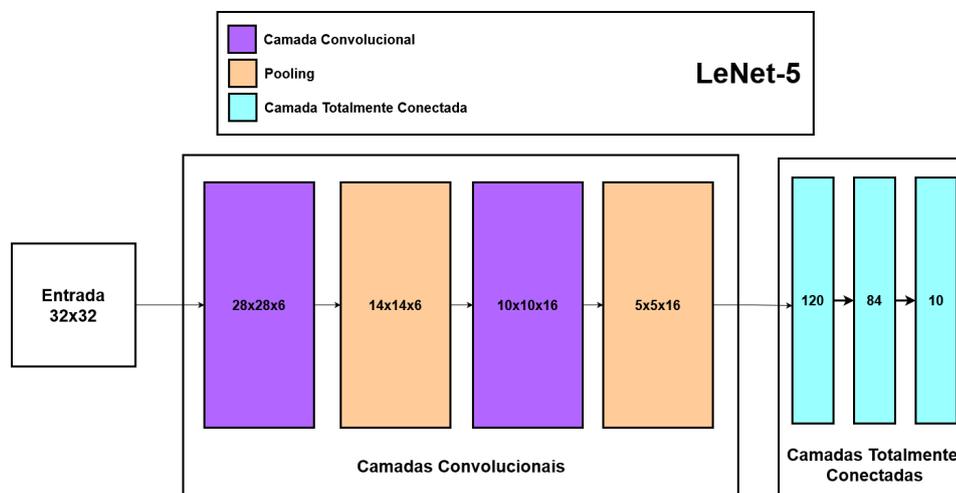


Figura 8 – Representação da arquitetura da LeNet-5
Fonte: Autor (2024)

Em uma imagem digital, os *pixels* próximos possuem uma forte correlação. Essa propriedade permite extrair características locais de pequenas sub-regiões da imagem, que podem ser combinadas em estágios subsequentes para detectar características mais complexas. As Redes Neurais Convolucionais baseiam-se nessa propriedade, utilizando conceitos como campos receptivos locais (*local receptive fields*), compartilhamento de pesos (*shared weights*), convolução e subamostragem (*subsampling* ou *pooling*) para processar eficientemente imagens e extrair informações significativas (BISHOP, 2006).

Na camada convolutiva, um campo receptivo local conecta regiões pequenas e localizadas da imagem de entrada a um neurônio da camada oculta (BISHOP, 2006). Cada neurônio da primeira camada oculta pode detectar características visuais elementares, como textura, bordas, cantos e formas. Essas características são combinadas nas camadas subsequentes para detectar características mais complexas, como bocas, rodas e bicos. O compartilhamento de pesos e vieses (*bias*) entre diferentes regiões da imagem, possibilita que padrões já identificados em determinada região, possam ser localizados em outras partes da imagem. Os pesos e vieses compartilhados são conhecidos como *Kernel* ou Filtro. Os neurônios de uma camada convolutiva são organizados em conjuntos chamados de mapas de características (*feature map*), onde cada mapa é responsável por detectar diferentes características locais da imagem (BEZERRA, 2016). Como exemplo, temos a LeNet-5, desenvolvida por LeCun, que utilizou 6 mapas de características, cada um associado a um campo receptivo local de 5×5 , para reconhecer dígitos da base de dados MNIST (LECUN et al., 1998).

Para exemplificar melhor o campo receptivo local, consideremos uma imagem digital em escala de cinza (contendo apenas um canal) com 28×28 *pixels*. Essa imagem deverá possuir 784 neurônios de entrada ($28 \times 28 \times 1$). Consideremos agora uma pequena região de 5×5 , que corresponde a um total de 25 *pixels* da nossa imagem. Essa pequena região, quando sobreposta sobre a nossa imagem, representa o campo receptivo local (NIELSEN, 2015).

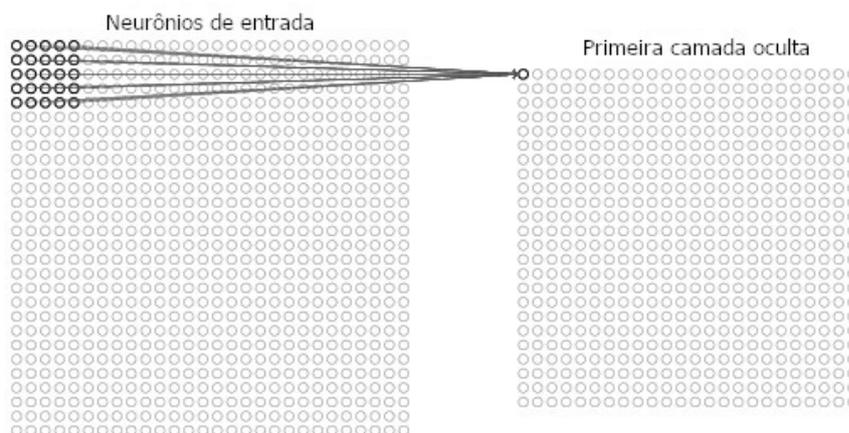


Figura 9 – Representação de uma imagem em escala de cinza de 28×28 *pixels* com um campo receptivo local de 5×5 conectado a um neurônio oculto na primeira camada oculta

Fonte: (NIELSEN, 2015)

O campo receptivo local se move pela imagem, iniciando no canto superior esquerdo, e para cada região, existe um neurônio oculto correspondente na primeira camada oculta (Figura 9). O tamanho da camada oculta depende do campo receptivo local e do tamanho do passo (*stride*), que determina quantos *pixels* o campo irá deslocar para baixo e para direita pela imagem (Figura 10). Durante todo o processo, os campos receptivos aplicam operações conhecidas como convolução, para identificar padrões de características específicas na região selecionada (NIELSEN, 2015).

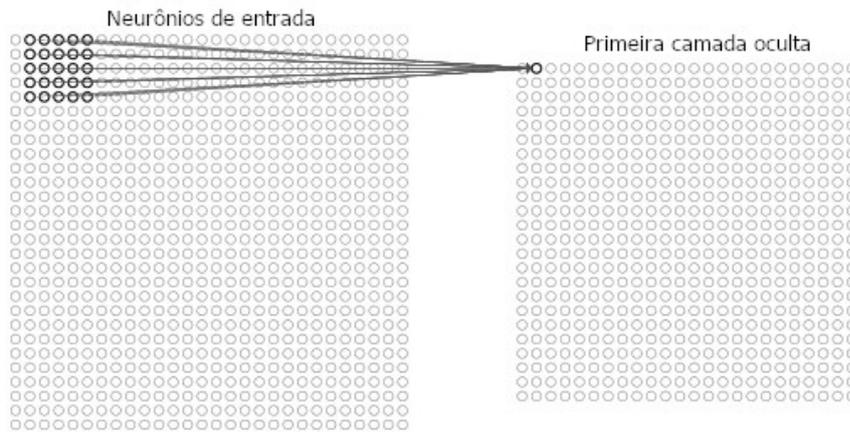


Figura 10 – Mostra o campo receptivo apresentado na Figura 9 deslizando nos *pixels* da imagem de entrada com um *stride* de tamanho 1
 Fonte: (NIELSEN, 2015)

A convolução, frequentemente representada pelo símbolo asterisco (*), pode ser definida como uma operação matemática linear que envolve duas funções, onde a primeira função é a entrada e a segunda função é conhecida por *kernel*. Essa operação é utilizada para computar o produto escalar entre os valores da entrada e os valores do *kernel*, atenuando a função de entrada e dando um peso maior às medições mais recentes. O resultado da operação de convolução é chamado de mapa de características (*feature map*) (Figuras 9 e 10). Em aplicações de AM, a entrada de uma convolução geralmente é uma matriz multidimensional de dados, também conhecida como tensor, obtida a partir de uma imagem digital, enquanto o *kernel* é uma outra matriz multidimensional contendo parâmetros (os pesos e vieses compartilhados) que são ajustáveis pelo algoritmo de aprendizagem (GOODFELLOW; BENGIO; COURVILLE, 2016; BOCHIE et al., 2020).

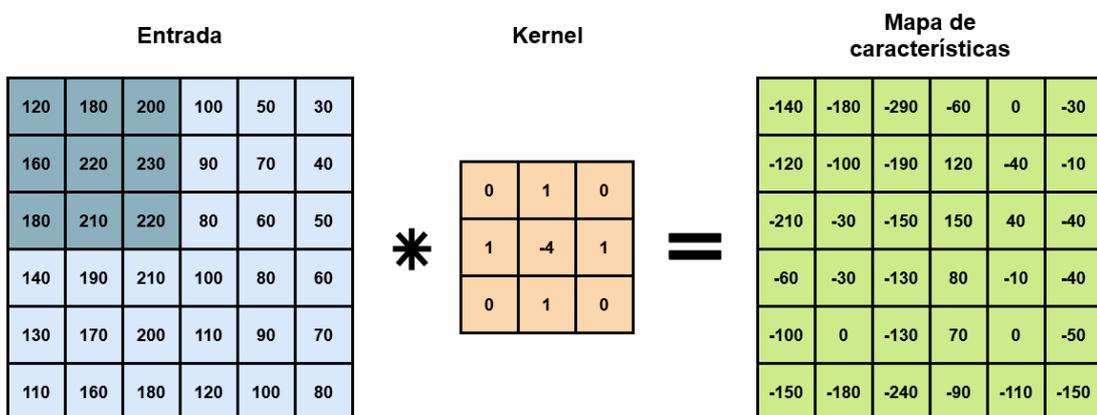


Figura 11 – Operação de convolução, onde o filtro de detecção de bordas, um *kernel* 3 × 3 é aplicado na entrada.
 Fonte: Autor (2024)

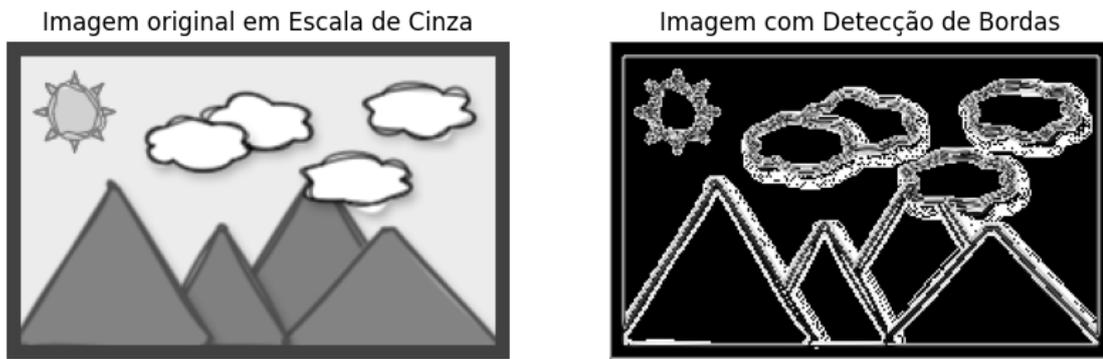


Figura 12 – Antes (esquerda) e depois (direita) de uma imagem que foi submetida a um filtro de detecção de bordas
 Fonte: Autor (2024)

A camada de *pooling*, também conhecida como camada de subamostragem, atua após a camada convolucional em uma RNC. Nessa camada, as saídas das unidades convolucionais da camada anterior são reduzidas e simplificadas. Cada unidade na camada de *pooling* recebe entradas de um campo receptivo no mapa de características correspondente da camada convolucional e realiza uma subamostragem (BISHOP, 2006). O *pooling* substitui os valores de saída em uma região específica por uma estatística resumida, ajudando a tornar as representações das características mais robustas, garantindo que pequenas mudanças na posição dos elementos da imagem não afetem demais os valores das saídas. Essa técnica é de muita importância na criação de invariância de translação a pequenos deslocamentos nas imagens de entrada. Dessa maneira, mesmo que uma imagem seja girada ou movida, ainda é possível reconhecer os mesmo padrões na imagem (GOODFELLOW; BENGIO; COURVILLE, 2016).

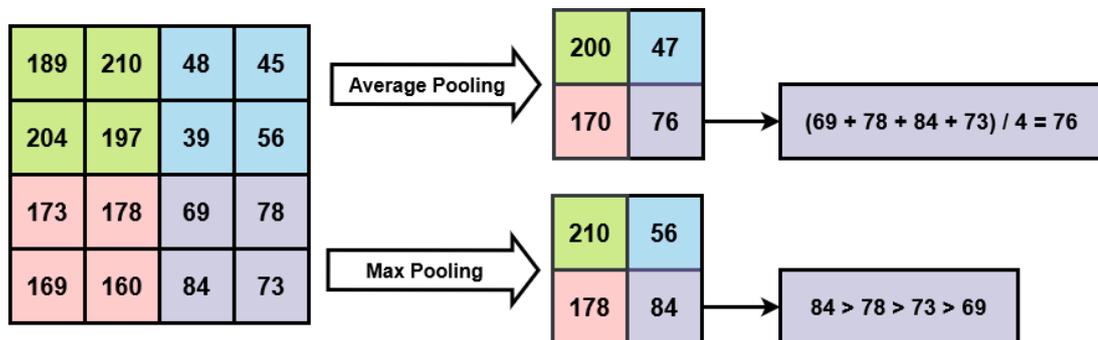


Figura 13 – Exemplo de operações de *Max Pooling* e *Average Pooling*
 Fonte: Autor (2024)

Existem várias abordagens para realizar a subamostragem, como o *Max Pooling* e o *Average Pooling* que podem ser tidos como os mais comuns, exemplificados na Figura 13. O *Max Pooling* seleciona o valor máximo em uma região, permitindo à rede identificar se um determinado recurso está presente em uma região da imagem, sem

se preocupar com sua localização exata, enquanto com o *Average Pooling*, calculamos a média dos valores em sub-regiões da entrada (BEZERRA, 2016).

Como as RNCs processam imagens como tensores, essa operação ajuda a reduzir o custo computacional, especialmente à medida que a profundidade dos tensores aumenta ao longo das camadas da rede (PONTI; COSTA, 2017). O *pooling* não só melhora a eficiência computacional da rede, mas também pode levar a uma melhor generalização do modelo, reduzindo a chance de *overfitting* e tornando o treinamento mais rápido e eficiente (GOODFELLOW; BENGIO; COURVILLE, 2016).

Após a LeNet, as RNCs ganharam um destaque maior na comunidade acadêmica, porém, ainda não eram predominantes na área de visão computacional e AM. Isso se dava ao fato de que mesmo com bons resultados para conjuntos pequenos de dados, ainda era incerta a eficácia da LeNet para conjuntos de dados maiores. Outro ponto bastante impactante, é que muitos dos modelos de redes neurais ainda eram superados por outros algoritmos e abordagens de AM. Apesar das dificuldades encontradas, alguns pesquisadores influentes no campo de AM acreditavam e defendiam que as características deveriam ser aprendidas de forma hierárquica, através de várias camadas, com cada uma detectando aspectos diferentes dos dados (ZHANG et al., 2023).

Impulsionados pelo projeto ImageNet, um enorme conjunto de dados de imagens voltado para o uso de pesquisadores como fonte de dados para treinamento de modelos de reconhecimento de padrões em grande escala (DENG et al., 2009), KRIZHEVSKY, SUTSKEVER e HINTON (2012), propuseram uma nova variante de RNC, a AlexNet, que venceu o Desafio de Reconhecimento Visual em Grande Escala ImageNet 2012, conquistando um resultado excelente. A AlexNet (Figura 14), nomeada assim por conta do nome do primeiro autor do artigo, Alex Krizhevsky, consiste em uma RNC, com 8 camadas, das quais 5 são camadas convolucionais seguidas de *max pooling*, e as 3 últimas são duas camadas totalmente conectadas e uma camada de saída totalmente conectada. As duas camadas totalmente conectadas que seguem a última camada convolucional, possuem 4096 neurônios de saídas, produzindo parâmetros de modelo de quase 1 GB. Além de utilizar o *Dropout* para controlar a complexidade do modelo nas suas camadas totalmente conectadas, a AlexNet na sua versão mais recente, mudou a sua função de ativação de uma função sigmóide, para uma função ReLU. Por fim, os dados da sua camada de saída, com 1000 neurônios, são fornecidos para uma função *Softmax* que irá fornecer as probabilidades das categorias.

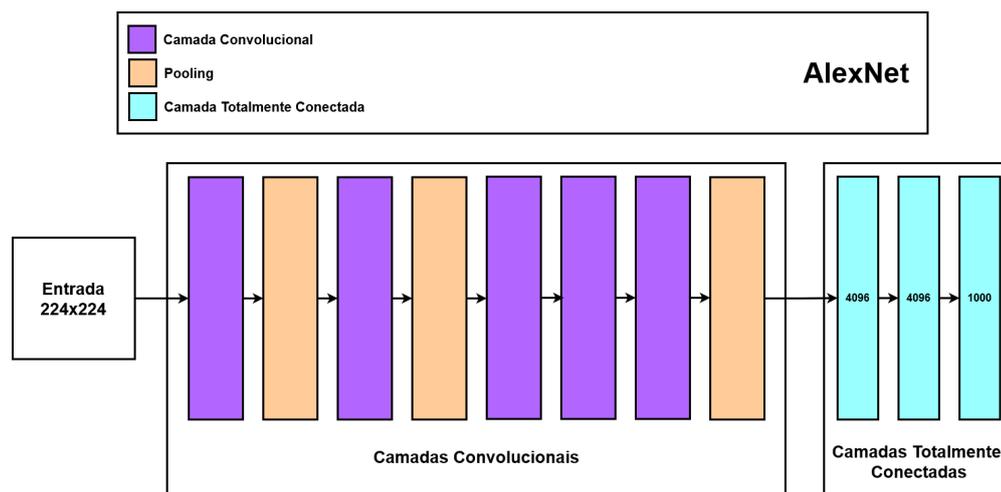


Figura 14 – Representação da arquitetura da AlexNet
Fonte: Autor (2024)

Esse grande custo computacional e a limitação do *hardware* das primeiras GPUs, fez com que a AlexNet original usasse um fluxo paralelo de dados, dividindo o processamento em duas GPUs. Atualmente, com o avanço do poder computacional das peças mais modernas de *hardware*, não é mais necessário separar o processamento em mais de uma GPU. Pode-se considerar que o uso da técnica de *Dropout*, a mudança para a função ReLU e o uso de técnicas de pré-processamento foram de extrema importância para que a AlexNet obtivesse um ótimo desempenho no campo da visão computacional (ZHANG et al., 2023).

Com o progresso na área de redes neurais, as arquiteturas passaram a ser cada vez mais abstratas, passando de neurônios individuais para camadas inteiras e, agora, para blocos de camadas. Essa ideia de usar blocos foi introduzida pelo Grupo de Geometria Visual (VGG) da Universidade de Oxford, com a sua rede VGG, na qual um bloco convolutivo da rede é composto por camadas convolucionais seguidas por *max pooling*, enfileiradas de maneira sequencial (ZHANG et al., 2023).

Em seu artigo *Very Deep Convolutional Networks for Large-Scale Image Recognition*, SIMONYAN e ZISSERMAN (2014) definiram a VGG original usando convoluções com *kernels* de 3×3 e *padding* (preenchimento) de tamanho 1 para manter as dimensões da imagem, além de *max pooling* de 2×2 com *stride* de 2 para reduzir a dimensionalidade em cada bloco. Inicialmente, a versão original possuía 5 blocos convolucionais, com os dois primeiros blocos contendo cada um uma camada convolutiva e os três últimos contendo cada um duas camadas convolucionais. Por possuir um total de 11 camadas, sendo 8 camadas convolucionais e 3 camadas totalmente conectadas, a rede VGG original também ficou conhecida como VGG-11. Atualmente a VGG possui outras variações, que diferem apenas no número de camadas convolucionais, sendo elas as de 11 (original), 13, 16 e 19 camadas, das quais as três últimas são

fixas e do tipo totalmente conectadas, com 4096, 4096 e 1000 neurônios, respectivamente, e as primeiras são camadas convolucionais para todas as variações. A versão da VGG utilizada para este trabalho será a VGG-16, que é a variação com 16 camadas, sendo 13 convolucionais e 3 totalmente conectadas, sendo distribuídas conforme a Figura 15.

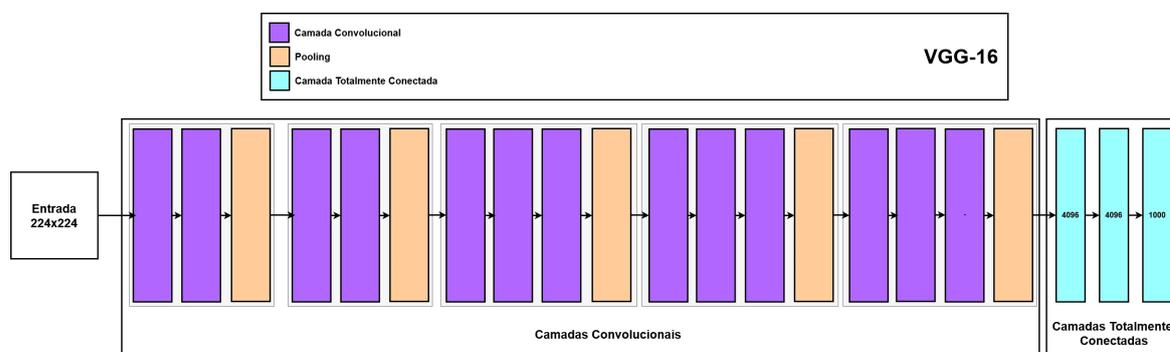


Figura 15 – Representação da arquitetura da VGG-16
Fonte: Autor (2024)

2.2.3 Aprendizado Profundo

Desafios de generalização enfrentados pelo AM tradicional foram os principais motivadores para o desenvolvimento dos algoritmos de Aprendizado Profundo (AP). Se destacando pelas suas redes de modelos robustos, compostas por múltiplas camadas de transformações de dados, o AP capacita sistemas a aprender com a experiência obtida dos dados através de técnicas poderosas, onde cada camada, fazendo uso do conhecimento adquirido na camada anterior, consegue aprender um nível maior de abstração da tarefa. Isso permite que os algoritmos de AP sejam capazes de executar tarefas consideradas complexas, como o reconhecimento de imagens, além de impulsionar avanços em áreas como a Visão Computacional e o Processamento de Linguagem Natural (ZHANG et al., 2023).

Surgindo como o principal mecanismo de construção de sistemas de IA, aproveitando o crescimento exponencial do poder computacional no início dos anos 2000, algoritmos de AP conquistaram competições importantes de AM, e o seu interesse continua em ascensão, sendo frequentemente mencionado e aplicada em soluções comerciais (Data Science Academy, 2021).

Com o surgimento de grandes volumes de dados (*Big Data*), impulsionados pela disseminação da Internet e avanços na tecnologia de armazenamento e computação, algoritmos antes considerados computacionalmente inviáveis tornaram-se relevantes, proporcionando um ambiente propício para o desenvolvimento e aplicação de técnicas de AP (ZHANG et al., 2023).

A performance de alguns dos algoritmos de AM mais antigos, como a Regres-

são Logística, muitas vezes pode estagnar, mesmo com maiores volumes de dados, ficando presa em mínimos locais na curva de aprendizado. Um melhor desempenho pode ser obtido, por exemplo, treinando Redes Neurais Artificiais grandes e com uma grande quantidade de dados. Considere uma RNA grande, como uma rede com um grande número de camadas ocultas. RNAs menores podem oferecer melhor desempenho em tarefas específicas, especialmente com conjuntos de dados pequenos, onde a engenharia manual de recursos pode ter mais impacto que a escolha do algoritmo. No entanto, para conjuntos de dados grandes, as RNAs tendem a superar os algoritmos tradicionais, devido à sua capacidade de lidar com uma maior quantidade de dados, como demonstrado na Figura 16 (NG, 2018).

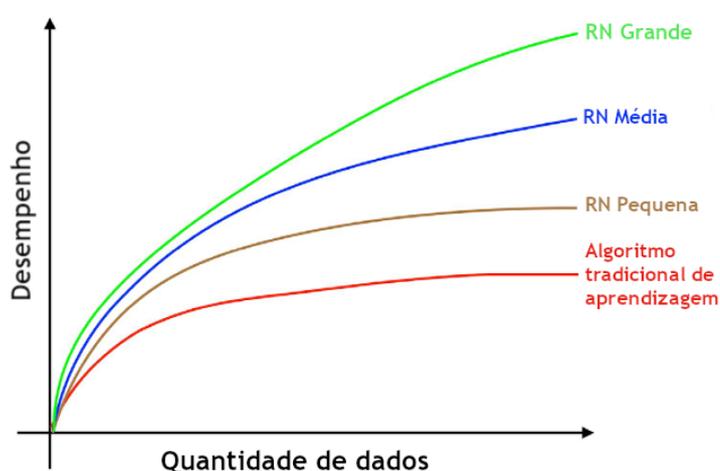


Figura 16 – Relação Performance x Quantidade de dados em algoritmos tradicionais de AM e RNA

Fonte: (NG, 2018)

O AP oferece um conjunto importante de métodos para análise de sinais, como áudio, fala, imagens e texto. Essas técnicas são aplicadas em diversas áreas, como o Processamento de Imagens e Computação Gráfica. Impulsionados por competições como o ImageNet, as Redes Neurais Convolucionais (RNC), *Deep Belief Networks* e outras arquiteturas são utilizadas como base para avanços em classificação e segmentação de imagens (PONTI; COSTA, 2017).

2.3 Comitê de Classificadores

Ao resolver problemas de classificação ou de regressão, podemos obter um desempenho mais aprimorado combinando vários modelos, em vez de usar apenas um único modelo isoladamente. Essa técnica é comumente conhecida por comitê (*ensemble*). Existem inúmeras maneiras de construir um comitê, porém, uma das mais simples, é fazer a média das previsões de um conjunto de modelos individuais (BISHOP, 2006).

Para entendermos melhor o conceito de comitê, imagine, por exemplo, que treinamos 5 RNAs distintas, com cada uma delas atingindo acurácias próximas a 99,6%. Apesar de todas as redes apresentarem acurácias semelhantes, é possível que cometam erros diferentes devido às inicializações aleatórias. Utilizando o método mencionado anteriormente, podemos realizar uma votação entre as 5 redes, o que pode gerar uma classificação superior à de qualquer uma das redes de modo individual (NIELSEN, 2015).

O *Bagging* (*Bootstrap Aggregating*) (Figura 17) é uma técnica para melhorar a acurácia de um modelo de aprendizado de máquina por meio da construção de um comitê de modelos. Em vez de usar apenas um modelo isolado, o *bagging* utiliza várias cópias do modelo treinadas em conjuntos de dados de inicialização diferentes. Cada modelo individual faz previsões e a previsão final é dada pela média das previsões de todos os modelos. Esta abordagem visa reduzir o erro geral do modelo, aproveitando as diferenças nas amostras de inicialização e a diversidade dos modelos. Embora essa técnica possa reduzir o erro médio de um modelo, a redução real do erro depende da correlação entre os erros individuais dos modelos. Uma outra variante importante do método do comitê, conhecida como *Boosting*, envolve o treinamento de múltiplos modelos em sequência, onde a função de erro usada para treinar um modelo depende do desempenho dos modelos anteriores (BISHOP, 2006).

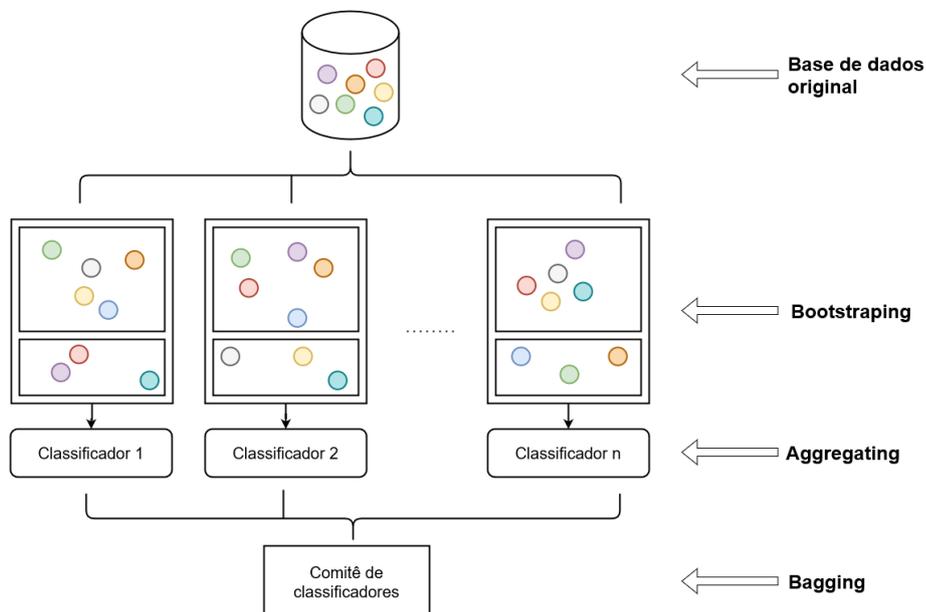


Figura 17 – Representação visual de um comitê de classificadores do tipo *Bagging*
Fonte: Autor (2024)

2.4 Transfer Learning

No campo do Aprendizado de Máquina, a escassez de dados em um domínio específico é um dos problemas mais comuns com o qual nos deparamos. O *Transfer Learning* (TL), traduzido para o português como Aprendizado por Transferência, é uma técnica fundamental que nos permite aproveitar o conhecimento adquirido em um domínio de origem (*Source Domain*) D^s , onde existem dados abundantes e modelos já treinados, para melhorar o desempenho de um modelo em um domínio de destino (*Target Domain*) D^t , onde os dados são limitados ou às vezes, até inexistentes (WANDERLEY, 2018). Essa estratégia se torna eficaz para resolver o desafio de treinar modelos em tarefas novas e similares, aproveitando o aprendizado prévio obtido em domínios relacionados, além de se mostrar uma abordagem especialmente útil, quando uma das classes de dados é escassa, pois treinar uma rede neural com poucas amostras geralmente não resulta em uma generalização adequada do modelo (YAQING et al., 2020).

2.4.1 Fine-Tuning

O *Fine-Tuning* (Ajuste Fino) é uma técnica bastante utilizada no TL e envolve congelar as camadas mais profundas de um modelo já treinado, modificando apenas as camadas mais superficiais do modelo, como a camada de saída. Dessa forma é possível treinar apenas as camadas novas e aproveitar o conhecimento prévio adquirido nas camadas mais profundas. O objetivo é extrair características gerais de imagens, que ajudem na identificação de bordas, texturas, formas e composição de objetos, que podem ser igualmente úteis em domínios e tarefas diferentes. O processo de *Fine-Tuning* consiste em quatro etapas (Figura 18) (ZHANG et al., 2023):

- Pré-treinar um modelo inicial em um conjunto de dados do domínio de origem D^s ;
- Criar um novo modelo replicando o modelo inicial, exceto pela camada final;
- Adicionar uma nova camada final ao novo modelo, no qual o tamanho da saída é equivalente ao número de categorias do domínio de destino D^t , e inicializar seus parâmetros aleatoriamente;
- Treinar o novo modelo com o conjunto de dados alvo, ajustando os parâmetros das camadas existentes e treinando a nova camada final.

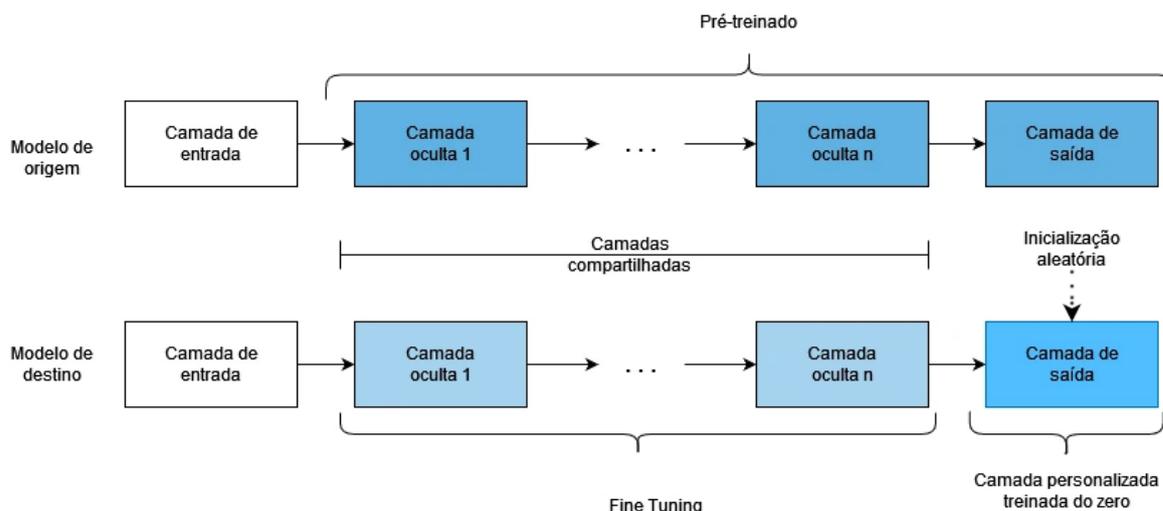


Figura 18 – Etapas do *Fine-Tuning*
 Fonte: (ZHANG et al., 2023)

2.5 Métodos de Amostragem e Seleção de Hiperparâmetros

Os métodos de amostragem e de seleção de hiperparâmetros desempenham um papel fundamental no desenvolvimento e avaliação de algoritmos de AM. Fornecendo uma forma adequada de dividir o conjunto de dados, os métodos de amostragem nos permitem obter uma avaliação mais confiável e precisa do modelo, além de nos ajudar a evitar problemas de *overfitting*. Os métodos de seleção de hiperparâmetros por sua vez, exercem um papel crucial no treinamento de algoritmos de AM, nos fornecendo uma maneira de escolher o hiperparâmetros que podem fazer com que o nosso modelo seja mais eficiente, impactando tanto na convergência como o desempenho final do modelo. Quando trabalhamos com comitês de classificadores, a seleção otimizada de hiperparâmetros tem um papel ainda mais evidente. Como visto anteriormente na seção 2.3, um comitê é formado por vários classificadores que de certa forma, colaboram para tomar decisões. Cada classificador individual pode ter seus próprios hiperparâmetros e a seleção precisa desses hiperparâmetros pode afetar a diversidade dos classificadores no comitê, impactando diretamente na sua capacidade de generalização e desempenho preditivo.

2.5.1 Divisão do conjunto de dados

A divisão do conjunto de dados em dois ou mais subconjuntos, é fundamental no aprendizado de máquina para avaliar e ajustar modelos adequadamente, garantindo que possam generalizar bem. Geralmente, o conjunto de dados costuma ser dividido entre três novos subconjuntos: o conjunto de treinamento, o de validação e o de teste. Os dados presentes no conjunto de treinamento são usados para treinar o modelo, os

dados do conjunto de validação são utilizados para comparar os modelos e hiperparâmetros, e os dados do conjunto de teste são utilizados para verificar o desempenho final. Essa divisão do conjunto de dados ajuda a evitar que o modelo se ajuste demais aos dados de treinamento, o que conhecemos como *overfitting*, e acabe prejudicando a sua generalização para dados novos (GOODFELLOW; BENGIO; COURVILLE, 2016). Por isso, separar outros subconjuntos independentes do processo de treinamento, com dados nunca vistos antes pelo modelo, como o conjunto de teste, é extremamente importante para garantir uma avaliação imparcial e um melhor desempenho (GÉRON, 2022).

2.5.2 Validação Cruzada

A Validação Cruzada (*K-fold Cross Validation*), representada na Figura 19, é uma técnica de seleção de modelos, especialmente útil em conjuntos de dados pequenos, onde os dados são divididos em k grupos (conhecidos como *folds*). Durante cada iteração, $k - 1$ grupos são usados para treinamento e o grupo restante é utilizado para teste. Essa abordagem permite que cada dado tenha a chance de ser validado (BISHOP, 2006). Por fim, a média de todas as avaliações do modelo é calculada, proporcionando uma avaliação mais precisa do modelo, mesmo com conjuntos de dados limitados. Porém, o tempo de treinamento pode aumentar proporcionalmente ao número de conjuntos de validação, o que pode ser caro computacionalmente (GÉRON, 2022).

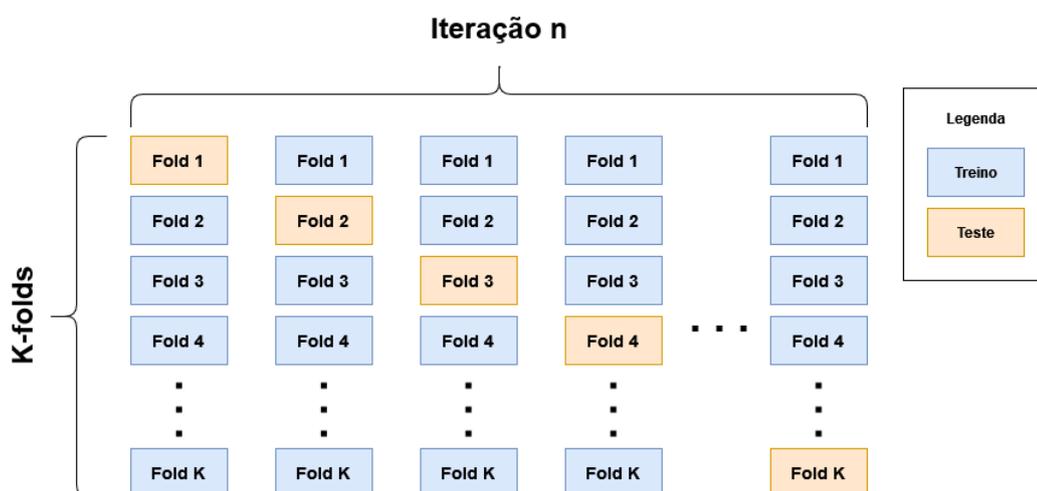


Figura 19 – Representação da Validação Cruzada
Fonte: Autor (2024)

2.5.3 Busca Aleatória

A Busca Aleatória, ou *Random Search*, é uma técnica muito utilizada no Aprendizado de Máquina para otimização e seleção de hiperparâmetros de modelos. Ela busca identificar combinações eficazes de hiperparâmetros, sem se restringir a padrões pré definidos e permite que uma ampla variedade de combinações de hiperparâmetros seja explorada. Os conjuntos de hiperparâmetros são escolhidos aleatoriamente e usados para treinar os modelos, geralmente com cada uma das combinações sendo avaliada por meio da Validação Cruzada. Dessa forma, a Busca Aleatória tende a ser mais simples de ser implementada e mais eficiente, quando comparada com outras abordagens, como a Busca em Grade, porém, pode ser limitada quando aplicada em problemas complexos (BERGSTRÄ; BENGIO, 2012).

2.6 Data Augmentation

Data Augmentation (Aumento de Dados) é uma técnica que amplia a diversidade dos conjuntos de dados de treinamento, realizando transformações nos exemplos de treinamento para gerar exemplos similares, porém diferentes. Essas transformações, geralmente aleatórias, ajudam a reduzir o sobreajuste ao introduzir variações nos dados de treinamento. Por exemplo, podemos recortar as imagens de diferentes maneiras, para que os objetos de interesse apareçam em diferentes posições, reduzindo a sensibilidade do modelo da posição de destino onde os objetos aparecem. Inverter a imagem da esquerda para a direita ou de cima para baixo, ou até mesmo rotacionar a imagem em um determinado grau, geralmente não irá afetar a categoria do objeto. Os métodos mais comuns incluem recorte, rotação, translação e alteração das cores das imagens. Ao aplicar essas transformações, o modelo se torna menos dependente de propriedades específicas dos dados, melhorando sua capacidade de generalização. Durante sua aplicação, é comum sobrepor vários dos métodos mencionados anteriormente e aplicá-los em cada um dos dados usando uma composição das técnicas. Normalmente, o DA é aplicado apenas ao conjunto de treinamento, evitando a sua aplicação com operações aleatórias durante a etapa de previsão, o que nos permite obter resultados definitivos durante a previsão (ZHANG et al., 2023). Esta técnica é particularmente útil em conjuntos de dados desbalanceados, pois pode aumentar virtualmente a quantidade de dados disponíveis, minimizando os efeitos da escassez de exemplos da classe minoritária durante o treinamento de uma rede neural (OLIVEIRA; LEITE, 2021). Na Figura 20 é possível visualizar exemplos de transformações aplicadas em uma imagem digital.



Figura 20 – Exemplo de transformações aplicadas com o uso do *Data Augmentation*
Fonte: (ZHANG et al., 2023)

2.7 Métodos de Avaliação de Desempenho

A Matriz de Confusão (Figura 21) é uma tabela que apresenta o número de previsões corretas e incorretas feitas pelo modelo para cada classe. Ela é útil para avaliar o desempenho do modelo em cada classe individualmente e calcular métricas como a Precisão (*Precision*), Revocação (*Recall*) e Pontuação F1 (*F1-score*). Para o presente trabalho, iremos assumir duas classes, a classe Melanoma, como sendo a classe Positiva e a classe Não Melanoma, como sendo a classe Negativa.

		Classe Predita	
		Positivo	Negativo
Classe Verdadeira	Positivo	VP	FN
	Negativo	FP	VN

Figura 21 – Matriz de Confusão.
Fonte: Autor (2024)

- **Verdadeiro Positivo (VP):** Indica os casos em que o modelo previu corretamente que uma instância pertencia à classe positiva. No caso deste trabalho: quando

a lesão pertence a classe melanoma e foi classificada corretamente como melanoma.

- **Falso Positivo (FP):** Representa os casos em que o modelo previu incorretamente que uma instância pertencia à classe positiva, quando na verdade ela pertencia à classe negativa. No caso deste trabalho: quando a lesão pertence a classe não melanoma e foi classificada de forma incorreta como pertencente a classe melanoma.
- **Falso Negativo (FN):** São os casos em que o modelo previu erroneamente que uma instância pertencia a uma classe negativa, quando na verdade ela pertencia à classe positiva. No caso deste trabalho: quando a lesão pertence a classe melanoma e foi classificada de forma errada como pertencente a classe não melanoma.
- **Verdadeiro Negativo (VN):** São os casos em que o modelo previu corretamente que uma instância pertencia à classe negativa. No caso deste trabalho: quando a lesão não pertence a classe melanoma e foi classificada de forma correta como pertencente a classe não melanoma.

Através dessas quatro categorias, a Matriz de Confusão nos ajuda a entender melhor como o modelo está se saindo em termos de detecção e classificação das diferentes classes no conjunto de dados.

2.7.1 Precisão, Revocação e Pontuação F1

A Precisão e a Revocação não são métricas de avaliação de desempenho únicas, pois fornecem dois números para avaliar o classificador. A Precisão (Equação 2.1) mede a proporção de Verdadeiros Positivos corretos em relação ao total de previsões positivas, sendo importante em casos com alto custo de Falsos Positivos (NG, 2018).

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.1)$$

Já a Revocação (Equação 2.2) mede a proporção de Verdadeiros Positivos em relação ao total de amostras positivas reais, sendo relevante quando há alto custo de Falsos Negativos (NG, 2018).

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (2.2)$$

Geralmente, tende a existir uma compensação entre alta precisão e alta revocação. Ao se trabalhar tanto com a precisão quanto a revocação, é interessante usar

uma maneira padrão de combiná-las em um único número. Uma opção é calcular a Pontuação F1 (*F1-score*), a qual se trata da média harmônica da precisão e da revocação, e é útil quando há um desequilíbrio entre as classes. Essa métrica fornece uma avaliação considerando ambas precisão e revocação (NG, 2018), e é calculada conforme demonstrado na Equação 2.3:

$$F1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (2.3)$$

Ter uma métrica de avaliação de desempenho facilita a tomada de decisões ao escolher entre vários classificadores. Isso resulta em uma classificação clara de preferência entre eles, fornecendo uma direção para o progresso. Portanto, para este trabalho, utilizaremos as medidas Precisão, Revocação e Pontuação F1 como métricas de avaliação de desempenho.

3 Trabalhos Relacionados

No presente capítulo abordamos os trabalhos relacionados que serviram como inspiração para o estudo e desenvolvimento do modelo proposto. Esses trabalhos empregam estratégias de Aprendizado de Máquina com o intuito de fornecer um auxílio ao diagnóstico de lesões dermatoscópicas por meio da análise de imagens digitais

O reconhecimento automático do melanoma enfrenta desafios devido à alta similaridade visual entre lesões de melanoma e não melanoma. Diante dessa problemática, [SANTOS et al. \(2017\)](#) propõem uma abordagem de Aprendizado Profundo utilizando Redes Neurais Convolucionais e métodos tradicionais do Aprendizado de Máquina, que seja capaz de extrair informações importantes das lesões de pele, possibilitando a classificação das lesões do tipo melanoma precisamente. Por usar um conjunto de dados relativamente pequeno, os autores fizeram uso de técnicas de *Data Augmentation* (DA) para aumentar artificialmente o número de amostras. Para a extração de características os autores treinaram as redes convolucionais AlexNet e VGG-F com as imagens do conjunto de dados. Os dados obtidos são classificados nas últimas camadas através dos algoritmos Máquina de Vetores de Suporte (SVM), K-Vizinhos mais Próximos (KNN) e Perceptron Multicamadas (MLP). A combinação da VGG-F utilizando *Batch Normalization* (Normalização em lote) com o classificador KNN alcançou o melhor resultado com uma acurácia de 91,5%, índice *kappa* de 0,83, revocação e precisão superiores a 0,9. Os autores optaram por não utilizar as redes AlexNet e VGG-F pré-treinadas, não fazendo uso de todo o conhecimento que poderia ser obtido da base de dados ImageNet.

Visando explorar o potencial das Redes Neurais Convolucionais e apresentar um modelo capaz de diagnosticar melanomas e nevos displásicos em imagens dermatoscópicas, [PESSANHA e CAMPOS \(2021\)](#), optaram por não desenvolver uma arquitetura própria, mas fazer o uso do *Transfer Learning*, adaptando a arquitetura Inception-ResNet-v2 para tentar aprimorar os seus resultados. A arquitetura utilizada foi baseada na família Inception com a adição de conceitos da família ResNet. Como forma de aprimoramento, foram adicionadas três camadas sequenciais às camadas da arquitetura Inception-ResNet-v2, sendo a primeira, uma camada de *Global Average Pooling 2D*, a segunda, uma camada de *Dropout* com valor de 0,5, e a terceira, uma camada de *Batch Normalization*. As duas primeiras camadas adicionais foram adicionadas visando diminuir o risco de *overfitting* do modelo, enquanto a última camada adicional, foi adicionada pretendendo obter valores mais estáveis para serem processados na camada de saída. A camada de saída conta com dois neurônios, um para a classe melanoma e outra para a classe não melanoma, usando uma função de ativação *softmax*.

Utilizando amostras obtidas através da base do ISIC, no conjunto de teste o modelo atingiu uma sensibilidade de 0,75, especificidade de 0,90 e ROC AUC (Área Sob a Curva da Característica de Operação do Receptor) de 0,86, destacando seu desempenho satisfatório, especialmente em comparação com o modelo o qual realizava uma análise comparativa.

OLIVEIRA e LEITE (2021), realizaram um estudo comparativo do uso de técnicas de técnicas de *Transfer Learning* e *Data Augmentation*, com o objetivo de avaliar o impacto dessas técnicas no aumento do desempenho das arquiteturas de RNCs escolhidas e a influência que uma base de dados desbalanceada poderia ter na capacidade preditiva dos modelos utilizados. Para o estudo foram utilizadas as arquiteturas AlexNet, GoogLeNet e ResNet. Duas bases distintas foram utilizadas para o treinamento das arquiteturas. Inicialmente, com o intuito de utilizar uma base consideravelmente grande para fazer com que durante o treinamento, as arquiteturas aprendessem o processo de extração de características em imagens, os autores utilizaram um subconjunto da base de dados ImageNet com mais de 1 milhão de imagens e distribuídas em 1000 classes. Em seguida, as arquiteturas foram treinadas, com a base de dados contendo as imagens de lesões de pele, obtidas através do ISIC *Challenge*. Após as etapas experimentais, cada arquitetura foi avaliada de forma individual, e a ResNet, ao aplicar as técnicas de TL e DA em conjunto, destacou-se com uma acurácia balanceada de 0,752, revocação de 0,752, especificidade de 0,753 e ROC AUC de 0,839. A eficácia combinada das técnicas resultou em um aumento na acurácia balanceada em mais de 25% para a ResNet, evidenciando sua utilidade conjunta. Os autores consideram que o desbalanceamento das classes teve um impacto significativo no desempenho, e que só foi possível verificar um aumento desse desempenho, após tentarem mitigar os problemas causados pelo desbalanceamento. Também consideraram que o uso de técnicas como comitês (*ensembles*) poderiam ajudar a reduzir a variabilidade das avaliações de cada uma das arquiteturas.

Aplicando técnicas de processamento de imagens e técnicas de TL com arquiteturas de RNCs, UTSCH (2018) propôs um classificador binário de imagens de lesões dermatológicas utilizando adaptações da rede GoogLeNet, especificamente a arquitetura Inception-v3. Para o seu projeto, o autor visou resolver o problema proposto pelo desafio ISIC 2017, que envolvia a classificação binária de imagens de melanomas ou ceratoses seborréicas contra nevos. Ao retirar algumas camadas da rede original e treinar os classificadores em camadas mais baixas, foi identificado melhorias no desempenho dos modelos. A substituição do *average pooling* por uma camada totalmente conectada nessas camadas baixas também contribuiu para resultar em um desempenho melhor. O classificador que demonstrou o melhor resultado, obteve uma média de curva ROC AUC de 0,877, e teve um desempenho considerado satisfatório pelo autor. Contrariando a literatura, foi observado que o uso de imagens pré-processadas redu-

ziu o desempenho do classificador, Utsch acredita que isso possa ter ocorrido devido ao uso do TL, já que a rede original foi treinada para imagens naturais. O autor sugere que o uso de imagens apenas normalizadas em cor pode melhorar o desempenho, e supõe que a técnica de *sharpening* que foi aplicada, pode ter modificado as imagens ao ponto de resultar na perda de informações cruciais. Apesar de aplicar variações na rede e nos hiperparâmetros da GoogLeNet, o autor poderia ter feito uso de técnicas de comitê (*ensemble*) com os classificadores implementados para reduzir a variabilidade dos resultados, mas optou por escolher apenas o classificador que obteve o melhor resultado.

Com o objetivo de desenvolver uma forma de automatização para o auxílio da detecção de câncer de pele, DELAZERI e STEVANI (2020) analisaram como as técnicas de *Data Augmentation* e *Transfer Learning* impactam o desempenho de RNCs. O autor fez uso de quatro modelos de RNC para classificar imagens de câncer de pele, sendo eles o VGG-16, ResNet-50, ResNet-101 e Inception-ResNet. As amostras utilizadas no projeto foram adquiridas da base de dados HAM10000, disponível no site *Kaggle*. A base de dados escolhida é desbalanceada, contendo 10.015 imagens dermatoscópicas rotuladas de sete categorias de lesões da pele. A abordagem escolhida foi utilizar TL com a base de imagens da ImageNet e DA para lidar com o desbalanceamento dos dados, a fim de melhorar o desempenho das RNCs. As quatro redes foram analisadas e tiveram os seus desempenhos obtidos em relação à precisão, sensibilidade e especificidade, a partir de cinco diferentes cenários desenvolvidos pelo autor. Através dos resultados obtidos, foi identificado que a rede VGG-16 obteve os melhores resultados em termos de precisão geral (82% no primeiro cenário) e métricas analisadas para cada classe separadamente (quinto cenário). Foi verificado que o uso de DA não ocasionou uma melhora significativa para a precisão geral. O autor considera que o refinamento das técnicas utilizadas aliado à validação em conjuntos de dados externos e testes em ambiente clínico podem contribuir para uma melhora do modelo proposto como ferramenta de auxílio diagnóstico.

Perante os trabalhos apresentados, o presente trabalho se dispõe a apresentar e explorar técnicas de Aprendizado de Máquina, propondo como diferencial, uma abordagem que através de técnicas de Comitês (*Ensembles*), combina a classificação de vários modelos, criados a partir de uma nova arquitetura de Redes Neurais Convolucionais, com a intenção de comparar o seu desempenho preditivo com os de arquiteturas já consolidadas e amplamente utilizadas na literatura, a AlexNet e a VGG-16.

4 Técnicas preditivas para auxílio no diagnóstico de melanomas via imagens

No capítulo 2, foram introduzidas técnicas e conceitos relevantes para o desenvolvimento deste trabalho. No presente capítulo, serão apresentados o modelo de comitê de Redes Neurais Convolucionais proposto e os modelos utilizados com a técnica de *Transfer Learning*. O objetivo é realizar um estudo exploratório dos modelos que permita comparar a capacidade de predição dos métodos que foram utilizados.

As Redes Neurais Convolucionais são consideradas um dos principais algoritmos para a classificação de imagens, pois, as suas camadas convolucionais conseguem através de operações de convolução extrair características das imagens, sem desconsiderar a sua estrutura espacial (NIELSEN, 2015).

Para este trabalho, foram utilizadas duas abordagens diferentes para a classificação das imagens dermatoscópicas envolvendo RNCs: a primeira, um comitê de classificadores, com o seu aprendiz base sendo definido a partir de um novo modelo de Rede Neural Convolutional criado e treinado do zero, chamado de *SkinLesionClassifier*; e o segundo, a classificação a partir de dois modelos que fazem uso de arquiteturas conhecidas e validadas na literatura, utilizando as redes AlexNet e VGG-16, com técnicas de *Transfer Learning*.

4.1 Comitê de Redes Neurais Convolucionais

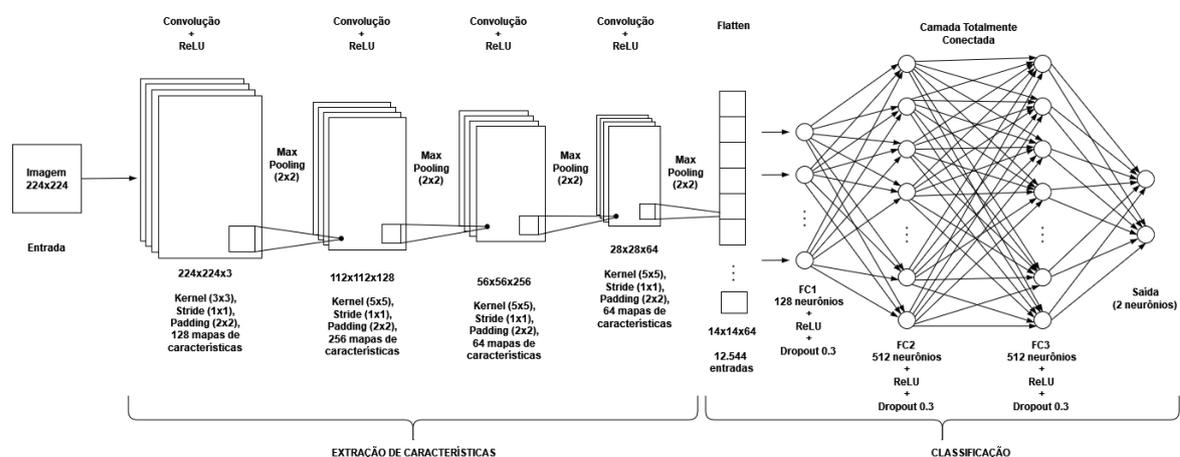


Figura 22 – Arquitetura da SkinLesionClassifier
Fonte: Autor (2024)

Estruturada por 2 grupos, sendo eles, o grupo de Extração de Características, composto por 4 camadas convolucionais (Convolução 2D + ReLU + *Max Pooling*), seguido pelo grupo de Classificação, composto por 3 camadas totalmente conectadas (Camada totalmente conectada + ReLU + *Dropout*) e uma camada de saída contendo 2 neurônios, a *SkinLesionClassifier* (Figura 22) é uma RNC que foi desenvolvida para servir como aprendiz base (*base learner*) para os estimadores utilizados no Comitê de classificadores proposto por este trabalho.

A *SkinLesionClassifier* recebe como entrada uma imagem digital, convertida em um Tensor. Esse Tensor atravessa as camadas convolucionais presentes no grupo de Extração de Características, onde cada camada irá identificar e abstrair aspectos importantes da imagem. As camadas seguintes são capazes de aprimorar o nível de abstração e extração de atributos. Ao passar por todas as camadas convolucionais, o Tensor multidimensional contendo as características obtidas é transformado em um Tensor unidimensional através de um processo conhecido como Achatamento (*Flatten*). Dessa forma, os dados podem ser direcionados para as camadas totalmente conectadas no grupo de Classificação, onde as camadas irão aprender com os dados, adquirindo conhecimento e o propagando adiante na rede. No final, será obtido um Tensor com dois valores, onde o primeiro corresponde a probabilidade da primeira classe, não melanoma, e o segundo, a probabilidade da segunda classe, melanoma.

Devido ao fato que redes únicas podem se tornar enviesadas e possuir um variância muito grande, o uso de técnicas de comitê, como o *Bagging*, ajudam a mitigar esses problemas, além de auxiliar no combate ao sobreajuste (*overfitting*). Considerando os benefícios do uso de técnicas de comitês, foi implementada a classe chamada *BaggingEnsembleClassifier*, que recebe alguns parâmetros na construção, como o aprendiz base (*base learner*) e o número de estimadores desejados, e aplica a técnica de *Bagging* através dos seus métodos *fit* (responsável pelo treinamento e validação) e *predict* (responsável pela predição dos estimadores). O pseudocódigo do Algoritmo 2 descreve o funcionamento dos métodos *fit* e *predict* da *BaggingEnsembleClassifier*.

Algorithm 2 BaggingEnsembleClassifier

```
1: Função fit
2:   Para  $i \leftarrow 1$  até Número de estimadores faça
3:     Obtenha os conjuntos de treino e validação para o estimador( $i$ )
4:     melhor_metrica  $\leftarrow 0.0$ 
5:     melhor_modelo  $\leftarrow$  Nulo
6:     Para cada época em épocas faça
7:       Treine o estimador( $i$ ) com os dados do conjunto de treino( $i$ )
8:       Valide o estimador( $i$ ) com os dados do conjunto de validação( $i$ )
9:       Se as métricas de validação do modelo  $>$  melhor_metrica então
10:        melhor_modelo  $\leftarrow$  modelo
11:     Fim Se
12:   Fim Para
13: Fim Para
14: Fim Função
15:
16: Função predict
17:   predicao_estimadores  $\leftarrow$  []
18:   Para cada estimador em estimadores faça
19:     Obtenha as predições do estimador
20:     Armazene as predições do estimador em predicao_estimadores
21:   Fim Para
22:   Obtenha todas as classes previstas em predicao_estimadores
23:   Determine a classe com mais votos e armazene a previsão final
24:   Retorne Predições do Comitê
25: Fim Função
```

O método *fit*, ilustrado no diagrama apresentado na Figura 23, realiza uma iteração sobre cada um dos estimadores, e para cada estimador é realizada uma subamostragem aleatória dos dados com reposição, onde uma parte dos dados (equivalente a 70% da base inicial) é separada para treinamento e uma parte (equivalente a 15% da base inicial) é separada para validação. Essa subamostragem acontece para cada um dos estimadores e todas as amostras possuem a mesma cardinalidade. Dessa maneira, mesmo que todos os estimadores sejam uma instância do aprendiz base, eles terão acesso aos dados de forma diferente, conseqüentemente, aprendendo de forma diferente. Durante a fase de validação, para cada estimador, o modelo que apresentar o melhor desempenho, considerando as métricas a serem consideradas, será salvo na lista final de estimadores, que será utilizada para realizar a predição através do método *predict*.

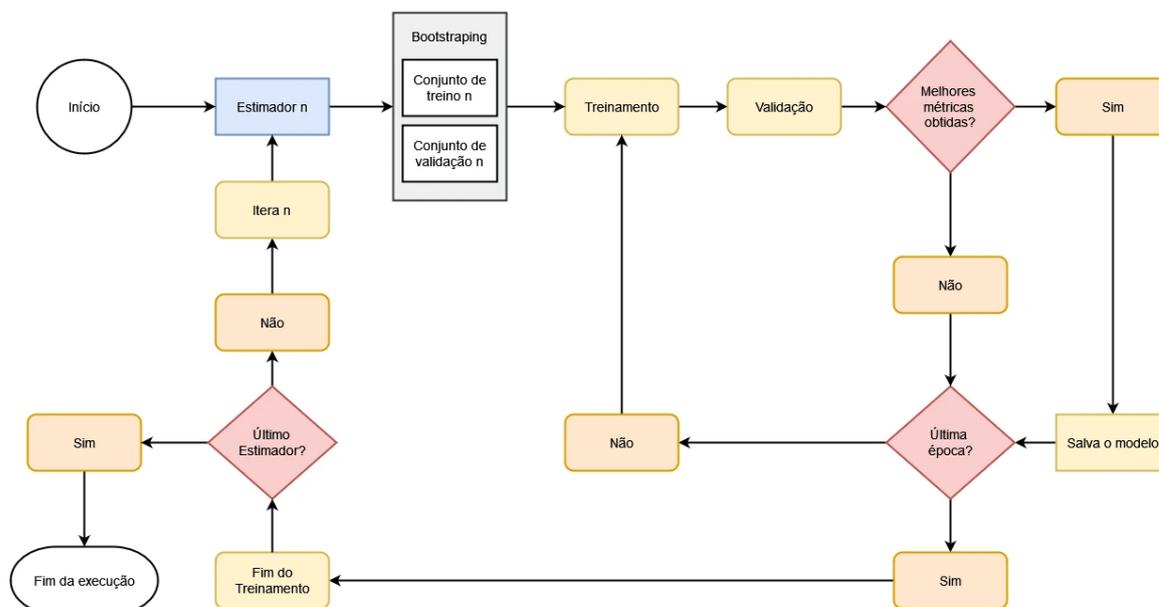


Figura 23 – Fluxo de execução do método *fit*
 Fonte: Autor (2024)

O método *predict*, ilustrado na Figura 24, por sua vez, faz com que todos os estimadores avaliem o mesmo conjunto de teste, como uma forma de garantir uma avaliação justa. Cada estimador irá classificar as imagens e o resultado obtido por cada um servirá como uma base, onde será realizada uma votação para verificar a classe predominante entre os resultados, sendo essa a classe definida pelo comitê.

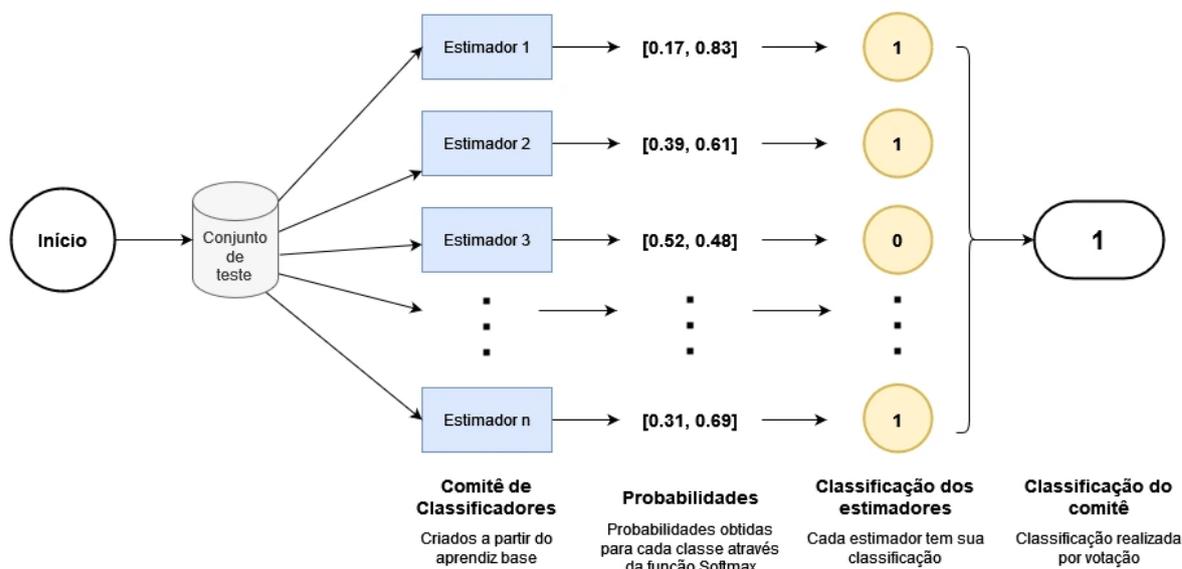


Figura 24 – Fluxo de execução do método *predict*
 Fonte: Autor (2024)

4.2 Redes Neurais Convolucionais e Transfer Learning

Através da aplicação de diversas abordagens e métodos como a utilização de camadas convolucionais e técnicas de *pooling*, as Redes Neurais Convolucionais foram se tornando uma das estruturas mais populares para classificação de imagens e objetos. Essa popularidade incentivou o desenvolvimento de inúmeras arquiteturas que se mostraram muito eficientes e algumas delas estão disponíveis para uso do público em geral. Para este trabalho, foi feito o uso de duas dessas arquiteturas: a AlexNet e a VGG-16.

A AlexNet possui 5 camadas convolucionais, todas seguidas de *max pooling*, essas camadas são responsáveis por realizar a extração de características das imagens através das operações de convolução e realizar o processo de *downsampling*, que é o procedimento utilizado para reduzir o tamanho da amostra para uma versão com uma resolução menor. Essas camadas convolucionais são seguidas por 3 camadas totalmente conectadas, onde as duas primeiras camadas possuem 4096 neurônios ocultos cada e a última, a camada de saída, possui 1000 neurônios, onde cada neurônio representa uma das classes do desafio do *ImageNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

A VGG faz uso do conceito de blocos de camadas convolucionais, onde cada bloco possui mais de uma camada convolucional que são agrupadas de maneira sequencial e são seguidas por *max pooling* (SIMONYAN; ZISSERMAN, 2014). A sua versão com 16 camadas, conhecida como VGG-16, possui 5 blocos convolucionais, onde cada um dos dois primeiros blocos possui 2 camadas convolucionais seguidas por *max pooling*, e cada um dos 3 últimos blocos, é composto por 3 camadas convolucionais seguidas por *max pooling*. Por fim, assim como todas as demais versões da VGG, possui 3 camadas totalmente conectadas, de maneira similar as 3 últimas camadas totalmente conectadas da AlexNet, com cada uma das duas primeiras camadas contendo 4096 neurônios e a última camada, a camada de saída, contendo 1000 neurônios, equivalentes às classes da *ImageNet*.

Alguns *frameworks* e bibliotecas disponibilizam versões de arquiteturas de RNCs para uso de forma gratuita, como é o caso do *PyTorch* (PyTorch, 2023), que possui implementações tanto da AlexNet como da VGG-16, com a opção de utilizá-las pré-treinadas na base de dados *ImageNet*. Isso nos permite explorar essas arquiteturas sem a necessidade de recriá-las do zero e fazer uso do conhecimento adquirido através de uma base de dados enorme. Esses fatores nos possibilitam utilizar essas redes com técnicas como o *Transfer Learning*, que consiste em aproveitar o conhecimento de um modelo que foi treinado em um domínio de origem, para detectar características de um domínio de destino, utilizando as informações e aspectos aprendidos no domínio de origem. Essa técnica é bastante útil quando lidamos com conjuntos de dados

pequenos ou que possuem classes desbalanceadas e quando possuímos pouco poder computacional, já que realizar um treinamento de uma RNC é altamente custoso.

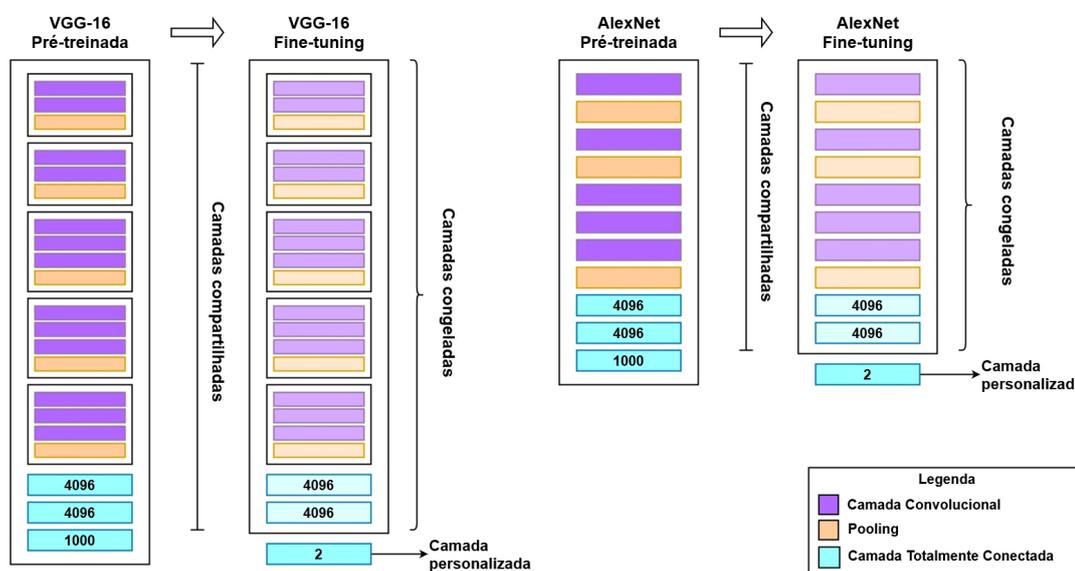


Figura 25 – Processo de *Fine-tuning* das redes VGG-16 e AlexNet
 Fonte: Autor (2024)

Dessa forma, é possível utilizar o conhecimento de um modelo pré-treinado, em uma base de amostras muito maior e usar desse conhecimento para obter informações de uma outra base, para uma outra tarefa. Quando se trata de TL e Aprendizado Profundo é comum o uso da técnica de *Fine-Tuning*, onde se congelam os parâmetros das camadas mais profundas do modelo, modificando apenas as camadas mais superficiais como a camada de saída, e treinando apenas essas novas camadas com os dados do domínio de destino para que a saída da rede se ajuste à nova tarefa. Para este trabalho, a técnica de *Fine-tuning* foi aplicada tanto na AlexNet como na VGG-16, ambas pré-treinadas na base de dados *ImageNet* e obtidas através do *framework PyTorch*. As duas redes tiveram suas camadas mais profundas e pesos congelados, e a camada de saída das redes foi substituída por uma nova camada com apenas dois neurônios, onde cada neurônio representa uma das duas classes da tarefa explorada neste projeto, a classe não melanoma e a classe melanoma (Figura 25). Essa abordagem visa utilizar o conhecimento das redes pré-treinadas que se encontra presente nas camadas congeladas, mais a camada de saída personalizada treinada com os dados do domínio de destino, para avaliar e classificar a tarefa de identificação de melanomas.

5 Experimentos

Neste capítulo, abordaremos a base de dados utilizada, assim como tudo relacionado a criação dos experimentos, como o ambiente utilizado para o desenvolvimento e o pré-processamento das imagens. Também faremos uma análise e breve discussão sobre os resultados alcançados.

5.1 Conjunto de Dados

Para o desenvolvimento deste trabalho, foi utilizada a base de dados do ISIC Challenge 2019 (Figura 26) (ISIC, 2019b), fornecida pela *International Skin Imaging Collaboration* (ISIC) (ISIC, 2019a) e disponibilizada pela *International Society for Digital Imaging of the Skin* (ISDIS) (ISDIS, 2019).

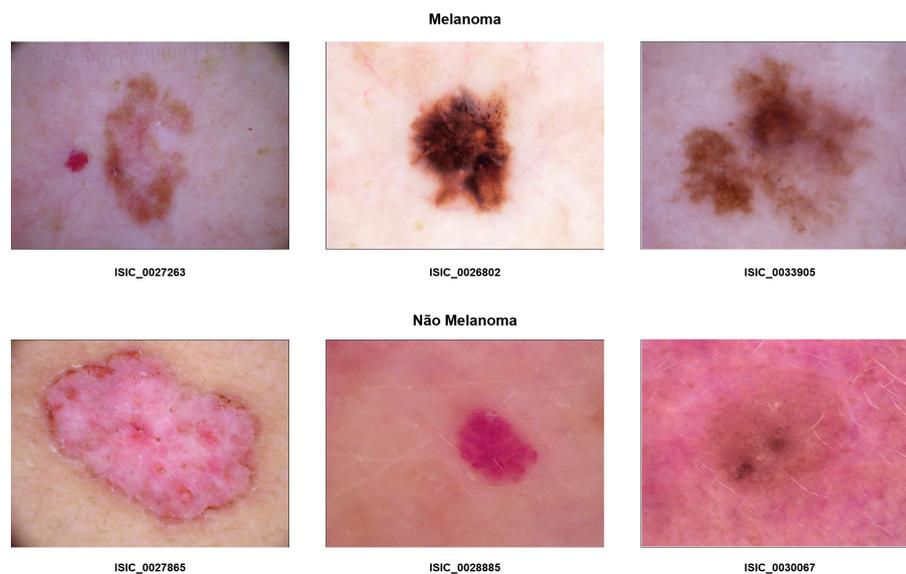


Figura 26 – Amostras da base de dados do ISIC Challenge 2019
Fonte: (ISIC, 2019b)

O ISIC Challenge é um desafio voltado para a comunidade da ciência da computação e do campo da Visão Computacional, que desde 2016 é patrocinado pelo ISIC. Inicialmente, os desafios tinham como objetivo classificar com precisão imagens de lesões dermatoscópicas como melanoma ou não melanoma. À medida que o desafio foi crescendo em complexidade, em número de participantes e os desempenhos apresentados nas precisões iam se superando, os desafios passaram a abordar classificações multiclases de lesões dermatoscópicas. Contendo milhares de imagens coletadas dos principais centros clínicos internacionais, o Arquivo ISIC conta atualmente com a maior coleção de imagens de lesões dermatoscópicas de alta qualidade disponi-

bilizadas de forma pública (ISIC, 2019b). Por se tratarem de bases de dados com alta confiabilidade e gratuitas, as bases dos desafios do ISIC *Challenge* se tornam ideais para serem utilizadas em experimentos (PESSANHA; CAMPOS, 2021; SANTOS et al., 2017; OLIVEIRA; LEITE, 2021).

Doença de pele	Quantidade de imagens
Melanoma	4522
Nevo melanocítico	12875
Carcinoma basocelular	3323
Ceratose actínica	867
Ceratose benigna	2624
Dermatofibroma	239
Lesão vascular	253
Carcinoma de células escamosas	628

Tabela 1 – Distribuição das imagens do ISIC *Challenge* 2019 de acordo com as suas classes.

Fonte: (ISIC, 2019b)

A base de dados do ISIC *Challenge* 2019 contém 25331 imagens dermatoscópicas que estão divididas em oito categorias diferentes, com cada categoria representando um tipo específico de lesão de pele. Na Tabela 1 é possível visualizar a forma como as imagens estão distribuídas nas oito categorias existentes. Cada categoria corresponde a uma classe do desafio de classificação. Quando nos concentramos na classificação das imagens como imagens de melanoma ou não melanoma, é possível identificar uma grande desproporcionalidade, considerando que apenas 4522 imagens (17,85%) pertencem à categoria de melanoma, enquanto que 20809 imagens (82,15%), pertencem a categorias de não melanoma, demonstrando um desbalanceamento na base de dados. Todas as imagens estão no formato JPEG e são do tipo RGB. A base possui 101 dimensões diferentes ao todo, tendo larguras no intervalo de 576 a 1024 *pixels*, e alturas no intervalo de 450 a 1024 *pixels* (ISIC, 2019b).

Para que fosse possível dar prosseguimento com os experimentos, houve a necessidade de realizar uma redução da base de dados devido a limitação dos recursos de hardware. Dessa maneira foi realizada uma subamostragem da base de dados, reduzindo-a para 9000 imagens. Na Tabela 2 é possível visualizar como ficou organizada a base de dados após a realização da subamostragem:

Doença de pele	Quantidade de imagens
Melanoma	4500
Nevo melanocítico	845
Carcinoma basocelular	845
Ceratose actínica	845
Ceratose benigna	845
Dermatofibroma	239
Lesão vascular	253
Carcinoma de células escamosas	628

Tabela 2 – Nova distribuição das imagens após a subamostragem de acordo com as suas classes.

Fonte: (ISIC, 2019b)

Para o desenvolvimento dos experimentos, por se tratar de uma tarefa de classificação binária, as amostras foram divididas em apenas duas classes, as imagens pertencentes a classe melanoma e as imagens das demais categorias, foram agrupadas em uma nova classe, a não melanoma, totalizando assim um total de 4500 imagens da classe melanoma e 4500 para a classe não melanoma (Tabela 3). A realização da subamostragem permitiu realizar um balanceamento da base, de maneira que agora cada categoria possui 50% dos dados total da base.

Classe	Quantidade de imagens
Melanoma	4500
Não Melanoma	4500

Tabela 3 – Distribuição final das imagens de acordo com as novas classes.

Fonte: (ISIC, 2019b)

Para execução dos experimentos, foi realizada uma divisão da base de dados, onde 15% dos dados são separados para o conjunto de teste e os demais 85% serão destinados para o treinamento e validação em uma nova base chamada de base temporária. Os dados do conjunto de teste foram utilizados tanto pelo Comitê de classificadores, como pela AlexNet e a VGG-16. Para a abordagem de *Transfer Learning* com os modelos AlexNet e VGG-16, a base temporária foi dividida de modo que o equivalente a 70% da base inicial foi destinada para treinamento e 15% da base inicial foi destinada para validação. Ambos os modelos utilizaram os mesmos conjuntos de dados. Para o Comitê de classificadores, devido a técnica de *Bagging*, foi necessário realizar uma subamostragem com reposição dos dados (*bootstrap*) contidos na base temporária para cada um dos estimadores, sendo uma parte dos dados (equivalente a 70% da base inicial) separada para treinamento e uma parte (equivalente a 15% da base inicial) separada para validação. Todas as divisões de amostras foram realizadas através da técnica de divisão de conjuntos aleatória de maneira embaralhada e estratificada, utilizando a função *train_test_split* da biblioteca *Scikit-Learn* (Scikit-Learn, 2023).

O conjunto de teste recebeu um total de 1350 imagens, enquanto que para cada um dos conjuntos de treinamento (para cada estimador do comitê e para o TL) foram destinadas 6300 imagens e para cada um dos conjuntos de validação (para cada estimador do comitê e para o TL) 1350 imagens. Após a divisão dos conjuntos, a porcentagem das amostras para cada classe foi preservada graças à estratificação.

5.2 Ambiente experimental

Para o desenvolvimento do projeto foi utilizada a linguagem de programação *Python* em sua versão 3.10.12 ([Python, 2024](#)). As bibliotecas *Pandas* (versão 2.0.3) ([Pandas, 2023](#)), *NumPy* (versão 1.25.2) ([NumPy, 2024](#)), *Scikit-Learn* (versão 1.3.2) ([Scikit-Learn, 2023](#)) e *PyTorch* (versão 2.3.1+cu121) ([PyTorch, 2023](#)), foram utilizadas para o tratamento dos dados, implementação e execução dos algoritmos de Aprendizado de Máquina.

O *Google Colaboratory (Colab)* foi usado como ambiente de desenvolvimento, por oferecer um ambiente com vários recursos para codificação em *Python* na nuvem, voltados para o AM, análise de dados e educação ([Google Colab, 2023](#)). Inicialmente, os experimentos foram executados utilizando os recursos gratuitos disponibilizados pelo *Colab*, que contam com memória RAM do sistema de 12.7GB, disco com 107.7GB. Ainda em sua versão gratuita, é disponibilizada a GPU no modelo GPU T4 com memória RAM do sistema de 12.7GB, memória RAM da GPU de 15GB e disco com 112.6GB. Porém, devido às limitações do ambiente e a necessidade do uso de mais recursos, foi necessário realizar a atualização do plano gratuito para o Colab Pro+, que disponibiliza mais recursos, como por exemplo, execução em segundo plano, mais memória RAM (até 83.5GB), mais memória RAM da GPU (até 40GB), mais disco (até 235.7GB) e modelos de GPU mais poderosos, como a NVIDIA A100 GPU e a NVIDIA L4 GPU. No total foram contratadas 1300 unidades de processamento do Colab, das quais foram utilizadas aproximadamente 1191 unidades de processamento até o fim dos experimentos.

5.3 Pré-processamento das imagens e técnicas de Data Augmentation

Para leitura das imagens no *Dataset* foi utilizada a biblioteca *Pillow* em sua versão 9.4.0 ([Pillow, 2024](#)) e para a manipulação foi utilizada a biblioteca *Torchvision* do *PyTorch* ([Python, 2024](#)). Segundo sua documentação, a biblioteca *Torchvision* consiste em um pacote de código aberto (*open source*) de conjuntos de dados populares, arquiteturas de modelos e de transformações de imagem comuns para visão computacional.

A biblioteca foi escolhida por ser parte do projeto *PyTorch* e possuir um módulo específico para transformações, o *torchvision.transforms*, que tem total compatibilidade com imagens PIL. Além de fornecer uma gama de transformações de imagens, o módulo *torchvision.transforms* permite que os usuários usem composição de transformações através do método *Compose*.

Como visto na seção 5.1, todas as imagens da base de dados são do tipo RGB, possuindo três canais de cores, e possuem dimensões diversas. Para todas as imagens foram aplicadas as transformações *Resize* e *ToTensor*. A transformação *Resize* redimensiona as imagens para a dimensão dada, no caso deste trabalho, todas as imagens foram redimensionadas para o tamanho 224×224 . A transformação *ToTensor* converte todas as imagens PIL que estão em um intervalo $[0, 255]$ para um Tensor de Imagem no intervalo $[0.0, 1.0]$. Um Tensor de Imagem é um tensor com forma (C, H, W) , onde o C representa o número de canais, H a altura W a largura da imagem. A quantidade de atributos de uma imagem é o resultado do produto de $C \times H \times W$, e os atributos das imagens da base original variam entre 810000 e 3145728 atributos. Com essas duas transformações foi possível normalizar os dados da base para que os algoritmos entendam todos os dados como igualmente importantes, além que após o redimensionamento, todas as imagens ficaram com a mesma dimensão, resultando em 150528 atributos ($3 \times 224 \times 224$) como entrada para as redes, uma redução de 81,42% a 95,21% no número de atributos.

Para o conjunto de treinamento ainda foram aplicadas as transformações *RandomHorizontalFlip* com a probabilidade de 50% da imagem ser invertida e o *RandomRotation* que rotaciona as amostras aleatoriamente em 45° (Figura 27). Essas transformações buscam remover ruídos e destacar algumas características importantes das imagens, fortalecendo a capacidade do modelo em reconhecer padrões de forma generalizada conforme diversifica a variedade dos dados de treinamento. Além das transformações utilizadas, existem inúmeras outras que possuem os mesmos objetivos, como as Transformações Geométricas (rotação, espelhamento, translação, corte, zoom), Transformações Fotométricas (Adição de ruído e transformações nos espaços de cores) e Transformações com Aprendizado Profundo (novas imagens geradas por modelos de *Generative Adversarial Network* ou como são mais conhecidos GAN) (CLARO, 2020).

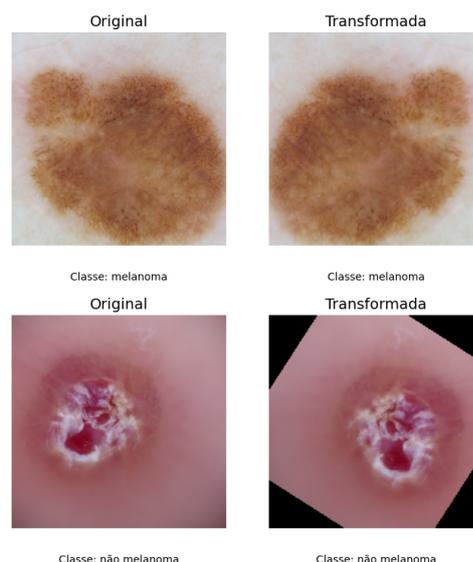


Figura 27 – Amostras de imagens antes e após aplicação das transformações
Fonte: (ISIC, 2019b)

5.4 Seleção de modelos

Os hiperparâmetros desempenham um papel importante na definição da arquitetura e no comportamento de um modelo, e sua escolha tem o poder de afetar de maneira positiva ou negativa o desempenho e a capacidade de generalização de um modelo.

Para este trabalho, foram apresentadas duas abordagens: um Comitê de Redes Neurais Convolucionais, onde o aprendiz base é uma arquitetura nova e a segunda abordagem se trata da utilização de técnicas de *Transfer Learning* através do uso das redes AlexNet e VGG-16.

A Busca Aleatória é uma técnica voltada para a otimização e seleção de hiperparâmetros, onde os conjuntos de hiperparâmetros são escolhidos de forma aleatória e são utilizados para treinar o modelo n vezes. Segundo BERGSTRÅ e BENGIO (2012), a Busca Aleatória tende a ser mais eficiente que a Busca em Grade e a Busca Manual, por permitir a exploração de um espaço de configuração maior e menos promissor. Como para muitos conjuntos de dados apenas alguns hiperparâmetros realmente são importantes, essa abordagem se torna altamente eficaz, permitindo que os melhores conjuntos de hiperparâmetros sejam encontrados em uma fração do tempo computacional que levaria em outras abordagens. Cada combinação é avaliada através da Validação Cruzada (*K-fold Cross Validation*) com o número de k definido, e o número de ajustes (*fit*) realizados para encontrar as melhores combinações é definido por $k \times n$. Para este trabalho, foi utilizado o valor de $k = 5$ e $n = 30$, totalizando 150 ajustes.

Hiperparâmetro	Distribuição de hiperparâmetros
Mapa de características (<i>feature map</i>) da Camada convolucional 1	Uniforme em {32, 64, 128, 256}
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 1	Uniforme em {3, 5, 7}
Mapa de características (<i>feature map</i>) da Camada convolucional 2	Uniforme em {32, 64, 128, 256}
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 2	Uniforme em {3, 5, 7}
Mapa de características (<i>feature map</i>) da Camada convolucional 3	Uniforme em {32, 64, 128, 256}
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 3	Uniforme em {3, 5, 7}
Mapa de características (<i>feature map</i>) da Camada convolucional 4	Uniforme em {32, 64, 128}
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 4	Uniforme em {3, 5, 7}

Tabela 4 – Distribuição de hiperparâmetros para as Camadas convolucionais.
Fonte: Autor (2024)

Hiperparâmetro	Distribuição de hiperparâmetros
Neurônios da Camada totalmente conectada 1	Uniforme em {128, 256, 512}
Neurônios da Camada totalmente conectada 2	Uniforme em {128, 256, 512}
Neurônios da Camada totalmente conectada 3	Uniforme em {128, 256, 512}
Taxa de abandono (<i>Dropout</i>)	Uniforme em {0.2, 0.3, 0.5}

Tabela 5 – Distribuição de hiperparâmetros para as Camadas totalmente conectadas e regularização.

Fonte: Autor (2024)

Hiperparâmetro	Distribuição de hiperparâmetros
Número de épocas	Uniforme em {50, 75, 100}
Tamanho do lote (<i>batch size</i>)	Uniforme em {32, 64}
Taxa de aprendizado (<i>learning rate</i>)	Uniforme em {0.0001, 0.001, 0.01}

Tabela 6 – Distribuição de hiperparâmetros para o treinamento.
Fonte: Autor (2024)

As Tabelas 4, 5 e 6 demonstram os hiperparâmetros e as distribuições utilizadas para compor a Busca Aleatória que resultou no conjunto de hiperparâmetros utilizados pela rede *SkinLesionClassifier*, a qual serviu como aprendiz base para o comitê de classificadores. Alguns hiperparâmetros da rede não foram otimizados, permanecendo fixos, como o número de estimadores para o comitê que permaneceu em 10. A função

de ativação utilizada na arquitetura da rede foi a ReLU, o tipo da subamostragem (*pooling*) foi o *Max Pooling* e o tamanho da janela de subamostragem utilizada teve seu tamanho fixado em 2×2 , por fim, o otimizador utilizado foi o Adam e a função de erro a Perda de Entropia Cruzada (*Cross Entropy Loss*).

Para a execução da Busca Aleatória foi utilizado o módulo *RandomizedSearchCV* da biblioteca *Scikit-Learn* (Scikit-Learn, 2023). Foi necessário o uso de uma outra biblioteca intermediária para utilizar o *Scikit-Learn* com o *PyTorch*, a *Skorch* (versão 1.0.0) (Skorch, 2024). Segundo a sua documentação, a *Skorch* é uma biblioteca para desenvolvimento de redes neurais que torna possível o uso de *PyTorch* com *Scikit-Learn*. A execução da Busca Aleatória é um processo custoso e que durante sua execução, consumiu praticamente todas as unidades de processamento do plano *Google Colab Pro*. Para o procedimento foi utilizada a NVIDIA A100-SXM4-40GB GPU e o processo com 150 ajustes durou 04 horas 47 minutos e 50 segundos. Nos primeiros testes realizados com a CPU inicial fornecida pelo *Google Colab*, a média para cada ajuste dos 10 primeiros ajustes foi de aproximadamente 01 hora e 10 minutos, o que levaria em torno de 07 dias para finalizar todo o processo dos 150 ajustes, fazendo assim necessário a contratação de um plano que com um hardware mais potente.

O *Transfer Learning* oferece uma série de benefícios, permitindo o uso de modelos pré-treinados em grandes conjuntos de dados, reduzindo a quantidade de dados rotulados necessária para treinar um novo modelo. Como o modelo já foi pré-treinado em uma tarefa anterior, ele já aprendeu características úteis, que facilitam a aplicação em novos domínios com ajustes mínimos, sendo necessário apenas o ajuste dos pesos para a nova tarefa, economizando recursos computacionais e tempo (PAN; YANG, 2010). Esses modelos pré-treinados podem ser facilmente ajustados a novas tarefas adicionando novas camadas finais ou realizando um ajuste fino (*fine-tuning*) completo ou parcial, substituindo apenas algumas camadas para a adaptação do modelo a nova tarefa, tornando o processo de desenvolvimento flexível e ágil para diferentes aplicações (YAQING et al., 2020).

Os hiperparâmetros desempenham um papel crucial no processo de *Transfer Learning*, principalmente em um ajuste fino (*fine-tuning*), onde algumas camadas do modelo pré-treinado são congeladas (para não serem treinadas novamente), enquanto outras são ajustadas. Para este trabalho, as camadas de saída dos modelos AlexNet e VGG-16, foram substituídas de modo que os modelos se adequassem à nova tarefa. As camadas de saída continham 1000 neurônios cada, sendo um neurônio para cada classe do *ImageNet*. Essas camadas foram substituídas por uma camada de 2 neurônios cada uma, para classificar as duas classes da nova tarefa. Para o treinamento a função de erro e o otimizador não foram otimizados e foram utilizados os mesmos que o comitê de classificadores. O número de épocas, a taxa de aprendizagem e o

tamanho do lote foram os mesmos obtidos pela Busca Aleatória para o comitê de classificadores.

5.5 Análise e Resultados

Quando nos referimos a classificação de lesões dermatoscópicas, os melanomas podem ter características visuais sutis e variar bastante em aparência. Um Comitê de classificadores pode capturar padrões que um único classificador poderia perder, reduzindo a sensibilidade a ruídos e *outliers* (dados que se destacam drasticamente dos demais) que podem estar presentes nas imagens, tornando a classificação mais consistente. Outro benefício do uso de comitês é que ao combinar diferentes modelos, o risco de sobreajuste aos dados (*overfitting*) diminui, pois existe uma melhora na capacidade dos sistemas de generalizar novos dados.

Já o aprendizado por transferência (*transfer learning*) e o ajuste fino (*fine-tuning*) são técnicas poderosas frequentemente utilizadas em problemas de classificação de imagens, especialmente quando se trabalha com conjuntos de dados limitados ou quando o problema em questão é complexo, como a detecção de melanomas. O uso de modelos pré-treinados em grandes bases de dados genéricas, como o *ImageNet*, que contém milhões de imagens de objetos comuns, permite capturar características visuais generalizadas, como bordas, texturas e padrões de cores e aplicá-las ao problema de identificar lesões cancerígenas em imagens dermatoscópicas.

As três arquiteturas (Comitê de RNCs, AlexNet e VGG-16) utilizadas nos experimentos foram treinadas durante 50 épocas cada. O número de épocas foi obtido através da otimização de hiperparâmetros do aprendiz base do comitê e utilizado para as três arquiteturas. A Curva de Aprendizado expressa a relação entre o erro do conjunto de validação e o número de exemplos de treinamento. Segundo NG (2018), a análise da curva de aprendizado pode evitar desperdícios de tempo na coleta de dados que não trazem benefícios e é um gráfico útil para estimar o impacto da adição de mais dados ao analisarmos o erro de treinamento. A Figura 28 ilustra a Curva de Aprendizado de cada arquitetura durante a fase de treinamento. Conforme mencionado anteriormente, o comitê de RNCs foi treinado por 50 épocas, isso significa que, cada estimador do comitê também foi treinado por 50 épocas. A Figura 29 mostra a Curva de Aprendizado de cada estimador individualmente. A curva do comitê foi calculada a partir da média dos erros de cada estimador.

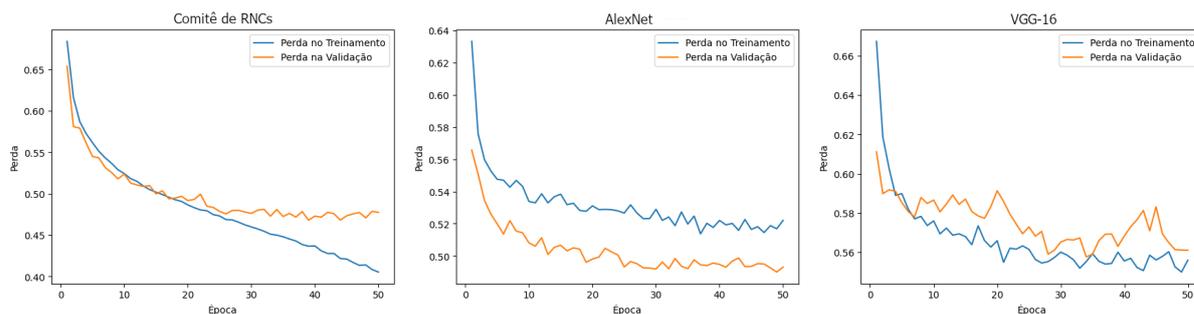


Figura 28 – Curva de Aprendizagem dos modelos
Fonte: Autor (2024)

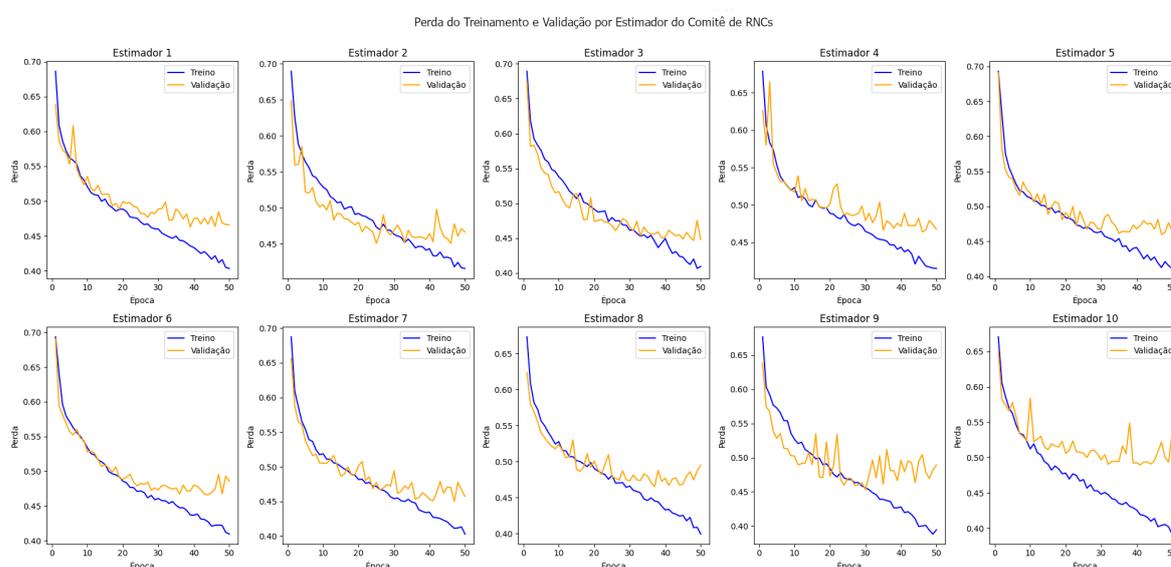


Figura 29 – Curva de Aprendizagem de cada estimador do comitê de RNCs
Fonte: Autor (2024)

No contexto da detecção de melanoma, é muito importante evitar o diagnóstico incorreto de lesões não melanoma como melanoma, pois uma classificação incorreta pode gerar procedimentos médicos desnecessários e custos emocionais para o paciente, porém, garantir que todos os casos de melanoma, que têm um prognóstico potencialmente muito mais grave, sejam identificados corretamente e de forma ágil é extremamente crítico, pois o diagnóstico precoce tem um impacto crucial nas chances de cura do paciente. Dadas essas condições, as métricas utilizadas neste projeto são a Precisão, a Revocação e a Pontuação F1. Para obter tais métricas, é necessário anteriormente gerar a Matriz de Confusão para cada uma das abordagens e seus modelos, visto que a matriz de confusão nos fornece dados para calcular essas e outras várias métricas de desempenho. As imagens 30, 31 e 32 apresentam as matrizes de confusão obtidas.

		Classe Predita	
		Melanoma	Não Melanoma
Classe Verdadeira	Melanoma	579	96
	Não Melanoma	186	489

Figura 30 – Matriz de confusão do Comitê de RNCs
 Fonte: Autor (2024)

		Classe Predita	
		Melanoma	Não Melanoma
Classe Verdadeira	Melanoma	519	156
	Não Melanoma	167	508

Figura 31 – Matriz de confusão da AlexNet com o uso do *Transfer Learning* e *Fine-tuning*

Fonte: Autor (2024)

		Classe Predita	
		Melanoma	Não Melanoma
Classe Verdadeira	Melanoma	501	174
	Não Melanoma	226	449

Figura 32 – Matriz de confusão da VGG-16 com o uso do *Transfer Learning* e *Fine-tuning*

Fonte: Autor (2024)

Embora a Precisão, que representa a proporção de lesões identificadas como melanoma que realmente são, seja importante, a Revocação é a métrica mais crucial e principal para garantir que as lesões cancerígenas sejam detectadas precocemente, permitindo um tratamento eficaz. Em diagnósticos médicos, a perda de um caso (FN) tem um custo elevado e pode ter consequências graves. Nesse contexto, a Pontuação F1 pode ser utilizada como métrica secundária para equilibrar a necessidade de alta Revocação com a Precisão, assegurando que, mesmo com uma alta Revocação, a Precisão não seja excessivamente comprometida, evitando que o modelo seja considerado eficaz com base em métricas isoladas.

Como mencionado anteriormente, o mesmo conjunto de testes foi utilizado pelo comitê e pelos modelos na qual foram aplicadas as técnicas de *Transfer Learning*. O conjunto de testes possui 1350 amostras de imagens dermatoscópicas (15% do total da base de dados), onde, 50% das amostras (675 imagens) pertencem à classe não melanoma e os demais 50% (675 imagens) pertencem à classe melanoma. Ao analisarmos a matriz de confusão de cada abordagem e modelo, é visível que a quantidade de acertos (VP e VN) de cada abordagem foi superior a quantidade de erros, porém, a quantidade de erros (FP e FN) foi consideravelmente alta. Os FP ocorrem quando uma lesão é identificada como melanoma pelo modelo, mas na verdade é não melanoma, enquanto os FN ocorrem quando uma lesão que de fato é melanoma, é classificada erroneamente como não melanoma pelo modelo. Se um modelo otimizar a Revocação, a Precisão pode cair, gerando muitos FP. Por outro lado, se o modelo otimizar a Precisão, pode-se perder casos reais. A Pontuação F1 é a média harmônica entre Precisão e Revocação, e busca um equilíbrio entre essas duas métricas, sendo ideal

para situações em que queremos tanto identificar o máximo de VP quanto evitar FP.

Modelo	Precisão	Revocação	Pontuação F1
Comitê de RNCs	75.69%	85.78%	80.42%
AlexNet	75.66%	76.89%	76.27%
VGG-16	68.91%	74.22%	71.47%

Tabela 7 – Precisão, Revocação e Pontuação F1 das abordagens.
 Fonte: Autor (2024)

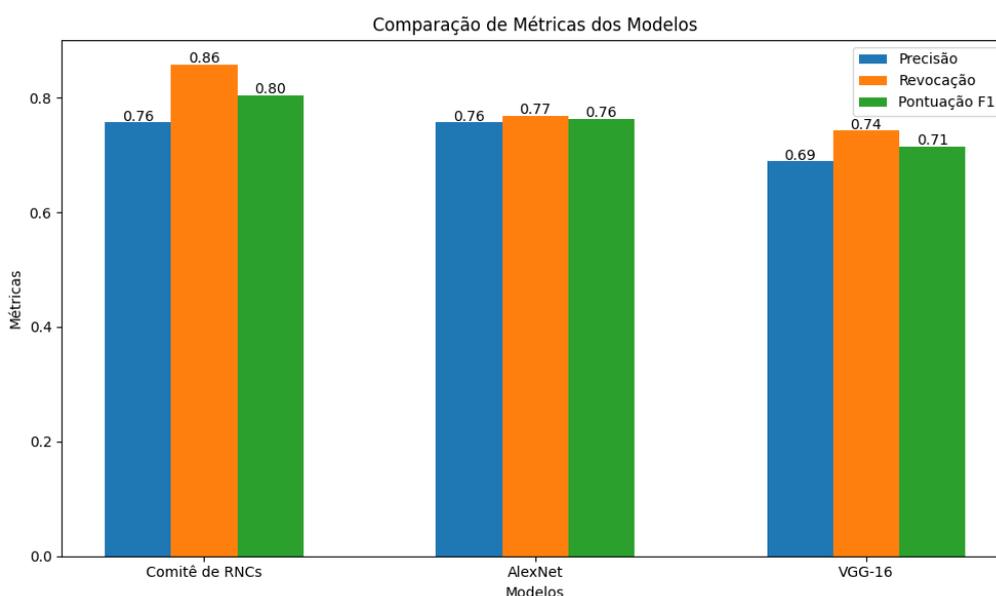


Figura 33 – Comparação gráfica das métricas obtidas pelas abordagens
 Fonte: Autor (2024)

Na Tabela 7 é possível visualizar os valores da Precisão, Revocação e Pontuação F1 das abordagens utilizadas. Os modelos apresentaram resultados bem próximos, especialmente na Precisão, onde o Comitê de RNCs e a AlexNet obtiveram quase o mesmo desempenho, com uma diferença mínima de 0,03%. Já para a Revocação e para Pontuação F1, o comitê de classificadores com a arquitetura de RNCs criada e treinada do zero, apresentou um desempenho geral melhor em comparação com os demais modelos (Figura 33). É possível notar que mesmo sendo uma rede mais profunda, a VGG-16 teve um desempenho inferior ao das demais redes.

Quando analisamos as métricas separadas por classes, vemos que os três modelos apresentaram melhor Precisão para a classe Não Melanoma (Tabela 8 e Figura 34), enquanto apresentaram melhor Revocação para a classe Melanoma (Tabela 9 e Figura 35). A AlexNet demonstrou uma Revocação superior aos outros modelos para casos Não Melanoma. O Comitê por sua vez, obteve o melhor desempenho na Precisão para casos Não Melanoma, enquanto apresentou o melhor resultado na Revocação para a classe Melanoma.

Classe	Precisão	Revocação	Pontuação F1
Comitê de RNCs	84%	72%	78%
AlexNet	77%	75%	76%
VGG-16	72%	67%	69%

Tabela 8 – Precisão, Revocação e Pontuação F1 das abordagens para a classe Não Melanoma.

Fonte: Autor (2024)

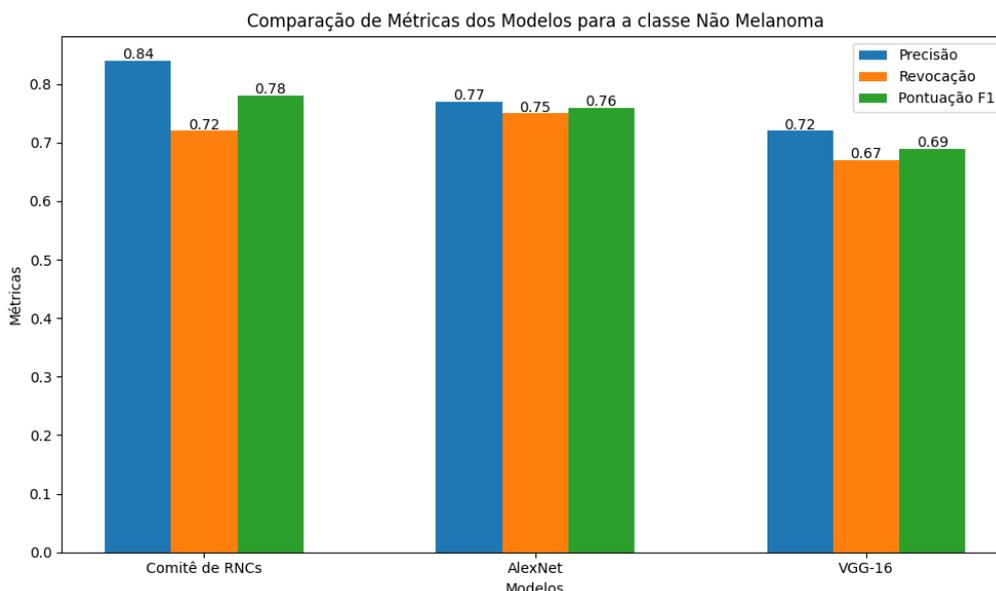


Figura 34 – Comparação gráfica das métricas obtidas pelas abordagens para a classe Não Melanoma.

Fonte: Autor (2024)

Classe	Precisão	Revocação	Pontuação F1
Comitê de RNCs	76%	86%	80%
AlexNet	76%	77%	76%
VGG-16	69%	74%	71%

Tabela 9 – Precisão, Revocação e Pontuação F1 das abordagens para a classe Melanoma.

Fonte: Autor (2024)

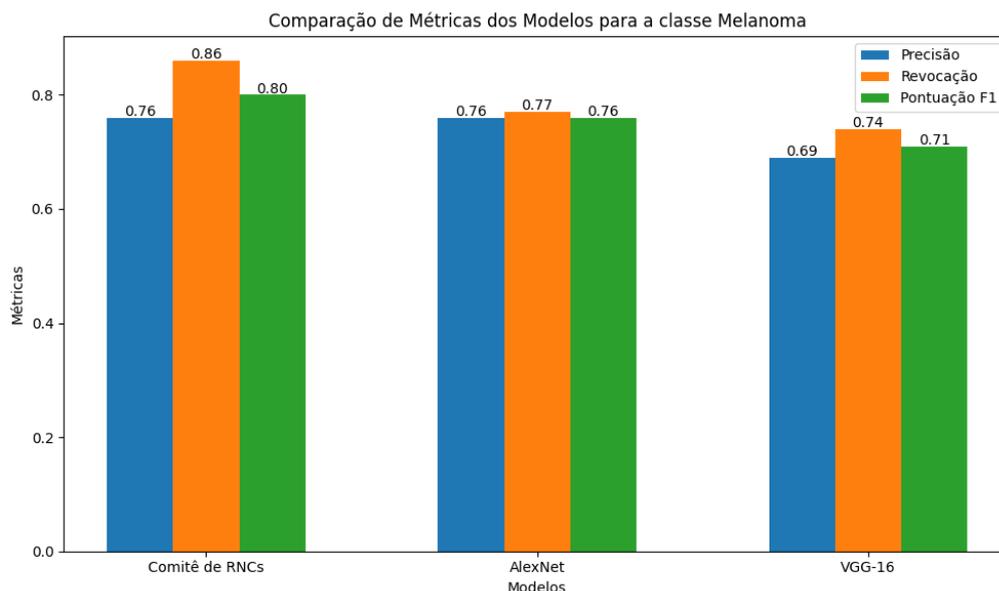


Figura 35 – Comparação gráfica das métricas obtidas pelas abordagens para a classe Melanoma.

Fonte: Autor (2024)

Quando comparamos os resultados apenas da classe Melanoma dos modelos (Tabela 9 e Figura 35), vemos que os resultados se assemelham aos obtidos no geral (Tabela 7 e Figura 33), onde o modelo de comitê também obteve um desempenho melhor, seguido pela AlexNet e com o VGG-16 em último.

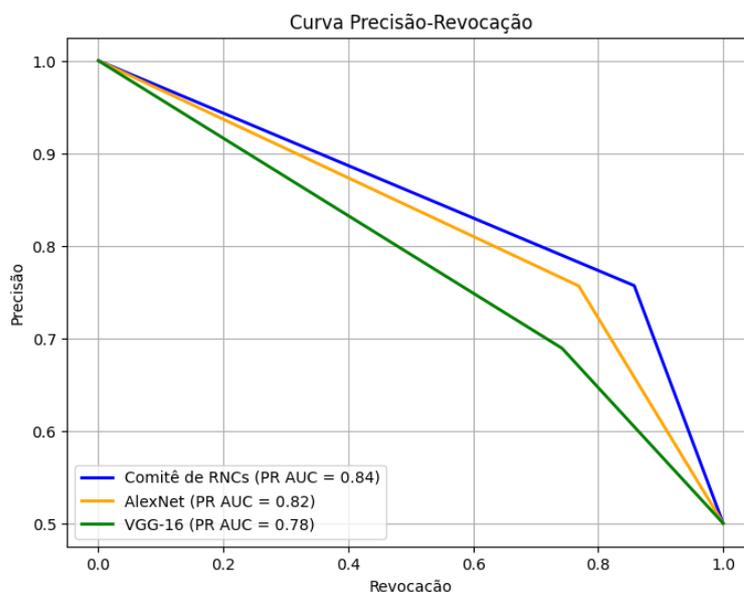


Figura 36 – Curva PR-AUC dos modelos

Fonte: Autor (2024)

A Curva Precisão-Revocação (PR), do inglês *Precision-Recall Curve*, é uma ferramenta valiosa para avaliar o desempenho de classificadores em tarefas de detecção,

especialmente em contextos onde as classes positivas são escassas, como o melanoma. Ela traça a relação entre a Precisão e a Revocação em diferentes limiares de decisão, permitindo que os usuários ajustem o modelo para otimizar a identificação de eventos de interesse. Uma das métricas derivadas da Curva PR é a Área Sob a Curva (PR-AUC), que fornece uma medida geral da qualidade do classificador. Quanto maior a Área Sob a Curva, melhor é o desempenho do modelo em detectar eventos positivos, equilibrando a taxa de verdadeiros positivos e a taxa de falsos positivos. Além disso, a PR-AUC, juntamente com a Pontuação F1, permite resumir o desempenho do classificador em um único número, facilitando comparações entre diferentes modelos e estratégias. Essa abordagem é particularmente útil em cenários onde a Precisão e a Revocação precisam ser balanceadas, como em aplicações médicas, onde decisões incorretas podem ter consequências graves (GOODFELLOW; BENGIO; COURVILLE, 2016). É possível visualizar a Curva PR-AUC de cada modelo na Figura 36. No contexto da detecção de melanomas, a PR-AUC de 0,78 obtida pela VGG-16 pode ser considerada aceitável. Os valores de 0,82 e 0,84 obtidos pela AlexNet e pelo Comitê de RNCs respectivamente, refletem uma maior confiabilidade e eficácia na identificação de melanomas.

Na Tabela 10, é possível verificar a média e o desvio padrão das métricas dos estimadores do comitê, que nos indicam que as predições do comitê são consistentes, generalizando bem em diferentes amostras de dados e têm pouca variabilidade entre os estimadores que o compõem.

Métrica	Média	Desvio Padrão
Precisão	74.55%	0.010640
Revocação	85.17%	0.020554
Pontuação F1	79.48%	0.006148

Tabela 10 – Média e Desvio Padrão obtidos dos resultados dos estimadores do comitê.
Fonte: Autor (2024)

O comitê de classificadores composto por estimadores criados a partir da *Skin-LesionClassifier*, apresentou o melhor resultado em todas as métricas, sugerindo que os estimadores conseguiram captar bem as características relevantes das imagens de melanoma e não melanoma. Um comitê costuma ser mais robusto e tende a generalizar melhor, já que combina a diversidade das predições de múltiplos estimadores, mitigando o impacto de modelos individuais que poderiam estar super ou subajustados. A Busca Aleatória permitiu encontrar configurações de hiperparâmetros mais adequadas para a tarefa específica (Tabelas 11, 12 e 13). O treinamento do zero, combinado com a Busca Aleatória, proporcionaram uma melhor adaptação da rede à tarefa de classificação de lesões de pele em termos de desempenho e generalização.

Hiperparâmetro	Distribuição de hiperparâmetros
Mapa de características (<i>feature map</i>) da Camada convolucional 1	128
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 1	3
Mapa de características (<i>feature map</i>) da Camada convolucional 2	256
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 2	5
Mapa de características (<i>feature map</i>) da Camada convolucional 3	64
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 3	5
Mapa de características (<i>feature map</i>) da Camada convolucional 4	64
Tamanho do filtro (<i>kernel size</i>) da Camada convolucional 4	5

Tabela 11 – Hiperparâmetros selecionados para as Camadas convolucionais da *SkinLesionClassifier*.

Fonte: Autor (2024)

Hiperparâmetro	Distribuição de hiperparâmetros
Neurônios da Camada totalmente conectada 1	128
Neurônios da Camada totalmente conectada 2	512
Neurônios da Camada totalmente conectada 3	512
Taxa de abandono (<i>Dropout</i>)	0.3

Tabela 12 – Hiperparâmetros selecionados para as Camadas totalmente conectadas e regularização da *SkinLesionClassifier*.

Fonte: Autor (2024)

Hiperparâmetro	Distribuição de hiperparâmetros
Número de épocas	50
Tamanho do lote (<i>batch size</i>)	64
Taxa de aprendizado (<i>learning rate</i>)	0.0001

Tabela 13 – Hiperparâmetros selecionados para o treinamento da *SkinLesionClassifier*.

Fonte: Autor (2024)

O *Transfer Learning* e *Fine-tuning* foram aplicados nas arquiteturas AlexNet e VGG-16 pré-treinadas no *ImageNet*. As camadas internas foram congeladas, e o *fine-tuning* foi realizado apenas na camada de saída, sem otimização adicional dos hiperpa-

râmetros, como os das redes do comitê, tendo apenas as suas camadas de saída ajustadas para classificar as classes não melanoma e melanoma. Apesar da AlexNet ter apresentado um desempenho melhor que a VGG-16, ambos ficaram atrás do comitê. Mesmo sem otimização específica dos hiperparâmetros e com apenas a camada de saída ajustada, os modelos ainda conseguem resultados próximos ao do comitê, mostrando a força do *Transfer Learning*. O uso de redes pré-treinadas economiza tempo e recursos computacionais, pois aproveita o conhecimento previamente adquirido em um grande conjunto de dados como o *ImageNet*. As redes AlexNet e VGG-16 foram projetadas e pré-treinadas em um conjunto de dados diferente e o simples ajuste da camada de saída pode não ser suficiente para capturar as características específicas de lesões de pele. O congelamento das camadas internas limita a capacidade dessas redes de se adaptarem totalmente à nova tarefa, afetando as redes ao ponto no qual elas não conseguem captar detalhes finos importantes para a classificação de melanoma.

6 Conclusões

O melanoma é o tipo mais grave de câncer de pele, e embora seja menos comum do que outros tipos de câncer de pele, é mais agressivo e tem maior probabilidade de se espalhar para outros órgãos se não for detectado e tratado precocemente. É extremamente importante que sua detecção seja feita em seus estágios iniciais, pois, dessa forma, as chances de cura do paciente aumentam significativamente. Sendo assim, este trabalho tem como objetivo utilizar técnicas de Aprendizado de Máquina para propor um modelo capaz de classificar uma lesão de pele como melanoma ou não melanoma, que seja boa suficiente para ser utilizada como apoio aos médicos especialistas.

Foram utilizadas duas abordagens distintas para este trabalho, a primeira, um comitê de classificadores, onde o aprendiz base foi um modelo de Redes Neurais Convolucionais criado e treinado do zero, e a segunda abordagem, foi o uso de técnicas de *Transfer Learning* com as arquiteturas AlexNet e VGG-16, que são arquiteturas de Redes Neurais Convolucionais amplamente reconhecidas pelo seu desempenho e eficiência. O modelo proposto, o comitê de Redes Neurais Convolucionais foi comparado com as duas arquiteturas mencionadas, através da comparação das métricas de avaliação de desempenho Precisão, Revocação e Pontuação F1. Embora os resultados gerais tenham sido semelhantes, o comitê se mostrou superior, apresentando resultados melhores para cada uma das métricas de avaliação, com precisão de 75,69%, revocação de 85,78% e pontuação F1 de 80,42%. Apesar de obter um resultado consideravelmente satisfatório, devido a importância da tarefa de classificação de melanomas, o comitê ainda não atingiu um nível ideal para seu propósito e pode ser considerado insuficiente para o uso clínico. Porém, o modelo se mostrou promissor, já que mesmo se tratando de um projeto em fase inicial, foi capaz de obter um resultado melhor que os demais modelos baseados em *Transfer Learning*.

Treinar múltiplos estimadores do zero exige tempo e um maior poder de processamento, especialmente no ajuste dos hiperparâmetros, no qual foi utilizada a Busca Aleatória, uma técnica com um custo elevado. Ao longo do processo de otimização dos hiperparâmetros e do treinamento, o modelo se mostrou computacionalmente caro, se fazendo necessário a realização de uma subsamostragem da base de dados e a redução nos intervalos dos hiperparâmetros utilizados para compor a Busca Aleatória. No entanto, o uso do *Transfer Learning* se mostrou mais eficiente em termos de tempo e recursos, e seu desempenho pode ser melhorado com ajustes mais completos e otimizações. Mesmo sem otimização específica dos hiperparâmetros e com apenas a camada de saída ajustada, a AlexNet e a VGG-16 ainda conseguiram resultados próxi-

mos do modelo proposto, que foi desde o início otimizado para a tarefa de classificação de melanomas, mostrando a força do *Transfer Learning* e o motivo de ser uma técnica tão adotada.

Quando comparado aos trabalhos relacionados que fizeram uso de uma das bases do ISIC *Challenge* (original ou derivada), o comitê de Redes Neurais proposto se destaca pela sua Revocação e Pontuação F1, indicando sua forte capacidade de identificar corretamente os casos positivos (melanomas), essencial em diagnósticos médicos sensíveis. Em seu trabalho, embora tenham utilizado uma base de dados menor, [SANTOS et al. \(2017\)](#) apresentaram um modelo eficaz tanto na identificação de casos positivos quanto na minimização de falsos positivos, obtendo uma alta Revocação e Precisão de 91%. Já [PESSANHA e CAMPOS \(2021\)](#) utilizaram a métrica ROC AUC, que mede a capacidade geral de distinção entre as classes, onde obtiveram um valor de 86%, mas a Revocação de 75% sugere uma menor capacidade de identificar melanomas. O trabalho de [OLIVEIRA e LEITE \(2021\)](#) apresentou uma Revocação de 75,2% e uma ROC AUC de 83,9%, indicando um bom equilíbrio na discriminação de classes, mas desempenho inferior em comparação a Revocação do comitê proposto. A métrica PR AUC utilizada no comitê sugere uma eficácia superior em cenários onde há desbalanceamento de classes, sendo mais robusto na detecção da classe positiva. A ausência de métricas como a Precisão em alguns dos trabalhos dificulta a análise completa sobre falsos positivos, limitando comparações mais abrangentes de desempenho.

Novos ajustes de hiperparâmetros, aumentar o número de iterações na Busca Aleatória e técnicas de regularização como o *Batch Normalization* (Normalização em lote) para acelerar a convergência e melhorar a estabilidade do treinamento, seriam refinamentos e atualizações necessárias para se alcançar uma melhora no desempenho do comitê. Como a base inicial do ISIC *Challenge* 2019 é desbalanceada, a técnica de *Data Augmentation*, que foi aplicada apenas com transformações para aumentar a diversidade dos dados e diminuir a sensibilidade do modelo as imagens, poderia ser usada também para aumentar digitalmente o volume de dados das classes menores, sendo considerada uma forma de mitigar o desbalanceamento da base, evitando a necessidade de realizar uma subamostragem dos dados. Tendo em vista que as limitações de *hardware* também tiveram um impacto grande no desempenho do modelo, mais recursos computacionais e técnicas de otimização e performance seriam necessárias para que o modelo possa obter um desempenho capaz de atingir o propósito do uso clínico. Mesmo com todas as limitações e desafios encontrados, o comitê de Redes Neurais Convolucionais ainda conseguiu captar bem as características relevantes do melanoma.

Utilizar a base de dados do ISIC *Challenge* 2019 completa, além de expandir

a tarefa de uma classificação binária, onde o objetivo inicial era identificar apenas melanomas e não melanomas, para a identificação de todas as 8 classes da base inicial, são possibilidades a serem exploradas para trabalhos futuros.

Referências

ACS. *Tests for Melanoma Skin Cancer*. 2019. Disponível em: <<https://www.cancer.org/cancer/melanoma-skin-cancer/detection-diagnosis-staging/how-diagnosed.html>>. Acesso em: 10/04/2022. Citado na página 16.

ACS. *What Is Melanoma Skin Cancer?* 2019. Disponível em: <<https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html>>. Acesso em: 10/04/2022. Citado na página 15.

ANWAR, S. M. et al. Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, n. 42, 2018. Citado na página 18.

BERGSTRA, J.; BENGIO, Y. Yoshua. random search for hyperparameter optimization. *Journal of machine learning research*, 2012. Citado 2 vezes nas páginas 40 e 58.

BEZERRA, E. Introdução à aprendizagem profunda. In: OGASAWARA, E.; VIEIRA, V. (Org.). *Tópicos em Gerenciamento de Dados e Informações*. 1. ed. Porto Alegre: SBC, 2016. v. 1, p. 57–86. Citado 4 vezes nas páginas 25, 27, 28 e 32.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. 1. ed. New York: Springer, 2006. ISBN ISBN 978-0-387-31073-2. Citado 5 vezes nas páginas 28, 31, 35, 36 e 39.

BOCHIE, K. et al. Aprendizado profundo em redes desafiadoras: Conceitos e aplicações. In: GOMES, D. G.; MORAES, I. M.; CAMPISTA, M. E. M. (Org.). *Minicursos do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Porto Alegre: SBC, 2020. p. 140–189. Citado 2 vezes nas páginas 27 e 30.

CHAPPUIS, P. et al. Dermoscopy, a useful tool for general practitioners in melanoma screening: a nationwide survey. *Br J Dermatol*, n. 175(4), p. 744–750, 2016. Citado na página 17.

CLARO, M. e. a. Utilização de técnicas de data augmentation em imagens: Teoria e prática. In: TELES, A. et al. *Minicursos da ERCEMAPI 2020*. 8. ed. SBC, 2020. Disponível em: <<https://books-sol.sbc.org.br/index.php/sbc/catalog/view/48/217/455>>. Citado na página 57.

CORONA, R. et al. Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. *J Clin Oncol*, n. 14(4), p. 1218–1223, 1996. Citado na página 17.

CUI, X. et al. Assessing the effectiveness of artificial intelligence methods for melanoma: A retrospective review. *J Am Acad Dermatol*, n. 81(5), p. 1176–1180, 2019. Citado na página 18.

Data Science Academy. *Deep Learning Book*. 2021. Disponível em: <<https://www.deeplearningbook.com.br>>. Acesso em: 22/11/2023. Citado 5 vezes nas páginas 23, 24, 25, 27 e 34.

- DELAZERI, A. V.; STEVANI, E. S. *Classificação de câncer de pele usando redes neurais convolucionais*: uma análise do desempenho de classificação em um conjunto de dados desbalanceado. Curitiba: [s.n.], 2020. Citado na página 46.
- DENG, J. et al. *ImageNet: A Large-Scale Hierarchical Image Database*. 2009. Citado na página 32.
- ELLIOTT, T. M. et al. Estimated healthcare costs of melanoma in australia over 3 years post-diagnosis. *Appl Health Econ Health Policy*, n. 15(6), p. 115–118, 2017. Citado na página 17.
- ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, n. 542(7639), p. 115–118, 2017. Citado na página 18.
- FARMER, E. R.; GONIN, R.; HANNA, M. P. Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists. *Human Pathology*, v. 27, n. 6, p. 528–531, 1996. ISSN 0046-8177. Citado na página 17.
- FURTADO, M. I. V. *Redes Neurais Artificiais*: Uma abordagem para sala de aula. [S.l.]: Atena Editora, 2019. Citado 3 vezes nas páginas 23, 24 e 25.
- GIAVINA-BIANCHI, M.; CORDIOLI, E.; MACHADO, B. S. Melanoma: implicações da falha diagnóstica e perspectivas. *Einstein*, São Paulo, n. 19, 2021. Citado na página 18.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. 1. ed. [S.l.]: Blucher, 2008. 4–5 p. Citado na página 20.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. Disponível em: <<https://www.deeplearningbook.org>>. Acesso em: 15/12/2023. Citado 6 vezes nas páginas 22, 30, 31, 32, 39 e 68.
- Google Colab. *Google Colab*. 2023. Disponível em: <<https://research.google.com/colaboratory>>. Acesso em: 22/11/2023. Citado na página 56.
- GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*: Concepts, tools, and techniques to build intelligent systems. [S.l.]: O'Reilly Media, Inc, 2022. ISBN 978-1098125974. Citado na página 39.
- HADIMI, R.; PENG, D.; COCKBURN, M. Efficacy of skin self-examination for the early detection of melanoma. *Int J Dermatol*, n. 49(2), p. 126—134, 2010. Citado na página 17.
- HAENSSLE, H. A. et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*, n. 29(8), p. 1836–1842, 2018. Citado na página 18.
- HOFF, P. M. G. *Tratado de oncologia*. 1. ed. São Paulo: Atheneu, 2013. 2829 p. Citado na página 16.
- INCA. *Câncer de pele*: saiba como prevenir, diagnosticar e tratar. 2020. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/2020/dezembro/cancer-de-pele-saiba-como-prevenir-diagnosticar-e-tratar>>. Acesso em: 16/09/2024. Citado na página 15.

- INCA. *Câncer de pele melanoma*. 2022. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/pele-melanoma>>. Acesso em: 16/09/2024. Citado na página 15.
- INCA. *Câncer de pele não melanoma*. 2022. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/pele-nao-melanoma>>. Acesso em: 16/09/2024. Citado na página 15.
- ISDIS. *ISDIS*. 2019. Disponível em: <<https://isdis.org/>>. Acesso em: 11/06/2022. Citado na página 53.
- ISIC. *International Skin Imaging Collaboration*. 2019. Disponível em: <<https://challenge.isic-archive.com>>. Acesso em: 11/06/2022. Citado na página 53.
- ISIC. *ISIC Challenge*. 2019. Disponível em: <<https://challenge.isic-archive.com>>. Acesso em: 12/06/2022. Citado 5 vezes nas páginas 17, 53, 54, 55 e 58.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012. Citado 2 vezes nas páginas 32 e 51.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 28.
- MARQUES FILHO, O.; VIEIRA NETO, H. *Processamento Digital de Imagens*. 1. ed. Rio de Janeiro: Brasport, 1999. 21 p. ISBN 8574520098. Citado 2 vezes nas páginas 20 e 21.
- NG, A. *Machine Learning Yearning - Technical Strategy for AI Engineers, In the Era of Deep Learning*. [S.l.]: DeepLearning.AI Project, 2018. Citado 5 vezes nas páginas 23, 35, 42, 43 e 61.
- NIELSEN, M. A. *Neural Networks and Deep Learning*. Determination Press, 2015. Disponível em: <<http://neuralnetworksanddeeplearning.com>>. Acesso em: 13/12/2023. Citado 7 vezes nas páginas 24, 26, 27, 29, 30, 36 e 47.
- NumPy. *NumPy*. 2024. Disponível em: <<https://numpy.org>>. Acesso em: 19/03/2024. Citado na página 56.
- OLIVEIRA, M. G. C.; LEITE, W. M. D. *Comparação de redes neurais convolucionais para o diagnóstico de melanoma por meio de transfer learning e data augmentation em imagens dermatoscópicas com classes desbalanceadas*. Brasília: [s.n.], 2021. Citado 4 vezes nas páginas 40, 45, 54 e 72.
- PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, n. 22(10), p. 1345–1359, 2010. ISSN 1558-2191. Citado na página 60.
- Pandas. *Pandas*. 2023. Disponível em: <<https://pandas.pydata.org>>. Acesso em: 17/11/2023. Citado na página 56.
- PESSANHA, G. R. G.; CAMPOS, E. Um modelo de inteligência artificial para detecção de melanoma via redes neurais convolucionais. In: *ESCOLA REGIONAL DE COMPUTAÇÃO APLICADA À SAÚDE (ERCAS)*. 8. ed. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 46–49. Citado 3 vezes nas páginas 44, 54 e 72.

- Pillow. *Pillow*. 2024. Disponível em: <<https://python-pillow.org>>. Acesso em: 02/03/2024. Citado na página 56.
- PONTI, M. A.; COSTA, G. B. P. Como funciona o deep learning. In: VIEIRA, V.; RAZENTE, H. L.; BARIONI, M. C. N. (Org.). *Tópicos em Gerenciamento de Dados e Informações*. 1. ed. Uberlândia: SBC, 2017. p. 63–93. Citado 2 vezes nas páginas 32 e 35.
- Python. *Python*. 2024. Disponível em: <<https://www.python.org>>. Acesso em: 17/04/2024. Citado na página 56.
- PyTorch. *PyTorch*. 2023. Disponível em: <<https://pytorch.org>>. Acesso em: 22/11/2023. Citado 2 vezes nas páginas 51 e 56.
- SANTOS, A. et al. Uma abordagem de classificação de imagens dermatoscópicas utilizando aprendizado profundo com redes neurais convolucionais. In: *SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE (SBCAS)*. 17. ed. Porto Alegre: Sociedade Brasileira de Computação, 2017. Citado 3 vezes nas páginas 44, 54 e 72.
- Scikit-Learn. *Scikit-Learn*. 2023. Disponível em: <<https://scikitlearn.org>>. Acesso em: 21/11/2023. Citado 3 vezes nas páginas 55, 56 e 60.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks For Large-Scale Image Recognition*. 2014. Disponível em: <<https://arxiv.org/abs/1409.1556v1>>. Citado 2 vezes nas páginas 33 e 51.
- Skorch. *Skorch*. 2024. Disponível em: <<https://skorch.readthedocs.io>>. Acesso em: 17/04/2024. Citado na página 60.
- SWEIGART, A. *Automatize Tarefas Maçantes com Python*. 1. ed. [S.l.]: Novatec, 2015. 483–485 p. ISBN 978-85-7522-446-5. Citado 2 vezes nas páginas 20 e 21.
- TSCHANDL, P. et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol*, n. 20(7), p. 938–947, 2019. Citado na página 18.
- UTSCH, K. G. *Uso de Redes Neurais Convolucionais para classificação de imagens digitais de lesões de pele*. Espírito Santo: [s.n.], 2018. Citado na página 45.
- WANDERLEY, M. D. S. Transferindo conhecimento de textos para imagens através da aprendizagem das características semânticas. *Programa de Pós-Graduação em Ciência da Computação, CIN-UFPE*, 2018. Citado na página 37.
- WELZEL, J.; SCHUH, S. Noninvasive diagnosis in dermatology. *JDDG*, n. 15(10), p. 999–1016, Outubro 2017. Citado na página 16.
- YAQING, W. et al. *Generalizing from a Few Examples: A survey on few-shot learning*. New York, NY, USA: Association for Computing Machinery, 2020. ISSN 0360-0300. Citado 2 vezes nas páginas 37 e 60.
- ZHANG, A. et al. *Dive Into Deep Learning*. Cambridge University Press, 2023. Disponível em: <<https://d2l.ai>>. Acesso em: 22/12/2023. Citado 8 vezes nas páginas 22, 32, 33, 34, 37, 38, 40 e 41.

APÊNDICE A - Código Fonte Disponível no GitHub

O código fonte utilizado neste trabalho está disponível no repositório do GitHub. Neste repositório, você encontrará todas as implementações, dados utilizados e instruções para a reprodução dos experimentos.

Link para o repositório: <https://github.com/jmontejr/tecnicas-preditivas-para-auxilio-no-diagnostico-de-melanomas-via-images>