

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
PROJETO DE GRADUAÇÃO



VINÍCIUS BREDÁ ALTOÉ

APLICAÇÃO DE ÍNDICE DE MARROM,  
PROGRAMAÇÃO GENÉTICA E VISÃO  
COMPUTACIONAL PARA DETECÇÃO DE CÂNCER  
DE PELE

VITÓRIA-ES

MARÇO/2025

Vinícius Breda Altoé

**APLICAÇÃO DE ÍNDICE DE MARROM,  
PROGRAMAÇÃO GENÉTICA E VISÃO  
COMPUTACIONAL PARA DETECÇÃO DE CÂNCER  
DE PELE**

Parte manuscrita do Projeto de Graduação do aluno Vinícius Breda Altoé, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Vitória-ES

Março/2025

Vinícius Breda Altoé

# **APLICAÇÃO DE ÍNDICE DE MARROM, PROGRAMAÇÃO GENÉTICA E VISÃO COMPUTACIONAL PARA DETECÇÃO DE CÂNCER DE PELE**

Parte manuscrita do Projeto de Graduação do aluno Vinícius Breda Altoé, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Aprovado em 19 de março de 2025.

**COMISSÃO                      EXAMINADORA:**

---

**Prof. Dr. Patrick Marques Ciarelli**  
Universidade Federal do Espírito Santo  
Orientador

---

**Prof. Dr. Jacques Facon**  
Universidade Federal do Espírito Santo  
Examinador

---

**Msc. Edmilson Queiroz dos Santos**  
**Filho**  
Examinador

Vitória-ES

Março/2025

## AGRADECIMENTOS

Primeiramente, agradeço a Deus por me acompanhar durante toda a minha vida e por me educar sobre o que é importante.

Aos meus pais, Viviane e André, pelo incentivo e amor incondicional ao longo dos anos, e à minha família, por todo o apoio e por sempre me ajudar na realização dos meus sonhos, especialmente aos meus irmãos Eduardo e Mateus.

À minha namorada, Laysa Madeira, e à sua família, pelo apoio e incentivo à minha profissionalização e realização pessoal.

Ao meu orientador, Prof. Dr. Patrick Marques Ciarelli, agradeço pela sua disponibilidade, orientação e dedicação ao desenvolvimento deste trabalho.

Aos meus amigos de faculdade, que estiveram ao meu lado nos momentos difíceis e com quem compartilhei muitos risos e momentos descontraídos ao longo dessa jornada. Em especial, agradeço aos amigos Arthur Macedo, Catarina Sastre, Gabriel Boasquives e Gabriel Falcone.

À banca examinadora, pela aceitação do convite e pelo tempo investido na leitura e avaliação deste trabalho.

Finalmente, agradeço à Universidade Federal do Espírito Santo pela minha formação.

## RESUMO

O câncer de pele é o tipo mais comum de câncer no Brasil, representando mais de 30% dos casos, com o melanoma sendo a forma mais letal. A detecção precoce é essencial, pois aumenta consideravelmente as chances de sobrevivência dos pacientes. No entanto, devido à semelhança com lesões comuns, o diagnóstico é desafiador, até mesmo para especialistas. Diante disso, há uma crescente demanda por ferramentas de diagnóstico automatizado. Embora as soluções atuais baseadas em redes neurais profundas sejam precisas, elas enfrentam desafios relacionados à complexidade e ao custo de processamento. O uso de índices de marrom, tradicionalmente utilizados na indústria alimentícia, surge como uma abordagem inovadora na detecção de câncer de pele, mas ainda não foi amplamente explorado nessa área. Este estudo avaliou a aplicação desses índices combinados com redes neurais profundas para a classificação binária de imagens de melanoma. Além disso, um algoritmo de programação genética foi utilizado para otimizar os índices, buscando melhorar a acurácia e eficácia na classificação. Os testes, realizados com os conjuntos de dados ISIC 2019 e 2020, mostraram que os índices otimizados por programação genética resultaram em melhorias na acurácia e no F1 Score, em comparação ao uso de redes neurais profundas com imagens no espaço de cor RGB. Em particular, os índices simples otimizados aumentaram o F1 Score em 0,81% na rede MobileNetV2, com uma precisão de 95,77% e sensibilidade de 90,12%. O modelo NasNetMobile obteve uma precisão de 95,39% e sensibilidade de 89,99%, mostrando que os índices de marrom têm potencial para melhorar a classificação de melanoma. Esses resultados destacam o potencial dos índices de marrom como uma ferramenta promissora para a detecção de melanoma, com aplicações em sistemas de diagnóstico assistido por computador.

**Palavras-chave:** Câncer de pele; Aprendizado profundo; Índice de marrom; Programação Genética.

## ABSTRACT

Skin cancer is the most common type of cancer in Brazil, accounting for more than 30% of all cancer cases, with melanoma being the most lethal form. Early detection is essential, as it significantly increases patients' chances of survival. However, due to its similarity to common lesions, diagnosis is challenging, even for specialists. As a result, there is a growing demand for automated diagnostic tools. Although current solutions based on deep neural networks are accurate, they face challenges related to complexity and processing costs. The use of browning indices, traditionally employed in the food industry, emerges as an innovative approach in skin cancer detection but has not been widely explored in this field. This study evaluated the application of these indices combined with deep neural networks for binary melanoma image classification. Furthermore, a genetic programming algorithm was used to optimize the indices, aiming to improve accuracy and effectiveness in classification. The tests, conducted using the ISIC 2019 and 2020 datasets, showed that indices optimized by genetic programming resulted in improvements in accuracy and F1 Score, compared to using deep neural networks with RGB color space images. Specifically, the optimized simple indices increased the F1 Score by 0.81% in the MobileNetV2 model, with a precision of 95.77% and a recall of 90.12%. The NasNetMobile model achieved a precision of 95.39% and a recall of 89.99%, demonstrating that browning indices have the potential to improve melanoma classification. These results highlight the potential of browning indices as a promising tool for melanoma detection, with applications in computer-aided diagnostic systems.

**Keywords:** Skin cancer; Deep learning; Browning indices; Genetic Programming.

## LISTA DE FIGURAS

Figura 1 – Estrutura básica de um <i>perceptron</i> . . . . .	15
Figura 2 – Arquitetura de uma Rede Neural Multicamadas. . . . .	16
Figura 3 – LeNet-5: Um dos primeiros modelos de redes neurais convolucionais. . .	18
Figura 4 – Exemplo de indivíduo representado por árvore de sintaxe. . . . .	22
Figura 5 – Exemplo de indivíduo representado por árvore de sintaxe para a expressão $[(2 \times (R + G)) - B]^2$ . . . . .	22
Figura 6 – Exemplo de indivíduo representado por árvore de sintaxe para a expressão $\sqrt{(1 - R) \times (R + G)}$ . . . . .	23
Figura 7 – Matriz de confusão. . . . .	25
Figura 8 – Fluxograma da metodologia proposta. . . . .	29
Figura 9 – Imagens com melanoma à esquerda e outras lesões de pele, não melanoma, à direita. . . . .	35
Figura 10 – Imagens de melanoma do banco de imagens ISIC 2016 e suas segmentações feitas por profissionais médicos. . . . .	35
Figura 11 – Exemplo de imagens com adesivos e com sombras provocadas por equipamentos dermatológicos que foram retirados do treinamento. . . .	36
Figura 12 – Segmentação obtida utilizando cada índice. . . . .	43
Figura 13 – Segmentação obtida utilizando cada índice. . . . .	44
Figura 14 – Imagens originais e as mesmas com os índices aplicados nas camadas <i>RGB</i> . . . . .	45

## LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
RNA	Rede Neural Artificial
CNN	<i>Convolutional Neural Network</i>
DL	<i>Deep Learning</i>
GPU	<i>Graphics Processing Unit</i>
IC	Índice de Cor
IF	<i>Isolation Forest</i>
IM	Índice de Marrom
PG	Programação Genética
SVM	<i>Support Vector Machine</i>
UFES	Universidade Federal do Espírito Santo



# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>1.1</b>	<b>Objetivos</b>	<b>12</b>
1.1.1	Objetivo Geral	12
1.1.2	Objetivos Específicos	13
<b>1.2</b>	<b>Estrutura do Texto</b>	<b>13</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>14</b>
<b>2.1</b>	<b>Aprendizado de Máquina</b>	<b>14</b>
2.1.1	Redes Neurais	15
2.1.2	Redes CNNs Pré-treinadas	18
2.1.2.1	<i>Transfer Learning</i>	19
2.1.2.2	MobileNetV2	19
2.1.2.3	ResNet50	19
2.1.2.4	NasNetMobile	20
2.1.2.5	Comparação das Arquiteturas	20
2.1.3	Programação Genética	21
<b>2.2</b>	<b>Métricas de Avaliação</b>	<b>24</b>
2.2.1	Matriz de Confusão	24
2.2.2	Acurácia	25
2.2.3	Precisão	25
2.2.4	Sensibilidade	26
2.2.5	<i>F1 Score</i>	26
<b>2.3</b>	<b>Índices de Cor (IC) e Índice de Marrom (IM)</b>	<b>26</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>29</b>
<b>3.1</b>	<b>Descoberta de Novos Índices por Programação Genética</b>	<b>30</b>
<b>3.2</b>	<b>Conjuntos de Dados</b>	<b>33</b>
<b>3.3</b>	<b>Pré-processamento das Imagens</b>	<b>34</b>
<b>3.4</b>	<b>Avaliação de Redes Pré-Treinadas</b>	<b>36</b>
<b>4</b>	<b>RESULTADOS</b>	<b>40</b>
<b>4.1</b>	<b>Índices Obtidos pela Programação Genética</b>	<b>40</b>
4.1.1	Índices com Operações Simples	40
4.1.2	Índices com Maior Nível de Complexidade	41
4.1.3	Resultados da Segmentação	41
4.1.4	F1 Score dos Índices Encontrados por Programação Genética	41

4.2	<b>Resultados da Classificação</b>	43
4.2.1	Resultados com as Redes Pré-Treinadas e Imagens Originais	46
4.2.2	Resultados com os Índices de Marrom da Literatura	46
4.2.3	Resultados com os Índices Obtidos	48
4.3	<b>Discussão dos Resultados</b>	49
5	<b>CONCLUSÃO E TRABALHOS FUTUROS</b>	51
5.1	<b>Conclusões</b>	51
5.2	<b>Trabalhos Futuros</b>	52
5.2.1	Métodos para Evitar o <i>Overfitting</i> da Programação Genética	52
5.2.2	Aprimoramento da Programação Genética	52
5.2.3	Utilização do Índice de Jacard como Alternativa ao F1 Score	53
5.2.4	Exploração de Novos Índices de Cor	53
5.2.5	Aprimoramento da Aplicação dos Índices	53
5.2.6	Avaliação de Desempenho em Outros Conjuntos de Dados	53
5.2.7	Aprimoramento das Arquiteturas de Redes Neurais	54
5.2.8	Implementação em Ambiente Clínico	54
5.2.9	Utilização de Bancos de Dados para Comparação em Trabalhos Futuros	54
	<b>REFERÊNCIAS</b>	56

# 1 INTRODUÇÃO

O câncer de pele é o tumor maligno mais comum no Brasil, representando expressivos 31,3% do total de casos de câncer no país, de acordo com dados do Instituto Nacional do Câncer (INCA) (INCA, 2022). O melanoma, uma forma particularmente letal do câncer de pele, compõe uma parcela desses casos e, globalmente, é reconhecido como a forma mais mortal de câncer de pele devido à sua alta capacidade de causar metástases, sendo responsável por aproximadamente 75% das mortes associadas ao câncer de pele (LI; SHEN, 2018). Isso destaca a importância crucial da detecção precoce para aumentar as chances de sobrevivência. O desafio é ainda maior devido à semelhança marcante entre lesões cancerígenas e condições dermatológicas comuns, o que torna o diagnóstico difícil, até mesmo para especialistas. Segundo a Organização Mundial da Saúde (OMS), a detecção precoce do melanoma é essencial, pois pode melhorar significativamente os resultados do tratamento e as taxas de sobrevivência (SAÚDE, 2022).

A urgência na detecção rápida é ainda mais evidente quando são consideradas as taxas de sobrevivência associadas ao melanoma cutâneo. Em estágios iniciais, as chances de sobrevivência atingem 99%, destacando a eficácia do diagnóstico precoce. No entanto, essas taxas decaem drasticamente para 14% em estágios mais avançados (ESTEVA, 2017), sublinhando a necessidade crítica de estratégias de detecções ágeis para otimizar o tratamento e a sobrevida dos pacientes.

Diante desse cenário, a busca por soluções inovadoras e automatizadas de diagnóstico torna-se imperativa. A detecção de anomalias em dados médicos envolve a aplicação de diversos métodos de aprendizado de máquina, cada um com suas características distintas. As Redes Neurais Convolucionais (do inglês *Convolutional Neural Networks* - CNNs) são amplamente utilizadas em imagens dermatoscópicas, oferecendo a capacidade de identificar padrões visuais complexos associados a lesões cutâneas. A flexibilidade das CNNs as torna eficazes na extração de recursos relevantes, contribuindo para a precisão na classificação de imagens (OMARI, 2022).

Shah et al. (2023) exploram duas abordagens automatizadas, Redes Neurais Artificiais (RNAs) e Redes Neurais Convolucionais (CNNs), que empregam técnicas de aprendizado profundo na detecção de câncer de pele. Os resultados indicam que ambas as abordagens demonstraram eficácia na detecção precoce de câncer de pele, utilizando conjuntos de dados diversos e modelos híbridos. O artigo enfatiza a inovação trazida pelo uso de técnicas de aprendizado profundo na detecção de câncer de pele e destaca a urgência de sistemas automatizados para o reconhecimento de lesões cutâneas. Da mesma forma, os estudos de Smith, Lee e Zhang (2021), que aplicaram CNNs na análise de imagens dermatoscópicas,

e de Li, Wang e Zhou (2022), que exploraram modelos híbridos combinando aprendizado de máquina e imagens médicas, enfatizam o crescente papel da IA neste campo. Além disso, o trabalho recente de Zhang, Xu e Li (2023) demonstrou o potencial do aprendizado por transferência para a detecção de melanoma, mostrando resultados promissores com modelos pré-treinados e conjuntos de dados menores. Juntos, esses estudos ressaltam o potencial dos sistemas baseados em IA para melhorar a precisão diagnóstica e reduzir a sobrecarga dos profissionais de saúde.

A utilização do aprendizado profundo na detecção automática de câncer de pele oferece diversas vantagens fundamentais. Essa abordagem proporciona uma análise mais rápida e precisa de grandes conjuntos de dados, superando as limitações dos métodos tradicionais de aprendizado de máquina. Embora desafios como desequilíbrio de dados, limitações de conjuntos de dados e adaptação de domínio ainda persistam, as técnicas de aprendizado profundo demonstram ser soluções promissoras para a classificação eficaz do câncer de pele, como observado em (WU et al., 2022).

A seleção da metodologia apropriada depende das características específicas do problema e da disponibilidade de dados rotulados. A validação clínica e a colaboração com especialistas médicos, a partir de uma boa rotulação de imagens, são fundamentais para garantir que os resultados sejam interpretáveis e clinicamente relevantes, promovendo avanços significativos na detecção precoce e diagnóstico preciso do câncer de pele.

Embora existam abordagens precisas disponíveis atualmente, como aquelas baseadas em redes neurais profundas, essas enfrentam obstáculos consideráveis em termos de complexidade e custo de processamento. Assim, continua-se a busca por métodos que atendam a um nível alto de acurácia a partir de um menor custo computacional. Nesse contexto, surge a oportunidade de explorar abordagens inovadoras, que utilizam diferentes métodos para segmentação e classificação das lesões de pele.

O índice de marrom, tradicionalmente utilizado na indústria alimentícia, tem como principal aplicação a avaliação do processo de caramelização e o monitoramento da qualidade de alimentos, como pães, carnes e produtos assados. Esse índice é um parâmetro importante na determinação da cor e qualidade desses produtos, sendo útil para garantir padrões de produção e otimizar processos industriais. Através da análise das mudanças nas cores durante o aquecimento, o índice de marrom fornece uma maneira rápida e eficiente de avaliar a quantidade de compostos responsáveis pela alteração da cor dos alimentos (SILVA; SOUZA; COSTA, 2020). Neste contexto, este trabalho visa avaliar a viabilidade e eficácia do uso de índices combinados com redes profundas como uma ferramenta de diagnóstico de câncer de pele.

Este índice ainda não foi bem explorado para a segmentação de lesões de pele, e há uma escassez de estudos que estabeleçam uma relação direta entre o índice de marrom e o tema em questão. Contudo, os resultados podem ser promissores, considerando que a pele humana e suas possíveis lesões podem atingir diferentes tons de marrom, como já demonstrado por Filho et al. (2023) em sua pesquisa. Filho et al. (2023) utilizaram o índice de marrom para segmentar lesões de melanoma em imagens dermatoscópicas, evidenciando que esse índice pode realçar as variações de cor que ajudam na identificação das lesões. O estudo, ao combinar os índices de marrom Aimonino, Fetuga e Lunadei com a rede neural U-Net, obteve resultados promissores, com melhorias significativas nas métricas de segmentação, como o F1 Score e o Índice de Jaccard, quando comparado ao uso convencional de imagens RGB. Embora a aplicação de índices de marrom na segmentação de lesões de melanoma ainda seja um campo emergente, os resultados de (FILHO et al., 2023) sugerem que essa abordagem pode oferecer uma solução eficaz para a detecção precoce de câncer de pele. O foco deste trabalho está na adaptação e otimização desses índices por meio de testes extensivos em um sistema de visão computacional, com o intuito de avaliar e identificar a efetividade de tal abordagem na área de detecção de câncer.

Além disso, a Programação Genética surge como uma técnica poderosa para otimizar índices e adaptá-los a diferentes características das imagens dermatológicas. Ao empregar algoritmos evolutivos, a programação genética pode identificar as melhores combinações de operações matemáticas para gerar índices mais eficazes na segmentação de lesões, ajustando automaticamente os parâmetros para maximizar a precisão do diagnóstico. A utilização de programação genética tem se mostrado promissora em diversas áreas da computação, incluindo o processamento de imagens e a análise de dados médicos, onde sua capacidade de explorar grandes espaços de soluções e otimizar funções complexas é uma vantagem significativa. Estudos como o de Koza (1992), que introduziu a programação genética, e trabalhos mais recentes, como os de Ferreira, Silva e Souza (2018) e outros estudos na área de aprendizado de máquina, demonstram como a programação genética pode ser eficaz na criação de modelos para tarefas de diagnóstico médico, como a detecção de melanoma.

## 1.1 Objetivos

### 1.1.1 Objetivo Geral

Este trabalho tem como objetivo geral avaliar a viabilidade e eficácia do uso de índices de marrom, amplamente empregados na indústria alimentícia, combinados com redes neurais

profundas e programação genética, como uma abordagem inovadora para a classificação e segmentação de imagens de melanomas. A programação genética será utilizada para otimizar os índices de marrom, visando melhorar a segmentação das lesões de pele, que servirá como uma fonte de informação para a tarefa de classificação. Embora existam muitos outros tipos de índices na literatura, o foco deste trabalho é explorar especificamente os índices de marrom e aqueles gerados por programação genética, dada sua aplicabilidade e potencial na detecção de melanoma.

### 1.1.2 Objetivos Específicos

- Realizar testes nos conjuntos de dados dos desafios ISIC (2019) e ISIC (2020) para avaliar o desempenho do uso dos índices de marrom (Aimonino, Lunadei e Fetuga) na tarefa de segmentação e classificação de lesões de pele.
- Utilizar programação genética para encontrar um índice de marrom otimizado que possa fornecer melhores resultados na segmentação das lesões de pele, visando aprimorar a classificação de lesões malignas e benignas.

## 1.2 Estrutura do Texto

A estrutura deste trabalho está organizada da seguinte maneira:

- **Introdução:** neste capítulo inicial é contextualizado o tema do estudo, assim como a motivação e os objetivos a serem alcançados;
- **Referencial teórico:** este capítulo aborda os conceitos teóricos fundamentais para a compreensão do trabalho;
- **Metodologia:** neste capítulo é realizada a descrição das etapas do desenvolvimento do projeto, de forma a detalhar os recursos utilizados para atingir os objetivos estabelecidos;
- **Resultados:** neste capítulo são apresentados os resultados obtidos a partir da metodologia adotada;
- **Conclusão e Projetos Futuros:** este capítulo apresenta a conclusão do trabalho, a partir da discussão dos resultados adquiridos. Também é sugerida uma alternativa de continuação do estudo a partir de outra abordagem.

## 2 REFERENCIAL TEÓRICO

### 2.1 Aprendizado de Máquina

O Aprendizado de Máquina (AM) é uma subárea da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos computacionais capazes de aprender a partir de dados. Em vez de serem explicitamente programados para realizar uma tarefa específica, esses algoritmos são projetados para aprender e melhorar seu desempenho à medida que são expostos a mais dados. De acordo com Chandrashekar e Sahin (2014), o Aprendizado de Máquina é empregado para organizar e compreender grandes volumes de dados, suprimindo a limitação humana nessas atividades. Em essência, o AM capacita as máquinas a organizar, aprender e tomar decisões com base em dados, possibilitando, inclusive, a realização de previsões com base em novos dados inseridos.

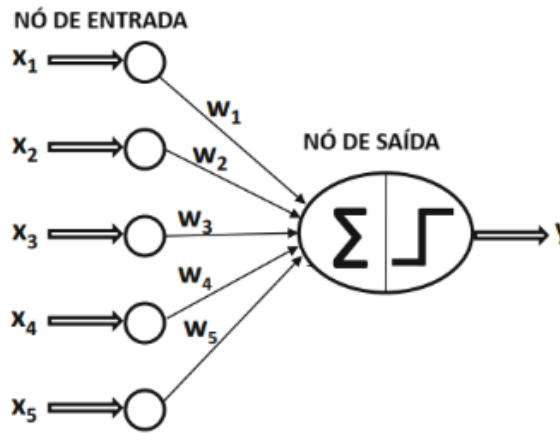
Os modelos de AM podem ser classificados em três categorias principais: supervisionado, não supervisionado e semissupervisionado. No aprendizado supervisionado, o modelo é treinado com um conjunto de dados rotulado, onde as entradas estão associadas a saídas conhecidas. O objetivo é aprender a mapear as entradas para as saídas, possibilitando a previsão de saídas para novas entradas não vistas durante o treinamento. Já no aprendizado não supervisionado, o modelo é treinado com um conjunto de dados não rotulado, onde não há informações explícitas sobre as saídas desejadas. O foco está em identificar padrões, estruturas ou relações intrínsecas nos dados, como agrupamentos ou redução de dimensionalidade. A abordagem semissupervisionada combina elementos do aprendizado supervisionado e não supervisionado. Ela utiliza conjuntos de dados que contêm tanto exemplos rotulados quanto não rotulados. O objetivo é aproveitar a informação limitada dos rótulos disponíveis, ao mesmo tempo que explora a estrutura dos dados não rotulados para melhorar o desempenho do modelo (MÜLLER; GUIDO, 2016).

Na detecção de anomalias médicas, a escolha do modelo de aprendizado de máquina geralmente depende do tipo de dados disponíveis, do contexto clínico e das características específicas do problema em questão. Embora não exista um modelo único que seja amplamente aceito como o mais adequado para todas as situações, alguns métodos de aprendizado de máquina, como redes neurais e redes neurais profundas, são frequentemente explorados devido à sua capacidade de identificar padrões complexos nas imagens e dados médicos.

### 2.1.1 Redes Neurais

As redes neurais artificiais são compostas por vários neurônios organizados em camadas, sendo o neurônio mais simples chamado de *perceptron*. A estrutura básica do *perceptron* é mostrada na Figura 1. Considere uma situação em que uma rede neural contenha somente um neurônio *perceptron*, e cada instância de treinamento é da forma  $(X, y)$ , onde cada  $X = [x_1, \dots, x_d]$  contém  $d$  características e  $y \in \{-1, +1\}$  contém o valor observado da variável de classe binária (AGGARWAL, 2018).

Figura 1 – Estrutura básica de um *perceptron*.



Fonte: Adaptado de Aggarwal (2018)

A camada de entrada, representada pelos círculos na figura, contém  $d$  entradas que transmitem as características  $X = [x_1, \dots, x_d]$  para o neurônio de saída, com as arestas de peso  $W = [w_1, \dots, w_d]$ . A camada de entrada não realiza nenhuma computação por conta própria. A função linear  $W \cdot X = \sum_{i=1}^d w_i x_i$  é calculada no nó de saída. O valor resultante passa por uma função de ativação, como **ReLU** ou **sigmóide**, para introduzir não linearidade na predição.

A predição  $\hat{y}$  é calculada como:

$$\hat{y} = \sigma(W \cdot X) = \sigma \left( \sum_{j=1}^d w_j x_j \right), \quad (2.1)$$

onde  $\sigma$  é uma função de ativação, como a **sigmóide** ( $\sigma(x) = \frac{1}{1+e^{-x}}$ ) ou **ReLU** ( $\sigma(x) = \max(0, x)$ ), dependendo do modelo e do problema em questão.

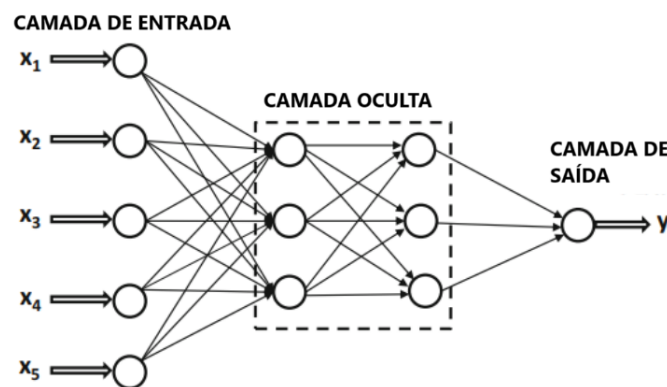
Redes neurais com uma única camada, como o exemplo do *perceptron*, são simples, mas suas capacidades são limitadas, pois só conseguem aprender decisões lineares, o que as torna eficazes apenas para problemas linearmente separáveis.



Para superar essa limitação, foram desenvolvidas as **Redes Neurais Multicamadas**, que são compostas por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Essas redes são capazes de aprender representações hierárquicas e complexas, com cada camada oculta extraíndo características mais abstratas à medida que os dados passam pela rede. A Figura 2 mostra a arquitetura de uma rede neural multicamadas.

As **Redes Neurais Profundas**, por sua vez, são uma extensão das redes neurais multicamadas, contendo um número significativamente maior de camadas ocultas, tipicamente da ordem de dezenas a centenas. Esse aumento no número de camadas permite que as redes neurais profundas aprendam representações ainda mais complexas e detalhadas dos dados. Redes como as **ResNets** de (HE et al., 2015) utilizam conexões residuais para melhorar o treinamento, permitindo que elas aprendam de forma mais eficaz, especialmente em tarefas que envolvem grandes volumes de dados complexos (HE et al., 2015).

Figura 2 – Arquitetura de uma Rede Neural Multicamadas.



Fonte: Adaptado de Aggarwal (2018)

A organização das camadas de uma rede neural multicamadas é definida pela sua arquitetura, sendo a *feed-forward* uma das mais comuns para tarefas de classificação. Nessa arquitetura, todos os nós de uma camada estão conectados aos nós da próxima camada, formando uma rede completamente conectada. Assim, a estrutura da rede é amplamente definida quando se especificam o número de camadas e o número de neurônios em cada camada (RUMELHART; HINTON; WILLIAMS, 1986).

A arquitetura *feed-forward* impacta diretamente os resultados da rede neural. À medida que a informação percorre as camadas, cada camada oculta extrai características mais complexas e abstratas. O processo de propagação para a frente permite que a rede aprenda representações hierárquicas dos dados, o que capacita a rede a identificar padrões mais sofisticados e a resolver tarefas complexas. Isso contribui para a capacidade das redes neurais multicamadas em aprender automaticamente as características mais relevantes do problema em questão.

Redes neurais convolucionais são uma categoria especializada de redes neurais projetadas para processar dados com estrutura de grade, como imagens. Elas são particularmente eficazes na extração de características espaciais e padrões visuais. As CNNs introduzem camadas convolucionais que aplicam filtros locais a regiões específicas da entrada, permitindo a detecção de características como bordas, texturas e padrões locais (AGGARWAL, 2018).

Na arquitetura de uma CNN, cada camada da rede é tridimensional, com uma extensão espacial e uma profundidade correspondente ao número de características. A profundidade de uma única camada em uma CNN é distinta do conceito de profundidade em termos do número de camadas. Na camada de entrada, essas características correspondem aos canais de cor, e nas camadas ocultas, essas características representam mapas de características ocultas que codificam vários tipos de formas na imagem (AGGARWAL, 2018).

As **Redes Neurais Convolucionais (CNNs)** consistem em dois tipos principais de camadas: camadas de **convolução** e camadas de **subamostragem** (*pooling*). De acordo com Aggarwal (2018), nas camadas de convolução, uma operação de convolução é definida, usando um filtro para mapear as ativações de uma camada para a próxima. Essa operação é realizada em todas as posições possíveis para definir a próxima camada. As conexões em uma CNN são esparsas, ou seja, nem todas as unidades em uma camada estão conectadas a todas as unidades na camada seguinte, pois qualquer ativação em uma camada é uma função de uma pequena região espacial na camada anterior. As camadas, exceto o conjunto final de duas ou três camadas, mantêm sua estrutura espacial o que significa que as dimensões espaciais das imagens (largura e altura) são preservadas até as últimas camadas convolucionais ou de *pooling*.

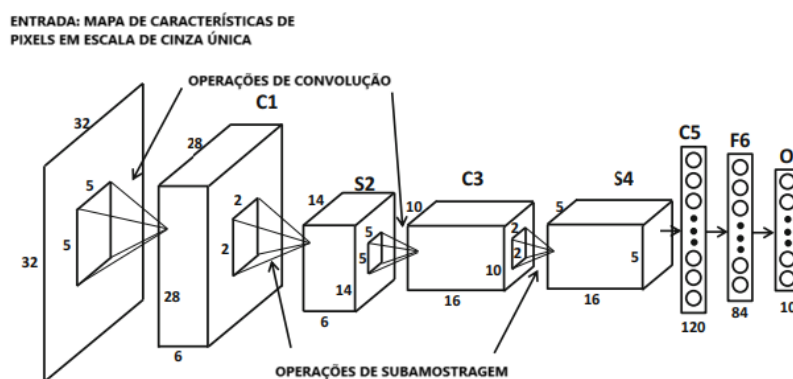
Além disso, as **camadas completamente conectadas** (*fully connected layers*) desempenham um papel crucial na parte final das CNNs. Após as camadas convolucionais e de *pooling*, as ativações extraídas das imagens são representadas na forma de um vetor unidimensional e passadas para uma ou mais camadas completamente conectadas. Em uma camada completamente conectada, cada unidade está conectada a todas as unidades da camada anterior. Essas camadas são responsáveis por realizar a classificação final das características extraídas, transformando as ativações em uma saída, como a probabilidade de uma imagem pertencer a uma classe específica. As camadas completamente conectadas são essenciais para a tomada de decisão no processo de aprendizado, pois combinam as informações extraídas pelas camadas anteriores para formar a resposta final (LECUN; BENGIO; HINTON, 2015).

As camadas de subamostragem (*pooling*), por sua vez, têm como objetivo reduzir as

dimensões espaciais da entrada, comprimindo os dados e, ao mesmo tempo, mantendo as informações mais relevantes para a tarefa. Em muitas implementações, o *pooling* é realizado utilizando uma janela deslizante, como  $2 \times 2$  ou  $3 \times 3$ , e uma operação de agregação, como máximo (*max pooling*) ou média (*average pooling*). No *max pooling*, o valor da janela é substituído pelo valor máximo encontrado dentro da janela, enquanto no *average pooling*, a média dos valores na janela é calculada (GOODFELLOW; BENGIO; COURVILLE, 2016).

Por exemplo, uma janela de tamanho  $2 \times 2$  comprime a saída de uma camada anterior por um fator de 4, reduzindo a resolução espacial dos dados. Essas operações ajudam a reduzir a complexidade computacional e a prevenir o overfitting, mantendo as características mais relevantes. Isso pode ser observado na arquitetura do LeNet-5, mostrada na Figura 3, onde o *pooling* é utilizado para reduzir a dimensionalidade das representações extraídas pelas camadas convolucionais (LECUN et al., 1998).

Figura 3 – LeNet-5: Um dos primeiros modelos de redes neurais convolucionais.



Fonte: Adaptado de Aggarwal (2018)

É interessante notar que CNNs pré-treinadas muitas vezes estão disponíveis para uso imediato em outras aplicações e conjuntos de dados. Os pesos pré-treinados nas camadas iniciais são úteis, pois aprendem vários tipos de formas nas imagens, sendo aplicáveis a virtualmente qualquer tipo de aplicação de classificação baseadas em imagens.

### 2.1.2 Redes CNNs Pré-treinadas

Para a classificação de imagens dermatoscópicas, foram utilizadas três arquiteturas de CNNs pré-treinadas conhecidas na literatura: **MobileNetV2** (HOWARD et al., 2017), **ResNet50** (HE et al., 2015) e **NasNetMobile** (ZOPH et al., 2018). Essas redes foram escolhidas por sua capacidade de extração de características, eficiência computacional e desempenho na tarefa de classificação de imagens médicas.

### 2.1.2.1 *Transfer Learning*

O treinamento de CNNs do princípio requer uma grande quantidade de dados anotados e recursos computacionais significativos. Para contornar esse problema, foi empregada a técnica de *Transfer Learning*, na qual são utilizadas redes neurais que foram previamente treinadas com um conjunto de dados. No caso deste trabalho, as redes neurais foram treinadas com o conjunto de dados **ImageNet**.

O ImageNet (Jong (1975)) é um dos maiores e mais comuns conjuntos de dados para treinamento de redes neurais baseadas na CNN. ImageNet contém milhões de imagens rotuladas para mais de mil categorias, o que permite que essas redes aprendam uma variedade de características visuais gerais. Após o treinamento com a ImageNet, essas redes podem ser adaptadas para tarefas específicas (como a detecção de melanoma) usando a técnica de *fine-tuning*, onde os parâmetros das camadas finais são ajustados.

### 2.1.2.2 MobileNetV2

A **MobileNetV2** é uma arquitetura leve e otimizada para dispositivos móveis, projetada para equilibrar eficiência computacional e precisão (SANDLER et al., 2018). Seu design se baseia em convoluções separáveis em profundidade (*depthwise separable convolutions*), que reduzem significativamente o número de parâmetros e operações computacionais, mantendo um desempenho competitivo.

Além disso, MobileNetV2 introduz blocos de convolução invertida com conexões residuais (*inverted residual blocks*), permitindo preservar informações detalhadas mesmo em redes compactas. No presente estudo, a MobileNetV2 foi utilizada com *Transfer Learning*, onde as camadas finais foram substituídas por um classificador especializado na detecção de melanoma.

### 2.1.2.3 ResNet50

A **ResNet50** (*Residual Network*) é uma rede composta por 50 camadas profundas projetada para resolver o problema do desaparecimento do gradiente por meio de conexões residuais (*skip connections*) (HE et al., 2016). Essas conexões permitem que a informação flua diretamente entre camadas mais profundas, facilitando o aprendizado e melhorando a estabilidade da rede.

A ResNet50 é amplamente utilizada em diversas tarefas de visão computacional, devido à sua capacidade de aprender representações ricas e hierárquicas. No presente estudo, a ResNet50 foi utilizada com *Transfer Learning*, congelando as camadas iniciais e ajustando as camadas finais para se especializar na classificação de lesões cutâneas.

#### 2.1.2.4 NasNetMobile

A **NasNetMobile** (*Neural Architecture Search Network*) é uma arquitetura gerada automaticamente por meio da busca de arquitetura neural (*Neural Architecture Search* - *NAS*) (ZOPH et al., 2018). Essa abordagem permite encontrar configurações de redes neurais otimizadas para maximizar o desempenho em tarefas específicas.

A NasNetMobile foi projetada como uma versão compacta da NasNet, equilibrando eficiência computacional e precisão. Seu design permite ajustar dinamicamente a disposição dos blocos convolucionais para capturar melhor as características das imagens. No presente estudo, a NasNetMobile foi utilizada com *Transfer Learning*, permitindo sua adaptação à detecção de lesões dermatológicas.

#### 2.1.2.5 Comparação das Arquiteturas

Cada uma dessas redes possui características distintas que influenciam seu desempenho e aplicabilidade na tarefa de detecção de melanoma. A Tabela 1 apresenta uma comparação entre as três arquiteturas utilizadas no estudo em termos de tamanho e velocidade de inferência.

Tabela 1 – Comparação das arquiteturas CNNs utilizadas.

Modelo	Nº de Parâmetros	Tamanho	Velocidade de Inferência
MobileNetV2	≈ 3.4M	Pequeno	Rápida
ResNet50	≈ 25.6M	Médio	Moderada
NasNetMobile	≈ 5.3M	Pequeno	Moderada

A análise dessas três redes em diferentes cenários permitirá avaliar a influência dos índices na classificação de imagens de lesões cutâneas. O uso de *Transfer Learning* possibilita o aproveitamento do conhecimento prévio dessas arquiteturas, tornando a adaptação para a detecção de melanoma mais eficiente e robusta.

Para as Redes Neurais Convolucionais (CNNs) pré-treinadas, a escolha foi motivada pela necessidade de aproveitar o aprendizado prévio sobre grandes volumes de dados,

o que possibilita uma adaptação eficiente em tarefas específicas, como a detecção de melanoma. Essas redes foram treinadas previamente em grandes conjuntos de dados como o ImageNet, que contém milhões de imagens rotuladas, permitindo que as redes aprendessem representações gerais e universais de padrões visuais. Ao utilizar as CNNs pré-treinadas, foi possível obter um desempenho robusto mesmo com um conjunto de dados limitado de imagens dermatoscópicas. A técnica de Transfer Learning foi aplicada, ajustando as camadas finais das redes para a tarefa de classificação de lesões cutâneas. As redes pré-treinadas, como MobileNetV2, ResNet50 e NasNetMobile, foram escolhidas pela sua capacidade de extração de características, eficiência computacional e desempenho consolidado em tarefas de classificação de imagens médicas, proporcionando uma solução eficiente tanto em termos de precisão quanto de custo computacional.

### 2.1.3 Programação Genética

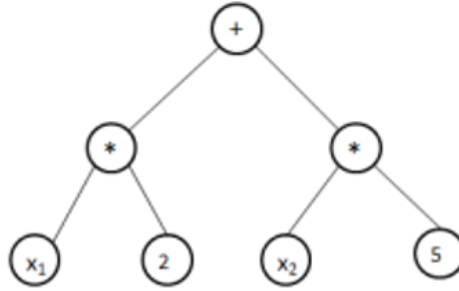
A Programação Genética (PG) é uma técnica para gerar e evoluir automaticamente programas computacionais. Ela foi introduzida por Koza (1994) com o objetivo de permitir que computadores aprendam a resolver problemas sem serem explicitamente programados para isso. Diferentemente de algumas outras técnicas evolucionárias, a PG fornece soluções na forma de programas computacionais interpretáveis. A abordagem é bastante genérica e não exige grandes adaptações para lidar com diferentes problemas.

De acordo com Medeiros (2019), na PG os programas são representados como árvores de sintaxe abstrata, compostas por nós, onde cada nó representa um operador ou um operando. Os nós internos geralmente representam operadores, enquanto os nós folha representam operandos. Cada caminho da raiz até um nó folha representa uma expressão completa. Cada função e terminal representam um gene do indivíduo, e cada indivíduo é uma possível solução para o problema em questão.

A Figura 4 ilustra um exemplo de indivíduo representado por árvore de sintaxe, de modo a corresponder à expressão  $2x_1 + 5x_2$ . Nessa representação, as operações aritméticas são representadas por nós internos da árvore, enquanto as variáveis e constantes são representadas nas folhas. A árvore de sintaxe correspondente à expressão  $2x_1 + 5x_2$  é organizada de maneira hierárquica, onde o nó raiz realiza a operação de adição (+) entre dois subelementos. O primeiro subelemento é o produto de 2 e  $x_1$ , enquanto o segundo subelemento é o produto de 5 e  $x_2$ . Assim, a estrutura da árvore reflete a forma como as operações são agrupadas e realizadas.

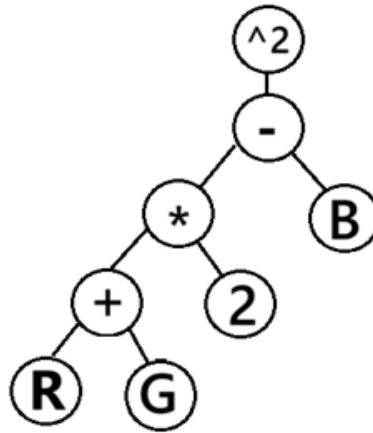
A Figura 5 ilustra um exemplo de indivíduo representado por árvore de sintaxe, de

Figura 4 – Exemplo de indivíduo representado por árvore de sintaxe.



Fonte: Medeiros (2019)

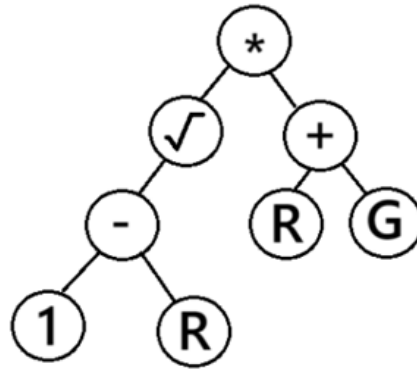
modo a corresponder à expressão  $[(2 \times (R + G)) - B]^2$ . Nessa representação, as operações aritméticas são representadas por nós internos da árvore, enquanto as variáveis e constantes são representadas nas folhas. A árvore de sintaxe correspondente à expressão  $[(2 \times (R + G)) - B]^2$  é organizada de maneira hierárquica, onde o nó raiz realiza a operação de exponenciação ( $^2$ ) sobre o resultado da subtração ( $-$ ) entre o produto de 2 e  $(R + G)$ , e a constante  $B$ . A estrutura da árvore reflete a forma como as operações são agrupadas e realizadas.

Figura 5 – Exemplo de indivíduo representado por árvore de sintaxe para a expressão  $[(2 \times (R + G)) - B]^2$ .

Fonte: Autor

A Figura 6 ilustra um exemplo de indivíduo representado por árvore de sintaxe, de modo a corresponder à expressão  $\sqrt{(1 - R) \times (R + G)}$ . Nessa representação, as operações aritméticas são representadas por nós internos da árvore, enquanto as variáveis e constantes são representadas nas folhas. A árvore de sintaxe correspondente à expressão  $\sqrt{(1 - R) \times (R + G)}$  é organizada de maneira hierárquica, onde o nó raiz realiza a operação de raiz quadrada ( $\sqrt{\phantom{x}}$ ) sobre o produto ( $\times$ ) entre  $(1 - R)$  e  $(R + G)$ . A estrutura da árvore reflete a forma como as operações são agrupadas e realizadas.

Figura 6 – Exemplo de indivíduo representado por árvore de sintaxe para a expressão  $\sqrt{(1 - R) \times (R + G)}$ .



Fonte: Autor

O processo evolutivo inicia-se com a geração de uma população inicial, seguida por avaliação conforme uma função de aptidão. Os melhores indivíduos têm mais chances de serem selecionados para operações de cruzamento, reprodução e mutação, simulando processos análogos à evolução natural.

A PG é utilizada em diversos contextos, como regressão simbólica de funções, aprendizado de máquina e inteligência artificial. Sua capacidade de representar soluções na forma de programas de computador, que podem ser hierarquicamente compostos por funções primitivas e terminais, torna-a uma ferramenta poderosa para resolver problemas complexos (KOZA, 1994).

Neste trabalho, a PG é utilizada para gerar um novo índice específico para tratar da segmentação das lesões de pele, que será subsequentemente usada para a classificação.

A biblioteca Python utilizada para implementar a Programação Genética neste trabalho foi o *Distributed Evolutionary Algorithms in Python* (DEAP) (FORTIN et al., 2012). O DEAP é uma biblioteca popular e eficiente para a implementação de algoritmos evolutivos, incluindo a Programação Genética. Ela oferece uma plataforma flexível e fácil de usar para criar algoritmos evolucionários, como seleção, cruzamento, mutação e avaliação de indivíduos. A DEAP fornece uma estrutura intuitiva para definir a população inicial, a função de aptidão e as operações de evolução, permitindo que a programação genética seja aplicada de maneira eficiente para otimizar as expressões matemáticas necessárias para a segmentação das lesões de pele. Além disso, a biblioteca oferece suporte a paralelismo, o que facilita o processamento em grande escala, tornando o processo de otimização mais rápido e eficiente.



## 2.2 Métricas de Avaliação

As métricas de avaliação referem-se a métodos utilizados para medir o desempenho de modelos de aprendizado de máquina. Essas métricas são essenciais para entender quão bem um modelo está realizando suas previsões em tarefas específicas. Elas são fundamentais para orientar o ajuste do modelo, identificar áreas de melhoria e tomar decisões informadas sobre o uso do modelo em diferentes contextos. GÉron (2019) enfatiza a importância de escolher as métricas adequadas com base no tipo de problema e nos requisitos específicos da aplicação.

### 2.2.1 Matriz de Confusão

A matriz de confusão é uma ferramenta essencial na avaliação do desempenho de modelos de classificação. Ela é uma tabela que descreve o desempenho de um modelo na classificação de instâncias em diferentes classes. Ela inclui as seguintes métricas:

- Verdadeiro Positivo (VP): Número de instâncias positivas corretamente classificadas como positivas.
- Falso Positivo (FP): Número de instâncias negativas erroneamente classificadas como positivas.
- Falso Negativo (FN): Número de instâncias positivas erroneamente classificadas como negativas.
- Verdadeiro Negativo (VN): Número de instâncias negativas corretamente classificadas como negativas.

Na Figura 7, a Classe Verdadeira refere-se à classe real ou verdadeira, enquanto a Classe Predita refere-se à classe que o modelo previu. Portanto, um Verdadeiro Positivo ocorre quando o modelo prevê corretamente uma instância positiva, um Falso Positivo ocorre quando o modelo erroneamente prevê uma instância negativa como positiva, um Falso Negativo ocorre quando o modelo erroneamente prevê uma instância positiva como negativa, e um Verdadeiro Negativo ocorre quando o modelo prevê corretamente uma instância negativa.

A utilização da matriz de confusão é fundamental para entender melhor o desempenho do modelo em diferentes situações e para tomar decisões informadas sobre ajustes e melhorias no modelo de aprendizado de máquina (COSTA et al., 2007).

Figura 7 – Matriz de confusão.

		Classe Predita	
		Positivo	Negativo
Classe Verdadeira	Positivo	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Costa et al. (2007)

### 2.2.2 Acurácia

A acurácia é uma métrica fundamental utilizada na avaliação de modelos de classificação, sendo uma medida simples e intuitiva do desempenho geral do modelo. Ela representa a porcentagem de predições corretas em relação ao número total de instâncias no conjunto de dados (COSTA et al., 2007). A Equação 2.2 apresenta o seu cálculo:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}. \quad (2.2)$$

### 2.2.3 Precisão

A precisão é uma métrica que mede a exatidão das predições positivas corretas feitas por um modelo em relação ao total de predições positivas feitas, incluindo as corretas e as incorretas. É particularmente relevante em problemas onde o foco está na minimização de falsos positivos, ou seja, na redução do número de instâncias que foram incorretamente classificadas como positivas (GÉRON, 2019). A Precisão é calculada conforme a Equação (2.3):

$$\text{Precisão} = \frac{VP}{VP + FP}. \quad (2.3)$$

### 2.2.4 Sensibilidade

Sensibilidade mede a capacidade do modelo em capturar todas as instâncias positivas. É a proporção de instâncias positivas corretamente classificadas em relação ao total de instâncias realmente positivas. Essa métrica é importante quando os falsos negativos são críticos (GÉRON, 2019). O Sensibilidade é calculado conforme a Equação (2.4):

$$\text{Sensibilidade} = \frac{VP}{VP + FN}. \quad (2.4)$$

### 2.2.5 F1 Score

O *F1 Score* é uma métrica de avaliação que combina precisão e sensibilidade em uma única medida. Essa métrica é especialmente útil quando há um desequilíbrio entre as classes do problema de classificação, ou seja, quando uma classe ocorre com mais frequência do que a outra (GÉRON, 2019):

$$F1Score = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}. \quad (2.5)$$

## 2.3 Índices de Cor (IC) e Índice de Marrom (IM)

Os índices de cor referem-se a medidas quantitativas que descrevem as características de cores em uma imagem. Eles são frequentemente usados em visão computacional e processamento de imagens para extrair informações sobre a distribuição de cores em uma cena (GONZÁLEZ; WOODS; EDDINS, 2009).

Esses índices são utilizados para extrair, efetivamente, características significativas em aplicações como classificação de imagem, reconhecimento de padrões e processamento de imagens em geral. A escolha do índice depende do contexto da aplicação e das características específicas que se deseja extrair da imagem.

A coloração de matérias orgânicas é resultante de uma interação complexa de alterações químicas, bioquímicas, microbianas e físicas. A prática rotineira de mensurar a cor dos produtos alimentícios tem-se mostrado uma abordagem valiosa para avaliar de forma indireta diversos atributos de qualidade, tais como sabor e teor de pigmentos. Isso se deve à sua natureza simples, rápida e à sua correlação efetiva com outras propriedades físico-químicas (PATHARE; OPARA; AL-SAID, 2013).

Na indústria alimentícia, o índice de marrom é um dos indicadores mais comuns para verificar a qualidade de algum produto alimentício no seu processo de apodrecimento, pois são observados diferentes tons de marrom, como comentado por Lunadei et al. (2010).

Alguns índices estipulados pela literatura são:

- **Índice de marrom no espaço RGB de Aimonino et al. (2015):**

$$IM_{Aimonino} = kR - G - B, \quad (2.6)$$

sendo  $k \in [2, 4]$  uma constante representando o ganho aplicado ao canal de cor vermelha.

- **Índice de marrom no espaço RGB de Lunadei et al. (2010):**

$$IM_{Lunadei1} = \frac{(R - B)}{(R + B)}, \quad (2.7)$$

$$IM_{Lunadei2} = (R - B), \quad (2.8)$$

$$IM_{Lunadei3} = \frac{B}{R}. \quad (2.9)$$

- **Índice de marrom no espaço LAB de Ebrahimi et al. (2015):**

Para utilizar o índice de marrom no espaço  $LAB$ , é necessário antes converter as imagens do espaço  $RGB$  para o espaço  $LAB$ . O espaço  $LAB$  foi desenvolvido para aproximar a percepção visual humana das cores e é dividido em três componentes:  $L^*$  (luminância),  $a^*$  (verde a vermelho) e  $b^*$  (azul a amarelo). Para realizar a conversão para o espaço  $LAB$ , utilizam-se os componentes  $XYZ$ , que representam os valores absolutos das cores de referência.

Os componentes  $X$ ,  $Y$  e  $Z$  são os valores do espaço  $XYZ$ , com  $X$  representando a quantidade de cor vermelha,  $Y$  representando a luminância (brilho) e  $Z$  representando a quantidade de cor azul. Esses valores são calculados a partir da conversão do espaço  $RGB$  para o espaço  $XYZ$ . Uma vez no espaço  $LAB$ , o índice de marrom de Ebrahimi et al. (2015) é dado pela seguinte fórmula:

$$IM_{Ebrahimi} = \frac{100 \times (x - 0.31)}{0.17}, \quad (2.10)$$

sendo:

$$x = \frac{(4L^* + 15X^* + 1)}{(5.645L^* + a^* - 3.012b^*)}, \quad (2.11)$$

com  $L^*$ ,  $a^*$  e  $b^*$  definidos pelas equações:

$$L^* = 116 \cdot S(Y/Y_n) - 16 \quad (2.12)$$

$$a^* = 500 \left( S \left( \frac{X}{X_n} \right) - S \left( \frac{Y}{Y_n} \right) \right) \quad (2.13)$$

$$b^* = 200 \left( S \left( \frac{Y}{Y_n} \right) - S \left( \frac{Z}{Z_n} \right) \right), \quad (2.14)$$

sendo:

$$S(t) = \begin{cases} \sqrt[3]{t}, & \text{se } t > \delta^3, \\ \frac{t}{\delta^2} + \frac{4}{29}, & \text{caso contrário,} \end{cases}$$

com  $\delta = \frac{6}{29}$  e  $(X_n, Y_n, Z_n)$  correspondendo à luminância D65 com  $Y_n = 100$ .

• **Índice de marrom no espaço LAB de Fetuga, Tomlins e Henshaw (2014):**

$$IM_{Fetuga} = 100 - L^*, \quad (2.15)$$

sendo  $L^*$  definido pela Equação 2.12.

Assim, observado que tanto a pele humana quanto as lesões melanômicas geralmente exibem uma tonalidade de marrom, Filho et al. (2023) verificaram a usabilidade de índices de marrom para a área médica, especificamente na detecção e análise de lesões cutâneas, obtendo bons resultados na diferenciação entre pele saudável e melanoma. Como essa abordagem ainda foi pouco explorada na área dermatológica, especialmente no diagnóstico automatizado de câncer de pele por meio de imagens dermatoscópicas, este projeto tem como objetivo realizar uma análise mais aprofundada da aplicação de índices na identificação precoce de melanoma, com foco na melhoria da acurácia e na viabilidade da utilização desses índices para assistentes clínicos automatizados.

Será descrito na seção a seguir a metodologia proposta com o uso dos índices de marrom, abordando os detalhes sobre a aplicação e os procedimentos adotados para a análise e validação dos resultados obtidos.

### 3 METODOLOGIA

A metodologia proposta visa otimizar o uso de índices de cor e redes neurais convolucionais (CNNs) para a classificação e segmentação de lesões de pele, com o objetivo de aprimorar a detecção precoce de melanoma. O processo de metodologia foi ilustrado na Figura 8, que mostra as etapas principais do estudo, desde a organização dos dados até a avaliação dos resultados.

Figura 8 – Fluxograma da metodologia proposta.



Fonte: Autor.

De acordo com o fluxograma, a metodologia segue as seguintes etapas principais:

- **Organização dos dados:** O conjunto de dados é preparado e segmentado para treinamento e avaliação, garantindo que as imagens estejam devidamente rotuladas como benignas ou malignas.
- **Identificação de índices de cor por Programação Genética (PG):** Índices de marrom são extraídos a partir de operações sobre os canais de cor das imagens para aprimorar a diferenciação entre melanoma e pele saudável.
- **Transfer learning sem índices de cor:** Redes CNNs pré-treinadas são aplicadas às imagens sem o uso de índices de marrom para avaliar a eficácia da rede sem modificações adicionais.
- **Aplicação dos índices de marrom da literatura:** Índices previamente descritos na literatura são aplicados às imagens para verificar se há melhoria no desempenho da rede.
- **Aplicação dos índices de cor encontrados pela PG:** Novos índices gerados por programação genética são aplicados às imagens, e a eficácia deles é comparada com os índices da literatura.
- **Avaliação dos resultados:** A eficácia dos diferentes índices é comparada com base nas métricas de performance, como o F1-score, para determinar qual abordagem oferece a melhor segmentação e classificação.

A metodologia segue a estrutura descrita acima, com ênfase na aplicação de índices de marrom derivados de programação genética (PG) para segmentação de lesões cutâneas

e subsequente classificação. A comparação entre diferentes abordagens permite verificar como a utilização desses índices pode melhorar a detecção de melanomas.

A metodologia proposta teve como objetivo avaliar a eficácia de índices na classificação de imagens dermatoscópicas utilizando redes neurais convolucionais (CNNs) pré-treinadas. Para isso, o estudo foi conduzido em duas etapas principais descritas nas próximas seções.

### 3.1 Descoberta de Novos Índices por Programação Genética

Inicialmente, foi empregada a programação genética para encontrar novos índices que melhor segmentem as características das lesões de pele, para, em seguida, serem usados na tarefa de classificação. Esses índices foram gerados a partir de combinações matemáticas das camadas de cor das imagens, explorando operações como soma, subtração, multiplicação, divisão, raiz quadrada e potenciação. O objetivo foi identificar uma equação que otimize a segmentação das lesões ao maximizar a diferenciação entre áreas afetadas e saudáveis.

A programação genética foi executada seis vezes com o objetivo de gerar seis equações distintas, sendo três para cada grupo de complexidade das operações consideradas. Cada execução do PG resultou em um índice específico, o qual foi avaliado a partir da conversão das imagens RGB para uma imagem em nível de cinza. Essa conversão foi realizada aplicando o índice gerado pela PG às imagens, permitindo a segmentação das lesões de pele. O desempenho de cada índice foi então analisado com base em sua capacidade de segmentar corretamente as áreas de interesse, sendo a segmentação avaliada por meio de métricas como o *F1 Score*, garantindo que os índices mais eficazes fossem selecionados para a tarefa de diagnóstico.

Os índices encontrados foram agrupados em dois tipos distintos:

- **Índices com operações simples:** equações que utilizam operações matemáticas básicas, como soma, subtrações, multiplicações e divisões entre os canais de cor (RGB) e valores constantes no intervalo  $[0, 1]$ . Os canais RGB foram previamente normalizados para o intervalo  $[0, 1]$  por meio da divisão por 255;
- **Índices com operações não lineares:** equações que envolvem as operações matemáticas raiz quadrada e potência, que são não lineares, além das quatro operações básicas e os valores constantes no intervalo  $[0, 1]$ . Isto permite uma modelagem mais detalhada e expressiva das características das lesões cutâneas.

Assim, os índices estipulados foram agrupados três a três, três índices que utilizam somente operações simples e três que utilizam operações não lineares, de modo que a imagem em nível de cinza gerada por cada um dos índices seja utilizada em um dos canais coloridos *RGB* da imagem utilizada na classificação. Então, a imagem a ser utilizada para classificação terá seus canais *RGB* alterados pelo resultado dos índices aplicados a eles.

A análise comparativa entre esses dois grupos permitiu avaliar se a introdução de não linearidades, como raiz e exponenciação, nas equações melhora a segmentação das lesões e, consequentemente, a eficácia das redes pré-treinadas na detecção de melanomas.

A descoberta dos índices foi conduzida utilizando um algoritmo de programação genética, onde cada indivíduo da população representa uma equação matemática candidata ao índice. O desempenho de cada índice gerado foi avaliado de acordo com sua capacidade de melhorar a segmentação das imagens quando aplicado ao algoritmo de K-Means (MACQUEEN, 1967), um método de agrupamento amplamente utilizado na segmentação de imagens médicas.

A métrica utilizada para avaliar o desempenho dos indivíduos da programação genética foi o F1-score médio, calculado a nível de pixel após a aplicação do K-Means sobre as imagens transformadas pelos índices gerados. O cálculo do F1-Score na segmentação de imagens considera a comparação entre os pixels atribuídos a diferentes classes (por exemplo, melanoma e pele saudável). Nesse contexto, o F1-Score é calculado utilizando a fórmula clássica de F1-Score, que é a média harmônica entre a precisão e o sensibilidade, mas de forma adaptada para uma análise pixel a pixel. A precisão e a sensibilidade (ou sensibilidade) são obtidos considerando a correspondência entre os pixels segmentados como pertencentes ao melanoma (positivos) e aqueles corretamente classificados como melanoma.

Embora o K-Means seja um algoritmo de agrupamento e não de classificação, ele pode ser adaptado para segmentação de imagens. O K-Means agrupa os pixels da imagem em grupos com base em características semelhantes (como cor ou textura), e os grupos resultantes são então mapeados para as classes de interesse, como melanoma e pele saudável. O F1-Score é então calculado comparando os grupos formados pelo K-Means com as áreas reais de melanoma, que são rotuladas na imagem, usando a segmentação de referência (máscara). Dessa forma, a programação genética busca equações que maximizem o F1-score médio do banco de imagens, garantindo que os índices resultantes sejam eficazes na distinção entre regiões de melanoma e pele saudável.

Para fazer esse mapeamento entre os grupos e as classes de interesse, foi adotada uma



abordagem baseada na localização das lesões. Como o banco de imagens para segmentação é composto de imagens em que a lesão de pele (melanoma) está centralizada, foi realizado um processo de verificação. Nesse processo, o grupo localizado na região central da imagem foi associado ao melanoma, e o outro grupo foi associado à pele saudável. Isso foi possível porque a lesão de melanoma está predominantemente localizada no centro da imagem, o que facilita a identificação do grupo relevante.

Além disso, foi limitado o número de grupos a dois ( $k = 2$ ), garantindo que apenas duas regiões fossem formadas: uma representando o melanoma e outra representando a pele saudável. Este método assegura que a segmentação se concentrasse apenas nas duas classes principais de interesse, minimizando confusões entre outras possíveis variações de cor e textura na imagem.

O processo de otimização foi conduzido iterativamente utilizando a biblioteca **DEAP** (Distributed Evolutionary Algorithms in Python) para implementar o algoritmo de programação genética (PG). As etapas envolvidas foram as seguintes:

1. **Inicialização da população:** foi gerada uma população inicial de equações matemáticas candidatas, representadas como indivíduos em uma população evolutiva. Cada indivíduo foi representado como uma árvore de expressão matemática.
2. **Avaliação de *fitness*:** o desempenho de cada indivíduo foi avaliado utilizando o F1-score médio, calculado após a aplicação do K-Means nas imagens transformadas pelos índices gerados.
3. **Seleção:** a seleção dos indivíduos foi feita com base na sua aptidão (*fitness*), priorizando aqueles que maximizaram o F1-score médio.
4. **Crossover e Mutação:** foram realizadas operações de *crossover* e mutação para gerar novas soluções, combinando e alterando as equações para explorar o espaço de soluções de forma mais eficaz.

Os hiperparâmetros utilizados no algoritmo de programação genética foram os seguintes:

- **Tamanho da população:** 200 indivíduos
- **Probabilidade de *crossover*:** 0.7
- **Probabilidade de mutação:** 0.4
- **Número de gerações:** 100

- **Taxa de elitismo:** 0.1 (proporção de indivíduos melhores que são mantidos inalterados para a próxima geração)

A escolha desses parâmetros foi baseada nas recomendações de Eiben e Smith (2003), que indicam que o número de gerações e o tamanho da população desempenham um papel crucial no equilíbrio entre a exploração do espaço de soluções e a convergência para soluções de alta qualidade em algoritmos evolutivos. Segundo Eiben e Smith (2003), uma população maior pode proporcionar maior diversidade nas soluções, evitando a convergência prematura para um ótimo local. No entanto, o aumento no número de indivíduos também implica em maior custo computacional. Por outro lado, o número de gerações deve ser suficientemente grande para permitir que o algoritmo explore e refine as melhores soluções, mas sem comprometer a eficiência computacional. O valor de 100 gerações foi escolhido para garantir que o algoritmo tivesse tempo suficiente para convergir, enquanto a população de 200 indivíduos foi considerada adequada para equilibrar a diversidade e a eficiência do processo.

A biblioteca DEAP foi escolhida devido à sua flexibilidade e eficiência na implementação de algoritmos evolutivos, permitindo o controle preciso sobre os operadores genéticos e a fácil definição de métricas de avaliação.

Por fim, os resultados obtidos com os índices de marrom da literatura e os novos índices gerados foram comparados com os resultados das imagens originais, possibilitando uma análise detalhada da contribuição dessa métrica para a classificação de melanomas. Essa abordagem buscou verificar se a aplicação destes novos índices pode aprimorar a detecção de padrões relevantes para o diagnóstico de câncer de pele.

### 3.2 Conjuntos de Dados

O desempenho de técnicas de AM está intimamente ligado à qualidade do conjunto de dados utilizado. Assim, para o estudo sobre melanoma, foi feita uma análise detalhada dos conjuntos de dados ISIC 2016 (ISIC, 2016), ISIC 2019 (ISIC, 2019) e ISIC 2020 (ISIC, 2020), fornecidos pela *International Skin Imaging Collaboration* (ISIC). Esses conjuntos de dados são amplamente utilizados nas pesquisas de diagnóstico de câncer de pele e dermatologia, fornecendo imagens dermatoscópicas de alta resolução e anotações clínicas para o desenvolvimento de algoritmos de inteligência artificial.

Para este estudo, não foi utilizado o conjunto de dados ISIC 2020 de forma isolada, mas

sim uma combinação entre os conjuntos de dados ISIC 2019 e ISIC 2020. Foram extraídas as imagens de melanoma do ISIC 2019 e as imagens malignas do ISIC 2020 para compor o conjunto de dados de melanoma. Além disso, no conjunto de dados do ISIC 2020, cada imagem está associada a um identificador único de paciente. Portanto, as imagens malignas do ISIC 2020 associadas às imagens benignas do mesmo paciente foram extraídas para compor o conjunto de dados de nevus.

Optou-se por não utilizar o ISIC 2016, pois, embora este conjunto contenha imagens de melanoma, ele foi especificamente projetado para tarefas de segmentação, e não de classificação. O foco principal do ISIC 2016 é fornecer imagens com anotações de segmentação (máscaras), o que não é necessário para a tarefa de classificação binária que estamos abordando, onde as imagens precisam ser rotuladas apenas como melanoma ou nevus. Além disso, as imagens do ISIC 2016 estão presentes nos conjuntos de dados ISIC 2019 e ISIC 2020, que já foram selecionados, garantindo a inclusão das imagens mais recentes e com a classificação associada.

Esse procedimento permitiu a construção de um conjunto de dados balanceado e estruturado, garantindo que as amostras benignas e malignas estejam mais alinhadas ao contexto clínico real, onde múltiplas imagens de um mesmo paciente podem estar disponíveis.

O conjunto de imagens ISIC 2016, por sua vez, é composto por imagens de melanoma de tamanhos variados, acompanhadas de suas respectivas segmentações. Esse conjunto de dados possibilita uma análise detalhada das características das lesões de pele, sendo especialmente útil para experimentos que envolvem técnicas de segmentação e extração de padrões em imagens médicas.

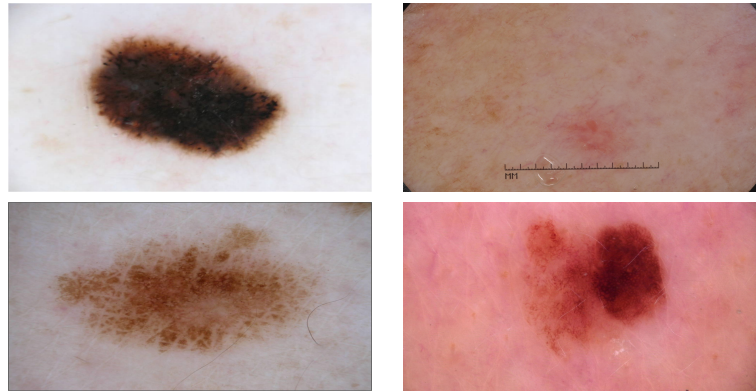
Para este estudo, as imagens do conjunto de dados combinado ISIC 2019-2020 foram utilizadas no treinamento das redes neurais, enquanto as imagens do ISIC 2016 foram empregadas para encontrar um novo índice por meio de programação genética.

Nas Figuras 9 e 10 é possível observar exemplos de imagens de lesões de pele com melanoma e sem melanoma e lesões com suas segmentações feitas por especialistas, respectivamente.

### 3.3 Pré-processamento das Imagens

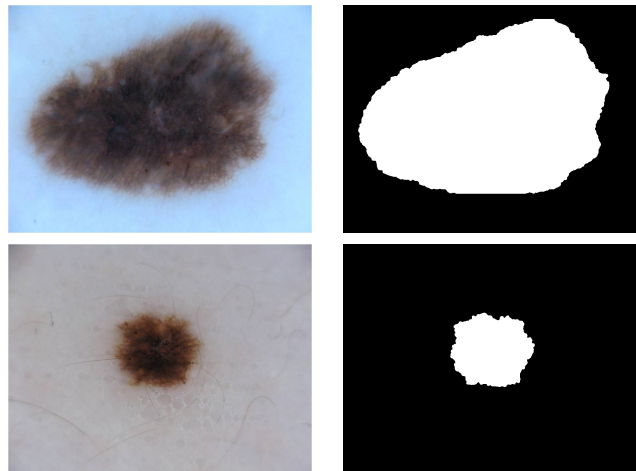
Após obter o conjunto de dados combinado entre ISIC 2019 e ISIC 2020, foi realizado um pré-processamento para organizar e padronizar as imagens antes do treinamento das redes neurais. Primeiramente, as imagens foram separadas em duas classes, **benignas** e

Figura 9 – Imagens com melanoma à esquerda e outras lesões de pele, não melanoma, à direita.



Fonte: ISIC (2019) e ISIC (2020)

Figura 10 – Imagens de melanoma do banco de imagens ISIC 2016 e suas segmentações feitas por profissionais médicos.



Fonte: ISIC (2016)

**malignas**, e distribuídas em três conjuntos distintos: **treino (70%)**, **validação (15%)** e **teste (15%)**, garantindo uma divisão equilibrada e reproduzível através da definição de uma semente aleatória.

As imagens foram redimensionadas para o tamanho de  $224 \times 224$  pixels, o que é um requisito comum para redes neurais convolucionais (CNNs). Para garantir que todas as imagens tivessem as mesmas dimensões, foi adicionado um fundo preto nas imagens menores. Essa abordagem de redimensionamento com adição de fundo preto foi adotada para evitar deformações nas imagens, preservando suas proporções originais. Caso a imagem não tenha o tamanho exato de  $224 \times 224$ , ela foi centralizada e margens pretas foram adicionadas nas laterais ou nas extremidades da imagem. Essa técnica assegura que as características visuais das lesões de pele não sejam distorcidas, proporcionando um conjunto de dados padronizado e adequado para a etapa de treinamento.

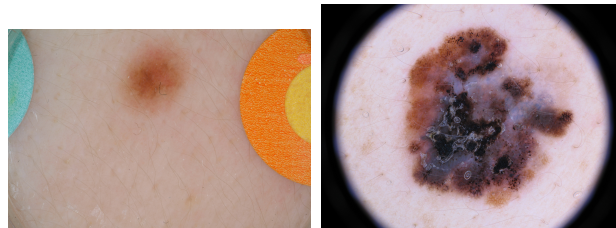
A distribuição das imagens nos conjuntos de dados ficou da seguinte forma:

- **Conjunto de Teste:** 952 imagens benignas / 767 imagens malignas.
- **Conjunto de Validação:** 951 imagens benignas / 765 imagens malignas.
- **Conjunto de Treinamento:** 4440 imagens benignas / 3574 imagens malignas.

Essa organização permitiu um balanceamento adequado entre as classes, garantindo que os modelos treinados fossem capazes de aprender padrões relevantes para a detecção de melanomas e sua diferenciação em relação a lesões benignas.

Além do conjunto de dados combinado ISIC 2019-2020, foi utilizado o conjunto de dados ISIC 2016 para otimizar a identificação de índices por meio de programação genética. No entanto, algumas imagens do ISIC 2016 foram removidas do conjunto de dados, pois apresentavam **adesivos coloridos** ou **regiões escuras** resultantes de equipamentos dermatológicos. Essa filtragem foi realizada com o objetivo de otimizar o processo de segmentação baseado no algoritmo K-Means, garantindo que a extração dos padrões de cor fosse mais precisa e reduzindo interferências que poderiam comprometer a identificação de características relevantes nas lesões. A Figura 11 mostra duas imagens que foram removidas do conjunto de imagens devido a essas características.

Figura 11 – Exemplo de imagens com adesivos e com sombras provocadas por equipamentos dermatológicos que foram retirados do treinamento.



Fonte: ISIC (2016)

### 3.4 Avaliação de Redes Pré-Treinadas

Na segunda etapa, foram testadas redes CNNs pré-treinadas aplicadas diretamente sobre as imagens originais do banco de dados, bem como sobre imagens transformadas pelos índices mais relevantes descritos na literatura. O desempenho das redes foi medido utilizando a métrica *F1-score*, permitindo uma comparação entre a abordagem tradicional e a modificação baseada nos índices.

Para a avaliação, foram utilizadas três arquiteturas de redes neurais convolucionais amplamente conhecidas e eficientes na extração de características de imagens médicas:

- **MobileNetV2:** É uma rede neural convolucional otimizada para dispositivos móveis, que foi projetada para equilibrar eficiência computacional e precisão. Ela utiliza uma abordagem de convolução separável em profundidade (*depthwise separable convolutions*), que reduz significativamente o número de parâmetros e operações computacionais, enquanto mantém um desempenho competitivo. Além disso, a MobileNetV2 introduz blocos de convolução invertida com conexões residuais (*inverted residual blocks*), que ajudam a preservar informações detalhadas, mesmo com a redução de parâmetros. Esses blocos tornam a rede mais eficiente em termos de recursos computacionais, sem comprometer o desempenho. Essa arquitetura foi projetada para ser leve e adequada a dispositivos com restrições de hardware, o que a torna ideal para aplicações em tempo real em dispositivos móveis e sistemas embarcados (HOWARD et al., 2017).
- **ResNet50:** É uma arquitetura profunda composta por 50 camadas, que resolve o problema de redes neurais muito profundas, como o desaparecimento do gradiente, por meio do uso de conexões residuais (*skip connections*). Essas conexões permitem que as informações fluam diretamente entre as camadas mais profundas, facilitando o aprendizado e melhorando a estabilidade da rede. A ResNet50 é amplamente utilizada em várias tarefas de visão computacional devido à sua capacidade de aprender representações hierárquicas e ricas. Com a introdução dessas conexões, a ResNet50 permite que redes mais profundas sejam treinadas sem a degradação do desempenho, o que a torna uma excelente escolha para problemas de classificação de imagens complexas, como a detecção de melanoma (HE et al., 2015).
- **NasNetMobile:** A arquitetura NasNetMobile foi criada usando a técnica de *Neural Architecture Search* (NAS), que explora automaticamente diferentes configurações de redes neurais para encontrar as mais eficientes, balanceando desempenho e custo computacional. A versão utilizada neste estudo foi a NasNetMobile, que é uma versão otimizada da arquitetura NasNet, voltada para dispositivos móveis e com limitações de recursos computacionais. A arquitetura combina camadas convolucionais com normalização e operações de redução de parâmetros para aumentar a eficiência, tornando-a adequada para dispositivos com capacidade computacional limitada. O modelo foi implementado utilizando o *TensorFlow Hub* para carregar a camada pré-treinada do NasNetMobile (ZOPH et al., 2018).

Cada uma das redes foi avaliada em três cenários distintos: (i) utilizando as imagens originais (no espaço de cor RGB) do conjunto de dados, (ii) aplicando os índices de marrom provenientes da literatura nas imagens, e (iii) utilizando os novos índices obtidos por meio de programação genética. Dessa forma, foi possível analisar a influência desses

índices na classificação das lesões cutâneas e verificar se a modificação das imagens contribuiu positivamente para o desempenho das Redes Neurais Convolucionais (CNNs).

Na metodologia adotada, foram utilizados os índices de marrom recomendados por Filho et al. (2023), que foram aplicados aos canais de cor RGB das imagens dermatoscópicas. Cada canal (R, G, B) foi modificado com base em um índice de marrom específico, conforme descrito abaixo:

- O canal *R* foi modificado utilizando o índice de *BI Aimonino*, com a constante *k* igual a 2. Este índice foi escolhido para realçar características relacionadas à tonalidade marrom nas imagens, especialmente nas regiões associadas ao melanoma.
- O canal *G* foi modificado com o índice *BI Lunadei2*, que também é um índice de marrom, mas com um foco diferente, destacando variações específicas de tonalidade que são relevantes para a diferenciação entre pele saudável e lesões.
- O canal *B* foi modificado utilizando o índice *BI Fetuga*, outro índice de marrom que atua na realce das áreas de interesse na imagem, facilitando a visualização das lesões de pele.

Esses índices de marrom, comumente utilizados na indústria alimentícia para destacar características específicas, foram adaptados para melhorar a segmentação das imagens médicas. O uso desses índices permitiu que as áreas com maior concentração de marrom fossem mais facilmente destacadas, o que facilita a identificação das lesões cutâneas. Os resultados obtidos com a aplicação desses índices foram comparados com os resultados obtidos pelos novos índices desenvolvidos por meio de programação genética, permitindo uma análise mais profunda da eficácia de cada abordagem na classificação das lesões de pele.

As redes neurais foram treinadas utilizando os seguintes hiperparâmetros:

- **Tamanho do batch:** 32 imagens por iteração.
- **Número de épocas:** 50.
- **Otimizador:** Adam, com taxa de aprendizado de  $3 \times 10^{-4}$ .
- **Função de perda:** *Binary Crossentropy*.
- **Métricas:** *F1-score*, Acurácia, Precisão, Sensibilidade.

- **Callback de Early Stopping:** Para evitar *overfitting*, o treinamento é interrompido se não houver melhoria no *F1-score* da validação após 10 épocas.
- **Callback de Redução da Taxa de Aprendizado:** Se o *F1-score* da validação não melhorar por 5 épocas consecutivas, a taxa de aprendizado é reduzida em 10%.

Além disso, antes de alimentar as imagens nas redes neurais, os valores dos pixels foram normalizados para o intervalo  $[0, 1]$  dividindo cada valor de pixel por 255. Isso é uma prática comum no treinamento de redes neurais, pois facilita o aprendizado, garantindo que os valores estejam em uma escala adequada para a entrada nas redes convolucionais.

Neste processo, as primeiras camadas das redes foram mantidas congeladas para evitar o ajuste dos seus pesos, permitindo que o modelo aproveitasse o aprendizado previamente adquirido, enquanto as camadas superiores (camadas densas finais) foram re-treinadas para aprender padrões específicos das imagens dermatoscópicas. A camada densa final consistiu em 1 neurônio com ativação sigmoide, e a função de perda utilizada foi a *binary crossentropy*, apropriada para problemas de classificação binária. Essa abordagem permite aproveitar representações genéricas extraídas de imagens naturais e adaptá-las para o diagnóstico de lesões cutâneas, acelerando a convergência do modelo e melhorando seu desempenho.



## 4 RESULTADOS

### 4.1 Índices Obtidos pela Programação Genética

Após a execução da programação genética, foram obtidos os índices com foco nas equações com operações simples e operações com um maior nível de complexidade. A heurística utilizada para encontrar os índices foi baseada na combinação de operações matemáticas entre os canais de cor das imagens, como soma, subtração, multiplicação, divisão, raiz quadrada e potência. A programação genética explorou essas operações em diferentes configurações para gerar equações que otimizassem a segmentação das lesões cutâneas. As melhores soluções foram selecionadas com base em sua capacidade de maximizar a diferenciação entre as áreas de melanoma e as áreas saudáveis, utilizando o F1-score como métrica de avaliação. Nas subseções seguintes, estão os índices encontrados pela programação genética que alcançaram os melhores resultados de segmentação do conjunto de dados ISIC 2016.

#### 4.1.1 Índices com Operações Simples

- **IC1:**  $2R - 3B + G$ : este índice combina as componentes de cor vermelha (R), azul (B) e verde (G) com coeficientes lineares. A fórmula destaca a relação entre as cores, em que o vermelho é multiplicado por 2, o azul por -3 e o verde por 1, o que ajuda a enfatizar as características das áreas com maior intensidade de cor marrom.
- **IC2:**  $\frac{2R+1,47}{R+0,73}$ : este índice ajusta a relação entre as intensidades do vermelho (R) e inclui uma constante, visando destacar a presença de marrom nas regiões da imagem. A fórmula sugere que uma combinação de R com uma constante pode ajudar a identificar as regiões relevantes.
- **IC3:**  $\frac{0,73}{2R-B+1,46}$ : aqui a relação entre o vermelho (R) e o azul (B) é novamente enfatizada. O valor constante (0.73) serve para ajustar a escala e melhorar a precisão do índice, enquanto a fórmula destaca a diferença de intensidade entre o vermelho e o azul.

Esses índices foram aplicados às imagens para se obter o valor dos canais *RGB* da imagem a ser usada no treinamento e teste da rede neural. Desta forma, na aplicação de índices

simples, foram montadas imagens onde a camada  $R$  era representada pelo resultado do índice  $IC1$ , a camada  $G$  pelo resultado do índice  $IC2$ , e a camada  $B$  pelo resultado do índice  $IC3$ .

#### 4.1.2 Índices com Maior Nível de Complexidade

- **IC4:**  $\sqrt{|0,5964 + R - B + G|}$ : este índice aplica uma transformação utilizando a raiz quadrada, combinando as componentes  $R$ ,  $B$  e  $G$ . A inclusão da raiz quadrada visa suavizar os resultados, destacando de forma mais eficiente os tons de marrom nas imagens, aumentando a sensibilidade nas áreas de transição entre diferentes tons de cor.
- **IC5:**  $\sqrt{\sqrt{|(2R + 0,44 - B)|}}$ : um índice mais complexo que utiliza raízes quadradas duplas. A fórmula realça a relação entre o vermelho e o azul, com um ajuste adicional de 0.44. Esse índice pode ser particularmente útil para identificar áreas com contraste acentuado entre o vermelho e o azul, típicos em lesões de pele com tons de marrom.
- **IC6:**  $\sqrt{\left|2R + 0,11 \cdot \left(0,95 \cdot \sqrt{|B - G| + 0,07} \cdot \sqrt{0,29 + R + 0,71}\right)\right|}$ : este índice envolve múltiplas operações complexas, incluindo a manipulação das diferenças entre as componentes de cor. Ele utiliza raízes quadradas e a função de valor absoluto para melhorar a precisão da detecção de marrom. A combinação dos coeficientes e a estrutura complexa visam refinar a sensibilidade do índice a nuances mais sutis de marrom nas imagens.

A mesma lógica de aplicação dos índices simples foi realizada para estes índices mais complexos para o treinamento e teste da rede neural.

#### 4.1.3 Resultados da Segmentação

A seguir, são apresentados os resultados da segmentação utilizando cada um dos índices descritos na Seção 4.1.

#### 4.1.4 F1 Score dos Índices Encontrados por Programação Genética

Os valores médios de  $F1\ Score$  obtidos por cada índice encontrado durante o processo de programação genética são apresentados na Tabela 2. Os resultados foram calculados a

partir da combinação das camadas de cor *RGB*, gerando um único canal de cor para a segmentação das imagens utilizando o algoritmo *K-means*, ou seja, as imagens usadas no *K-means* foram em nível de cinza.

Tabela 2 – *F1 Score* obtido para cada índice durante o processo de programação genética.

Índices	F1 Score
IC1	0,7081
IC2	0,7116
IC3	0,7229
IC4	0,7256
IC5	0,7267
IC6	0,7192

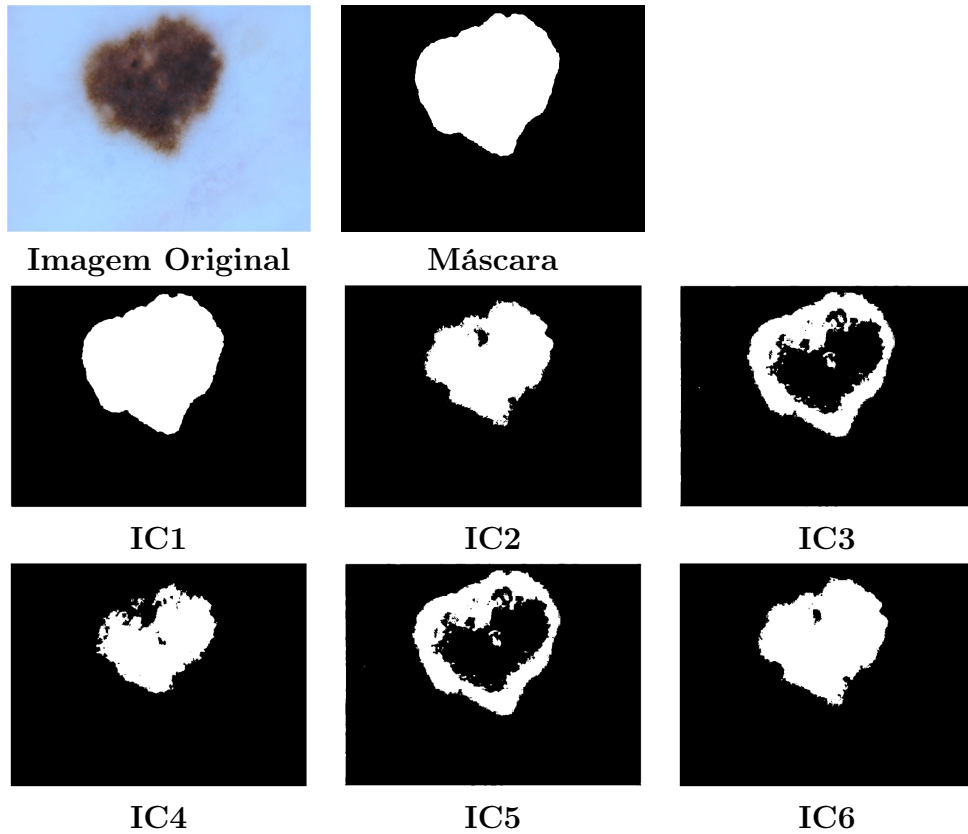
Observa-se que os índices IC5 e IC4 apresentaram os melhores resultados, com *F1 Scores* de 0,7267 e 0,7256, respectivamente, indicando que esses índices foram mais eficazes na diferenciação entre áreas de melanoma e pele saudável. Por outro lado, o índice IC1 teve o menor *F1 Score* de 0,7081, sugerindo que ele foi menos eficaz em capturar as características relevantes das lesões cutâneas. Em geral, os resultados indicam que índices mais complexos, como IC4, IC5 e IC6, apresentaram um melhor desempenho, o que pode refletir a maior capacidade desses índices em modelar as nuances de cor associadas ao melanoma. Isso sugere que a combinação de operações mais complexas pode ser vantajosa na segmentação de lesões dermatológicas, como o melanoma.

As segmentações obtidas utilizando os índices IC1 a IC6 são apresentadas na Figura 12, em uma imagem exemplo do conjunto de dados ISIC (2016). Na primeira linha, à esquerda, é exibida a imagem original do melanoma e, à direita, é exibida a máscara correspondente, usada como referência para a segmentação correta da lesão. Nas demais linhas, as imagens segmentadas por cada índice são mostradas: IC1, IC2, IC3, IC4, IC5 e IC6.

A técnica de binarização utilizada para a geração das máscaras foi baseada no algoritmo de limiarização, onde o valor do pixel na imagem transformada por cada índice foi comparado com um valor de limiar específico. Os pixels que apresentaram valores acima do limiar foram considerados pertencentes à lesão (melanoma), enquanto os abaixo do limiar foram considerados como pertencentes à região saudável. Essa técnica de binarização foi aplicada a cada imagem segmentada, gerando uma máscara binária que representa a área da lesão de forma destacada.

É possível observar, neste exemplo, que os índices IC1, IC2 e IC6 apresentam uma segmentação mais precisa, com a área de melanoma melhor destacada. Por outro lado, os índices IC3 e IC5 tendem a gerar segmentações mais imprecisas, com áreas no interior da lesão segmentadas como pele saudável.

Figura 12 – Segmentação obtida utilizando cada índice.



Fonte: (ISIC, 2016) adaptado.

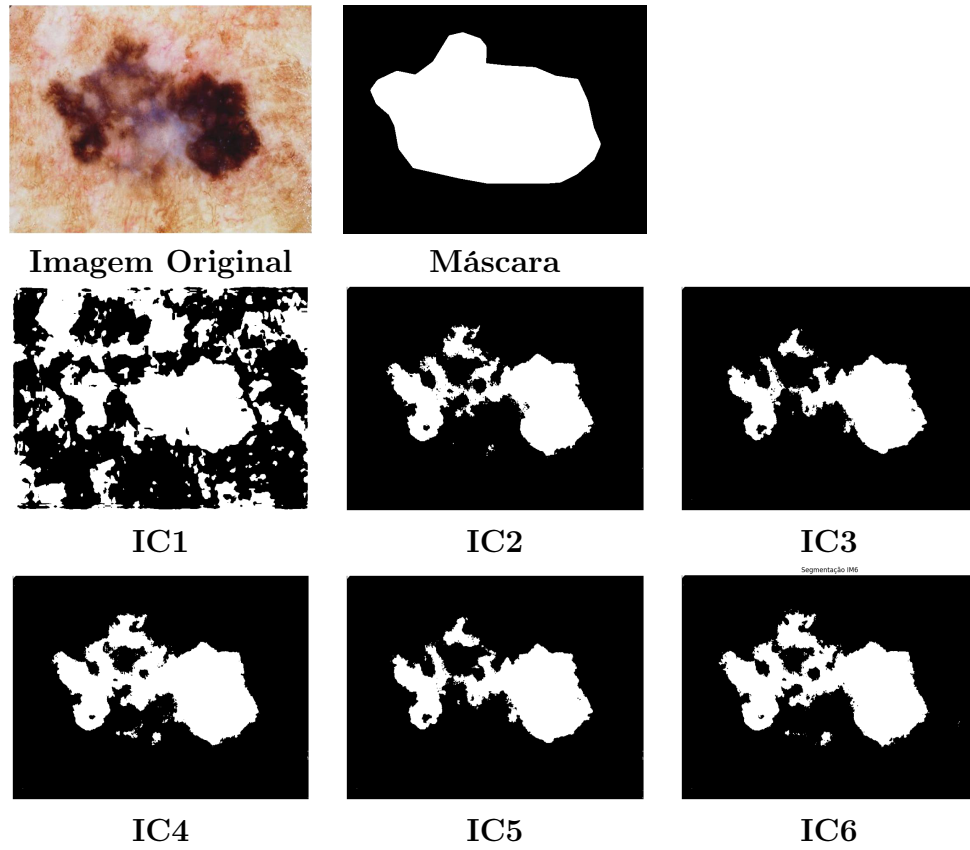
Outro exemplo pode ser observado na Figura 13. Neste caso, observa-se que a segmentação obtida a partir do IC1 não condiz com a máscara disponível, diferente do exemplo anterior. Não somente, é possível observar que a máscara disponibilizada pelo profissional médico não possui uma delimitação com grande resolução, de modo que nenhuma das imagens obtidas a partir dos índices atendeu de forma satisfatória, visualmente, a máscara.

## 4.2 Resultados da Classificação

As Subseções 4.2.1, 4.2.2 e 4.2.3 apresentam, respectivamente, os resultados dos testes realizados com as imagens em *RGB*, com os índices da literatura e com os novos índices encontrados pela programação genética. Os resultados foram obtidos após 10 execuções de cada rede para cada tipo de imagem.

Na Figura 14 é apresentado um exemplo de aplicação dos conjuntos de índices para uma imagem de melanoma. A imagem “Original” é a imagem no espaço de cor *RGB* sem alterações. A imagem “IMs Literatura” é a imagem usando os índices da literatura, na qual o canal *R* é o índice de *BI Aimonino*, o canal *G* é o índice de *BI Lunadei2* e o canal

Figura 13 – Segmentação obtida utilizando cada índice.



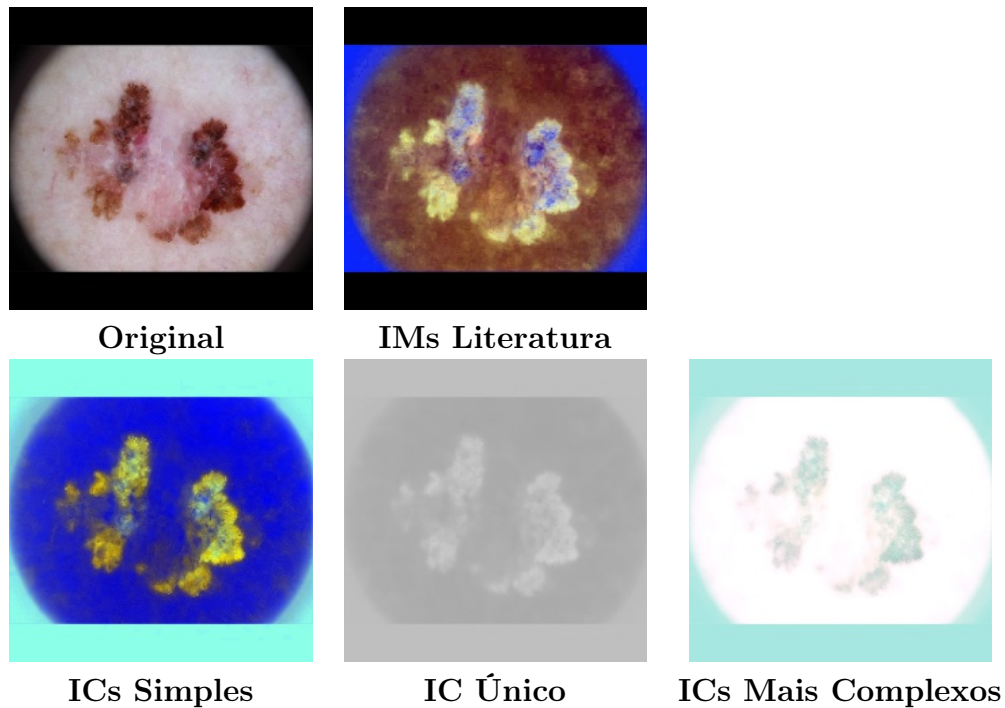
Fonte: (ISIC, 2016) adaptado.

$B$  é o índice de *BI Fetuga*. A imagem “ICs Simples” é a imagem usando os índices simples encontrados, na qual o canal  $R$  é o índice IC1, o canal  $G$  é o índice IC2 e o canal  $B$  é o índice IC3. A imagem “IC Único” é uma imagem cujos três canais da imagem usaram o mesmo índice, no caso o índice IC3, para verificar se haveria muita perda de informação ao ser utilizado apenas um índice. Por fim, a imagem “ICs Mais Complexos” é a imagem usando os índices mais complexos encontrados, na qual o canal  $R$  é o índice IC4, o canal  $G$  é o índice IC5 e o canal  $B$  é o índice IC6.

Como é possível observar na Figura 14, cada conjunto de índices resulta em uma modificação distinta da imagem original, e as diferenças são visíveis nos resultados:

- **Imagem Original:** a imagem sem qualquer modificação mostra o melanoma com suas características naturais, incluindo as cores originais. Nesta imagem é possível observar as áreas de interesse em termos de tonalidade e contrastes.
- **IMs Literatura:** neste caso, a aplicação dos índices de marrom realça áreas específicas do melanoma, como os pontos de maior contraste de cor marrom. Esse índice pode ser útil para destacar a região do melanoma em comparação com o fundo, mas

Figura 14 – Imagens originais e as mesmas com os índices aplicados nas camadas *RGB*.



pode ter dificuldades em distinguir áreas com características semelhantes em cor, levando a uma segmentação um pouco borrada.

- **ICs Simples:** Os índices simples aplicados nos canais *R*, *G* e *B* podem resultar em uma imagem com realce nas áreas de marrom, como o melanoma, mas de uma forma mais generalizada. A aplicação desses índices pode resultar em uma maior saturação das cores, como visto nas áreas azuis e amarelas que destacam regiões específicas. Esses índices podem ser mais sensíveis a variações de cor, mas podem ser mais imprecisos em regiões com tonalidades semelhantes.
- **IC Único (*IC3* aplicado em todos os canais):** O uso do mesmo índice (*IC3*) para todos os canais *R*, *G* e *B* tem como efeito uma modificação uniforme em todas as camadas de cor. O resultado tende a ser uma imagem mais homogênea, com menor contraste e com maior ênfase na cor predominante, gerando uma segmentação mais simplificada e com menos nuances comparada aos ICs aplicados de maneira independente. O *IC3* foi escolhido pois foi o índice que apresentou o melhor resultado de segmentação entre os índices mais simples.
- **ICs Mais Complexo:** A aplicação de índices mais complexos (*IC4*, *IC5* e *IC6*) apesar da grande quantidade de brilho para observação de um ser humano, gera uma imagem onde as diferenças de cor e contraste são mais evidentes. As operações matemáticas avançadas, como raízes quadradas e potências, permitem uma modelagem mais expressiva, melhorando a definição das áreas com características específicas,

como o melanoma. A imagem resultante, com mais contraste e detalhes, destaca melhor as regiões de interesse, como o melanoma, com mais precisão.

As diferenças entre as imagens refletem como a aplicação de diferentes índices pode influenciar a visualização das características do melanoma.

#### 4.2.1 Resultados com as Redes Pré-Treinadas e Imagens Originais

Primeiramente, foram testados os modelos de redes neurais pré-treinadas utilizando imagens originais em *RGB*, sem nenhuma modificação. Os modelos utilizados foram MobileNet, ResNet e NasNet, que foram avaliados em termos de acurácia, precisão, sensibilidade, *F1 Score* e Desvio Padrão do *F1 Score* (DP).

Os resultados das redes pré-treinadas com as imagens originais estão apresentados na Tabela 3. O Mobilenet obteve a melhor acurácia (0,9315) e precisão (0,9471), embora sua sensibilidade (0,8965) fosse ligeiramente menor, indicando que, apesar de ser preciso, o modelo teve dificuldades em identificar todos os casos positivos. A ResNet, embora com desempenho um pouco inferior em relação ao Mobilenet, obteve boa precisão (0,9346) e sensibilidade (0,8782). O NasNet também alcançou um desempenho relevante, com uma acurácia de 0,9302 e uma sensibilidade de 0,9053, o que o coloca como o modelo com melhor capacidade de identificar casos positivos entre os três.

#### 4.2.2 Resultados com os Índices de Marrom da Literatura

Em seguida, os índices de *BI Aimonino*, *BI Lunadei2* e *BI Fetuga*, sugeridos por Filho et al. (2023), foram aplicados aos canais *R*, *G* e *B*, respectivamente, das imagens de melanoma, e os resultados obtidos para cada rede neural são apresentados na Tabela 3.

Todos os modelos alcançaram um ligeiro aumento em sensibilidade e precisão ao serem aplicados em imagens com o IM da Literatura aplicado, exceto a Nasnet que obteve um ligeiro decremento na métrica sensibilidade. Como consequência, houve uma ligeira melhora na medida *F1 Score* de todas as redes neurais.

Os desvios padrão do *F1 Score* apresentados na 3 indicam a variabilidade dos resultados em função dos testes realizados. Eles foram calculados a partir de 10 execuções distintas de cada modelo, fornecendo uma medida de dispersão das pontuações de *F1* em relação à

Tabela 3 – Resultados médios dos modelos para diferentes tipos de imagens. Os melhores resultados para cada métrica e tipo de rede estão realçados em negrito.

Modelo	Acurácia	Precisão	Sensibilidade	F1 Score	DP F1 Score
<b>Original</b>					
Mobilenet	0,9315	0,9471	0,8965	0,9211	0,0051
Resnet	0,9182	0,9346	0,8782	0,9055	0,0146
Nasnet	0,9302	0,9369	<b>0,9053</b>	0,9206	0,0081
<b>IM Literatura</b>					
Mobilenet	0,9330	0,9484	0,8986	0,9228	0,0028
Resnet	0,9225	0,9391	0,8836	0,9105	0,0040
Nasnet	0,9327	0,9478	0,8986	0,9225	0,0036
<b>IC Simples</b>					
Mobilenet	<b>0,9381</b>	<b>0,9577</b>	<b>0,9012</b>	<b>0,9285</b>	<b>0,0017</b>
Resnet	<b>0,9247</b>	<b>0,9450</b>	<b>0,9027</b>	<b>0,9128</b>	0,0044
Nasnet	<b>0,9359</b>	<b>0,9539</b>	0,8999	<b>0,9261</b>	0,0031
<b>IC Único</b>					
Mobilenet	0,9220	0,9385	0,8874	0,9122	0,0030
Resnet	0,9112	0,9172	0,8816	0,8987	0,0075
Nasnet	0,9280	0,9490	0,8939	0,9172	0,0051
<b>IC Mais Complexo</b>					
Mobilenet	0,9311	0,9501	0,8926	0,9204	0,0040
Resnet	0,9196	0,9400	0,8759	0,9067	<b>0,0028</b>
Nasnet	0,9319	0,9487	0,8960	0,9216	<b>0,0029</b>

média obtida. De forma geral, os desvios padrão são relativamente baixos, o que indica que os modelos apresentaram um desempenho estável e consistente em todas as execuções. O modelo Mobilenet, por exemplo, teve, na maioria dos casos, os menores desvios padrão nas métricas *F1 Score*, o que sugere uma maior confiabilidade em seus resultados ao longo dos testes. Em contrapartida, o modelo Resnet apresentou um leve aumento nos desvios padrão em comparação com o Mobilenet, indicando uma variação um pouco maior nos resultados.

A Tabela 4 mostra o aumento relativo no *F1 Score* quando os índices de marrom foram aplicados, com destaque para a Resnet, que obteve um aumento de **0,56%**. O NasNet também apresentou um aumento, com **0,21%**, enquanto o Mobilenet teve um aumento de **0,19%**.

Tabela 4 – Aumento Relativo na Utilização dos Índices de Marrom em Comparação à Não Utilização de Índice

Modelo	IM Literatura(%)	IC Simples(%)	IC Único(%)	IC Mais Complexo(%)
Mobilenet	0,19	0,81	-0,96	-0,07
Resnet	0,56	0,80	-0,75	0,14
Nasnet	0,21	0,60	-0,37	0,11

Para medir a melhoria no desempenho dos índices, foi calculado o **Aumento Relativo**



do F1 Score, utilizando a seguinte fórmula:

$$AumentoRelativo(\%) = \left( \frac{F1ScoreMedio\_IC - F1ScoreMedio}{F1ScoreMedio} \right)$$

Onde:

- *F1ScoreMedio\_IC* representa o F1 Score médio obtido utilizando os índices de cor (IC).
- *F1ScoreMedio* representa o F1 Score médio obtido utilizando os índices originais ou padrão.

Essa fórmula foi aplicada para calcular a melhoria no desempenho das segmentações das lesões de pele, permitindo uma análise detalhada de como os índices gerados por Programação Genética (PG) influenciam a acurácia da classificação de melanomas em comparação aos índices tradicionais.

#### 4.2.3 Resultados com os Índices Obtidos

Na sequência, foram aplicados os índices obtidos por meio de programação genética a fim de verificar se essas modificações poderiam gerar resultados ainda melhores. A aplicação desses índices teve um efeito positivo na segmentação das lesões de melanoma, com um aumento no F1 *Score* para a maioria de suas utilizações, quando comparadas às imagens sem a utilização dos índices.

O conjunto de índices simples (resultados de *IC1*, *IC2* e *IC3* para os canais *RGB*, respectivamente) obteve os melhores resultados, com um aumento do F1 *Score* em todos os modelos, com destaque para o Mobilenet, que obteve um aumento de **0,81%**. O Resnet também apresentou um aumento, com **0,80%**, enquanto o Nasnet teve um aumento de **0,60%**. O Mobilenet se destaca com precisão de 0,9577 e sensibilidade de 0,9012, o que sugere que o modelo é bem eficaz em identificar tanto as instâncias positivas quanto as negativas, com F1 *Score* de 0,9285. O Nasnet, por sua vez, obteve precisão de 0,9539 e sensibilidade de 0,8999, um desempenho bom comparado aos outros modelos.

Para as imagens onde foram aplicados o IC Único (*IC3*) para resultar nas três camadas *RGB*, Nasnet apresentou a maior acurácia (0,9280), e Mobilenet manteve um bom desempenho

com precisão de 0,9385 e sensibilidade de 0,8874. O Resnet teve um desempenho mais modesto em comparação com os outros, com uma sensibilidade de 0,8816 e precisão de 0,9172. O experimento obteve um resultado inferior, já que os valores encontrados para o *F1 Score* foram menores que aqueles encontrados nos experimentos sem a utilização de índice, como mostra a Tabela 4, onde os valores relativos de aumento para Mobilenet, Resnet e Nasnet, foram **-0,96%**, **-0,75%** e **-0,37%**, respectivamente.

Finalmente, no caso das imagens onde foi aplicado o conjunto com IC mais complexo, Mobilenet e Nasnet apresentaram resultados muito semelhantes, com acurácia de 0,9311 e 0,9319, respectivamente, e *F1 Score* em torno de 0,9204 a 0,9216. O Resnet apresentou uma acurácia de 0,9196, que ainda é boa, mas um pouco inferior aos outros dois modelos. A sensibilidade para Resnet também é o mais baixo entre os três. Quando observadas as relações de *F1 Score* da Tabela 4, o Mobilenet obteve uma leve redução de **-0,07%** no *F1 Score*, enquanto o NasNet e o ResNet apresentaram aumentos modestos de **0,11%** e **0,14%**, respectivamente.

### 4.3 Discussão dos Resultados

De forma geral, foi observado que existe uma ligeira melhora na classificação em relação às imagens em RGB ao aplicar os índices de marrom da literatura nas imagens, e uma melhora maior ao aplicar os índices simples obtidos pela programação genética.

Embora os índices obtidos pela programação genética tenham melhorado a segmentação, a melhoria foi mais significativa nos índices simples e menos complexos. Isso sugere que os índices mais simples podem ser mais eficazes para a detecção de melanoma, provavelmente devido à sua maior sensibilidade e melhor relação entre precisão e custo computacional. Além disso, é possível que os índices mais complexos tenham sofrido de *overfitting*, se adaptando demais a detalhes e ruídos da base de dados ISIC (2016) e não conseguindo uma boa generalização para as imagens das bases de dados ISIC (2019) e ISIC (2020).

Os índices de marrom da literatura e os encontrados possuem algumas semelhanças, principalmente no uso do canal vermelho (*R*). O canal *R* está presente em todos os índices apresentados, seja diretamente ou indiretamente, pois ele é crucial na definição das regiões com tonalidades marrom. O uso do canal *R* é uma característica comum em muitos dos índices estudados, já que o vermelho é uma cor predominante na identificação de melanomas e outras características cutâneas em imagens dermatológicas.

A principal diferença entre os índices encontrados e os da literatura é que os índices

literários apresentam variações de complexidade, utilizando diferentes combinações entre os canais de cor (como  $R$ ,  $G$  e  $B$ ) e outros parâmetros (como  $L^*$ ,  $a^*$  e  $b^*$  no espaço de cor  $Lab^*$ ), proporcionando uma maior expressividade para modelar a cor marrom de forma mais precisa. Já os índices encontrados que retornaram os melhores resultados possuem uma aplicação mais direta nos canais  $RGB$ , com operações mais simples e focadas principalmente nas operações com os canais  $R$  e  $B$ .

## 5 CONCLUSÃO E TRABALHOS FUTUROS

### 5.1 Conclusões

Este trabalho teve como foco explorar o uso de índices de marrom combinados com redes neurais profundas para a tarefa de detecção de câncer de pele. Além disso, foi proposto o uso de programação genética para encontrar índices de cor que sejam mais adequados para esta tarefa.

Nos experimentos foram utilizados três conjuntos de dados públicos de competição internacional de câncer de pele, a ISIC, e três redes neurais profundas previamente treinadas com a base de dados ImageNet: Mobilenet, Resnet e Nasnet.

As redes neurais que utilizaram as imagens com os índices mais simples que foram descobertos obtiveram melhorias consistentes no *F1 Score*, indicando que essa abordagem é eficaz para uma pequena melhoria na classificação. Por outro lado, quando utilizados índices mais complexos, que envolvem raízes e exponenciações, como o conjunto *IC Mais Complexo*, a melhora não foi consistente e nem tão boa quanto os índices mais simples.

Também foi possível observar que a utilização de um mesmo índice para todos os canais *RGB* provoca uma perda de informação e, conseqüentemente, uma redução do *F1 Score*. Esse resultado pode ser explicado pela perda de nuances e variações importantes nos diferentes canais de cor. Ao aplicar o mesmo índice para os três canais, a imagem resultante tende a se tornar mais homogênea, com as características distintas das diferentes áreas da lesão e da pele saudável sendo suavizadas. Isso pode fazer com que a segmentação da lesão fique menos precisa, já que a diferenciação entre melanoma e áreas saudáveis fica mais difícil. Além disso, pode-se inferir que, nesse caso, a imagem se torna mais "transparente", perdendo informações cruciais que seriam capturadas se os índices fossem aplicados de maneira específica para cada canal de cor, mantendo a integridade das características visuais em cada uma das componentes RGB.

A partir dos resultados obtidos, é possível destacar que o Mobilenet demonstrou ser o modelo mais consistente, apresentando bons resultados em diversas métricas (precisão, sensibilidade, *F1 Score*), especialmente nas imagens onde foram aplicados os índices de cor mais simples obtidos pela programação genética. O Nasnet também alcançou resultados bons em precisão e *F1 Score*, especialmente em imagens que utilizaram de

conjuntos de índices encontrados pela programação genética. O Resnet, embora eficiente, teve dificuldades em algumas métricas como sensibilidade e precisão quando comparado aos outros dois modelos.

## 5.2 Trabalhos Futuros

Apesar dos resultados promissores obtidos com a utilização dos índices em modelos de redes neurais para a classificação de imagens dermatológicas, diversos aspectos podem ser aprimorados em pesquisas futuras.

### 5.2.1 Métodos para Evitar o *Overfitting* da Programação Genética

Foi observado que o uso de operadores mais complexos aumentou a suscetibilidade de ocorrerem *overfitting* nos índices. Técnicas podem ser empregadas para evitar este problema, como o uso de conjuntos de dados para treino e validação, limitação da profundidade das árvores construídas pela Programação Genética, entre outros.

### 5.2.2 Aprimoramento da Programação Genética

A Programação Genética utilizada neste trabalho pode ser aprimorada. Uma possibilidade seria a utilização de outras operações matemáticas, como funções logarítmicas e operações mais complexas (ex. operações baseadas em transformações geométricas ou topológicas), para aumentar a capacidade de modelagem e expressividade das equações geradas. Essas operações poderiam permitir uma maior flexibilidade na modelagem das relações entre as camadas de cor das imagens, resultando em índices mais robustos para a segmentação e classificação de lesões.

No entanto, é importante refletir sobre os custos associados ao aumento da complexidade das equações. Embora a introdução de operações mais complexas possa melhorar a expressividade e a capacidade de modelagem dos índices, ela também pode levar a um aumento significativo no custo computacional, similar ao que ocorre com redes neurais mais complexas. Redes mais profundas, por exemplo, tendem a exigir mais poder de processamento e mais tempo de treinamento. De forma análoga, a inclusão de operações matemáticas mais sofisticadas na Programação Genética pode gerar índices que exigem mais recursos computacionais para a sua avaliação e aplicação, o que pode não ser vantajoso

em cenários de tempo real ou quando recursos computacionais são limitados. Portanto, é necessário avaliar se o ganho em desempenho justifica o aumento no custo computacional, especialmente em aplicações práticas que exigem rapidez e eficiência.

### 5.2.3 Utilização do Índice de Jacard como Alternativa ao F1 Score

O índice de Jacard é uma métrica alternativa ao F1 Score para avaliação da segmentação, calculando a similaridade entre dois conjuntos baseados na interseção e união deles. No contexto de segmentação de imagens, o índice de Jacard mede a sobreposição das áreas segmentadas e reais, sendo útil para avaliar a precisão da segmentação. Ao contrário do F1 Score, que combina precisão e recall, o índice de Jacard oferece uma avaliação direta da sobreposição das regiões de interesse, sendo especialmente eficaz em casos de desequilíbrio entre as classes. Como métrica complementar, o índice de Jacard pode fornecer uma análise mais robusta do desempenho do modelo de segmentação.

### 5.2.4 Exploração de Novos Índices de Cor

Uma possível direção seria a exploração de outros índices de cor além do índice de marrom, como os índices de cor baseados nas características espectrais da pele, que podem melhorar a segmentação de lesões e melanomas. A comparação entre diferentes índices de cor e suas combinações pode fornecer *insights* valiosos para otimizar o desempenho do modelo.

### 5.2.5 Aprimoramento da Aplicação dos Índices

Embora os índices tenham mostrado resultados positivos em alguns tipos de imagens, eles não foram eficazes para todos os tipos de imagem, especialmente para imagens mais complexas e com características mais sutis. Assim, o aprimoramento da aplicação dos índices, como a sua combinação com outros métodos de segmentação ou técnicas de aumento de dados, pode melhorar ainda mais os resultados.

### 5.2.6 Avaliação de Desempenho em Outros Conjuntos de Dados

Para garantir que os resultados obtidos sejam generalizáveis, seria interessante testar os modelos e as técnicas propostas em outros conjuntos de dados de imagens dermatológicas,

que podem ter características diferentes em relação ao conjunto de dados utilizado neste estudo. Isso ajudaria a validar a robustez dos modelos e das técnicas de processamento de imagem.

### **5.2.7 Aprimoramento das Arquiteturas de Redes Neurais**

Futuras investigações podem explorar o aprimoramento das arquiteturas de redes neurais utilizadas, como o uso de modelos mais avançados de redes de convolução profundas. Também seria interessante investigar o desempenho de redes neurais treinadas desde o princípio com imagens nos índices de marrom ou nos índices identificados neste trabalho para a tarefa de classificação dermatológica.

### **5.2.8 Implementação em Ambiente Clínico**

Por fim, a implementação dos modelos treinados em um ambiente clínico real, com validação e teste em imagens de pacientes reais, seria uma aplicação prática e crucial para o uso de redes neurais na detecção de melanoma e outras doenças dermatológicas. A colaboração com dermatologistas e médicos especializados seria fundamental para garantir que as ferramentas de inteligência artificial possam ser aplicadas de forma eficaz e segura no diagnóstico clínico.

Essas abordagens podem contribuir para o avanço contínuo da utilização de inteligência artificial na área de diagnóstico médico, proporcionando ferramentas mais precisas e acessíveis para os profissionais da saúde.

### **5.2.9 Utilização de Bancos de Dados para Comparação em Trabalhos Futuros**

Uma possível melhoria e extensão do presente trabalho envolve a utilização de bancos de dados adicionais para a comparação dos índices gerados com a Programação Genética. Atualmente, o modelo foi treinado e avaliado utilizando um conjunto específico de dados, mas para validar a robustez e a generalização dos índices propostos, seria interessante aplicá-los a outros bancos de dados amplamente utilizados na literatura.

Em trabalhos futuros, propõe-se a utilização dos seguintes bancos de dados para comparação:

- **HAM10000** (Codella et al. (2018)): Um banco de dados com imagens dermatoscópicas rotuladas, incluindo uma variedade de lesões benignas e malignas. Este conjunto de dados é amplamente utilizado em *benchmarks* de classificação de melanoma, proporcionando uma base sólida para comparar o desempenho dos índices de marrom propostos com outros modelos existentes.
- **Dermnet** (Zealand (2016)): Um banco de dados que contém uma grande variedade de imagens dermatoscópicas e macroscópicas, com foco em diferentes tipos de lesões cutâneas. A comparação com este banco de dados poderia ampliar as possibilidades de avaliação, considerando uma diversidade ainda maior de tipos de lesões.

Esses bancos de dados fornecem imagens de qualidade com anotações precisas, o que permite uma avaliação detalhada e uma comparação objetiva com modelos de estado da arte. Com a aplicação desses dados adicionais, será possível avaliar se os índices gerados pela Programação Genética têm desempenho consistente em diferentes contextos e se podem ser usados para melhorar a classificação e segmentação de lesões cutâneas em larga escala.



## REFERÊNCIAS

- AGGARWAL, C. C. Neural Networks and Deep Learning: A Textbook. [S.l.]: Springer, 2018. ISBN 978-3319944623. Citado 4 vezes nas páginas 15, 16, 17 e 18.
- AIMONINO, D. R.; BARGE, P.; COMBA, P. G.; OCCELLI, A.; TORTIA, C. Computer vision for laboratory quality control on frozen fruit. In: Chemical Engineering Transactions. [S.l.: s.n.], 2015. v. 44. Citado na página 27.
- CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. Computers & Electrical Engineering, v. 40, n. 1, p. 16–28, 2014. ISSN 0045-7906. 40th-year commemorative issue. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0045790613003066>>. Citado na página 14.
- CODELLA, N. C. F.; VACAVANT, A.; ZHOU, B.; LIU, M.; TANG, S.; SUN, L.; HWANG, M.; NGUYEN, H.; MANOGARAN, A.; MAHAPATRA, D. Skin lesion analysis toward melanoma detection: A challenge at the IEEE international symposium on biomedical imaging (ISBI) 2018. IEEE Transactions on Medical Imaging, v. 37, n. 12, p. 5103–5115, 2018. Citado na página 55.
- COSTA, E.; LORENA, A.; CARVALHO, A.; FREITAS, A. A review of performance evaluation measures for hierarchical classifiers. AAAI Workshop - Technical Report, 01 2007. Citado na página 25.
- EBRAHIMI, M. A.; MOHTASEBI, S. S.; RAFIEE, S.; HOSSEINPOUR, S. Using online image processing technique for measurement the browning in during drying (a new and automatic method). Agric Eng Int: CIGR Journal, v. 15, 2015. Citado na página 27.
- EIBEN, A. E.; SMITH, J. E. A survey of evolutionary algorithms for optimization. Computational Intelligence, v. 19, n. 1, p. 1–19, 2003. Citado na página 33.
- ESTEVA, A. e. a. Dermatologist-level classification of skin cancer with deep neural networks. Nature, v. 542, p. 115–118, 2017. Citado na página 10.
- FERREIRA, C.; SILVA, D.; SOUZA, F. Genetic programming for medical image processing. International Journal of Machine Learning and Computing, IEEE, v. 8, n. 5, p. 433–438, 2018. Citado na página 12.
- FETUGA, G.; TOMLINS, K.; HENSHAW, F. Effect of variety and processing method on functional properties of traditional sweet potato flour (“elubo”) and sensory acceptability of cooked paste (“amala”). Dissertação (Mestrado) — Federal University Of Agriculture, Abeokuta, 2014. Citado na página 28.
- FILHO, E. Q. S.; SALLES, E. O. T.; FACON, J.; CIARELLI, P. M. Browning index: An image processing technique to aid in the segmentation of skin lesions on dermatoscopic images. CBIC, CBIC, 2023. Citado 4 vezes nas páginas 12, 28, 38 e 46.
- FORTIN, F. et al. Deap: Distributed evolutionary algorithms in python. Journal of Machine Learning Research, v. 13, p. 2171–2176, 2012. Citado na página 23.

- GONZÁLEZ, R. C.; WOODS, R. E.; EDDINS, S. L. Digital Image Processing. [S.l.]: Pearson Prentice Hall, 2009. Citado na página 26.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge, MA: MIT Press, 2016. ISBN 978-0262035613. Citado na página 18.
- GÉRON, A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. [S.l.]: O'Reilly Media, 2019. Citado 3 vezes nas páginas 24, 25 e 26.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, p. 770–778, 2015. Citado 3 vezes nas páginas 16, 18 e 37.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [s.n.], 2016. p. 770–778. Disponível em: <<https://arxiv.org/abs/1512.03385>>. Citado na página 19.
- HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; CHEN, L.; DAVID, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017. Citado 2 vezes nas páginas 18 e 37.
- INCA. Estimativa 2023 - Incidência de Câncer no Brasil. 2022. Disponível em: <<https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil>>. Acesso em: 02 dez. 2023. Citado na página 10.
- ISIC, I. S. I. C. ISIC 2016 Challenge Dataset. 2016. Accessed: 04/12/2023. Disponível em: <<https://challenge.isic-archive.com/landing/2016/>>. Citado 7 vezes nas páginas 33, 35, 36, 42, 43, 44 e 49.
- ISIC, I. S. I. C. ISIC 2019 Challenge Dataset. 2019. Accessed: 04/12/2023. Disponível em: <<https://challenge.isic-archive.com/landing/2019/>>. Citado 4 vezes nas páginas 13, 33, 35 e 49.
- ISIC, I. S. I. C. ISIC 2020 Challenge Dataset. 2020. Accessed: 04/12/2023. Disponível em: <<https://challenge.isic-archive.com/landing/2020/>>. Citado 4 vezes nas páginas 13, 33, 35 e 49.
- JONG, K. A. D. Analysis of the behavior of a class of genetic adaptive systems. University of Michigan Technical Report, 1975. Citado na página 19.
- KOZA, J. R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. [S.l.]: MIT Press, 1992. Citado na página 12.
- KOZA, J. R. Genetic programming as a means for programming computers by natural selection. Statistics and computing, Springer, v. 4, n. 2, p. 87–112, 1994. Citado na página 21. Citado 2 vezes nas páginas 21 e 23.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. Nature, v. 521, p. 436–444, 2015. Citado na página 17.

- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE. [S.l.: s.n.], 1998. v. 86, n. 11, p. 2278–2324. Citado na página 18.
- LI, X.; WANG, J.; ZHOU, Y. A hybrid model for melanoma detection using machine learning and medical imaging. Medical Image Analysis, Elsevier, v. 73, p. 102124, 2022. Citado na página 11.
- LI, Y.; SHEN, L. Skin lesion analysis towards melanoma detection using deep learning network. Sensors, MDPI, 2018. Citado na página 10.
- LUNADEI, L.; GALLEGUILLOS, P.; DIEZMA, B.; LIE, L. Evaluation of enzymatic browning in fresh-cut apple slices applying a multispectral vision system. In: International Conference on Agricultural Engineering. [S.l.: s.n.], 2010. Citado na página 27.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, v. 1, p. 281–297, 1967. Citado na página 31.
- MEDEIROS, G. B. de. Programação Genética Aplicada na Construção da Base de Regras em Redes Neuro-Fuzzy Tipo Neo-Fuzzy-Neuron. Tese (Doutorado) — CEFET-MG, 2019. Citado 2 vezes nas páginas 21 e 22.
- MÜLLER, A. C.; GUIDO, S. Introduction to Machine Learning with Python: A Guide for Data Scientists. 1st. ed. Sebastopol, CA: O'Reilly Media, 2016. ISBN 978-1-491-96229-9. Citado na página 14.
- OMARI, M. Use of CNNs for the Classification of Medical Images. GRIN Verlag, 2022. ISBN 9783346653963. Disponível em: <<https://books.google.com.br/books?id=-jN0EAAAQBAJ>>. Citado na página 10.
- PATHARE, P. B.; OPARA, U. L.; AL-SAID, F. A. Colour measurement and analysis in fresh and processed foods: A review. Food Bioprocess Technol, v. 6, p. 36–60, 2013. Disponível em: <<https://doi.org/10.1007/s11947-012-0867-9>>. Citado na página 26.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. Nature, v. 323, p. 533–536, 1986. Citado na página 16.
- SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [s.n.], 2018. p. 4510–4520. Disponível em: <<https://arxiv.org/abs/1801.04381>>. Citado na página 19.
- SAÚDE, O. M. da. Melanoma: Prevenção e Detecção Precoce. 2022. Acesso em: 25 mar. 2025. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/melanoma>>. Citado na página 10.
- SHAH, A.; SHAH, M.; PANDYA, A.; SUSHRA, R.; SUSHRA, R.; MEHTA, M.; PATEL, K.; PATEL, K. A comprehensive study on skin cancer detection using artificial neural network (ann) and convolutional neural network (cnn). Clinical eHealth, v. 6, 2023. ISSN 2588-9141. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2588914123000205>>. Citado na página 10.

- SILVA, A.; SOUZA, L.; COSTA, P. Brown index: Application in the food industry for quality monitoring. Food Quality and Safety, Oxford University Press, v. 7, n. 1, p. 45–57, 2020. Citado na página 11.
- SMITH, J.; LEE, Y.; ZHANG, H. Convolutional neural networks for skin cancer detection: A comparative study of approaches. IEEE Access, IEEE, v. 9, p. 67890–67898, 2021. Citado na página 10.
- WU, Y.; CHEN, B.; ZENG, A.; PAN, D.; WANG, R.; ZHAO, S. Skin cancer classification with deep learning: A systematic review. Frontiers in Oncology, 2022. Citado na página 11.
- ZEALAND, D. N. DermNet: Online Dermatology Resource. 2016. <<https://www.dermnetnz.org/>>. Citado na página 55.
- ZHANG, L.; XU, Q.; LI, F. Transfer learning for melanoma detection: A comparative evaluation of pre-trained models. Journal of Artificial Intelligence in Medicine, Elsevier, v. 45, p. 67–81, 2023. Citado na página 11.
- ZOPH, B.; VASUDEVAN, V.; SHLENS, J.; LE, Q. V. Learning transferable visual models from natural language supervision. Proceedings of the IEEE conference on computer vision and pattern recognition, p. 1–9, 2018. Citado 3 vezes nas páginas 18, 20 e 37.