

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

WAGNER MORENO SCHMITZ

**COMPARATIVO DE MODELOS DE REDES NEURAIAS
CONVOLUCIONAIS NO DIAGNOSTICO E CLASSIFICAÇÃO DE
LESÕES DE CÂNCER DE PELE ATRAVÊS DE IMAGENS
DERMATOSCOPICAS**

DOIS VIZINHOS

2023

WAGNER MORENO SCHMITZ

**COMPARATIVO DE MODELOS DE REDES NEURAI
CONVOLUCIONAIS NO DIAGNOSTICO E CLASSIFICAÇÃO DE
LESÕES DE CÂNCER DE PELE ATRAVÊS DE IMAGENS
DERMATOSCOPICAS**

**COMPARISON OF CONVOLUTIONAL NEURAL NETWORK MODELS
IN THE DIAGNOSIS AND CLASSIFICATION OF SKIN CANCER
LESIONS USING DERMATOSCOPIC IMAGES**

Trabalho de Conclusão de Curso de Especialização apresentado como requisito para obtenção do título de Especialista em Ciência de Dados do Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Jefferson Tales Oliva

DOIS VIZINHOS

2023



[4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

WAGNER MORENO SCHMITZ

**COMPARATIVO DE MODELOS DE REDES NEURAIS
CONVOLUCIONAIS NO DIAGNOSTICO E CLASSIFICAÇÃO DE
LESÕES DE CÂNCER DE PELE ATRAVÉS DE IMAGENS
DERMATOSCOPICAS**

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de aprovação: 21 de dezembro de 2023

Jefferson Tales Oliva
Doutorado
Universidade Tecnológica Federal do Paraná

Marcelo Teixeira
Doutorado
Universidade Tecnológica Federal do Paraná

Dalcimar Casanova
Doutorado
Universidade Tecnológica Federal do Paraná

DOIS VIZINHOS

2023

AGRADECIMENTOS

Agradeço imensamente a meus pais e irmão pelo inestimável suporte na educação e apoio ao longo de toda a jornada. Expresso minha gratidão aos amigos e colegas da especialização que compartilharam comigo cada passo nesse processo. Um agradecimento especial aos professores e à instituição que desempenharam papéis cruciais em meu aprendizado. Obrigado por tudo.

RESUMO

O propósito fundamental deste trabalho é desenvolver e comparar o resultado de técnicas de CNN na classificação e reconhecimento do melanoma, um tipo de câncer de pele, através de imagens dermatoscópicas. O câncer de pele é uma condição de saúde grave, onde o diagnóstico precoce é crucial para o sucesso do tratamento. Nas técnicas tradicionais o diagnóstico se baseia na observação visual de médicos, o que pode apresentar limitações. Para superar essas limitações, o uso de técnicas de aprendizado profundo com redes neurais convolucionais para criar um modelo computacional eficiente na classificação de forma automatizada de melanoma em imagens dermatoscópicas é uma alternativa. A metodologia adotada consiste na coleta de uma base de dados (HAM10000) e na aplicação de técnicas de pré-processamento e aumento de dados para otimizar a quantidade e qualidade dos dados disponíveis em cada uma das classes no treinamento do modelo. A implementação é feita utilizando cinco modelos de CNNs com ajustes específicos para a classificação de imagens dermatoscópicas. A avaliação dos modelos ocorre por meio de métricas de desempenho como acurácia, matriz de confusão e com testes de Friedman e Nemenyi a fim de garantir a eficiência e comparar a viabilidade dos modelos desenvolvidos. Como resultado foram encontradas diferenças significativas entre os modelos comparados, com destaque para os modelos MobileNet e InceptionV3 que tiveram diferenças estatisticamente significativas com todos os modelos.

Palavras-chave: aprendizado de máquina; câncer de pele; cnn; aprendizado profundo; melanoma.

ABSTRACT

The fundamental purpose of this work is to develop and compare the outcomes of CNN techniques in the classification and recognition of melanoma, a type of skin cancer, through dermoscopic images. Skin cancer is a serious health condition where early diagnosis is crucial for successful treatment. In traditional techniques, the diagnosis relies on visual observation by medical professionals, which may have limitations. To overcome these limitations, the use of deep learning techniques with convolutional neural networks to create an efficient computational model for automated classification of melanoma in dermoscopic images is an alternative. The adopted methodology involves collecting a dataset (HAM10000) and applying data preprocessing and augmentation techniques to optimize the quantity and quality of data available for each class in the model training. The implementation is carried out using five CNN models with specific adjustments for the classification of dermoscopic images. Model evaluation occurs through performance metrics such as accuracy, confusion matrix, and Friedman and Nemenyi tests to ensure efficiency and compare the feasibility of the developed models. As a result, significant differences were found between the compared models, with emphasis on the MobileNet and InceptionV3 models, which showed statistically significant differences with all other models.

Keywords: machine learning; skin cancer; cnn; deep learning; melanoma.

LISTA DE FIGURAS

Figura 1 – Operação Convolutacional.	16
Figura 2 – Análise da base de dados.	21
Figura 3 – Matriz de Confusão	23
Figura 4 – Treinamento e Validação	24
Figura 5 – Acurácia de Treino e Teste	25
Figura 6 – Fração de Previsões Incorretas	26
Figura 7 – Erro e Taxa de Acerto Durante o Treinamento	27

LISTA DE TABELAS

Tabela 1 – Teste de Nemenyi 29

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Problema de Pesquisa	11
1.2	Objetivos	11
1.2.1	Objetivo Geral	11
1.2.2	Objetivos Específico	12
1.3	Organização do Trabalho	12
2	REVISÃO DE LITERATURA	14
2.1	Câncer de Pele	14
2.1.1	Carcinoma Basocelular	14
2.1.2	Melanoma	14
2.2	Diagnóstico	14
2.3	Dermatoscopia	15
2.4	Redes Neurais Convolucionais	15
2.5	Arquitetura das Redes Neurais Convolucionais	16
2.5.1	Resnet50	16
2.5.2	MobileNet	17
2.5.3	VGG16 e VGG19	17
2.5.4	InceptionV3	17
2.6	Testes Não-Paramétricos	17
2.6.1	Teste de Friedman	18
2.6.2	Pós-teste de Nemenyi	18
3	MATERIAIS E MÉTODOS	19
3.1	Ferramentas	19
3.2	Pré-Processamento de Dados	20
3.3	Aplicação dos Modelos de Redes Neurais Convolucionais	20
3.4	Análise dos Resultados	20
4	RESULTADOS	21
4.1	Tratamento de Dados	21
4.2	Treinamento de Modelos	22

4.3	Análise dos Resultados	23
4.3.1	Matriz de Confusão	23
4.3.2	Perda de Treinamento e Validação	24
4.3.3	Acurácia de Treino e Teste	25
4.3.4	Fração de Previsões Incorretas	26
4.3.5	Erro e Taxa de Acerto Durante o Treinamento	26
4.4	Testes Não-Paramétricos	27
4.4.1	Teste de Friedman	28
4.4.2	Pós-teste de Nemenyi	28
5	CONCLUSÃO	30
	REFERÊNCIAS	31

1 INTRODUÇÃO

A utilização do aprendizado de máquina na área da saúde tem se tornado amplamente difundido como uma das principais aplicações à classificação de imagens. Essa tarefa é realizada por meio do uso do aprendizado profundo (*deep learning*), uma subárea do aprendizado e máquina (*machine learning*) que utiliza de redes neurais convolucionais (CNN, do inglês *Convolutional Neural Networks*) para realizar tarefas complexas de aprendizado e classificação, permitindo a identificação de características presentes nos dados e facilitando a análise de imagens (RGUIBI; HAJAMI; DYA, 2021).

O processo de aprendizado e classificação do *Deep Learning* é feito através de camadas de neurônios artificiais que emulam o funcionamento de neurônios biológicos, formando assim redes neurais que realizam o processamento de dados, que podem ser documentos, imagens, áudios, vídeos entre outros (GOODFELLOW; BENGIO; COURVILLE, 2016).

O *Deep Learning* tem demonstrado ser cada vez mais utilizado na área da saúde, podendo-se dizer que faz parte da saúde eletrônica, que é o uso de tecnologias da comunicação e informação voltadas para a saúde (World Health Organization, 2023). Além disso, ele contribui para a identificação de lesões e auxiliar médicos no diagnóstico de diversos problemas. Uma das aplicações destacadas para a identificação de padrões em exames médicos é o uso de CNNs para a análise de imagens.

As CNN são uma classe de algoritmos de aprendizado profundo projetadas para o processamento e análise de imagens de forma eficiente. Elas são inspiradas na organização visual do córtex cerebral, permitindo a extração de características específicas das imagens e com isso a detecção de padrões complexos, como os encontrados nas imagens de exames dermatoscópicos (FACELI; AL., 2021).

A dermatoscopia É uma técnica usada por médicos para examinar a pele de forma detalhada e ampliada. Esse exame permite a visualização de características não visíveis a olho nu, como estruturas pigmentadas, vasos sanguíneos e padrões específicos (A.C. Camargo Cancer Center, 2023). A dermatoscopia desempenha um papel fundamental na detecção precoce e no diagnóstico de diferentes condições, principalmente no contexto do câncer de pele.

Segundo a sociedade brasileira de dermatologia, o câncer de pele corresponde a 33% de todos os diagnósticos desta doença no Brasil, sendo registrado a cada ano pelo Instituto Nacional de Câncer (INCA) cerca de 185 mil novos casos. Nos estágios mais avançados, o câncer de pele do tipo melanoma forma lesões mais profundas e espessas, o que aumenta a chance de espalhar pelos órgãos conhecido como metástase e conseqüentemente diminui a chance de cura (Sociedade Brasileira de Dermatologia, 2023).

Dessa forma a combinação de técnicas de aprendizado profundo atreladas a modelos de arquiteturas de CNN na análise de imagens dermatoscópicas, junto da validação de médico especialistas, pode ser ferramenta promissora para o diagnóstico rápido e eficiente de doenças

de pele, possibilitando início do tratamento no tempo adequado. Nesse contexto, é notado um crescente interesse na pesquisa científica envolvendo técnicas de CNNs e saúde (COSTA, 2017).

1.1 Problema de Pesquisa

A análise de exames dermatoscópicos em muitos casos é feita por médicos especialistas que averíguam os padrões de pigmentação, tamanho e formato dos sinais e manchas da pele. Esse processo pode ser custoso e passível de erros devido a pouca experiência do examinador e seu treinamento sobre os critérios dermatoscópicos, cujo bom emprego também é fundamental (REZZE; SÁ; NEVES, 2006).

Segundo afirma Frange, Arruda e Daldon (2009), o exame dermatoscópico envolve mais de cem estruturas que podem ser identificadas. Dessa forma o uso da rede neural convolucional pode ser benéfico para auxiliar o médico especialista na análise rápida das informações, não descartando a importância do profissional para a validação desse processo.

Diversos modelos de arquitetura de CNN podem ser utilizados para auxiliar nessa análise. Espera-se que o comparativo desses modelos tragam soluções rápidas e eficientes na análise e classificação de exames dermatoscópicos

A rede neural convolucional é um tipo de arquitetura de aprendizado profundo especializada em processamento de dados que incluem arquivos de imagem. Essa arquitetura é vista com bastante sucesso em aplicações práticas, afirma Goodfellow, Bengio e Courville (2016)

Dessa forma, o uso de uma arquitetura robusta e reconhecida em meio a aplicações práticas pode ser benéfica com a apresentação de resultados positivos junto a identificação de padrões em imagens, otimizando o trabalho dos profissionais e reduzindo a dependência de classificação manual que ocorre na maioria dos casos em exames dermatoscópicos, possibilitando um diagnóstico através de uma abordagem simples e acessível.

O comparativo de arquiteturas diferentes contribui para a análise e implementação do modelo trazendo maior efetividade na compreensão das diferenças de desempenho existentes entre as arquiteturas.

1.2 Objetivos

1.2.1 Objetivo Geral

Desenvolver abordagens de análise de imagens dermatoscópicas utilizando modelos de arquiteturas de CNN com a finalidade de contribuir para o diagnóstico de doenças de pele de forma precisa e eficiente.

1.2.2 Objetivos Específico

Para alcançar os resultados do objetivo geral os seguintes objetivos específicos serão buscados:

- Realizar o pré-processamento das imagens presentes na base de dados utilizada para aprimorar a qualidade e e eficácia na extração de características.
- Implementar o algoritmo de classificação de imagens dermatoscópicas utilizando a base de dados HAM10000 para treinamento e teste do algoritmo.
- Utilizar modelos de arquiteturas diferentes e avaliar a eficácia dos resultados obtidos através dos modelos de classificação desenvolvido utilizando para essa análise dados estatísticos.
- Comparar e analisar os modelos de arquiteturas desenvolvidos através de testes estatísticos de hipótese.

1.3 Organização do Trabalho

Introdução: Tem como objetivo estabelecer o contexto, delimitar o tema e apresentar os objetivos de pesquisas que serão desenvolvidos no trabalho. Além disso, é destacada a importância do tema escolhido e detalhado os materiais e métodos que serão empregados no trabalho.

Fundamentação Teórica: São apresentados os conceitos teóricos relacionados a dermatoscopia, melanoma e o aumento do número de casos de câncer de pele. Abordando também os conceitos de aprendizagem de máquina, aprendizado profundo, e CNN para a análise de imagens.

Bibliografia: Nesse capítulo os trabalhos consultados relacionados ao tema abordado no trabalho são apresentados. Indicando estudos na mesma área ou áreas similares como também informações de livros e especialistas tanto da área de dermatologia como também da área de aprendizado de máquina e aprendizado profundo.

Metodologia: O processo utilizado para alcançar o resultado do trabalho será descrito nesse capítulo de forma detalhada. Informações como as amostras utilizadas e origem das imagens dermatoscópicas da pesquisa e como é feito o pré-processamento e extração de características para a implementação da rede neural convolucional e o aumento de dados estarão presentes na metodologia.

Resultados e Discussão: Os resultados alcançados com o modelo de rede neural convolucional serão apresentados nesse capítulo. Após a apresentação análises estatísticas de acurácia e a interpretação dos resultados fornecerá informações referentes as possíveis limitações e melhorias futuras.

Conclusão: É feita um resumo das principais descobertas e resultados alcançados através da pesquisa realizada ao longo do trabalho, passando por uma recapitulação dos objetivos seguindo pela apresentação e contextualização dos resultados junto de suas possíveis melhorias futuras, limitações encontradas durante o desenvolvimento e possíveis direções futuras.

Referência Bibliográficas: São listados todos os trabalhos consultados como referências diretas e indiretas durante todo o desenvolvimento do trabalho.

Anexos: São apresentados os possíveis dados complementares que foram utilizados ao longo do trabalho, podendo ser tabelas, imagens, gráficos, entre outros.

2 REVISÃO DE LITERATURA

2.1 Câncer de Pele

O Câncer de pele É uma neoplasia que se desenvolve na pele e É caracterizada pelo crescimento anormal e de forma descontrolada de células da pele (Ministério da Saúde, 2023). Sua aparência se assimila muito com pintas e manchas, tendo como os principais tipos o carcinoma basocelular e o melanoma.

2.1.1 Carcinoma Basocelular

O Carcinoma basocelular é o tipo mais comum de câncer de pele menos agressivo, considerado cerca de 80% dos casos. Ele é conhecido por ter um crescimento lento e baixas chances de gerar metástase, que são células cancerígenas que se soltam do tumor original e se espalham para outras partes do corpo, informa Hospital Sírio-Libanês (2023a). Devido ao crescimento lento e as baixas chances de metástase, é visto como um tipo de tumor com altos percentuais de cura se diagnosticado de forma precoce.

2.1.2 Melanoma

O câncer de pele do tipo melanoma é considerado um dos tumores mais graves devido ao seu potencial de gerar metástases, podendo atingir órgãos e gânglios diversos. Esse tipo de tumor ocorre com mais frequência em áreas do corpo com exposição ao sol, mas também em unhas, solas dos pés, palmas das mãos, mucosas e olhos. Devido ao alto risco, o diagnóstico precoce é essencial para um tratamento eficaz (Hospital Sírio-Libanês, 2023b).

2.2 Diagnóstico

O número de casos de câncer de pele tem crescido mundialmente, sendo o câncer mais frequente no Brasil e no mundo (Ministério da Saúde, 2023). Por se tratar de uma neoplasia com um nível de agressividade bastante elevado e de difícil tratamento em estágio mais avançado, é fundamental um diagnóstico precoce para a cura.

A dermatoscopia é um exame que surgiu com o papel fundamental para a realização do diagnóstico precoce e ampliar a acurácia dos diagnósticos de melanoma (REZZE; SÁ; NEVES, 2006).

2.3 Dermatoscopia

Conforme indica (VAZ, 2023), a detecção precoce de câncer aumenta as chances de sucesso do tratamento, com isso o caminho ideal é visto na busca por um diagnóstico em estágio inicial. Além disso (VAZ, 2023) afirma que, na última década, a aplicação do aprendizado.

A dermatoscopia é um método de diagnóstico não invasivo que auxilia na avaliação de lesões de manchas e sinais na pele através do uso do dermatoscópio, um instrumento que visualiza as lesões da pele e aumenta a precisão no diagnóstico de suspeitas de câncer de pele tanto do tipo carcinoma quanto melanoma, explica (CASAGRANDE, 2023).

Através da dermatoscopia digital, utiliza-se um aparelho com software que permite armazenar imagens de diferentes pintas ou sinais para acompanhamento e comparação através do tempo. Essas imagens podem ser usadas junto de tecnologias de inteligência artificial para o auxílio em um diagnóstico preciso e precoce, indica (CASAGRANDE, 2023).

Conforme indica a (Sociedade Brasileira de Dermatologia, 2023), o diagnóstico convencional é feito através da análise do exame de dermatoscopia realizado por um médico, trabalho esse que pode exigir uma atenção maior do especialista. Procedimentos mais modernos utilizam da digitalização de imagens para facilitar a comparação entre elas em um curto intervalo de tempo. Processo esse que se combinado com o uso das redes neurais convolucionais pode agilizar o processo de comparação e a assertividade dos resultados.

2.4 Redes Neurais Convolucionais

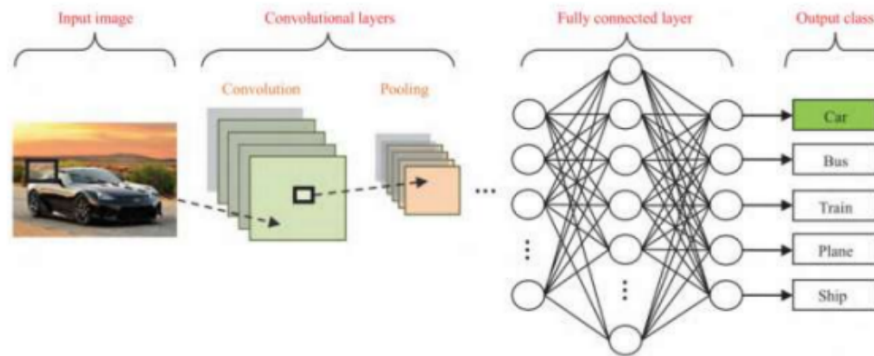
Redes neurais artificiais são modelos computacionais inspirados no funcionamento do sistema nervoso biológico através de diversas camadas, formada por neurônios artificiais, interconectadas seguindo o caminho da camada de entrada até a camada de saída, conhecido como padrão *feed-forward*. Em alguns modelos, as redes neurais convolucionais procuram imitar o processo de extração de características visuais que ocorre no córtex visual, onde neurônios individuais respondem a estímulos do campo receptivo, uma coleção desses campos se sobrepõe para cobrir toda a área visual (GOODFELLOW; BENGIO; COURVILLE, 2016). Dessa forma cérebro executa regularmente tarefas de reconhecimento perceptivo em um tempo bastante rápido (HAYKIN, 2001).

Vistas como ótimas ferramentas de classificação, as redes neurais convolucionais são um tipo de arquitetura que considera a estrutura espacial da imagem para realizar sua análise, assim sendo uma das principais arquiteturas utilizadas com esse objetivo (JÚNIOR, 2019).

Na rede neural convolucional, a imagem de entrada é propagada pela rede neural para gerar uma saída. O erro entre a saída prevista e a saída desejada é calculado. Esse erro é retropropagado através da rede para calcular gradientes em relação aos pesos. Os pesos são ajustados usando esses gradientes para reduzir o erro. Esse processo é repetido iterativamente até que a

rede atinja um desempenho desejado, permitindo que a rede aprenda automaticamente a partir dos dados de treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 1 – Operação Convolutacional.



Fonte: autor.

Fonte: (RAWAT; WANG, 2017).

Dessa forma a emulação do funcionamento do cérebro dos humanos para a análise de imagens otimizou o aperfeiçoamento de técnicas em nível de tempo e gasto computacional, alega (HAYKIN, 2001), possibilitando assim o uso dessa tecnologia também em análises médicas.

Desde então, as redes neurais convolucionais têm sido empregadas em uma variedade de domínios, abrangendo o reconhecimento de imagens, de fala, de padrões de escrita, a visão computacional e muitos outros (GOODFELLOW; BENGIO; COURVILLE, 2016). Acredita-se que essas redes também possam oferecer benefícios significativos na área da dermatoscopia e saúde, contribuindo para diagnósticos mais precisos e eficazes

2.5 Arquitetura das Redes Neurais Convolucionais

Existem diversas formas de aplicação das redes neurais convolucionais, essas aplicações são um tipo específico de modelos de rede neural convolucional pré-treinados projetados para processar dados como imagens. A aplicação de diferentes modelos pré-treinados podem resultar em níveis de acurácia diferentes de acordo com seu objetivo e configurações. Alguns desses modelos são: Resnet50, MobileNet, VGG16, VGG19 e InceptionV3.

2.5.1 Resnet50

A ResNet-50 é um modelo de rede neural convolucional pré-treinado que foi vencedor do desafio ImageNet em 2015. Esse modelo permite o treinamento de redes neurais com mais de 150 camadas. Ela foi projetada para lidar com o problema de degradação do desempenho à medida que a rede se torna mais profunda através de conexões de atalho para facilitar o fluxo de informação (HE *et al.*, 2016).

2.5.2 MobileNet

A arquitetura MobileNet foi especialmente concebida para eficiência em dispositivos móveis, incorporando convoluções separáveis e a capacidade de ajuste na largura. Essas características são estratégicas para garantir um equilíbrio adequado entre eficiência e desempenho em contextos com recursos restritos, como em smartphones e sistemas embarcados (HOWARD *et al.*, 2017)

2.5.3 VGG16 e VGG19

A arquitetura VGG (Visual Geometry Group) é uma família de CNN desenvolvida pelo Visual Geometry Group, da Universidade de Oxford. Duas variantes notáveis são a VGG16 e a VGG19, conhecidas por sua simplicidade e profundidade (SIMONYAN; ZISSERMAN, 2014).

A VGG16 é uma arquitetura composta por 16 camadas, incluindo camadas convolucionais e totalmente conectadas. Cada bloco convolucional consiste em camadas convolucionais 3x3 com ativação ReLU, seguidas por camadas de *pooling* máximas (MASCARENHAS; AGARWAL, 2021).

A VGG19 é uma extensão da VGG16, com 19 camadas. A principal diferença é a presença de camadas convolucionais adicionais, proporcionando uma maior profundidade à rede. Essas camadas extras ajudam a aprender características mais complexas (MASCARENHAS; AGARWAL, 2021).

2.5.4 InceptionV3

O InceptionV3 é uma arquitetura de rede neural convolucional desenvolvida pelo Google que foi treinada no conjunto de dados ImageNet. Ela alcançou um resultado com precisão superior a 78,1% nesse conjunto de dados. O objetivo principal do InceptionV3 é aprimorar o desempenho e a eficiência computacional em relação às iterações anteriores da arquitetura Inception (Google Developers, 2023).

A comparação das diversas arquiteturas é essencial para validar a variabilidade dos resultados, sendo os testes não paramétricos uma abordagem ideal para essa validação estatística.

2.6 Testes Não-Paramétricos

Testes não paramétricos são uma abordagem estatística que não depende de suposições sobre a distribuição dos dados, sendo valiosos para validar resultados de maneira robusta (CONTADOR; SENNE, 2016).

2.6.1 Teste de Friedman

Uma dessas abordagens é o teste de Friedman, um teste estatístico de hipótese não paramétrico que analisa se existem diferenças estatisticamente significativas nos resultados de experimentos repetidos (SIEGEL; JR, 2006).

Caso o resultado apresentado pelo teste de Friedman manifestar diferenças estatisticamente significativas, pode ser necessário realizar um pós-teste para compreender onde estão presentes essas diferenças.

2.6.2 Pós-teste de Nemenyi

O pós-teste de Nemenyi é uma técnica de análise de variância não-paramétrica utilizada para comparar múltiplos grupos em uma amostra. Ele compara todos os possíveis pares de grupos e determina se há diferenças estatisticamente significativas entre eles, indica (CAMPOS, 2002).

Este procedimento é empregado em cenários, nos quais análises, como teste de Friedman, revelam diferenças entre grupos, porém não identificam quais pares específicos apresentam diferenças estatisticamente significantes (FACELI; AL., 2021).

3 MATERIAIS E MÉTODOS

3.1 Ferramentas

Foi utilizada a base de dados HAM10000 (*Human Against Machine with 10000 training images*) desenvolvida pela Universidade de Vienna e disponibilizada por Harvard (TS-CHANDL; ROSENDAHL; KITTLER, 2018). A base contém 10.015 imagens dermatoscópicas e conta com 8 colunas, sendo elas:

- lesion_id: fornece o id da lesão;
- image_id: fornece o id da imagem;
- dx: fornece a classificação das lesões;
- dx_type: indica como foi feita a classificação das lesões;
- age: fornece a idade da pessoa que apresentou as lesões;
- sex: indica o sexo da pessoa que apresentou as lesões;
- localization: indica onde a lesão está localizada.

A categorização das lesões ou coluna dx, conta com sete classes identificadas, sendo elas:

- nv: Melanocytic nevi;
- mel: Melanoma;
- bkl: Benign keratosis-like lesions;
- bcc: Basal cell carcinoma;
- akiec: Actinic keratoses;
- vas: Vascular lesions; e
- df: Dermatofibroma.

Dentre as classes, mel e bcc são consideradas tumores malignos, as quais exigem maior atenção e tratamento de médicos.

Técnicas de análise de dados são imprescindíveis para o desenvolvimento do projeto, tanto para o pré-processamento como também para a análise das informações presentes na base de dados. Para isso, foram utilizadas algumas ferramentas como a linguagem Python e também as bibliotecas *scikit learn*, utilizada para pré-processamento, construção e treinamento do

modelo e a biblioteca *tensorflow*, utilizada na avaliação e comparação dos modelos através do desenvolvimento de gráfico. Ferramentas essas vistas como opções em potencial para desenvolvimento do código.

3.2 Pré-Processamento de Dados

O processo é iniciado com a leitura e pré-processamento e análise inicial da base de dados, lidando com dados inválidos e fazendo o redimensionamento das imagens para 112x112, gerando uma padronização e se adequando ao poder computacional do computador. Após pré-processamento dos dados, a separação da base de teste e treino é feita realizando um balanceamento do número de imagens por tipo de característica, tendo assim uma base de treino equilibrada possibilitando o reconhecimento de padrões presentes em cada categoria.

3.3 Aplicação dos Modelos de Redes Neurais Convolucionais

Tendo a base de treino e teste separada, os modelos de CNN Resnet50, MobileNet, VGG16, VGG19 e InceptionV3 foram aplicados iniciando pela validação da base de teste, passando pelo registro dos resultados em variáveis e seguindo para a análise estatística dos resultados.

3.4 Análise dos Resultados

Com os dados apresentados, são feitas análises estatísticas de hipótese no resultado dos cinco modelos utilizados, aplicando as seguintes análises:

- Acurácia: medir a precisão de previsões corretas no conjunto de treino;
- Perda: indicar o quanto as previsões do modelo estão diferentes dos valores reais no conjunto de treino;
- Validação de acurácia: medir a precisão de previsões corretas no conjunto de validação;
- Validação de perda: indicar o quanto as previsões do modelo estão diferentes dos valores reais no conjunto de validação;

Após realizar as análises estatísticas de hipótese, foram aplicados testes não para métricas iniciando pelo teste de Friedman. Com resultados do teste de Friedman foi possível verificar se existiam diferenças significativas e com isso foi aplicada a técnica de análise de variância não-paramétrica do teste de Nemenyi para validar quais modelos apresentam diferenças significativas entre eles.

4 RESULTADOS

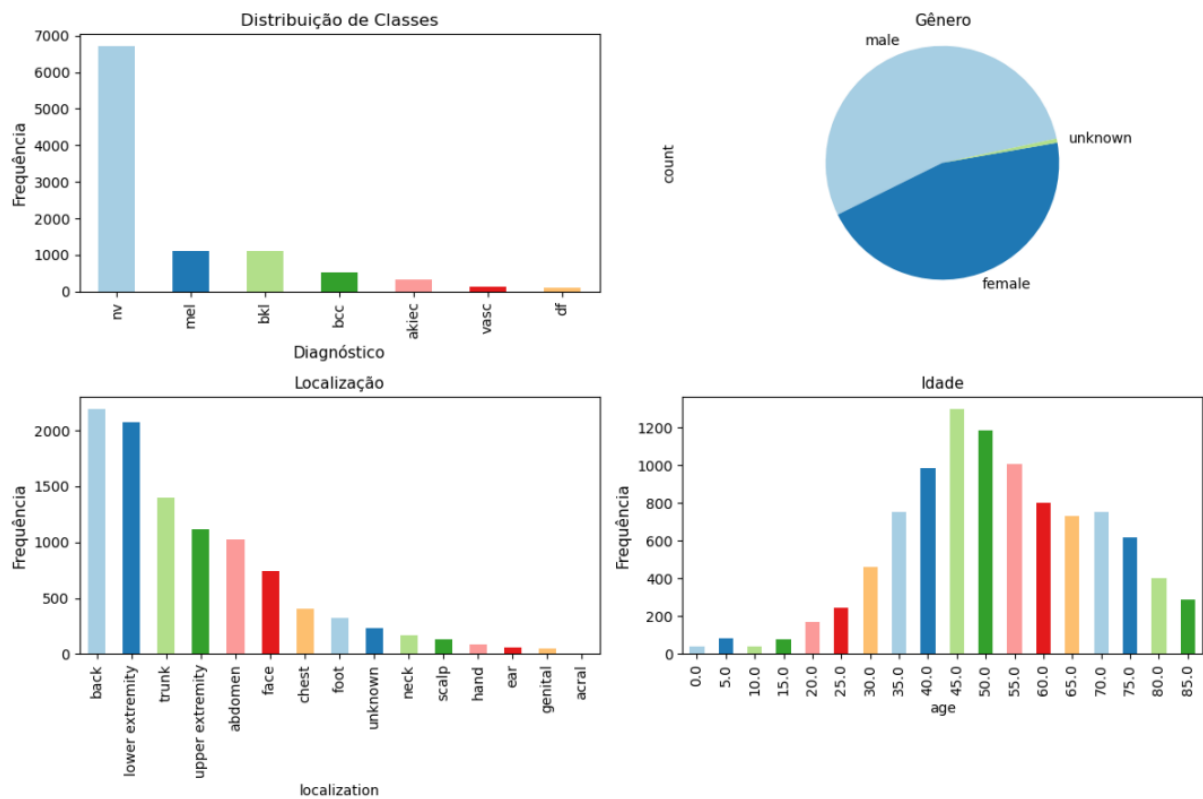
Neste capítulo serão apresentados os principais resultados obtidos através do uso da base de dados HAM10000 na análise de cada um dos modelos de arquiteturas avaliados para este trabalho. Será descrito a etapa de tratamento de dados, treinamento, em seguida é apresentada a análise dos resultados e comparativos dos testes.

4.1 Tratamento de Dados

Para realizar o tratamento de uma base de dados para um modelo *deep learning* é necessário algumas características. Uma dessas características utilizadas é a transformação dos rótulos de textos em numéricos. Esse processo foi realizado para a coluna dx, que comporta a categorização das imagens através da classe *LabelEncoder* pertencente a biblioteca *scikit-learn*. Os dados resultantes foram salvos em uma nova coluna chamada label.

Após a criação da coluna *label* foram gerados gráficos com o objetivo de entender a distribuição de informações dentro da base de dados, resultando na seguinte imagem:

Figura 2 – Análise da base de dados.



Fonte: autor.

Com base nesses dados é possível verificar que a classe *nv*, *melanocytic nevi*, contempla uma porcentagem desproporcional em relação as outras classes, sendo 66,9% dos dados. Dessa

forma torna-se necessário realizar um balanceamento para que os modelos de treinamento não aprendam as características de apenas uma classe.

Para realizar esse balanceamento, foi utilizada a função *resample* pertencente a biblioteca *scikit-learn* utilizando um número de 500 imagens. Com essa configuração, foram criados sete *datasets*, cada um deles contendo 500 imagens representando uma classe, gerando assim um *dataset* para cada classe.

Em seguida, todos esses *datasets* foram concatenados em um novo *dataset* chamado *data_pele_balanceado* que seria utilizado para a criação dos modelos de treinamento.

Com o *dataset* criado, as imagens do *dataframe* foram convertidas em um arranjo (*array*), onde X recebe a lista de imagens e Y recebe a coluna *label* (classe), seguindo para o uso da função *train_test_split*.

A função *train_test_split* foi usada para dividir o conjunto de dados em conjunto de treinamento e conjunto de teste, divisão essa essencial para avaliar o desempenho de um modelo, permitindo treinar esse modelo em com um conjunto de dados e testá-lo em outro conjunto com dados diferentes. Para essa divisão foi utilizado um tamanho do conjunto de treino de 25% e um conjunto de teste de 75%.

4.2 Treinamento de Modelos

Para a criação dos modelos de treinamento o código iniciou com a definição do modelo base utilizado. Seguindo pela definição das camadas adicionais da arquitetura CNN que extrairão características mais complexas das imagens, passando para as camadas convolucionais que aprenderão os padrões espaciais das imagens. Além disso, foram implementadas camadas de pool máximo, que reduzem as dimensões espaciais e camadas de dropout que auxiliam na diminuição da complexidade do modelo e na prevenção do overfitting. Em seguida, as camadas totalmente conectadas combinam as características extraídas em uma representação que pode ser usada pelo modelo para a classificação das imagens.

No treinamento algumas arquiteturas foram utilizadas, dentre elas a arquitetura ResNet50. Esse modelo de rede neural convolucional foi estruturado a partir de um objeto de modelo sequencial adicionando uma camada convolucional de 256 filtros com camada de para imagens tridimensionais com largura de 112x112x3 *pixels*. O modelo utiliza também uma camada de dropout com taxa de 0.3, descartando de forma aleatória 30% da camada anterior, prevenindo o superajuste. Para o treinamento foi utilizando o modelo pré-treinado ResNet50 como base.

O mesmo processo se seguiu para os outros quatro modelos de arquiteturas, utilizando suas arquiteturas como base para desenvolvimento dos modelos e resultados de treinamento. As arquiteturas utilizadas foram MobileNet, VGG16, VGG19 e InceptionV3.

O treinamento e validação de cada modelo foi realizado utilizando um número de sessenta épocas, calculando a acurácia e perda de validação em cada uma das sessenta épocas. Tendo

os modelos definidos, os dados foram salvos em arquivos *pickle* para análise de resultados posterior.

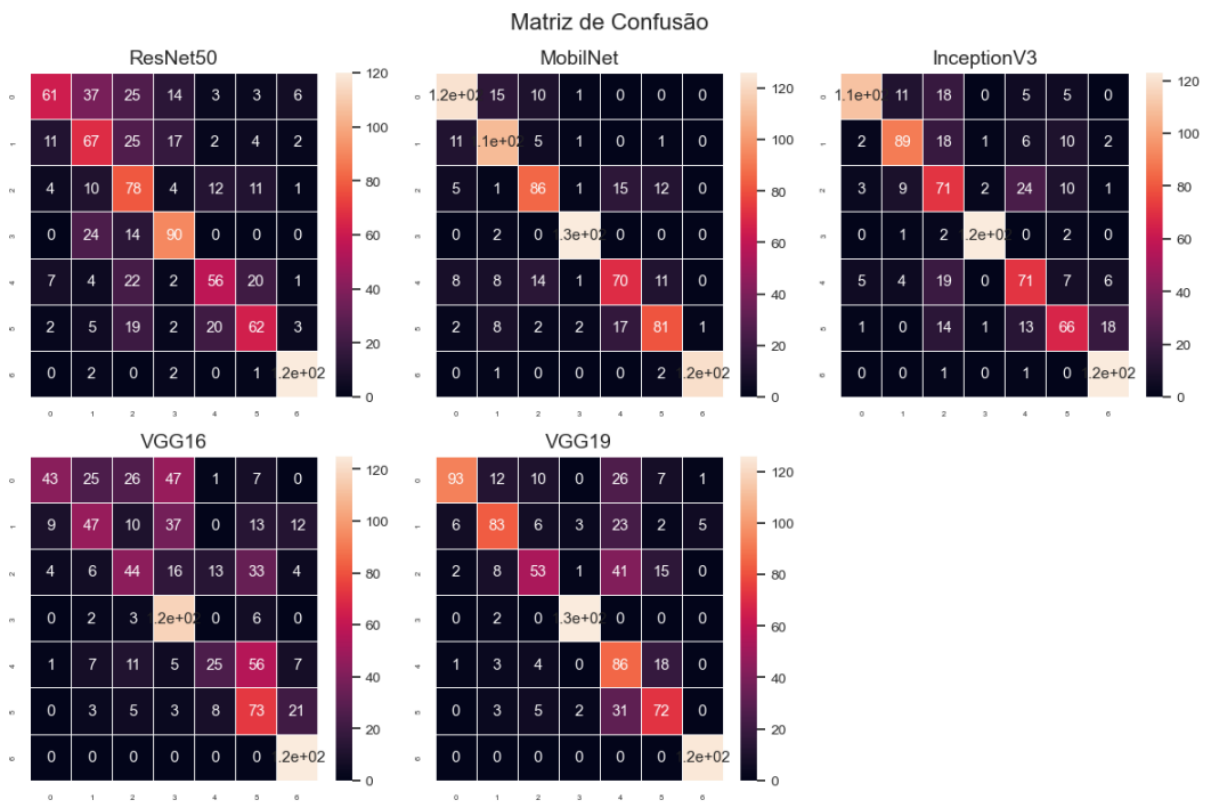
4.3 Análise dos Resultados

Análise dos resultados é o processo de avaliar o desempenho de modelos de aprendizado de máquina. Essa análise é importante para identificar possíveis inconsistências nos modelos e medir os resultados obtidos desse modelo.

4.3.1 Matriz de Confusão

Os modelos foram analisados por meio de gráficos, com a matriz de confusão sendo o primeiro gráfico utilizado. Essa ferramenta fornece uma visão geral do desempenho do modelo em cada classe, auxiliando na identificação de áreas que precisam ser melhoradas.

Figura 3 – Matriz de Confusão

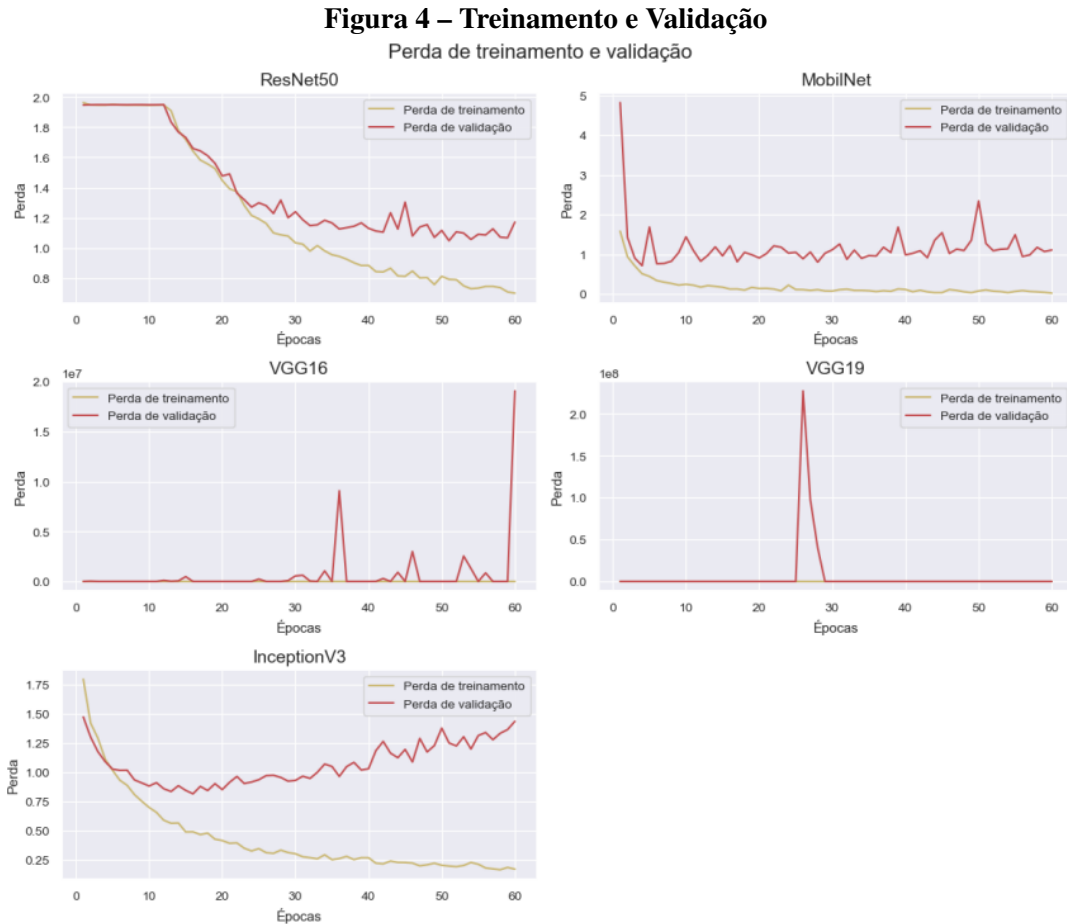


Fonte: autor.

Verificando os gráficos de matriz de confusão de cada um dos modelos é possível identificar que os modelos Resnet50 e o modelo VGG16 tiveram uma discrepância maior na acurácia da identificação dos dados de cada classe. Já o modelo MobileNet e InceptionV3 foram os modelos que tiveram uma melhor acurácia.

4.3.2 Perda de Treinamento e Validação

Esse gráfico demonstra a evolução da perda de treinamento e validação dos modelos através da evolução das épocas conforme Figura 4.



Fonte: autor.

Na análise dos gráficos é possível atestar que o treinamento e validação de cada modelo ocorreu de forma bastante diferente entre eles. Em relação ao treinamento, os modelos MobileNet, InceptionV3 e Resnet50 tiveram um comportamento similar, iniciando com um valor de perda alto, diminuindo perto do zero e estabilizando na mesma faixa de valor. Os modelos VGG16 e VGG19 tiveram um valor estável na faixa de zero de perda.

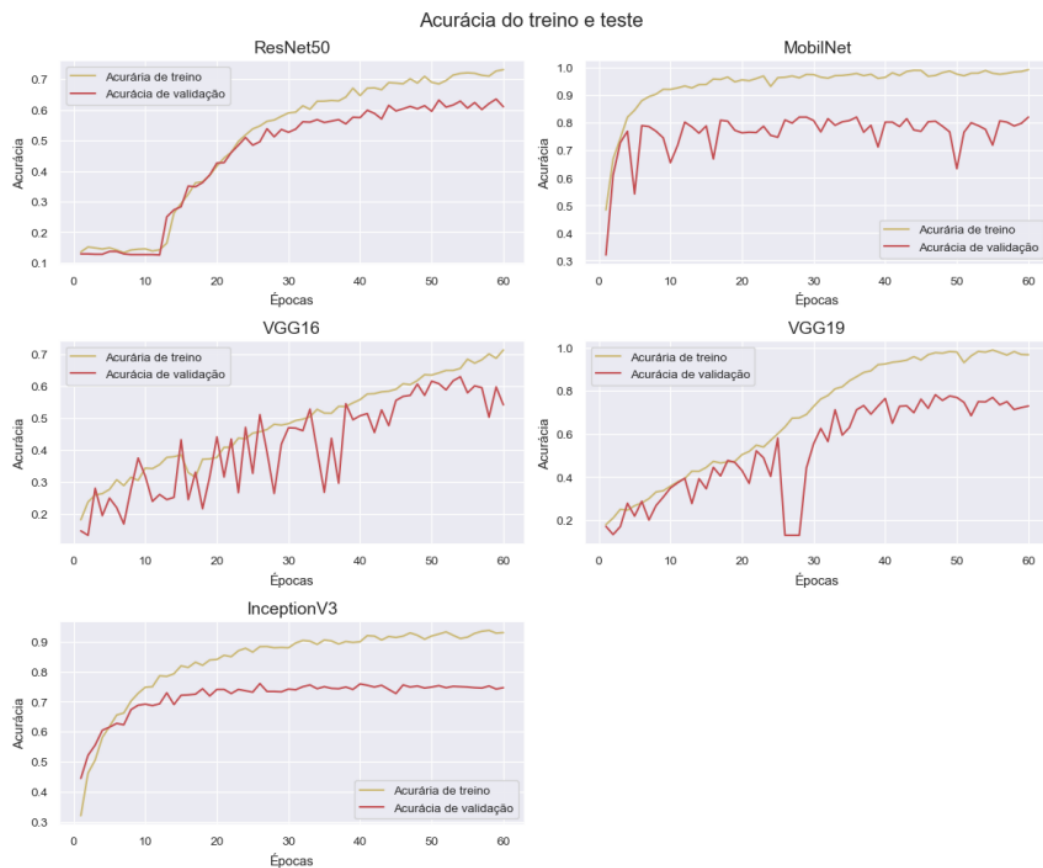
Em relação a validação, os modelos Resnet50 e MobileNet tiveram uma diminuição da perda de validação e estabilizaram em uma mesma faixa de valores, já InceptionV3 teve uma queda maior do índice de perda no início e posteriormente um crescimento. Os modelos VGG16 e VGG19 apresentaram uma validação quase que homogeneamente na faixa de zero de perda, oscilando com alguns picos e retornando para o valor zero.

4.3.3 Acurácia de Treino e Teste

Esse gráfico demonstra a evolução da acurácia de treinamento e teste dos modelos através da evolução das épocas conforme Figura 5.

Analisando os dados é possível ver que os modelos ResNet50, MobileNet, VGG19 e InceptionV3 alcançaram um resultado final de acurácia de treino próximo de 1, já o modelo VGG16 teve uma acurácia de 0.7 considerando a mesma quantidade de épocas empregadas. Os modelos Resnet50, MobileNet, InceptionV3 tiveram uma curva de crescimento similar, com um crescimento maior no início e uma estabilização posterior. Já os modelos VGG16 e VGG19 tiveram um crescimento incremental dos modelos com estabilização final no modelo VGG19.

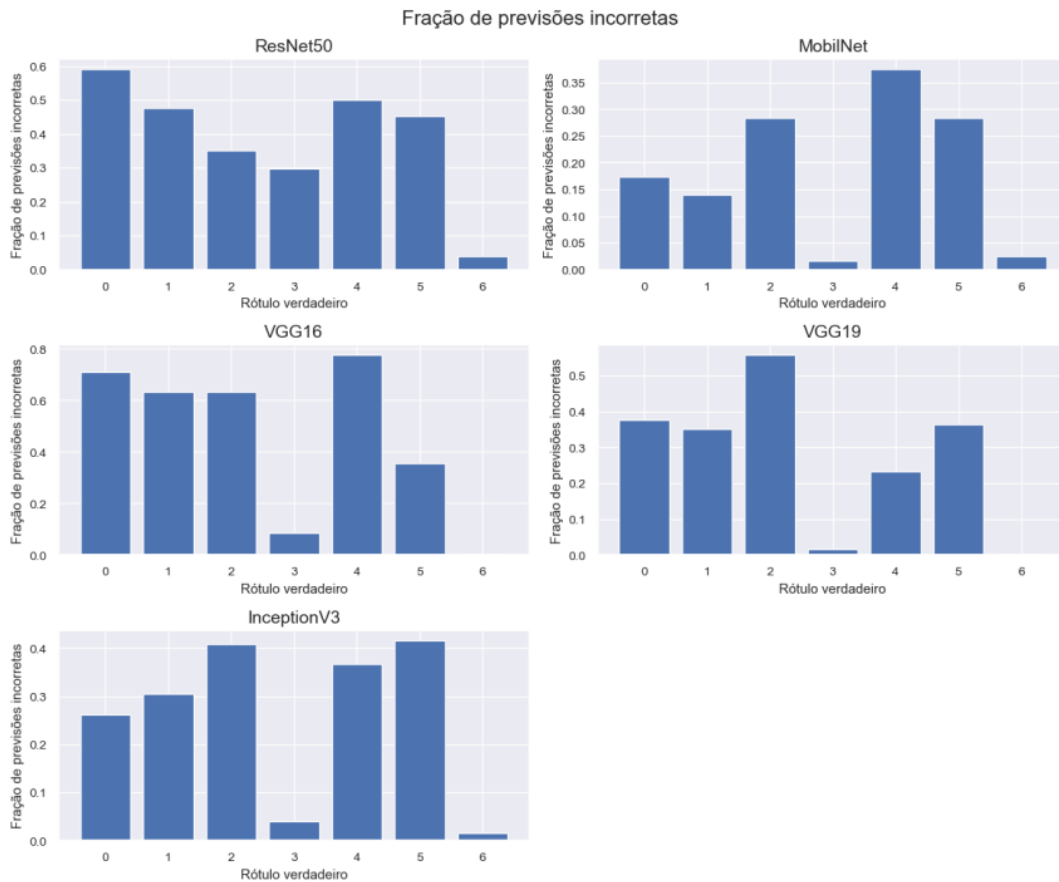
Figura 5 – Acurácia de Treino e Teste



Fonte: autor.

Em relação à acurácia de teste a curva de aprendizado de cada modelo foi similar ao formato da curva de treinamento considerando um resultado menor e uma instabilidade maior entre as épocas. Nos modelos MobileNet, VGG16 e VGG19 a instabilidade nos resultados foi maior.

Figura 6 – Fração de Previsões Incorretas



Fonte: autor.

4.3.4 Fração de Previsões Incorretas

Esse gráfico demonstra a evolução da fração de previsões incorretas dos modelos através da evolução das épocas conforme Figura 6.

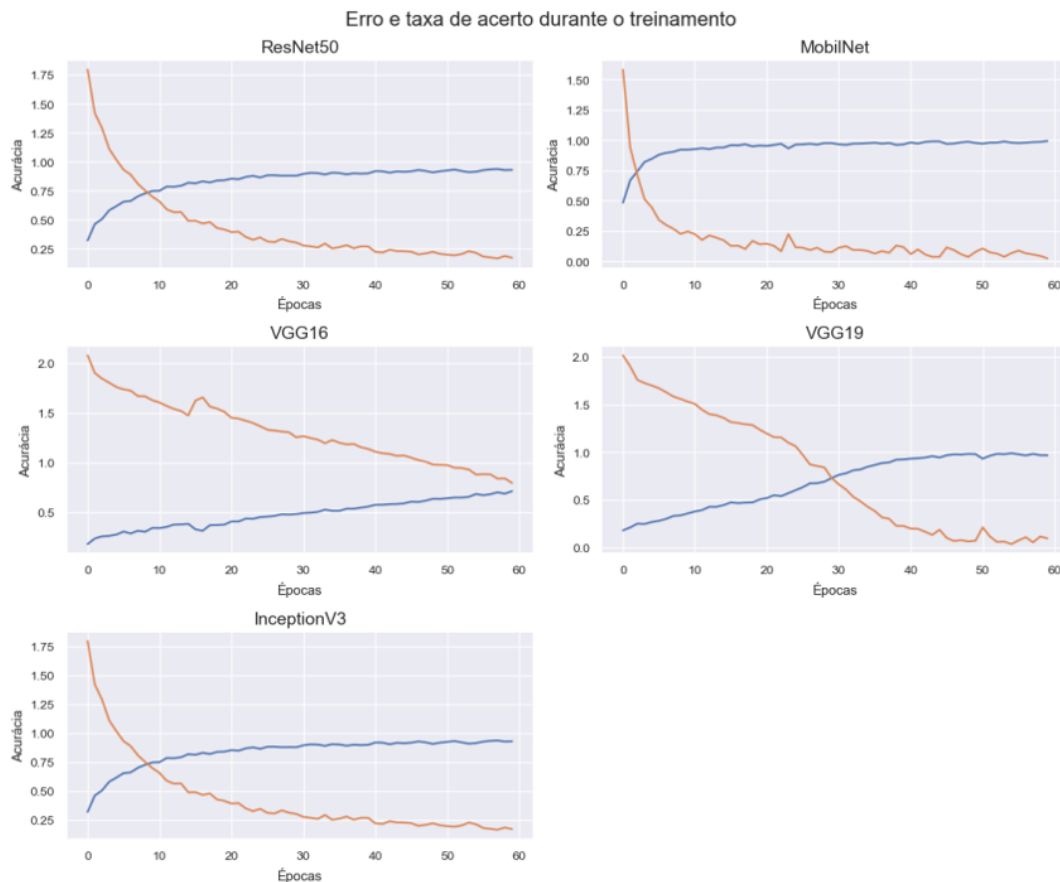
É possível verificar que em todos os modelos a categoria 6 teve o menor valor de fração de previsões incorretas, seguido pela categoria 3. O modelo VGG19 foi o modelo que teve o menor número de fração de previsões incorretas o modelo ResNet50 foi o que teve em sua totalidade o maior número de frações incorretas.

4.3.5 Erro e Taxa de Acerto Durante o Treinamento

Esse gráfico demonstra a evolução do erro e taxa de acerto durante o treinamento dos modelos através da evolução das épocas conforme Figura 7.

Nos modelos ResNet50, MobileNet e InceptionV3 a curva da taxa de erro e acerto durante o treinamento foram similares, tendo uma queda maior do erro (em amarelo) nas primeiras épocas seguido de uma estabilização e um crescimento maior no acerto (em azul) nas primeiras

Figura 7 – Erro e Taxa de Acerto Durante o Treinamento



Fonte: autor.

épocas e seguido de uma estabilização. Os modelos VGG16 e VGG19 tiveram um comportamento diferente, representando uma faixa de erro decrescente e um acerto crescente.

Com diferença de que no VGG16 o erro e acerto finalizam as épocas em valores próximos e no VGG19 o ponto de encontro é ultrapassado e a taxa de acerto passa a taxa de erro se tornando maior e a taxa de erro menor durante o treinamento.

4.4 Testes Não-Paramétricos

Os testes não-paramétricos são testes frequentemente usados para a comparação de duas ou mais amostras, podendo ser usado para analisar modelos de redes neurais convolucionais comparando o desempenho em um determinado conjunto de dados de teste. Para isso é preciso usar um teste de comparação de amostras como é o caso do teste de Friedman, que ajudará a identificar se existem diferenças estatisticamente significativas entre os experimentos repetidos que foram realizados (SIEGEL; JR, 2006).

4.4.1 Teste de Friedman

O teste de Friedman é um teste não-paramétrico utilizado raramente para analisar vários modelos de **deep learning**, uma das vantagens do seu trabalho é o uso do teste estatístico para a comparação de modelos. Ele funciona calculando uma estatística chamada qui-quadrado de Friedman onde caso o valor-p do teste de Friedman for menor que um nível determinado, conclui-se que existe uma diferença significativa entre os desempenhos dos modelos.

Com a aplicação do teste de Friedman na comparação entre os cinco modelos utilizado, alcançou-se um resultado P-value de $1.301789018141355e-40$ que é bastante próximo de zero, indica-se que há diferenças reais entre os modelos. Os testes foram aplicados considerando um nível de significância de 95.

Com os indicativos de diferença reais entre os modelos, a aplicação de outros testes como o teste de Nemenyi é vista como necessária para um aprofundamento na análise.

4.4.2 Pós-teste de Nemenyi

A aplicação do teste de Nemenyi foi realizada nos cinco modelos de redes neurais convolucionais utilizando a base de dados HAM10000. Os resultados foram tabulados para facilitar a comparação, gerando uma tabela correspondente a esses resultados que pode ser visualizada abaixo.

Analisando os resultados apresentados chega-se a seguinte constatação:

- Diferença significativa entre InceptionV3 e MobileNet (p-value: 0.034832714621342985)
- Diferença significativa entre InceptionV3 e ResNet50 (p-value: $6.453443664746892e-10$)
- Diferença significativa entre InceptionV3 e VGG16 (p-value: $8.35701216098028e-13$)
- Diferença significativa entre InceptionV3 e VGG19 (p-value: 0.00022485057843736356)
- Diferença significativa entre MobileNet e ResNet50 (p-value: $1.3131858689527199e-21$)
- Diferença significativa entre MobileNet e VGG16 (p-value: $8.082987204751365e-26$)
- Diferença significativa entre MobileNet e VGG19 (p-value: $1.0386243575702592e-12$)
- Não há diferença significativa entre ResNet50 e VGG16 (p-value: 0.9308305768079229)

Tabela 1 – Teste de Nemenyi

	InceptionV3	MobileNet	ResNet50	VGG16	VGG19
InceptionV3	1.000000e+00	3.483271e-02	6.453444e-10	8.357012e-13	2.248506e-04
MobileNet	3.483271e-02	1.000000e+00	1.313186e-21	8.082987e-26	1.038624e-12
ResNet50	6.453444e-10	1.313186e-21	1.000000e+00	9.308306e-01	2.496158e-01
VGG16	8.357012e-13	8.082987e-26	9.308306e-01	1.000000e+00	3.224843e-02
VGG19	2.248506e-04	1.038624e-12	2.496158e-01	3.224843e-02	1.000000e+00

Fonte: autor.

- Não há diferença significativa entre ResNet50 e VGG19 (p-value: 0.24961583312806282)
- Não há diferença significativa entre VGG16 e VGG19 (p-value: 0.03224843055711968)

Em resumo, com base nos dados e no teste de Nemenyi é possível afirmar que os modelos MobileNet e InceptionV3 tiveram diferenças estatisticamente significativas com todos os outros modelos. Já o modelo Resnet50 não teve diferença significativa em comparação com os modelos VGG16.

5 CONCLUSÃO

A utilização de diferentes arquiteturas de redes neurais convolucionais em conjunto com a base de dados HAM10000 levou à constatação de que a adoção de modelos diferentes pode ocasionar diferenças significativas nos resultados alcançados, exemplificado o caso dos modelos Mobilenet e InceptionV3, assim como também entre o modelo Resnet50 e os modelos VGG19 e VGG16.

Devido a limitação do dispositivo onde os modelos foram treinados algumas ações precisaram ser tomadas para a realização das análises como a diminuição do tamanho dos arquivos de imagens e também a delimitação do número de épocas do treinamento e teste do modelo. Com um poder computacional maior o número de modelos comparados também poderia ser aumentado.

A utilização do aumento de dados pode ser considerada uma perspectiva promissora para aprimoramento do projeto, conduzindo a um estudo mais aprofundado destinada à publicação acadêmica. A comparação entre os modelos revelou uma possível direção para estudo e aprofundamento, com o propósito de identificar o modelo de CNN mais apropriado para a investigação de imagens dermatoscópicas.

Assim, destacam não apenas os resultados desse trabalho como também apontam para um horizonte de investigações e avanços, gerando novas informações e trazendo progresso na interseção entre o uso de redes neurais convolucionais e a análise de imagens dermatoscópicas.

REFERÊNCIAS

- A.C. Camargo Cancer Center. **Digital health**. 2023. <https://accamargo.org.br/sobre-o-cancer/medicina-diagnostica/dermatoscopia>. Acesso em: 17 de setembro de 2023.
- CAMPOS, G. M. **Estatística prática para docentes e pós-graduandos**. [S.l.]: Faculdade de Odontologia de Ribeirão Preto da Universidade de São Paulo, 2002.
- CASAGRANDE, J. **Dermatoscopia, dermatoscopia digital e microscopia confocal, os exames que ajudam a diagnosticar o câncer de pele**. 2023. Disponível em: <https://accamargo.org.br/sobre-o-cancer/noticias/dermatoscopia-dermatoscopia-digital-e-microscopia-confocal-os-exames-que>.
- CONTADOR, J. L.; SENNE, E. L. F. Testes não paramétricos para pequenas amostras de variáveis não categorizadas: um estudo. **Gestão & Produção**, SciELO Brasil, v. 23, p. 588–599, 2016.
- COSTA, M. G. F. Redes neurais convolucionais na saúde. **Journal of Health Informatics**, v. 9, n. 4, 2017.
- FACELI, K.; AL., e. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 2. ed. Rio de Janeiro: Grupo GEN, 2021.
- FRANGE, V. M. N.; ARRUDA, L. H. F.; DALDON, P. É. C. Dermatoscopia: importância para a prática clínica. **Revista de Ciências Médicas**, v. 18, n. 4, 2009.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press, 2016.
- Google Developers. **Guia Avançado do Inception v3**. 2023. Acessado em 12 de novembro de 2023. Disponível em: <https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=pt-br>.
- HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2001.
- HE, K. *et al.* Deep residual learning for image recognition. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- Hospital Sírio-Libanês. **Carcinoma Basocelular: Saiba mais**. 2023. Acesso em: 24 de setembro de 2023. Disponível em: <https://hospitalsiriolibanes.org.br/blog/oncologia/carcinoma-basocelular-saiba-mais>.
- Hospital Sírio-Libanês. **O que é um melanoma cutâneo?** 2023. Acesso em: 24 de setembro de 2023. Disponível em: <https://hospitalsiriolibanes.org.br/blog/oncologia/o-que-e-um-melanoma-cutaneo>.
- HOWARD, A. G. *et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.
- JÚNIOR, M. R. S. **Redes neurais convolucionais e máquinas de vetores de suporte para classificação de imagens médicas**. 2019. Tese (Doutorado) — Universidade Estadual de Campinas, 2019.
- MASCARENHAS, S.; AGARWAL, M. A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. *In: IEEE. 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*. [S.l.], 2021. v. 1, p. 96–99.

- Ministério da Saúde. **Câncer de Pele**. 2023. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer-de-pele>.
- RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. **Neural computation**, MIT Press, v. 29, n. 9, p. 2352–2449, 2017.
- REZZE, G. G.; SÁ, B. C. S. d.; NEVES, R. I. Dermatoscopia: o método de análise de padrões. **Anais brasileiros de dermatologia**, SciELO Brasil, v. 81, p. 261–268, 2006.
- RGUIBI, Z.; HAJAMI, A.; DYA, Z. Explaining deep neural networks in medical imaging context. *In*: IEEE. **IEEE/ACS International Conference on Computer Systems and Applications**. [S.l.], 2021. p. 1–2.
- SIEGEL, S.; JR, N. J. C. **Estatística não-paramétrica para ciências do comportamento**. [S.l.]: Artmed Editora, 2006.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.
- Sociedade Brasileira de Dermatologia. **Câncer da pele**. 2023. <https://www.sbd.org.br/doencas/cancer-da-pele/>. Acesso em: 17 de setembro de 2023.
- TSCHANDL, P.; ROSENDAHL, C.; KITTLER, H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. **Scientific data**, Nature Publishing Group, v. 5, n. 1, p. 1–9, 2018.
- VAZ, G. C. **Aplicação de modelos de deep learning para qualificação da área da engenharia biomédica: um estudo de caso de visão computacional em imagens de raio-x da região torácica**. 2023. Tese (Doutorado) — [sn], 2023.
- World Health Organization. **Digital health**. 2023. Disponível em: <https://www.who.int/health-topics/digital-health#tab=tab1>.