



PUC
CAMPINAS
PONTIFÍCIA UNIVERSIDADE CATÓLICA

**Pontifícia Universidade Católica
de Campinas**

*Rua Prof. Dr. Euryclides de Jesus Zerbini, 1516 -
CEP 13087-571 - Campinas, SP, Brasil*

RELATÓRIO DE INICIAÇÃO CIENTÍFICA

Período de setembro de 2023 à Agosto de 2024

Uso de Arquitetura de Rede Neural Convolucional - GoogLeNet - para Diagnóstico de Lesões de Pele

Supervisor: Prof.º Dr. Ademar Takeo Akabane

Programa de Pós-Graduação em Gestão de Redes de

Telecomunicações

<http://lattes.cnpq.br/6781874728187325>

ademar.akabane@puc-campinas.edu.br

Candidato: Antônio Marcio Crepaldi Junior– **RA:** 21012485

Curso de Engenharia de

Computação **Ano de início:** 2021 – **Ano de conclusão:**

2025

<http://lattes.cnpq.br/608877999633546>

4 antonio.mcj@puccampinas.edu.br

Grupo de Pesquisa: Sistemas Inteligentes

Linha de Pesquisa: Tecnologia da Informação Aplicada a Serviços em Redes de Telecomunicações

Área de Tecnologias Prioritárias do MCTIC: Tecnologias para o Desenvolvimento Sustentável

*Esse Plano de Trabalho de Iniciação Científica está vinculado ao Plano de Trabalho de Pesquisa do docente pesquisador intitulado de **Sistemas de Auxílio ao Diagnóstico Médico Aplicando Redes Neurais Profundas para Classificação de Lesões de Pele***

Campinas

Agosto de 2024

Sumário

1	Introdução	1
2	Objetivos	2
3	Metodologia Específica	2
3.1	Atividades Previstas no Cronograma Original.....	3
3.2	Cronograma do Trabalho.....	4
3.3	Atividades Realizadas.....	4
4	Resultado Obtidos	6
4.1	Análise dos Dados Obtidos do Dataset.....	7
4.2	Pré-Processamento dos Dados.....	9
4.3	Segmentação e experimentação nas imagens.....	11
4.4	Implementação da GoogleNet.....	13
4.5	Resultados	15
5	Discussão	16
6	Conclusão	17
7	Referências	18

1 Introdução

De acordo com o Instituto Nacional de Câncer (INCA) são esperados 704 mil casos novos de câncer no Brasil para cada ano do triênio 2023-2025 [1], com destaque para as regiões Sul e Sudeste, que concentram cerca de 70% da incidência [2]. Neste mesmo estudo, destaca-se que o câncer de pele é o tipo mais comum no Brasil, representando cerca de 30% de todos os casos de câncer no país. Vale destacar que há vários tipos de câncer de pele, tais como, melanoma, carcinoma basocelular e carcinoma espinocelular, entre outros. Dos citados, o melanoma é considerado o mais letal, pois apesar de representar menos de 5% dos tumores de pele, é o responsável por mais de 90% das mortes [3, 4]. Apesar do melanoma ter altas taxas de letalidade, quando descoberto em fase inicial pode ser facilmente curado.

Normalmente, quando há suspeita de câncer de pele, a primeira etapa do diagnóstico é realizada por meio de inspeção visual de um especialista. Um ponto importante a ser destacado nessa etapa é a dificuldade de distinguir entre melanoma e nevos (pintas) melanocíticos, especialmente nos estágios iniciais da doença. Então, parte-se para a segunda etapa, que é uma análise dermatoscópica. Nessa etapa, realiza-se análises de imagens dermatoscópicas com intuito de aumentar a chance de descoberta de melanomas malignos de forma precoce. Para se ter uma ideia de números, a acurácia de diagnóstico por inspeção visual **não auxiliada** é de aproximadamente 60% [5] e aumentando para 75% com o auxílio de dermatoscopias [5, 6].

Por um lado, até mesmo por especialistas, realizar a classificação manualmente por meio de análise dermatoscópica das lesões de pele é muito custoso e difícil de ser realizado, devido a grande variedade de manchas que atualmente existem. Por outro lado, a rápida detecção é essencial para a cura, uma vez que a taxa de sobrevivência de mais de 95% nos estágios iniciais da doença cai para menos de 15% nos estágios mais avançados.

A *aprendizagem* é uma das funções mais importantes para os seres humanos, animais e sistemas inteligentes. Os humanos e os animais aprendem com a experiência adquirida, nesse sentido, a aprendizagem é um processo pelo qual o conhecimento é adquirido por meio de experiências passadas. Sistemas inteligentes aplicam técnicas de aprendizado de máquina para simular o aprendizado humano, permitindo que os computadores adquiram conhecimento por meio de algoritmos e dados do mundo real.

Aprendizado de máquina é uma área da Inteligência Artificial que tem como objetivo o desenvolvimento de técnicas computacionais que permitem a construção de sistemas com capacidade de adquirir conhecimento de forma automática a partir de dados de entrada não-aleatórios (ou exemplos). Vale destacar que essas técnicas são possíveis porque tais dados possuem um padrão e assim permitindo a extração de generalizações. Então com esses dados, é possível obter importantes aspectos sobre o conjunto de dados no qual o modelo foi treinado, veja a Figura 1.1.

Uma subárea de ML é *Deep Learning* - DL (aprendizado profundo) que está sendo aplicada

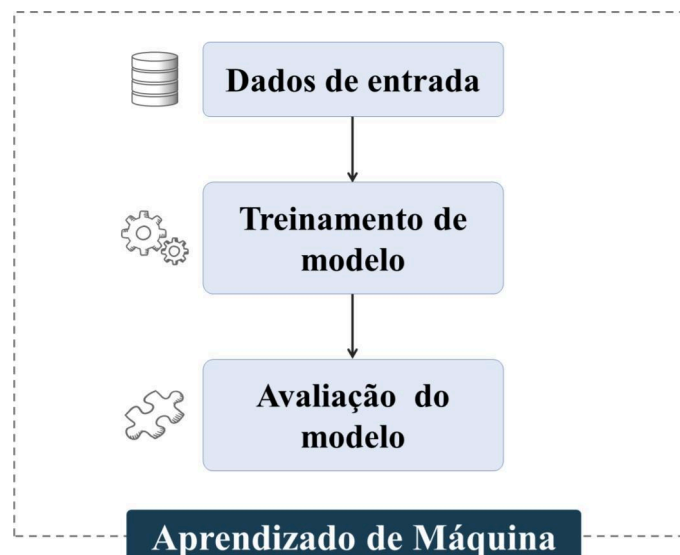


Figura 1.1: Arquitetura geral de aprendizado de máquina [7]

em diversas áreas do conhecimento, desde a área médica [8] até a área da engenharia civil [9]. Em DL os modelos organizados em camadas, onde cada camada utiliza os dados de saída da camada anterior, como dados de entrada. Dessa forma, uma camada posterior na rede é resultado de transformações lineares e não lineares das camadas anteriores a ela [10]. Para aplicações de visão computacional e de processamento de imagem, as arquiteturas de redes neurais profundas de maior uso são as *Convolutional Neural Networks* – CNN (redes neurais convolucionais), propostas por Yann LeCun e demais autores [11].

Os recentes avanços das CNN possibilitaram o uso dessas redes na automação de diagnósticos médicos por meio de imagens. Vale destacar que, que algumas aplicação de CNN em diagnósticos médicos chegou a ultrapassar o desempenho de médicos especialistas, por exemplo, diagnóstico de retinopatia diabética em imagens de retina [12, 13].

2 Objetivos

O presente Plano de Trabalho tem como objetivo geral implementar um classificador de reconhecimento de padrões em imagens dermatoscópicas para a detecção de melanoma utilizando a rede neural convolucional GoogLeNet. Além disso, pretende-se aplicar a técnica de *transfer learning* com intuito de promover um aumento no desempenho da rede utilizada.

3 Metodologia Específica

Esta seção foi dividida em três seções: Atividades Previstas no Cronograma Original (Seção 3.1), Cronograma do Trabalho (Seção 3.2) e Atividades Realizadas (Seção 3.3).

3.1 Atividades Previstas no Cronograma Original

Afim de atingir o objetivo proposto para esta pesquisa, foram propostas as seguintes atividades:

- **A1 - Verificação do estado da arte:** Estudo dos trabalhos existentes na literatura, por meio de *surveys*, *journals* e artigos. Pode-se destacar dois objetivos principais da revisão da literatura: (i) aprofundar os conhecimentos de técnicas e conceitos em rede neural convolucional GoogLeNet e (ii) identificar as lacunas em aberto. Além do mais, realizar reuniões periódicas com o supervisor;
- **A2 - Participação no Encontro de Iniciação Científica e Encontro de Iniciação em Desenvolvimento Tecnológico e Inovação:** A participação consolida a missão da Universidade de qualificar a graduação e proporcionar ao estudante seu engajamento na pesquisa científica;
- **A3 - Obtenção do conjunto de dados e estudo da arquitetura GoogLeNet:** Nesta etapa, tem como objetivo obter um conjunto de dados de imagens de lesão de pele, mas especificamente o conjunto de dados HAM-10000 e também o estudo desse conjunto de dados. Além disso, familiarizar-se com a arquitetura GoogLeNet;
- **A4 - Entrega do Relatório Parcial:** Escrita do Relatório Parcial detalhando as atividades realizadas até o momento da escrita;
- **A5 - Implementação do classificador:** Nesta etapa, tem como objetivo implementar um sistema que permita a detecção automatizada de melanoma em imagens dermatológicas digitais com o apoio da rede neural convolucional GoogLeNet;
- **A6 - Avaliação quantitativa e qualitativa:** Avaliação quantitativa e qualitativa serão realizadas com o objetivo de avaliar a solução proposta deste plano de pesquisa. Dessa forma, a escolha das características a serem avaliadas serão definidas e, também examinadas quais delas podem influenciar no desempenho da solução;
- **A7 - Escrita de artigo científico:** Os resultados obtidos durante o período da Iniciação Científica serão divulgados em eventos científicos. As melhorias das soluções ocorrerão a partir de *feedbacks* das reuniões periódicas, comentários dos revisores e da própria soluções do estado-da-arte;
- **A8 - Entrega do Relatório Final:** Escrita do Relatório Final detalhando todas as atividades desenvolvidas no período de Iniciação Científica.

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

3.2 Cronograma do Trabalho

Na Tabela 3.1 é apresentado o cronograma das atividades original do Plano de Pesquisa e que foram seguidas durante o decorrer do trabalho. Os círculos preenchidos se referem às atividades realizadas.

Tabela 3.1: Cronograma de trabalho da Iniciação Científica.

		2023		2024			
		Set.~Out.	Nov.~Dez.	Jan.~Fev.	Mar.~Abr.	Mai.~Jun.	Jul~Ag.
	A1	●	●	●	●	●	●
	A2	●					
	A3	●	●	●			
	A4			●	●		
	A5			●	●	●	
	A6				●	●	●
	A7					●	●
	A8						●

3.3 Atividades Realizadas

A1 - Foi concluída com sucesso, agregando conhecimento sobre o estado da arte no contexto de Redes Convolucionais para a classificação de lesões de pele . Vale destacar que nesse período, entre a A1 e a A3, foi realizada a Atividade 2 (A2). Isto é, esse trabalho foi apresentado na seção de pôster XXVIII Encontro de Iniciação Científica e XIII Encontro de Iniciação em Desenvolvimento Tecnológico e Inovação, no dia 24 de outubro de 2023.

- A3 - Envolvendo a obtenção e análise do conjunto de dados HAM-10000. Este dataset, composto por imagens de alta qualidade representando diferentes tipos de lesões cutâneas, foi fundamental para o desenvolvimento do projeto. Durante esta etapa, foi realizada uma avaliação detalhada das características das imagens, incluindo a diversidade das lesões de pele presentes.
- A4 - Focada na elaboração deste relatório parcial, foi finalizada com êxito.
- A5 - Implementação do classificador: Nesta etapa, tem como objetivo a implementação de um classificador com o auxílio da arquitetura GoogleNet e também as execuções dos experimentos;
- A6 - Avaliação quantitativa e qualitativa: Avaliação quantitativa e qualitativa serão

realizadas com o objetivo de avaliar a solução proposta deste plano de pesquisa. Dessa forma, a escolha das características a serem avaliadas serão definidas e, também examinadas quais delas podem influenciar no desempenho da solução;

- A7 - Escrita de artigo científico: Os resultados obtidos durante o período da Iniciação Científica serão divulgados em eventos científicos. As melhorias das soluções ocorrerão a partir de feedbacks das reuniões periódicas, comentários dos revisores e da própria soluções do estado-da-arte;
- A8 - Entrega do Relatório Final: Escrita do Relatório Final detalhando todas as atividades desenvolvidas no período de Iniciação Científica

4 Resultado Obtidos

Essa seção foi dividida em 5 subseções, análise dos dados contidos no dataset (Subseção 4.1), pré-processamento dos dados (Subseção 4.2), segmentação e experimentação nas imagens (Subseção 4.3), Implementação da GoogleNet (Subseção 4.4) e Resultados (Subseção 4.5).

4.1 Análise dos Dados Obtidos do Dataset

O conjunto de dados HAM-10000 (Human Against Machine with 10,000 images) constitui uma coleção abrangente de imagens dermatoscópicas, que são fundamentais para o estudo de lesões de pele. Esta seção resume as características principais do dataset, que foi analisado para embasar nosso projeto.

O dataset abrange uma diversidade de lesões cutâneas, incluindo categorias diagnosticadas como melanoma (mel), nevos melanocíticos (nv), lesões semelhantes a queratoses benignas (bkl), entre outras. Com aproximadamente 10.000 imagens, o dataset oferece uma amostra significativa para análise, possibilitando uma compreensão abrangente dos diferentes tipos de lesões de pele. Além das diferentes tipos de lesões o dataset também inclui uma série de dados do paciente, como sexo, idade e localidade da lesão.

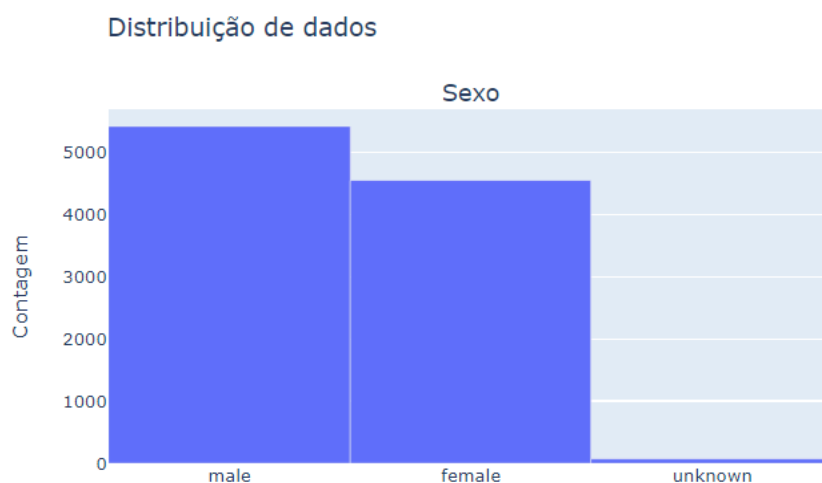


Figura 2: Distribuição de sexo das amostras

Conforme ilustrado na Figura 2, os dados revelam um discreto desbalanceamento de gênero, evidenciando um predomínio de amostras masculinas. Adicionalmente, a categoria "Unknown" sinaliza a presença de certos ruídos nos dados, indicando amostras cuja informação de gênero não está claramente definida ou é ausente, um fator a ser considerado na análise e no tratamento dos dados para garantir a precisão do modelo.

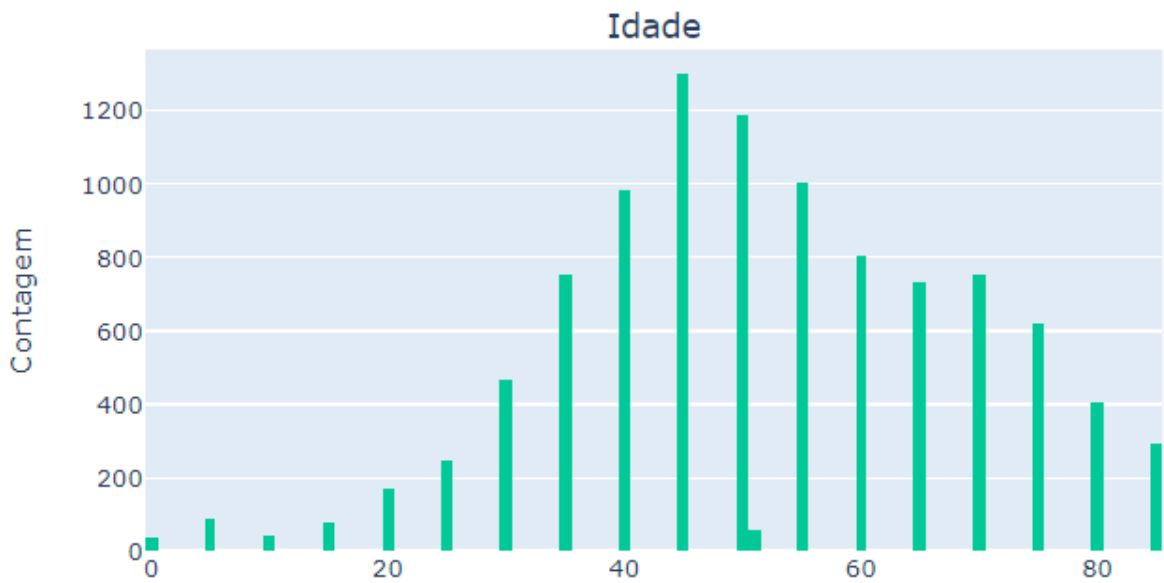


Figura 3: Distribuição de idade das amostras

A Figura 3 destaca a distribuição etária das amostras no conjunto de dados, revelando uma concentração significativa a partir dos trinta anos, com a moda situada em torno dos cinquenta anos. Tal distribuição etária, que inclui dados de um amplo espectro de idades, sublinha a robustez e a capacidade de generalização que este dataset pode conferir ao nosso modelo, refletindo uma diversidade que é crucial para o desenvolvimento de um sistema de diagnóstico abrangente e eficaz.

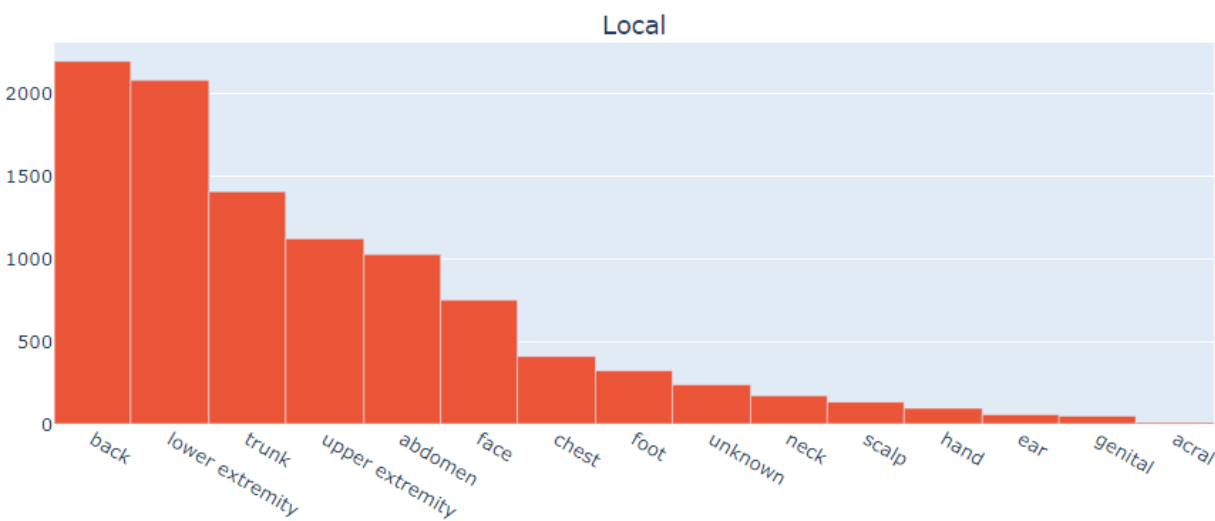


Figura 4: Distribuição da localidade das amostras

A Figura 4 apresenta a distribuição das lesões cutâneas conforme o local de incidência no corpo, abrangendo áreas diversas, que vão das costas até a região genital. Essa ampla dispersão

geográfica das lesões reforça a representatividade das amostras no conjunto de dados, evidenciando a sua abrangência e diversidade.

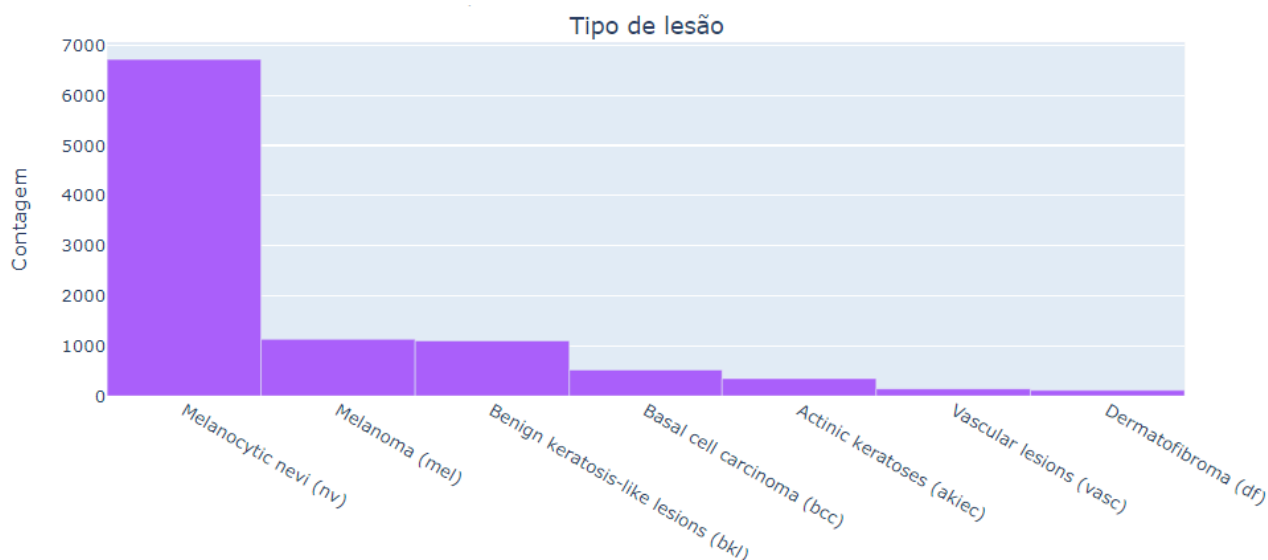


Figura 5: Distribuição dos tipos de lesões de pele das amostras

A Figura 5 ilustra a distribuição dos dados por tipo de lesão de pele no conjunto de dados, revelando a presença de 7 classes distintas, abrangendo lesões tanto benignas quanto malignas. O aspecto mais notável evidenciado pelo gráfico é a predominância de amostras classificadas como "Melanocytic nevi (nv)", conhecidas no Brasil como "marcas de nascença". Tal desbalanceamento acentuado é impróprio para a modelagem, exigindo medidas corretivas para assegurar a imparcialidade dos dados utilizados no treinamento do modelo, evitando assim o viés.

4.2 Pré-Processamento dos Dados

O processo de pré-processamento de dados desempenha um papel crucial na preparação para o treinamento eficiente de modelos de aprendizado de máquina. No contexto do nosso projeto, que envolve a classificação de lesões de pele, essa etapa é fundamental para assegurar que os dados sejam apresentados ao modelo de maneira equitativa e livre de vieses.

Uma das estratégias adotadas nessa fase consiste em simplificar a complexidade do problema de classificação, agrupando os diferentes tipos de lesões em duas categorias principais: benignas e malignas. Essa abordagem reduz o espaço de classificação de sete para apenas duas classes, o que é especialmente vantajoso para o treinamento do modelo. Ao limitar as opções de classificação, minimizamos a necessidade do modelo aprender a distinguir entre as nuances específicas de cada tipo de lesão, uma exigência que ultrapassa os objetivos do nosso projeto.

Concentrando-se em diferenciar lesões benignas de malignas, o modelo pode direcionar seu aprendizado para identificar os aspectos fundamentais que separam essas duas grandes categorias. Tal foco não apenas facilita o treinamento, como também potencializa a capacidade de generalização do modelo. Esse enfoque na generalização é essencial, pois permite ao modelo aplicar seu aprendizado a uma variedade mais ampla de casos, aumentando assim sua utilidade prática na identificação e diagnóstico de lesões de pele.

A divisão dos tipos está presente na Tabela 1 ilustrada abaixo:

Benignas	Malignas
Nevos Melanocíticos (nv)	Melanoma (mel)
Dermatofibroma (df)	Carcinoma Basocelular (bcc)
Lesões Semelhantes a Ceratose Benigna (bkl)	Ceratose Actínica (akiec)
Lesões Vasculares (vasc)	Lesões Vasculares (vasc)

Tabela 1: Resultados de diferentes filtros

A criação dessa tabela foi baseada no estudo “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions” publicado na plataforma Nature[14].

A classificação das lesões no dataset HAM-10000, inclui a categoria de Lesões Vasculares (vasc), que apresenta uma natureza ambígua podendo ser tanto benignas quanto malignas. Devido à falta de informações específicas no dataset que permitam distinguir claramente entre essas naturezas, optou-se por não considerar as Lesões Vasculares na análise.

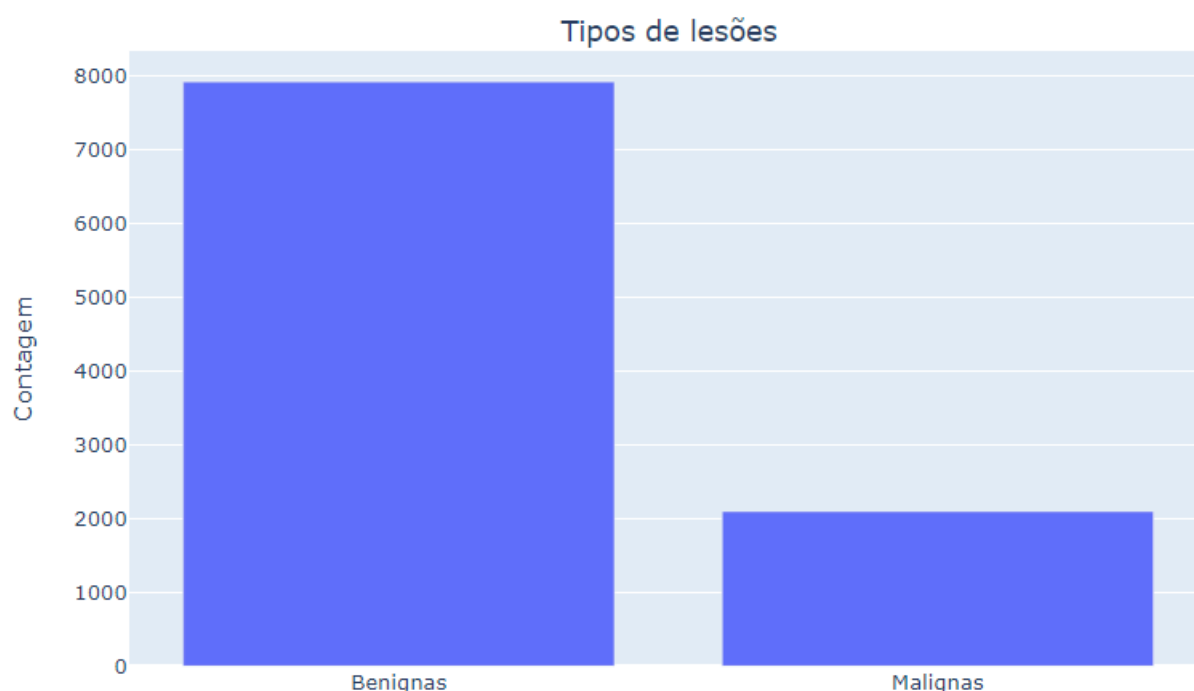


Figura 6: Distribuição binária entre lesões

A Figura 6 revela que, mesmo após a binarização das classes, o desbalanceamento entre as amostras persiste como um desafio. Para mitigar esse problema, adotou-se a técnica de *Data Augmentation*, gerando novas amostras artificialmente.

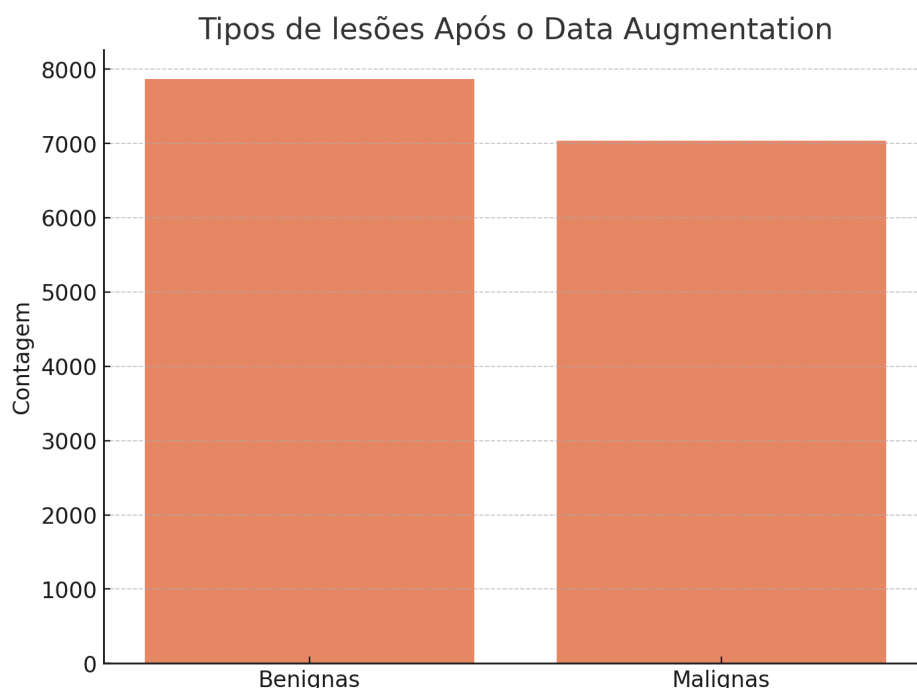


Figura 7: Resultado do Data Augmentation

A técnica de *Data Augmentation* foi aplicada apenas nas amostras malignas do conjunto de treinamento, com o intuito de balancear a proporção das classes. Essa abordagem tem duas principais vantagens: mitigar o viés de desbalanceamento de classe, que acontece quando o modelo tende a prever a classe majoritária com mais frequência, simplesmente porque foi treinado com mais exemplos dessa classe; e garantir maior robustez na sua classificação, visto que o modelo vai ser exposto a imagens de diferentes ângulos e rotações. O resultado da operação pode ser visto na Figura 7, com 7866 lesões benignas e 7040 lesões malignas.

4.3 Segmentação e experimentação nas imagens

As imagens foram segmentadas de duas formas distintas, denominadas aqui como Segmentação A e Segmentação B.

Na Segmentação A, as lesões foram recortadas exclusivamente na região de suas máscaras. Este processo envolveu isolar cada lesão, removendo completamente o tecido circundante e outras estruturas não relevantes. Como resultado, obtivemos imagens onde apenas as lesões estão presentes, permitindo uma análise focada e precisa das áreas afetadas. Esta abordagem pode ser observada na Figura 8, que ilustra claramente a precisão do recorte realizado na Segmentação A.



Figura 8: Segmentação A

Na Segmentação B, o recorte foi feito na região do bounding box da lesão. Diferentemente da Segmentação A, esse método resulta em imagens que contêm, além da lesão, uma porção do tecido circundante. Essa abordagem permite observar não apenas a lesão em si, mas também um pouco do contexto ao seu redor, o que pode ser útil para análises que dependem da interação da lesão com as estruturas próximas. Esta abordagem pode ser observada na Figura 9, que ilustra o recorte realizado na Segmentação B.

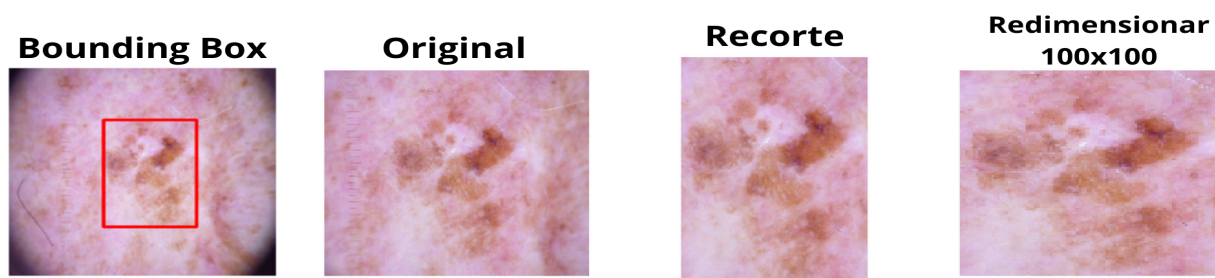


Figura 9: Segmentação B

Em seguida, aplicamos diferentes filtros de pré-processamento nas imagens com o objetivo de otimizar a performance do modelo. Para essa investigação, selecionamos cinco filtros distintos, cada um com propriedades únicas que poderiam realçar características importantes das imagens, facilitando o aprendizado do modelo. Os filtros escolhidos foram aplicados às imagens do nosso conjunto de dados, e os resultados dessa aplicação estão visualizados na Figura 10 abaixo.

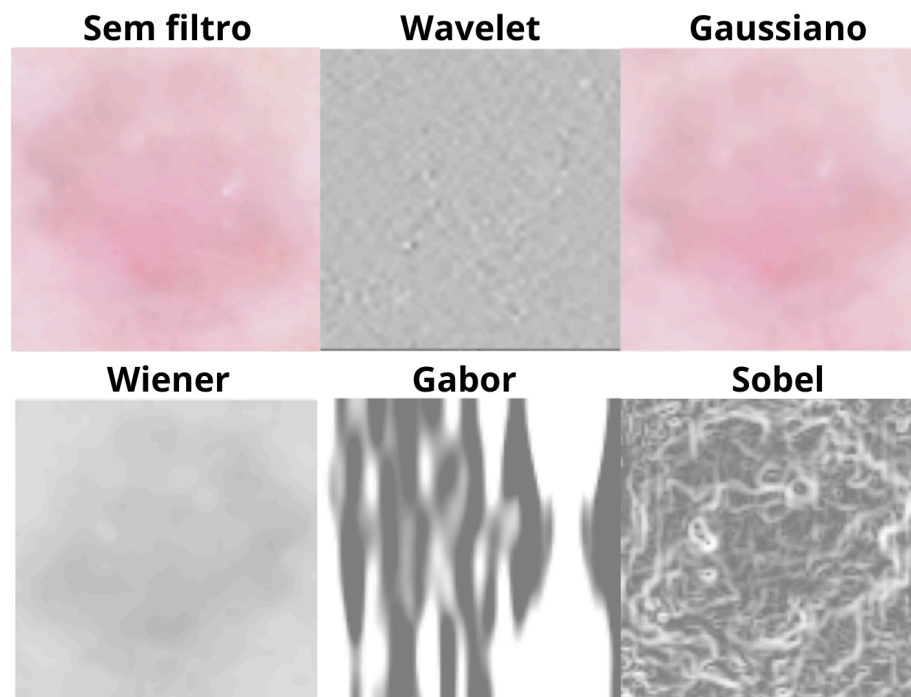


Figura 10: Filtros de pré-processamento

A seguir, apresentamos um resumo de cada filtro utilizado:

- **Wavelet:** Especialmente valioso para análise em diferentes escalas e direções, o filtro Wavelet preserva detalhes locais enquanto reduz ruídos. Esta característica o torna ideal para aplicações que exigem análise detalhada em múltiplas resoluções.
- **Gaussiano:** Reconhecido por sua capacidade de suavizar a imagem e reduzir ruído isotrópico, o filtro Gaussiano é particularmente útil em contextos onde a suavidade dos tons é preferível à nitidez dos detalhes, facilitando a modelagem de características gerais sem distrações de ruídos menores.
- **Wiener:** Utilizado para restaurar imagens degradadas por ruído ou desfoque, o filtro Wiener ajusta-se às variações locais da imagem, proporcionando uma recuperação mais precisa através da minimização do erro quadrático médio.
- **Gabor:** Este filtro é eficaz para detecção de bordas e texturas, aproveitando sua sensibilidade a orientações e frequências específicas. É amplamente usado em análise de texturas e segmentação de imagens, facilitando a identificação de padrões complexos.
- **Sobel:** Altamente eficaz na detecção de bordas verticais e horizontais, o filtro Sobel é crucial para aplicações que requerem alta precisão na identificação de bordas, como sistemas de visão computacional para navegação e reconhecimento de objetos.

A utilização desses diferentes filtros visa proporcionar uma comparação abrangente entre métodos que focam em realçar diferentes aspectos das imagens, permitindo ao modelo aprender características mais discriminativas e, assim, melhorar sua performance.

4.4 Implementação da GoogleNet

A arquitetura utilizada como base para a extração de features nesse modelo é a Inception V3,

atualização da arquitetura GoogLeNet. A GoogLeNet introduziu os módulos Inception em 2014, otimizando a eficiência das redes neurais sem aumentar o custo computacional. A Inception V2 melhorou isso com normalização de lotes e refinamento dos módulos Inception. A Inception V3 avançou ainda mais, utilizando fatorização para tornar as convoluções mais eficientes, melhorando a velocidade e a capacidade de aprendizado da rede. A evolução focou em aprimorar a profundidade e eficiência das redes sem aumentar a complexidade. O funcionamento do módulo Inception pode ser observado na Figura 11.

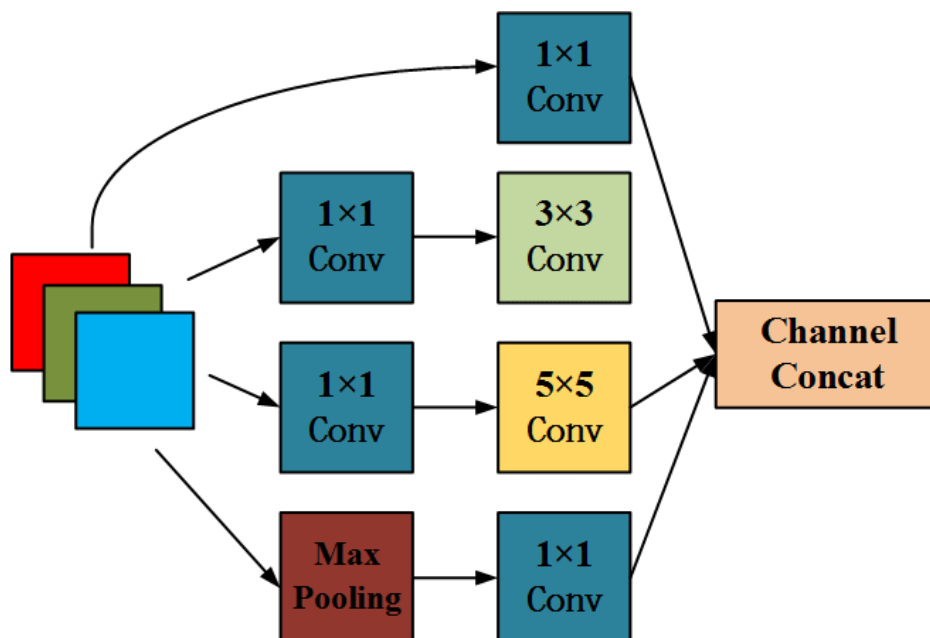


Figura 11: Ilustração da Inception

A principal vantagem do módulo Inception reside em sua habilidade de extrair características de variados tamanhos da imagem, identificando padrões que vão desde os mais discretos até os mais abrangentes. Essa capacidade única é viabilizada pela diversidade nos tamanhos dos filtros de convolução incorporados na arquitetura. Os filtros maiores capturam os padrões em larga escala, abordando as características macro da imagem, enquanto os filtros menores se concentram nos detalhes mais sutis, revelando os padrões micro. Tal estratégia assegura uma análise abrangente e detalhada da imagem.

Para interpretar os padrões capturados pela Inception V3, a arquitetura prossegue com a inclusão de uma camada de Max Pooling, que serve para reduzir a dimensionalidade dos dados preservando as características mais importantes. Seguindo essa etapa, a rede é estruturada com três camadas densas (fully connected), contendo 512, 256 e 128 neurônios, respectivamente. Essas camadas densas têm o papel de aprender padrões de alto nível a partir das características extraídas, funcionando como um mecanismo de decisão sobre os dados processados.

Para mitigar o risco de overfitting, que é a tendência do modelo de aprender demais a partir dos dados de treinamento a ponto de perder a generalização para novos dados, implementam-se taxas de dropout após cada uma dessas camadas densas. As taxas de 20%, 20% e 10% são aplicadas sucessivamente, o que ajuda a desativar aleatoriamente uma porção dos neurônios durante o treinamento, forçando a rede a aprender padrões mais robustos e generalizáveis.

A fase final da rede consiste em uma última camada densa que realiza a classificação binária, onde a decisão sobre o diagnóstico da lesão é tomada. Essa camada utiliza uma função

de ativação apropriada, como a sigmoid, para produzir a probabilidade da presença da lesão, permitindo assim uma interpretação clara e direta do resultado fornecido pela rede.

4.5 Resultados

Para avaliar o modelo, utilizamos quatro métricas distintas.

- **Precisão:** Esta métrica mede a proporção de previsões positivas corretas em relação ao total de casos classificados como positivos, permitindo avaliar a exatidão das previsões positivas.
- **Recall (Sensibilidade):** Avalia a capacidade do modelo de identificar todos os casos positivos reais, sendo crucial para garantir que todos os casos de interesse sejam detectados.
- **F1-Score:** Representa a média harmônica entre precisão e recall, e é especialmente importante em contextos médicos, pois equilibra os dois tipos de erros, fornecendo uma visão mais completa da performance do modelo.
- **Acurácia:** Indica a proporção total de previsões corretas, abrangendo tanto os casos positivos quanto os negativos, oferecendo uma visão geral da eficácia do modelo em prever corretamente ambos os tipos de casos.

Além disso, os resultados foram obtidos utilizando a técnica de Validação Cruzada de 3 Folds, que é empregada para reduzir o viés associado ao dataset. Essa metodologia divide o conjunto de dados em três partes iguais, utilizando cada uma delas, em turnos diferentes, como conjunto de teste, enquanto as demais servem como treinamento. Isso assegura que todos os dados são usados tanto para treinar quanto para testar o modelo, aumentando a generalização dos resultados.

Para mitigar o viés de inicialização dos pesos, o melhor modelo identificado na validação cruzada foi executado 10 vezes. Esse procedimento garante que os resultados não dependam de uma configuração específica de pesos iniciais, reforçando a robustez do modelo.

A Tabela 2 demonstra os resultados do modelo treinado utilizando as imagens originais e as segmentações A e B. Esta tabela apresenta uma comparação detalhada do desempenho do modelo em diferentes condições de segmentação, fornecendo uma visão clara sobre como cada método afeta a eficácia do modelo.

Accuracy	Precision	Recall	F1_Score	Tipo
0.73 +- 0.01	0.66 +- 0.00	0.72 +- 0.00	0.66 +- 0.00	Inteira
0.70 +- 0.01	0.64 +- 0.00	0.70 +- 0.01	0.64 +- 0.01	Segmentação A
0.74 +- 0.01	0.67 +- 0.00	0.73 +- 0.00	0.68 +- 0.01	Segmentação B

Tabela 2: Resultados de diferentes segmentações

Como pode ser observado, o fato da Segmentação B ser um meio-termo entre a imagem

inteira e a Segmentação A se reflete nas métricas do modelo. Por combinar o contexto com a redução de ruídos, a Segmentação B une o melhor dos dois mundos, proporcionando melhor desempenho.

A seguir, a Tabela 3 demonstra o resultado dos cinco filtros de pré-processamento, onde não foi praticado o protocolo de validação cruzada e 10 execuções, tratando-se de apenas 5 execuções em um treinamento convencional. Como podemos observar, nenhum filtro conseguiu melhorar o resultado comparado ao modelo sem filtro.

Accuracy	Precision	Recall	F1_Score	Filtro
0.70 +- 0.01	0.63 +- 0.00	0.68 +-0.01	0.64 +- 0.00	Sobel
0.68 +- 0.02	0.63 +- 0.01	0.68 +- 0.01	0.63 +- 0.01	Gabor
0.72 +- 0.02	0.65 +- 0.00	0.71 +- 0.00	0.66 +- 0.01	Wiener
0.50 +- 0.02	0.53 +- 0.00	0.54 +- 0.00	0.47 +- 0.01	Wavelet
0.72 +- 0.01	0.66 +- 0.00	0.73 +- 0.00	0.67 +- 0.00	Gaussiano
0.74 +- 0.01	0.67 +- 0.01	0.74 +- 0.00	0.68 +- 0.01	Sem filtro

Tabela 3: Resultados de diferentes filtros

Por fim, na Tabela 4, utilizamos os dois melhores resultados — Sem Filtro e Gaussiano — para o protocolo original de treinamento, com validação cruzada e 10 execuções. Além disso, foi feita uma alteração no modelo, desbloqueando parcialmente algumas camadas da InceptionV3 que originalmente estavam congeladas. Essa estratégia, conhecida como Fine Tuning, visa aprimorar a performance do modelo ajustando mais finamente os pesos das camadas pré-treinadas.

Accuracy	Precision	Recall	F1_Score	Filtro
0.74 +- 0.01	0.67 +- 0.00	0.73 +- 0.00	0.68 +- 0.00	B - Congelada
0.77 +- 0.01	0.70 +- 0.01	0.76 +- 0.00	0.71 +- 0.01	B - Fine tuning
0.74 +- 0.01	0.67 +- 0.00	0.73 +- 0.00	0.68 +- 0.00	B(Gauss) - Congelada
0.77 +- 0.01	0.70 +- 0.00	0.76 +- 0.01	0.71 +- 0.01	B(Gauss) - Fine tuning

Tabela 4: Resultados de diferentes filtros

Como pode ser observado, após refazer o teste com os dois tipos de imagem, a discrepância presente anteriormente não se mostrou presente, com os resultados se mostrando quase idênticos. O uso de Fine Tuning se mostrou extremamente eficaz, sendo o responsável pelo melhor resultado de F1-Score obtido, 71%.

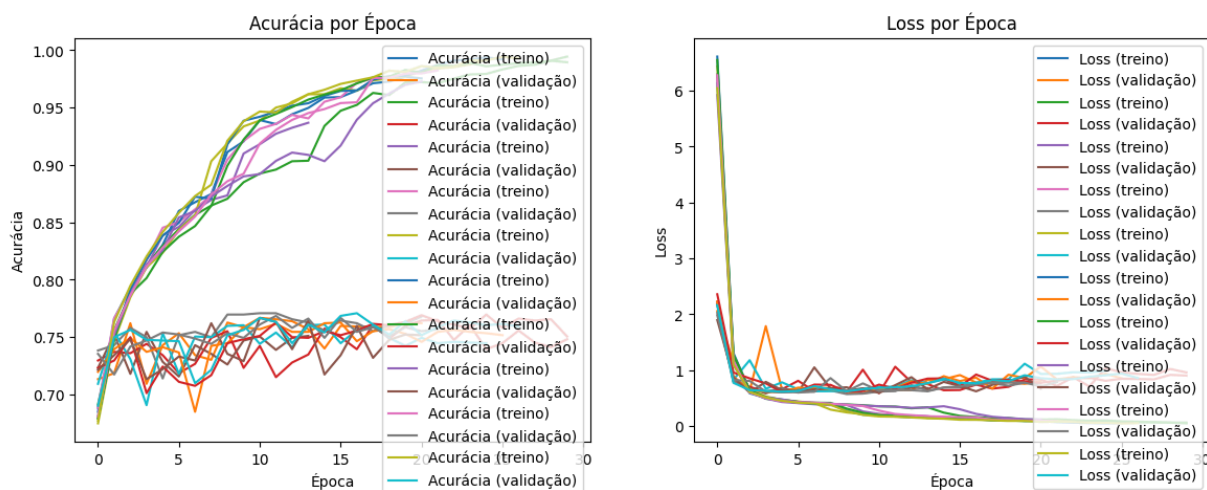


Figura 12: Ilustração da Inception

A Figura 12 apresenta os gráficos de acurácia por época e de loss por época para o modelo utilizando segmentação B com Fine Tuning. Observa-se um claro indício de overfitting, uma vez que, no primeiro gráfico, a acurácia no conjunto de treino é significativamente maior e mais constante do que no conjunto de validação. Além disso, o gráfico de loss por época mostra que, enquanto o loss no conjunto de treino diminui de forma suave e contínua, o loss no conjunto de validação é inconstante e, em alguns casos, chega a aumentar.

5 Discussão

O desempenho do modelo implementado utilizando a arquitetura GoogLeNet demonstrou resultados promissores na detecção de lesões malignas em imagens dermatoscópicas. No entanto, diversos aspectos podem ser explorados para melhorar ainda mais a eficiência e a precisão do modelo.

Em primeiro lugar, a técnica de fine-tuning aplicada às camadas da InceptionV3 provou ser eficaz, mas a presença de overfitting sugere que ajustes adicionais são necessários para melhorar a generalização do modelo. Uma abordagem para mitigar o overfitting poderia envolver o uso de técnicas mais avançadas de regularização, como a introdução de dropout nas camadas convolucionais ou o uso de data augmentation ainda mais diversificada durante o treinamento.

Outro ponto relevante é a comparação entre os diferentes filtros de pré-processamento. Embora os filtros aplicados não tenham superado o modelo sem filtros, o filtro Gaussiano apresentou resultados que se aproximaram dos melhores resultados obtidos. Isso sugere que um estudo mais aprofundado sobre a combinação de filtros ou a aplicação de técnicas de realce de contraste, como o CLAHE (Contrast Limited Adaptive Histogram Equalization), pode revelar melhorias significativas.

Além disso, os resultados sugerem que a segmentação B, que inclui contexto adicional ao redor das lesões, contribui para um melhor desempenho do modelo. Isso indica que o contexto visual das lesões pode ser crucial para a classificação precisa e sugere a exploração de redes neurais mais complexas que possam captar interações contextuais de forma mais eficaz.

Por fim, a implementação de um ensemble de modelos poderia ser uma abordagem promissora para melhorar a robustez e a precisão do diagnóstico. Ao combinar os resultados de múltiplos modelos treinados com diferentes arquiteturas ou pré-processamentos, é possível

mitigar as fraquezas de um único modelo e potencializar os pontos fortes de cada abordagem.

Esses pontos discutidos abrem caminhos para futuras pesquisas e desenvolvimento na área de diagnóstico automatizado de lesões de pele, com o objetivo de criar modelos mais precisos e confiáveis, que possam ser integrados em sistemas de apoio ao diagnóstico em ambientes clínicos.

6 Conclusão

Neste trabalho, foi implementado e avaliado um modelo de rede neural convolucional utilizando a arquitetura GoogLeNet, com o objetivo de diagnosticar lesões de pele a partir de imagens dermatoscópicas. A abordagem escolhida, incluindo a utilização de técnicas de fine-tuning e a aplicação de diferentes métodos de segmentação e pré-processamento, proporcionou resultados relevantes, demonstrando o potencial do modelo em classificar corretamente lesões malignas e benignas.

Os experimentos realizados indicaram que a segmentação B, que mantém parte do contexto ao redor das lesões, resultou em um melhor desempenho do modelo em comparação com a segmentação que isola totalmente a lesão. Além disso, a aplicação do fine-tuning sobre a InceptionV3 mostrou-se eficaz, embora o overfitting tenha sido um desafio, evidenciando a necessidade de técnicas adicionais de regularização.

Apesar de os filtros de pré-processamento testados não terem superado o desempenho do modelo sem filtros, o filtro Gaussiano apresentou resultados promissores. Esses achados sugerem que o modelo pode ser aprimorado por meio de abordagens mais sofisticadas de realce de contraste e processamento de imagem.

A conclusão mais significativa deste trabalho é que, embora o modelo tenha atingido um bom nível de acurácia e precisão, há espaço para melhorias que podem ser exploradas em trabalhos futuros. A implementação de técnicas como ensembles de modelos e o uso de dados adicionais sobre as características das lesões podem aumentar ainda mais a robustez e a eficácia do diagnóstico automatizado.

Finalmente, este estudo contribui para a literatura de diagnósticos médicos automatizados, oferecendo uma base sólida para futuros desenvolvimentos e pesquisas na utilização de redes neurais convolucionais no campo da dermatologia.

Referências

- [1] “Inca estima 704 mil casos de câncer por ano no brasil até 2025.” <https://www.gov.br/inca/pt-br/assuntos/noticias/2022/inca-estima-704-mil-casos-de-cancer-por-ano-no-brasil-ate-2025>. Acesso: 25/01/2023.
- [2] “Estimativa 2023 – incidência de câncer no brasil.” <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa>. Acesso: 25/01/2023.
- [3] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [4] S. Jain, N. Pise, *et al.*, “Computer aided melanoma skin cancer detection using image processing,” *Procedia Computer Science*, vol. 48, pp. 735–740, 2015.
- [5] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, “Diagnostic accuracy of dermoscopy,” *The lancet oncology*, vol. 3, no. 3, pp. 159–165, 2002.
- [6] M. Vestergaard, P. Macaskill, P. Holt, and S. Menzies, “Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting,” *British Journal of Dermatology*, vol. 159, no. 3, pp. 669–676, 2008.
- [7] E. Y. A. Rodríguez, “Técnicas de aprendizado de máquina para predição do custo da logística de transporte: uma aplicação em empresa do segmento de autopeças,” 2020.
- [8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [9] A. Zhang, K. C. Wang, Y. Fei, Y. Liu, S. Tao, C. Chen, J. Q. Li, and B. Li, “Deep learning-based fully automated pavement crack detection on 3d asphalt surfaces with an improved cracknet,” *Journal of Computing in Civil Engineering*, vol. 32, no. 5, p. 04018041, 2018.
- [10] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [11] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Back-propagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [12] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [13] J. Son, J. Y. Shin, H. D. Kim, K.-H. Jung, K. H. Park, and S. J. Park, “Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images,” *Ophthalmology*, vol. 127, no. 1, pp. 85–94, 2020.
- [14] [The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions | Scientific Data \(nature.co\)](#)