

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

PABLO MARTINS COELHO

***PERIOCULAR EFFICIENTNET: UM MODELO PROFUNDO PARA  
RECONHECIMENTO PERIOCULAR***

Ouro Preto, MG  
2024

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

PABLO MARTINS COELHO

***PERIOULAR EFFICIENTNET: UM MODELO PROFUNDO PARA  
RECONHECIMENTO PERIOULAR***

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

**Orientador:** Pedro Henrique Lopes Silva

Ouro Preto, MG  
2024



## FOLHA DE APROVAÇÃO

**Pablo Martins Coelho**

**Periocular EfficientNet: um modelo profundo para reconhecimento periocular**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 6 de Fevereiro de 2024.

Membros da banca:

Pedro Henrique Lopes Silva (Orientador) - Doutor - Universidade Federal de Ouro Preto  
Eduardo José da Silva Luz (Examinador) - Doutor - Universidade Federal de Ouro Preto  
Gabriel Bicalho Ferreira (Examinador) - Bacharel - PPGCC UFOP

Pedro Henrique Lopes Silva, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 6/02/2024.



Documento assinado eletronicamente por **Pedro Henrique Lopes Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 09/02/2024, às 09:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0665665** e o código CRC **91881E8F**.

*Dedico este trabalho a Deus, pois reconheço que sem Ele, nada seria possível. Dedico também à minha esposa e aos meus pais, em virtude do amor incondicional, apoio incansável e orientação fornecida ao longo desta jornada. Seus incentivos e presença constante foram fundamentais para alcançar essa conquista.*

# Agradecimentos

Agradeço a Deus, por minha vida, saúde, força, família e por permitir que meus sonhos se tornem realidade.

À minha esposa, Janice, pelo amor incondicional e pelo apoio incansável durante os momentos difíceis, bem como por estar sempre ao meu lado, encorajando-me.

Aos meus pais, Sidnei e Cristiane, pelo amor, incentivo constante, pelo apoio e valiosas orientações que foram essenciais para alcançar este momento. À minha mãe, em especial, agradeço pelo empenho diário, incentivo nas horas de desânimo e cansaço, além de me acolher em cada etapa dessa jornada.

Ao meu irmão, Rafael, pela companhia constante e apoio sempre presente quando necessário.

Ao meu orientador, Pedro Silva, por seus preciosos ensinamentos, paciência, disponibilidade e dedicação em auxiliar-me na elaboração deste trabalho.

“O sucesso é uma consequência e não um objetivo”,

Gustave Flaubert.

# Resumo

O reconhecimento periocular tem demonstrado notável evolução por meio de técnicas computacionais. No entanto, durante esse processo, várias dificuldades são encontradas, como variações nas condições de iluminação, resolução das imagens e oclusões parciais. Para superar esses desafios, é necessário desenvolver abordagens robustas que surpassem essas dificuldades. Para isso, neste trabalho, propõe-se o uso de redes neurais convolucionais, utilizando como modelo base a arquitetura EfficientNet-B0. Essa arquitetura é customizada por meio da adição de camadas fortemente conectadas com *Softmax* para a classificação. Os experimentos foram realizados na base de dados UFPR-Periocular, em que o modelo proposto obteve uma acurácia média de 99.41%, superior em 1.73% em relação ao estado-da-arte. Além disso, foram empregadas técnicas de multitarefa e interpretabilidade do modelo treinado.

**Palavras-chave:** Reconhecimento biométrico. Rede Neural Convolucional. Região periocular.

# Abstract

Periocular recognition has exhibited significant advancements through computational techniques. Nonetheless, this process presents several challenges, including variations in lighting conditions, image resolution, and partial occlusions. To address these obstacles, there is a need to devise robust approaches capable of overcoming such difficulties. In this study, we propose the utilization of convolutional neural networks, leveraging the EfficientNet-B0 architecture as the foundational model. This architecture is further customized by integrating densely connected layers with Softmax for classification purposes. Experiments were conducted using the UFPR-Periocular dataset, where the proposed model achieved an average accuracy of 99.41%, surpassing the state-of-the-art by 1.73%. Additionally, multitasking techniques and model interpretability were employed.

**Keywords:** Periocular region. Biometric recognition. Convolutional Neural Network.

# Lista de Ilustrações

Figura 2.1 – Ilustração do processo de digitalização facial para a extração de características distintas destinadas à aplicação em sistemas de biometria facial. . . . .	4
Figura 2.2 – Exemplo de curvas sub-ajustada (A), adequada (B) e sobre-ajustada (C), respectivamente. . . . .	7
Figura 2.3 – Exemplo de uma rede neural, onde é possível observar a camada de entrada ( <i>Input Layer</i> ), duas camadas ocultas ( <i>Hidden Layer 1</i> e <i>Hidden Layer 2</i> , respectivamente) e, por fim, a camada de saída ( <i>Output Layer</i> ). . . . .	8
Figura 2.4 – Exemplo ilustrativo de um neurônio com suas respectivas entradas, a função que é aplicada nas entradas e sua saída. . . . .	10
Figura 2.5 – Exemplo de uma rede neural com 3 camadas, sendo a camada da esquerda a camada de entrada, a camada do meio é a camada oculta e a camada da direita é a de saída. . . . .	10
Figura 2.6 – Gráfico da função de ativação ReLU. . . . .	11
Figura 2.7 – Exemplo de uma Rede Neural Convolutiva e suas respectivas camadas. . .	12
Figura 2.8 – Representação do funcionamento das operações de <i>pooling</i> . . . . .	13
Figura 2.9 – Diferenças entre as abordagens de aprendizado de máquina e o aprendizado por transferência. . . . .	14
Figura 2.10–Compartilhamento de parâmetros rígidos para aprendizado multitarefa em redes neurais profundas. . . . .	15
Figura 2.11–Compartilhamento de parâmetros suaves para aprendizado multitarefa em redes neurais profundas. . . . .	16
Figura 2.12–Exemplo da técnica <i>k-fold</i> com cinco subdivisões, ou seja, com $k$ igual a 5. .	17
Figura 2.13–Exemplo ilustrativo para melhor exemplificar as métricas de avaliação. . . .	17
Figura 2.14–Exemplo da obtenção da fórmula da acurácia. . . . .	18
Figura 2.15–Exemplo da obtenção da fórmula da precisão. . . . .	19
Figura 2.16–Exemplo da obtenção da fórmula da revocação. . . . .	19
Figura 2.17–Demonstração de uma saída da técnica de interpretabilidade usando o Grad-CAM para um problema de reconhecimento periocular. . . . .	20
Figura 2.18–Visualização das Saídas Intermediárias Produzidas por uma CNN Treinada. .	21
Figura 2.19–Exploração de Ativações nos Primeiros Filtros de Cada Camada Convolutiva da VGG-M. . . . .	22
Figura 2.20–Ilustração comparativa entre imagens pré e pós-aplicação do GradCAM, destacando, por meio de um mapa de calor, as regiões determinantes nas decisões do modelo. . . . .	23

Figura 2.21–Avaliação comparativa do desempenho do modelo proposto LDA-CNN em relação a outros modelos, utilizando diferentes modelos base, no conjunto de dados UFPR-Periocular. . . . .	26
Figura 2.22–Resultados de acurácias obtidas pela abordagem proposta em (Barcellos, 2022). . . . .	27
Figura 3.1 – Amostra de imagens da base de dados UFPR-Periocular, onde é possível observar uma grande diversidade nas imagens. . . . .	31
Figura 3.2 – Fluxograma do modelo descrevendo os estágios sequenciais, iniciando no processo de aquisição de dados e culminando nas previsões geradas pela abordagem proposta. . . . .	32
Figura 4.1 – Imagens geradas por meio de técnicas de interpretabilidade baseadas em Grad-CAM, onde a primeira coluna exibe a imagem inalterada e as colunas subsequentes exibem a imagem interpretativa. Nestas últimas colunas, tem-se uma redução na intensidade cromática, visando facilitar a visualização das características faciais. . . . .	39
Figura 4.2 – Imagens nas quais a íris foi ocultada por um bloco branco para avaliar a influência da íris nas decisões do modelo. . . . .	41

# Lista de Tabelas

Tabela 2.1 – Comparativo entre diferentes biometrias. . . . .	24
Tabela 2.2 – Avaliação de desempenho de cada método comparado nos diferentes <i>smartphones</i> e condições de iluminação da base de dados VISOB (Rattani <i>et al.</i> , 2016). . . . .	25
Tabela 2.3 – Resultados de <i>benchmark</i> no protocolo de mundo fechado para a tarefa de identificação. . . . .	28
Tabela 2.4 – Técnicas utilizadas pelos autores citados e as respectivas bases de dados utilizadas. . . . .	29
Tabela 4.1 – Resultados dos experimentos realizados com o modelo PEN em três cenários distintos. . . . .	36
Tabela 4.2 – Análise dos resultados experimentais conduzidos com o modelo PEN em três configurações específicas, empregando as métricas de avaliação: Precisão, Revocação e F-Score. . . . .	37
Tabela 4.3 – Resultados dos experimentos realizados com o modelo MPEN Multitarefas em três cenários distintos. . . . .	37
Tabela 4.4 – Análise dos resultados experimentais conduzidos com o modelo MPEN em três configurações específicas, empregando as métricas de avaliação: Precisão, Revocação e F-Score. . . . .	38
Tabela 4.5 – Apresentação dos resultados mais destacados no âmbito deste estudo, abrangendo as performances do modelo PEN, do MPEN, conforme os achados de Alahmadi, Hussain e Aboalsamh (2022), e os resultados delineados por Zanlorensi <i>et al.</i> (2022). Quadro comparativo dos melhores resultados obtidos na base de dados UFPR-Periocular. . . . .	39
Tabela 4.6 – Resultados obtidos a partir dos experimentos conduzidos com o modelo PEN nos três cenários, considerando a ocultação da íris em todas as imagens analisadas. . . . .	40

# Lista de Abreviaturas e Siglas

CNN	<i>Convolutional Neural Network.</i>
ECG	Eletrocardiograma.
EER	<i>Equal Error Rate.</i>
FN	Falso Negativo.
FP	Falso Positivo.
FRGC	<i>Face Recognition Grand Challenge.</i>
F1	<i>F1-Score.</i>
GPU	<i>Graphics Processing Unit.</i>
GradCAM	<i>Gradient-weighted Class Activation Mapping.</i>
HOG	<i>Histogram of Oriented Gradients.</i>
IA	Inteligência Artificial.
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge.</i>
LBP	<i>Local Binary Patterns.</i>
LDA	<i>Linear Discriminant Analysis.</i>
MCP	McCulloch-Pitts.
MPEN	Multitarefa Periocular EfficientNet-B0.
OOD	<i>Out-of-Distribution.</i>
PEN	Periocular EfficientNet-B0.
ReLU	<i>Rectified Linear Unit.</i>
SI	Sistema de Informação.
SIFT	<i>Scale-Invariant Feature Transform.</i>
SURF	<i>Speeded-Up Robust Features.</i>
SVM	<i>Support Vector Machine.</i>
TPU	<i>Tensioner Processing Unit.</i>

UFPR	Universidade Federal do Paraná.
uLBP	<i>Uniform Local Binary Pattern.</i>
VISOB	<i>Visible Light Mobile Ocular Biometric.</i>
ViT	<i>Visual Transformer.</i>
VN	Verdadeiro Negativo.
VP	Verdadeiro Positivo.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa	2
1.2	Objetivos	3
1.3	Organização do Trabalho	3
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>4</b>
2.1	Fundamentação Teórica	4
2.1.1	Biometria	4
2.1.2	Inteligência Artificial	5
2.1.3	<i>Machine Learning</i>	6
2.1.4	<i>Deep Learning</i>	8
2.1.4.1	Neurônio	9
2.1.4.2	Funções de Perda	11
2.1.5	<i>Convolutional Neural Networks - CNN</i>	12
2.1.6	<i>Transfer Learning</i>	13
2.1.7	Aprendizagem Multitarefa	14
2.1.8	Validação Cruzada - <i>k-fold</i>	15
2.1.9	Métricas de Avaliação de Modelos	16
2.1.10	Interpretabilidade	20
2.1.10.1	GradCAM	22
2.2	Trabalhos Relacionados	23
<b>3</b>	<b>Desenvolvimento</b>	<b>30</b>
3.1	Descrição da Base de Dados	30
3.1.1	Divisão dos Dados	31
3.2	Metodologia Proposta	32
3.2.1	Descrição do Pré-processamento dos Dados	32
3.2.2	Periocular EfficientNet-B0 (PEN)	33
3.2.3	Multitarefa Periocular EfficientNet-B0 (MPEN)	34
3.2.4	Treinamento dos Modelos	35
3.2.5	Interpretação do Aprendizado	35
<b>4</b>	<b>Experimentos e Resultados</b>	<b>36</b>
4.1	<i>Setup</i> de Experimentos	36
4.2	Resultados Obtidos	36
4.2.1	Treinamento do Modelo PEN	36
4.2.2	Treinamento do Modelo MPEN	37
4.2.3	Comparação com a Literatura	38
4.3	Análise da Interpretabilidade do Modelo PEN	39

<b>5 Considerações Finais . . . . .</b>	<b>42</b>
---	-----------

<b>Referências . . . . .</b>	<b>43</b>
------------------------------	-----------

# 1 Introdução

O reconhecimento biométrico é uma área tecnológica que tem ganhado destaque nos últimos anos. Seu objetivo principal é identificar ou verificar a identidade de uma pessoa por meio de características físicas ou comportamentais únicas (Barros, 2023). Essas características únicas podem vir, por exemplo, da impressão digital (Paoli; Dahia; Amparo, 2016), reconhecimento facial (Orvalho, 2019), voz (Sales; Sales; Silva, 2021), íris (Costa, 2009), região periocular (Zanlorensi *et al.*, 2022), entre outras.

Dessa forma, o reconhecimento biométrico apresenta um vasto potencial para ser aplicado na detecção e identificação de indivíduos, tanto para fins de localização quanto para controle de acesso em locais específicos. Ambas as finalidades destacadas são de elevada relevância no âmbito da segurança (Duarte *et al.*, 2021).

A crescente demanda por segurança em todo o mundo tem levado ao constante crescimento de equipamentos de vigilância, que fornecem maior segurança à população. Sistemas automatizados de segurança são cada vez mais necessários para auxiliar os agentes de segurança na localização de eventos de interesse em tempo real (Duarte *et al.*, 2021).

Nesse contexto, o reconhecimento periocular surge como uma técnica promissora de biometria que utiliza a análise da região ao redor dos olhos para identificação de indivíduos (Barcellos, 2022). Conforme descrito por Barcellos (2022), a técnica de reconhecimento por meio da região periocular tem sido amplamente utilizada na investigação acadêmica contemporânea, dado que ela fornece quantidades substanciais de informações discriminantes. Além disso, a análise da região periocular revela-se uma abordagem de menor invasividade e maior praticidade em comparação a outras técnicas biométricas que dependem da colaboração e interação direta do usuário, tais como a impressão digital e a palma da mão (Barcellos, 2022). Vale ressaltar ainda que tais métodos apresentam restrições sanitárias significativas, uma vez que há risco de contaminação associado a eles.

Apesar dos benefícios inerentes ao reconhecimento biométrico pela região periocular, ainda subsistem desafios a serem enfrentados. Esses desafios estão relacionados principalmente à garantia de uma alta precisão na aquisição das imagens, considerando as variações presentes em termos de iluminação, pose, expressão facial e características interindividuais (Frankiewicz, 2023).

Assim, para obter resultados mais promissores, é fundamental o uso de técnicas computacionais avançadas e modalidades biométricas específicas, como a biometria da região periocular. Além disso, diversos fatores exercem considerável influência nos resultados alcançados, abrangendo desde a coleta de dados de alta qualidade até o desenvolvimento de algoritmos robustos capazes de lidar eficientemente com as variações presentes em iluminação, pose e expressão facial.

Adicionalmente, destaca-se a aplicação de técnicas de aprendizado de máquina, como o uso de redes neurais convolucionais, ou CNN (*Convolutional Neural Network*) (Zanlorensi *et al.*, 2022), como estratégia para aprimorar a precisão do reconhecimento periocular. Consequentemente, torna-se viável a realização de tarefas de identificação e reconhecimento de indivíduos com um elevado grau de precisão.

No contexto dessas técnicas computacionais avançadas, o aprendizado multitarefa (em inglês *multitask*), uma vertente do campo de aprendizado de máquina, apresenta-se como uma abordagem promissora (Zhang; Yang, 2017). Essa abordagem visa tratar simultaneamente múltiplas tarefas distintas e se baseia na exploração das similaridades existentes entre elas. Esse enfoque não só pode otimizar a eficácia do processo de aprendizado, mas também desempenhar um papel relevante como um mecanismo regularizador, influenciando positivamente o processo de aprendizado de cada tarefa individualmente (Zhang; Yang, 2017).

Adicionalmente, será aplicada a técnica de interpretabilidade (Mesameki *et al.*, 2023) usando o Grad-CAM, com o intuito de aprofundar a compreensão acerca das decisões adotadas pelo modelo em estudo. Tal abordagem visa não apenas elucidar os critérios subjacentes às decisões tomadas pelo modelo, mas também extrair conclusões pertinentes sobre os resultados obtidos.

Sendo assim, o presente trabalho propõe investigar e analisar as técnicas e abordagens utilizadas no reconhecimento periocular, visando aprimorar a precisão e a eficácia desse processo. Espera-se que este estudo contribua para o desenvolvimento de sistemas automatizados de segurança mais confiáveis e eficientes, além de possibilitar a sua aplicação em diferentes áreas, como segurança pública e privada, identificação de indivíduos em locais públicos e controle de acesso a espaços restritos.

Com isso, o propósito deste trabalho é investigar se a abordagem de instruir um único modelo em múltiplas tarefas (abordagem multitarefa) pode aprimorar sua resistência e precisão. Assim, a pesquisa se concentra em explorar se o aprendizado de múltiplas tarefas por um único modelo pode resultar em melhorias substanciais na robustez e precisão desse modelo.

## 1.1 Justificativa

O uso de técnicas de biometria para a identificação de indivíduos tem se expandido significativamente em diversas áreas, como segurança pública, controle de acesso a edifícios, identificação de suspeitos em investigações criminais, entre outras (Duarte *et al.*, 2021). No entanto, a precisão dessas técnicas ainda enfrenta alguns desafios (Frankiewicz, 2023), os quais podem levar a erros na identificação de indivíduos.

A partir dessa perspectiva, a motivação central deste trabalho reside na relevância do reconhecimento biométrico, que abrange várias áreas do conhecimento. Além disso, vale ressaltar

que esse campo de estudo encontra-se em contínua evolução, o que propicia a realização de novas descobertas que contribuem para o avanço do estado da arte.

Portanto, com o intuito de destacar a importância desse tema, o trabalho irá identificar os desafios encontrados no processo de reconhecimento de indivíduos. Simultaneamente, esses obstáculos serão abordados por meio de técnicas eficazes para aprimorar a segurança do processo, com foco especial na região periocular.

## 1.2 Objetivos

O presente trabalho tem como objetivo utilizar uma CNN para construir um sistema biométrico baseado na região periocular, com foco no modo de identificação.

Para atingir o objetivo principal, alguns objetivos específicos foram definidos, tais como:

- Propor uma nova arquitetura para o reconhecimento periocular.
- Desenvolver e validar o método proposto.
- Aplicar a técnica de *transfer learning* na arquitetura proposta.
- Implementar uma abordagem multitarefas utilizando a base de dados em estudo.
- Conduzir a análise de interpretabilidade do modelo treinado.
- Comparar os resultados obtidos com o estado-da-arte.

## 1.3 Organização do Trabalho

Esta monografia está organizada de forma a permitir uma melhor compreensão dos resultados obtidos. O [Capítulo 2](#) apresenta a revisão bibliográfica, na qual são apresentados os trabalhos relacionados, os elementos conceituais e a fundamentação teórica que norteiam o desenvolvimento deste trabalho. No [Capítulo 3](#), é apresentada a metodologia empregada, incluindo a descrição dos experimentos realizados. O [Capítulo 4](#) expõe os resultados obtidos a partir do desenvolvimento e dos experimentos realizados.

Por fim, no [Capítulo 5](#), são apresentadas as conclusões alcançadas a partir dos resultados obtidos.

## 2 Revisão Bibliográfica

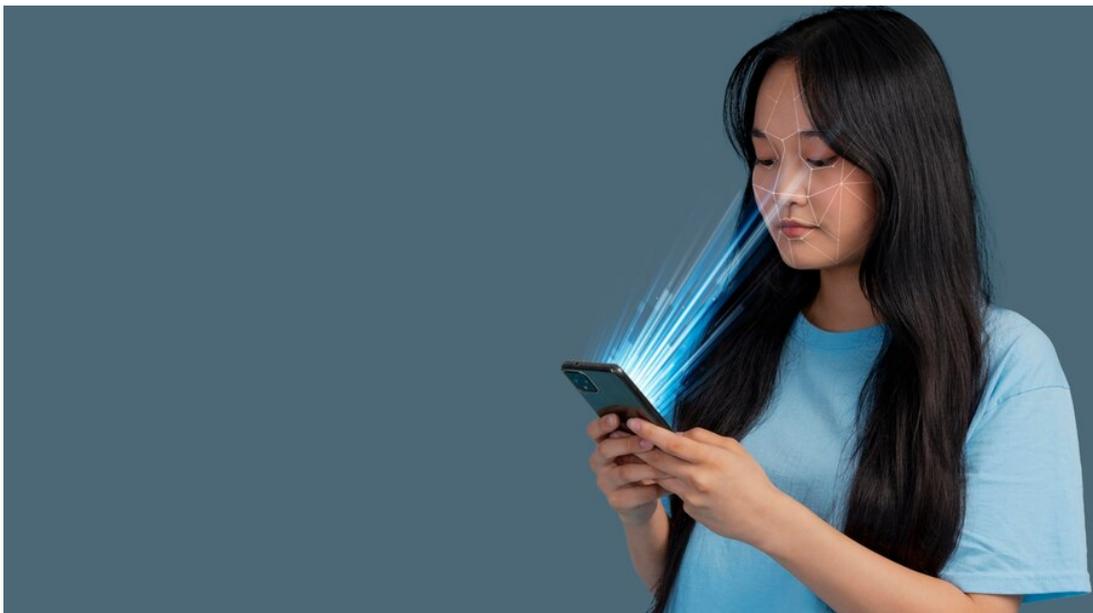
Neste capítulo, serão abordados os conceitos fundamentais relacionados ao presente estudo. A [Seção 2.1](#) apresentará a fundamentação teórica, na qual serão descritos alguns pontos-chave, tais como Inteligência Artificial, *Deep Learning* e Pré-Processamento de dados. Em seguida, na [Seção 2.2](#), serão discutidos os trabalhos relacionados que foram utilizados para análise e aquisição de conhecimento, permitindo assim compreender a evolução da área até o atual estado da arte.

### 2.1 Fundamentação Teórica

#### 2.1.1 Biometria

O conceito de biometria origina-se da junção etimológica das palavras gregas “*bios*” (vida) e “*metron*” (medida). No contexto da autenticação, a biometria alude à empregabilidade de características intrínsecas de um sujeito, para a realização de sua identificação e/ou verificação frente a um Sistema de Informação (SI) de uma entidade organizacional, conforme ilustrado na [Figura 2.1](#) e destacado por [Magalhaes e Santos \(2003\)](#).

Figura 2.1 – Ilustração do processo de digitalização facial para a extração de características distintivas destinadas à aplicação em sistemas de biometria facial.



Fonte: (FREEP!K, 2024).

A identificação de um indivíduo é conceituada como o desafio de determinar a identidade, inicialmente não conhecida, de um sujeito. Em contrapartida, a verificação de um indivíduo

envolve a validação ou refutação da alegação de identidade proposta pelo mesmo (Magalhaes; Santos, 2003).

A identificação e verificação de indivíduos podem ser contextualizadas em dois paradigmas distintos: o mundo aberto e o mundo fechado. No contexto do mundo aberto, o modelo é exposto a entradas de indivíduos não previamente observados durante o processo de treinamento. Nesse cenário, torna-se imperativo que o modelo possa discernir e sinalizar entradas que se caracterizem como pertencentes a uma classe inédita ou “fora da distribuição”. Por outro lado, no paradigma de mundo fechado, o modelo é limitado a receber somente entradas associadas às classes previamente estabelecidas e treinadas (Brauner, 2020).

### 2.1.2 Inteligência Artificial

De acordo com o dicionário Aurélio (Dicio, 2023), o termo “Inteligência Artificial” (IA) é definido como, “Mecanismo, programa de computador, software ou qualquer outro artefato que, produzido artificialmente, exibe uma inteligência similar à inteligência humana”. Em outras palavras, a Inteligência Artificial refere-se à capacidade de sistemas computacionais em desempenhar funções que se assemelham ao pensamento humano, permitindo que tais dispositivos possam “pensar” e “raciocinar” de maneira similar aos seres humanos. Essa capacidade abrange a percepção de variáveis, tomada de decisões, execução de tarefas e resolução de problemas, operando em uma lógica que se assemelha ao raciocínio humano (FIA, 2021).

O estudo da Inteligência Artificial teve início nos anos 50, quando os cientistas Hebert Simon e Allen Newell estabeleceram o primeiro laboratório de IA na Universidade de Carnegie Mellon (Silva, 2008). A aspiração de construir máquinas capazes de reproduzir a capacidade humana de pensar e agir remonta a tempos antigos. A área de Inteligência Artificial tem passado por um processo contínuo de evolução, impulsionado pelo avanço geral da computação, o que vem permitindo avanços significativos em diversos setores, como saúde e segurança.

No que diz respeito ao funcionamento de um sistema de Inteligência Artificial, seu processo é idealmente fundamentado na combinação de dados digitais com algoritmos de aprendizado de máquina. Essa combinação permite ao sistema realizar a leitura e interpretação de padrões e informações, o que resulta no aprendizado automático do sistema. Além disso, a IA também depende do poder de processamento, que se refere à capacidade operacional do sistema em processar essas informações (Amazon, 2023).

É possível observar que a Inteligência Artificial abrange um amplo espectro de conceitos, o que a torna um campo vasto. Para alcançar resultados satisfatórios, a IA se baseia em pilares fundamentais, como o Aprendizado de Máquina (*Machine Learning*).

### 2.1.3 *Machine Learning*

O Aprendizado de Máquina, também conhecido como *Machine Learning*, é um conjunto de técnicas que capacitam os sistemas a aprenderem de forma autônoma e evoluírem por conta própria (Neoway, 2020). Isso significa que a máquina é capaz de alterar seu comportamento com base em sua própria experiência.

De forma resumida, o processo de aprendizado de máquina começa com a coleta de dados relevantes para um problema específico, abrangendo um conjunto de informações relacionadas a esse problema. Esses dados podem incluir diversas informações, como características, preços, quantidades e outras variáveis pertinentes ao problema em questão. Em seguida, esses dados coletados passam por um processo de preparação e processamento, visando torná-los adequados para serem utilizados por um algoritmo de aprendizado de máquina. Nesse estágio, pode ocorrer a mineração de dados, ou seja, a separação dos dados com base em sua importância relativa (Camilo; Silva, 2009).

Sendo assim, o algoritmo de aprendizado de máquina é submetido a um processo de treinamento utilizando um conjunto de dados cuidadosamente dividido em dados de treinamento e teste. Os dados de treinamento são utilizados com o propósito de instruir o algoritmo e permitir que ele adquira conhecimento e habilidades relevantes para o problema em questão. Por outro lado, os dados de teste são empregados para avaliar a eficiência e o desempenho do algoritmo treinado.

O aprendizado de máquina pode ser realizado por meio de três principais abordagens: aprendizado supervisionado, aprendizado semi-supervisionado e aprendizado não supervisionado (Brunialti *et al.*, 2015). No aprendizado supervisionado, o algoritmo é treinado utilizando um conjunto de dados previamente rotulados, em que cada amostra é acompanhada de sua respectiva classe. Métodos de classificação, como Máquina de Vetores de Suporte (em inglês *Support Vector Machine* - SVM) ou regressão, são comumente empregados nesse tipo de abordagem (Silva, 2018). Nesse contexto, o algoritmo recebe dados de entrada juntamente com a resposta correta, permitindo que ele aprenda a mapear corretamente as características dos dados para suas respectivas classes.

Já no aprendizado semi-supervisionado, é fornecido um conjunto parcialmente rotulado de amostras, onde algumas possuem classes atribuídas e outras não. Utilizando as informações disponíveis no conjunto rotulado, o algoritmo estende seu aprendizado para as amostras não rotuladas, com base nas similaridades identificadas. Em alguns casos, quando existe um alto nível de confiança nas atribuições realizadas pelo algoritmo no conjunto não rotulado (por exemplo, uma confiança mínima de 99%), algumas amostras são adicionadas ao conjunto rotulado e o modelo é reajustado por meio de um novo treinamento (Silva, 2018).

Por fim, no aprendizado não supervisionado, o algoritmo é exposto a um conjunto de amostras e busca identificar padrões e estruturas intrínsecas nesses dados sem a utilização de

informações de classes predefinidas. Métodos de clusterização, como o algoritmo *K-means*, são frequentemente aplicados nessa abordagem (Silva, 2018). Nesse contexto, o algoritmo tenta agrupar as amostras com base em suas semelhanças, visando encontrar estruturas latentes nos dados sem a necessidade de informações prévias sobre as classes corretas das amostras.

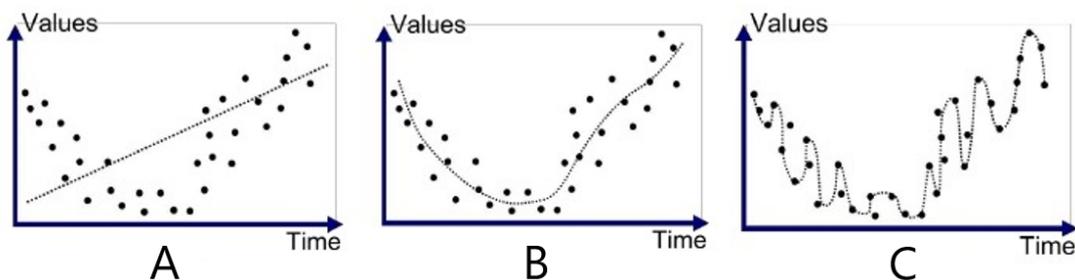
Essas diferentes formas de aprendizado de máquina proporcionam abordagens flexíveis e adaptáveis para lidar com diversos tipos de problemas e conjuntos de dados. Isso possibilita que os algoritmos sejam treinados e capazes de realizar tarefas como classificação, regressão e agrupamento, dependendo das características e das informações disponíveis em cada conjunto de dados específico.

Nas três abordagens mencionadas, o objetivo principal consiste no aprendizado do algoritmo e na geração de um modelo resultante (Silva, 2018). É fundamental que esse modelo alcance um equilíbrio adequado entre generalização e especificação. Dessa forma, evita-se que o modelo se torne excessivamente generalista, confundindo, por exemplo, uma vaca com um cavalo devido a algumas características similares. Ao mesmo tempo, busca-se evitar que o modelo seja excessivamente específico, a ponto de não reconhecer um Pastor Alemão como sendo um cão, apenas por ter aprendido que somente os vira-latas são considerados cães (Silva, 2018).

A busca pelo equilíbrio entre generalização e especificação visa evitar problemas de subajuste (*underfitting*) e sobreajuste (*overfitting*) do modelo (Branco, 2020). O *underfitting* ocorre quando o modelo não é capaz de aprender informações suficientes dos dados de treinamento, resultando em uma generalização excessiva das características. Por outro lado, o *overfitting* é o oposto, em que o modelo “decora” os dados de treinamento específicos, incapaz de generalizar para novos dados não vistos antes (Branco, 2020). Em outras palavras, ele se torna excessivamente específico durante o treinamento.

A seguir, na Figura 2.2 é exemplificado um modelo com alta generalização, um que alcança um equilíbrio entre generalização e especificidade, e, por fim, um com alta especificidade. Essas ilustrações ajudam a visualizar a importância de encontrar um ponto de equilíbrio adequado durante o treinamento e a geração do modelo.

Figura 2.2 – Exemplo de curvas sub-ajustada (A), adequada (B) e sobre-ajustada (C), respectivamente.



Fonte: (Branco, 2020).

Uma das inúmeras abordagens que se destaca na capacidade de aprendizagem e geração

de modelos é conhecida como aprendizado profundo, ou, em inglês, *Deep Learning* (Oracle, 2023).

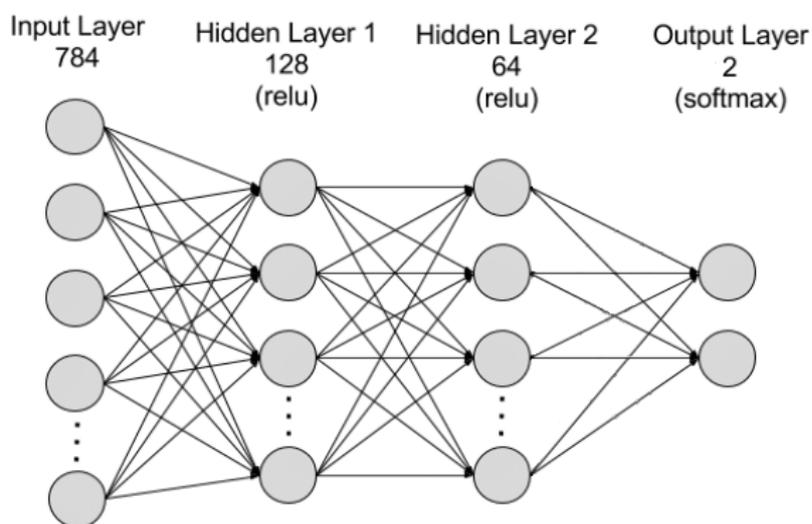
### 2.1.4 *Deep Learning*

A Aprendizagem Profunda, também conhecida como *Deep Learning*, é uma vertente do aprendizado de máquina que se dedica ao treinamento de redes neurais profundas para resolver problemas de aprendizado de máquina (Gaea, 2021). Essas redes neurais profundas são compostas por várias camadas de unidades de processamento, permitindo que aprendam a representar dados em níveis crescentes de abstração. No processo de treinamento, uma ampla quantidade de dados normalmente é utilizada (Amazon, 2023). Durante esse processo, os pesos e conexões entre as unidades de processamento da rede são ajustados para aprimorar suas capacidades de previsão.

As redes neurais profundas, também conhecidas como redes de aprendizado profundo, são compostas por múltiplas camadas que contêm milhões de neurônios artificiais interconectados. Essas conexões entre os neurônios são representadas por valores numéricos chamados de pesos, os quais determinam a influência de um nó em relação aos outros nós na rede (Amazon, 2023). Uma característica distintiva das redes neurais profundas é a presença de várias camadas de neurônios, em que cada camada recebe os dados de entrada da camada anterior, processa-os e os encaminha para a próxima camada.

Na Figura 2.3 é apresentado um exemplo simplificado de uma rede neural profunda, composta por uma camada de entrada, duas camadas ocultas e uma camada de saída.

Figura 2.3 – Exemplo de uma rede neural, onde é possível observar a camada de entrada (*Input Layer*), duas camadas ocultas (*Hidden Layer 1* e *Hidden Layer 2*, respectivamente) e, por fim, a camada de saída (*Output Layer*).



Fonte: (Amazon, 2023) - Adaptada pelo autor.

Uma Rede neural é um dos tipos de abordagens utilizada no campo do aprendizado de máquina, com o propósito de ensinar máquinas a processarem dados de forma inspirada

pelo funcionamento do cérebro humano (Amazon, 2023). No cérebro, as células conhecidas como neurônios formam uma rede altamente interconectada e se comunicam por meio de sinais elétricos, permitindo aos seres humanos processarem informações (Amazon, 2023). De maneira análoga, uma rede neural artificial é composta por neurônios artificiais que trabalham em conjunto para resolver problemas.

Uma rede neural artificial é composta por um conjunto de unidades de processamento, cujo funcionamento segue o que inicialmente foi proposto por McCulloch e Pitts (1943). Essa unidade pode ser resumida da seguinte maneira: sinais de entrada são apresentados à unidade, e cada sinal é multiplicado por um valor numérico chamado peso, o qual indica sua contribuição para a saída da unidade (Carvalho, 2019). Em seguida, ocorre uma soma ponderada dos sinais, resultando em um nível de ativação. Se esse nível de ativação ultrapassar um determinado limite, denominado *threshold*, a unidade produzirá uma resposta de saída específica (Carvalho, 2019).

A propriedade mais destacada das redes neurais é a sua capacidade de aprender a partir do ambiente em que estão inseridas, o que resulta em um aprimoramento do seu desempenho (Carvalho, 2019). Esse processo de aprendizado é alcançado por meio de ajustes iterativos aplicados aos pesos das unidades, conhecido como treinamento. A aprendizagem ocorre quando a rede neural adquire uma solução generalizada para uma classe de problemas específica (Carvalho, 2019).

#### 2.1.4.1 Neurônio

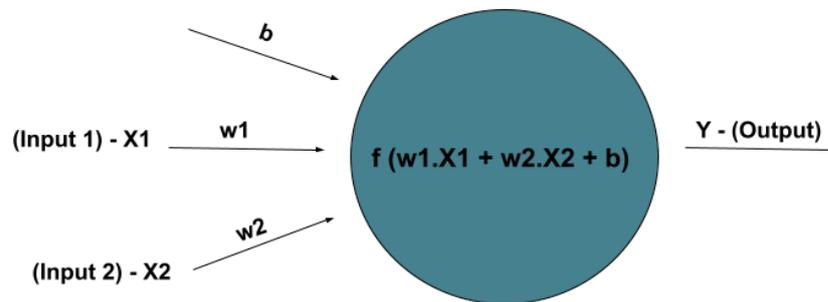
Os neurônios, componentes celulares fundamentais do sistema nervoso humano, desempenham um papel crucial ao estabelecer conexões sinápticas em resposta a estímulos provenientes do ambiente externo ou do próprio organismo (Varella, 2014).

Nas redes neurais, os neurônios constituem as unidades básicas de processamento de informações, buscando emular o funcionamento dos neurônios humanos. Essas unidades recebem entradas, processam-nas e geram saídas correspondentes. Além disso, esses neurônios são interconectados em uma estrutura em camadas, de forma análoga ao cérebro humano (Silva, 2018).

Nesse sentido, Rosenblatt (1958) foi o criador do “perceptron”, um modelo simples que precedeu o conceito atual de neurônio, e, exerceu uma influência significativa no desenvolvimento das redes neurais. Posteriormente, Aleksander e Morton (1990) propuseram um modelo mais sofisticado denominado McCulloch-Pitts (MCP). No entanto, o modelo final do neurônio corresponde a um MCP enriquecido com alguns pré-processamentos adicionais. Através da observação da Figura 2.4, é possível examinar o funcionamento de um neurônio de maneira ilustrativa.

O neurônio é considerado a unidade elementar fundamental de uma rede neural, possuindo a capacidade de realizar classificações binárias, isto é, responder de forma afirmativa ou negativa. Essa classificação se baseia essencialmente em uma regressão múltipla, expressa por meio de

Figura 2.4 – Exemplo ilustrativo de um neurônio com suas respectivas entradas, a função que é aplicada nas entradas e sua saída.



Fonte: (Produzida pelo próprio autor).

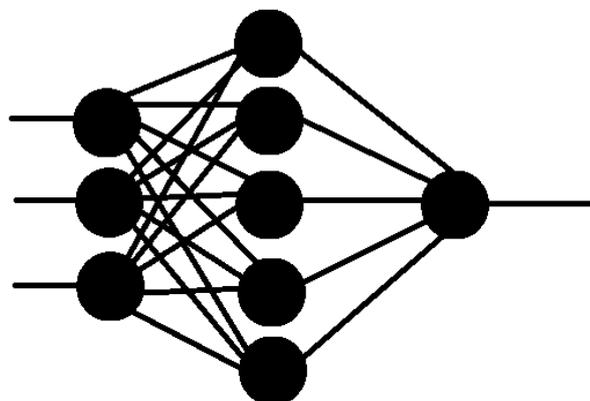
uma regressão ( $ax + b$ ), na qual os valores de entrada são processados para produzir uma saída correspondente (Jose, 2019).

Cada valor de entrada é associado a um peso específico, o qual é atribuído com base em sua relevância. O neurônio aplica uma função  $f$  à soma ponderada desses valores de entrada, conforme ilustrado na Figura 2.4.

De maneira concisa, cada entrada em um neurônio possui um peso ( $w$ ) associado que reflete sua importância ( $X$ ). Assim, o peso e a importância de cada entrada são multiplicados e somados com uma constante denominada “bias” (Jose, 2019). Essa constante é utilizada para ajustar o ponto de operação dos neurônios, permitindo que a rede seja ativada mesmo quando as entradas correspondentes são nulas (Ceccon, 2020b).

A Figura 2.5 representa uma rede neural com três camadas, na qual cada bolinha representa um neurônio. A primeira camada contém três neurônios, a segunda camada possui cinco neurônios e a terceira camada é composta por apenas um neurônio. A conexão entre os neurônios é claramente ilustrada na imagem (Alecrim, 2004).

Figura 2.5 – Exemplo de uma rede neural com 3 camadas, sendo a camada da esquerda a camada de entrada, a camada do meio é a camada oculta e a camada da direita é a de saída.

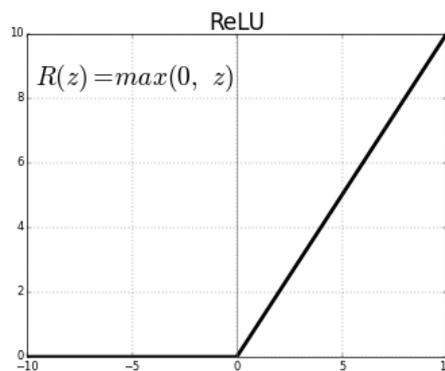


Fonte: (Produzida pelo próprio autor).

A saída de cada neurônio é aplicada uma função de ativação. Ela exerce o papel de

realizar uma transformação nas características recebidas. Em geral, são empregadas funções com propriedades não lineares, visando capturar de forma mais eficiente as relações complexas presentes nas características (Vargas; Paes; Vasconcelos, 2016). Entre essas funções, destaca-se a função de ativação *Rectified Linear Unit* (ReLU), amplamente adotada na prática. A ReLU é uma função não linear (Figura 2.6) que permite a propagação de gradientes retroativamente e a ativação de múltiplas camadas de neurônios de maneira eficaz. Ela produz valores no intervalo  $[0, \infty[$  (Ceccon, 2020a).

Figura 2.6 – Gráfico da função de ativação ReLU.



Fonte: (Silva; Siebert, 2020).

#### 2.1.4.2 Funções de Perda

No contexto da avaliação de eficiência do treinamento de uma arquitetura de *Deep Learning*, uma métrica amplamente utilizada é a função de perda, também conhecida como função de custo. Essa métrica é empregada para medir o nível de insatisfação com o desempenho do algoritmo, ou seja, quanto pior for o desempenho, maior será o valor da função de perda (Nishimoto, 2018).

Essencialmente, a função de perda compara as pontuações atribuídas à categoria correta com as pontuações atribuídas às demais categorias, permitindo assim determinar o grau de satisfação com o resultado obtido. Entre as funções de perda mais populares, destacam-se a *hinge loss* e a *cross-entropy* (Nishimoto, 2018).

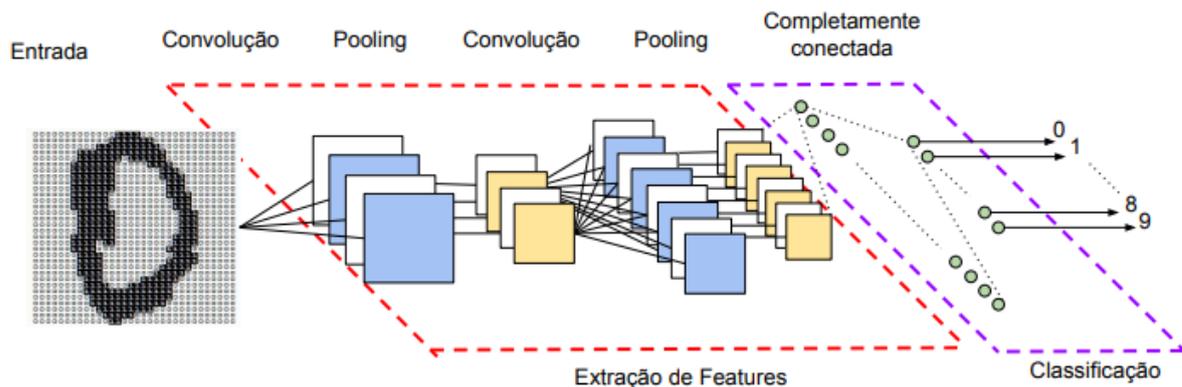
- A *hinge loss* é uma função de perda frequentemente utilizada em problemas de classificação em que se busca a separação linear das classes. Ela penaliza as classificações incorretas de forma proporcional à distância entre a pontuação da classe correta e a pontuação da classe incorretamente atribuída (Seidor, 2020).
- Já a *cross-entropy*, ou entropia cruzada, é uma função de perda comumente aplicada em problemas de classificação multi-classe. Ela mede a dissimilaridade entre as distribuições de probabilidade reais e as previstas pelo modelo, atribuindo uma penalidade maior às classificações incorretas com maior diferença no valor da confiança (Seidor, 2020).

A escolha entre a *hinge loss* e a *cross-entropy* depende do tipo de problema e das características dos dados. Ambas as funções de perda desempenham um papel fundamental na otimização de algoritmos de aprendizado de máquina no contexto de *Deep Learning*, uma vez que fornecem um indicativo do quão bem o modelo está se ajustando aos dados e realizando as classificações corretas.

### 2.1.5 Convolutional Neural Networks - CNN

Uma classe de modelos amplamente empregada e altamente eficaz no contexto do aprendizado profundo para imagens é a Rede Neural Convolutiva (CNN, do inglês *Convolutional Neural Network*). As CNNs são variantes das redes de Perceptrons de Múltiplas Camadas, mas foram especialmente concebidas com base nos princípios biológicos subjacentes ao processamento de dados visuais (Vargas; Paes; Vasconcelos, 2016). Na Figura 2.7 é possível observar um exemplo de uma rede neural convolutiva.

Figura 2.7 – Exemplo de uma Rede Neural Convolutiva e suas respectivas camadas.



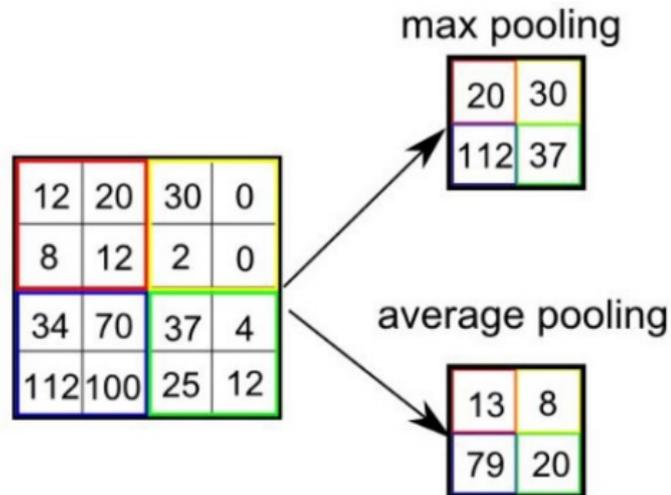
Fonte: (Vargas; Paes; Vasconcelos, 2016).

A CNN consiste em vários componentes, incluindo camadas de convolução, funções de ativação, camadas de agrupamento (*pooling*), camadas totalmente conectadas e funções de classificação (Vargas; Paes; Vasconcelos, 2016). Uma camada de convolução é composta por um conjunto de filtros responsáveis por extrair e mapear o conteúdo da imagem, convertendo-o em características. Esse processo envolve a utilização de filtros, que são pequenos blocos capazes de capturar informações de sub-regiões da imagem (Passos, 2021). Após a camada de convolução, é frequentemente aplicada uma função de ativação que introduz não linearidade ao modelo (Vargas; Paes; Vasconcelos, 2016).

Outra camada comumente utilizada após a camada de convolução e ativação é a camada de agrupamento (*pooling*) (Vargas; Paes; Vasconcelos, 2016). A camada de *pooling* recebe os blocos contendo as informações extraídas na etapa de convolução. Sua função consiste em simplificar a informação, resumindo os dados do sub-bloco da imagem em um único valor, como

pode ser observado na Figura 2.8, que é então transmitido para uma camada totalmente conectada (Passos, 2021).

Figura 2.8 – Representação do funcionamento das operações de *pooling*.



Fonte: (Rawat; Wang, 2017).

A camada totalmente conectada possui um número  $n$  de neurônios, onde todos estão conectados a todas as saídas da camada anterior. Essa camada recebe como entrada um vetor de características extraídas das camadas anteriores da rede e produz uma saída que é transformada de acordo também com uma função de ativação. Ela também é a camada usada para realizar a classificação das informações extraídas pela rede, onde a sua saída é transformada em uma distribuição de probabilidade sobre as classes ou em um valor contínuo, dependendo da tarefa em questão (Passos, 2021).

Com isso, após a camada totalmente conectada, o estágio final é a aplicação da função de classificação. Essa função desempenha um papel crucial no treinamento, pois influencia no aprendizado dos filtros e, conseqüentemente, no resultado alcançado pela rede (Vargas; Paes; Vasconcelos, 2016). Um exemplo de uma função de classificação é a função *Softmax*, que representa uma generalização da função *sigmoide* para casos não binários. A função *Softmax* produz valores no intervalo  $[0, 1]$  onde a soma de todas as probabilidades é 1 (Ceccon, 2020a).

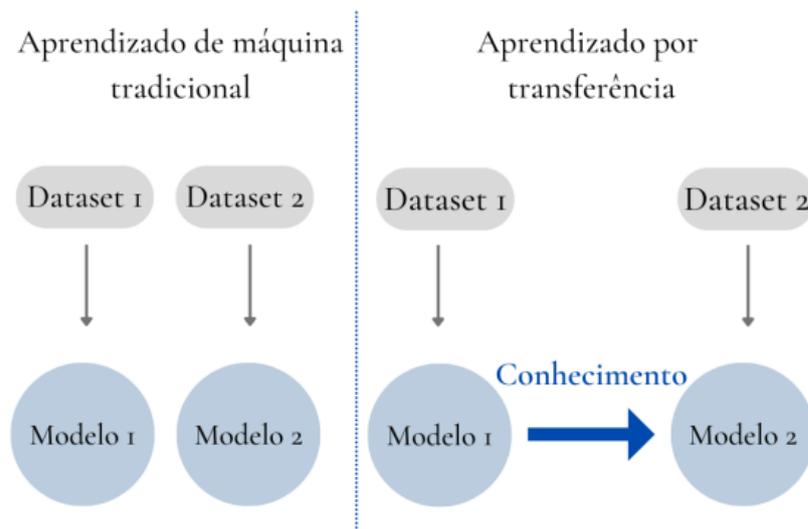
### 2.1.6 *Transfer Learning*

A etapa de treinamento de modelos de aprendizado profundo (*deep learning*), geralmente demanda uma quantidade significativa de dados e, conseqüentemente, recursos computacionais de alto nível, como GPUs ou TPUs. Isso resulta em um tempo prolongado de treinamento do modelo (Gayhardt *et al.*, 2022). No entanto, quando os recursos computacionais são limitados, é possível adotar uma abordagem com menor custo computacional e, portanto, mais rápida para a fase de treinamento do modelo.

Nesse contexto, surge a técnica conhecida como *Transfer Learning* (aprendizado por transferência) (Meazzini, 2019). Em resumo, o aprendizado por transferência envolve a reutilização de um modelo pré-treinado em um novo problema. Isso significa utilizar uma rede neural que tenha sido previamente treinada em um conjunto extenso de dados para resolver um novo problema relacionado àquele para o qual a rede neural foi inicialmente treinada.

A Figura 2.9 ilustra as diferenças entre as abordagens do aprendizado de máquina tradicional e o aprendizado de máquina por transferência.

Figura 2.9 – Diferenças entre as abordagens de aprendizado de máquina e o aprendizado por transferência.



Fonte: (Costa; Walter; Stadler, 2022).

Na abordagem do aprendizado por transferência, observa-se que um modelo é inicialmente treinado em um conjunto de dados específico. Após esse treinamento, o conhecimento adquirido é transferido para outro modelo, que combina essa transferência de aprendizado com o conhecimento obtido a partir de sua própria base de dados. Essa combinação de informações provenientes de diferentes fontes de aprendizado permite ao modelo aprimorar sua capacidade de generalização e desempenho em tarefas relacionadas.

### 2.1.7 Aprendizagem Multitarefa

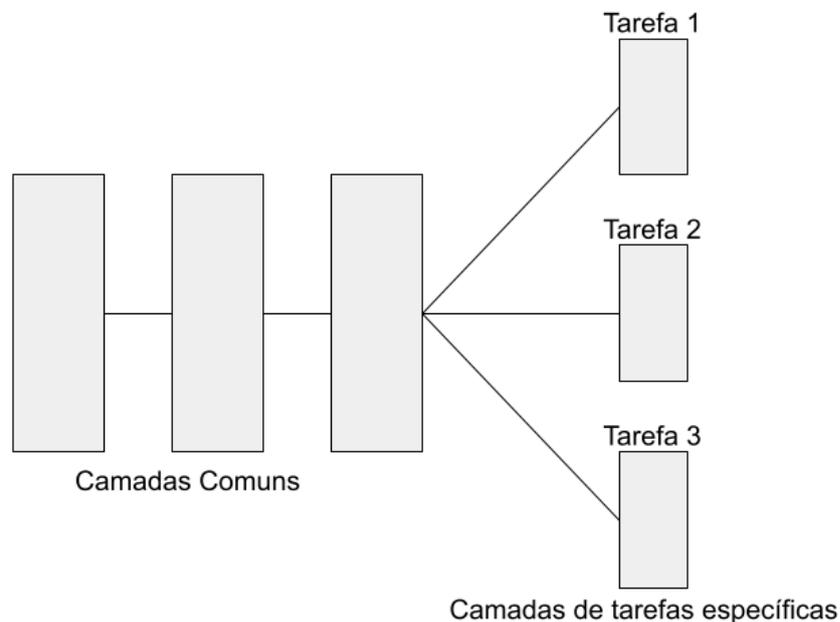
Normalmente, o treinamento de um modelo faz com que ele fique especializado em uma tarefa. Nesse contexto, pode ocorrer uma especialização da tarefa de tal forma a comprometer a generalização do modelo. Uma estratégia para evitar esse tipo de problema é variar o que é aprendido usando a técnica de multitarefa.

A aprendizagem multitarefa é caracterizada pelo aprendizado simultâneo em diversas tarefas, onde todas ou um subconjunto delas estão inter-relacionadas, embora não sejam idênticas. Nesse contexto, o propósito central é estabelecer conexões entre essas tarefas a fim de aprimorar

o processo de aprendizado de um modelo, utilizando o conhecimento contido nas respectivas tarefas (Zhang; Yang, 2017).

O paradigma de aprendizado multitarefa frequentemente emprega o compartilhamento de parâmetros, sejam eles rígidos ou suaves, nas camadas ocultas de uma Rede Neural Profunda. O compartilhamento de parâmetros rígidos é comumente implementado ao partilhar as camadas ocultas entre todas as tarefas, enquanto se preservam várias camadas de saída específicas para cada tarefa (COSTA, 2022), conforme ilustrado na Figura 2.10.

Figura 2.10 – Compartilhamento de parâmetros rígidos para aprendizado multitarefa em redes neurais profundas.



Fonte: (Produzida pelo próprio autor).

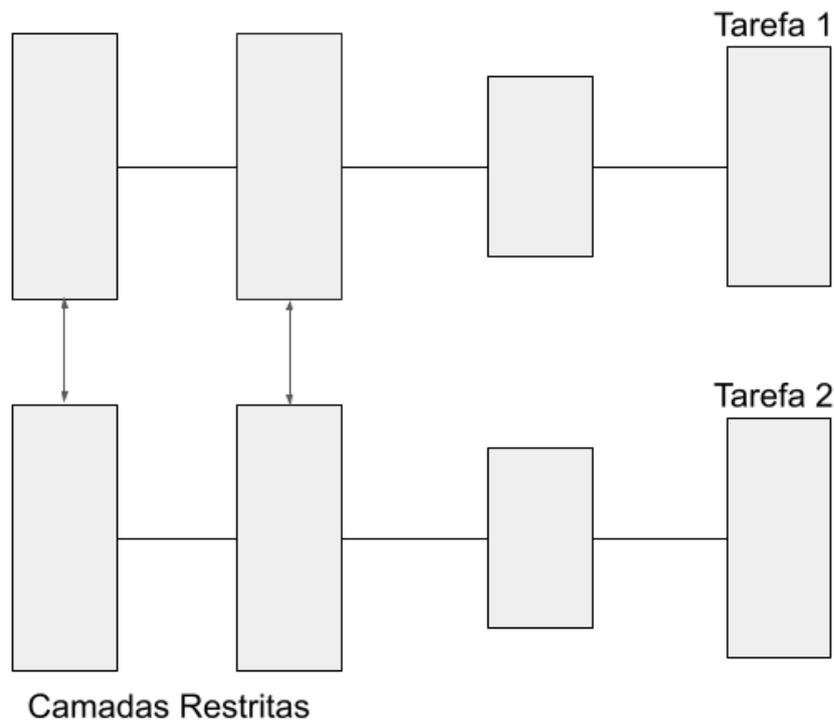
O compartilhamento rígido emerge como uma estratégia eficaz na mitigação do risco de *overfitting*. De maneira intuitiva, à medida que um maior número de tarefas é abordado simultaneamente, o modelo é compelido a buscar uma representação que englobe todas as tarefas, reduzindo, assim, a probabilidade de ajuste excessivo aos dados (COSTA, 2022).

No âmbito do Compartilhamento de Parâmetros Suaves, cada tarefa é dotada de seu próprio modelo, o qual é caracterizado por parâmetros distintos, conforme evidenciado na Figura 2.11. Nessa abordagem, a regularização da distância entre os parâmetros do modelo é implementada para promover semelhanças entre eles (COSTA, 2022).

### 2.1.8 Validação Cruzada - *k-fold*

A técnica de validação cruzada *k-fold* é um método direto e simples de validação amplamente utilizado na área de aprendizado de máquina (Leite, 2020). O termo “*K*” representa o número de subdivisões igualmente distribuídas em que o conjunto de dados é particionado, enquanto “*fold*” refere-se a cada um dos blocos resultantes dessas subdivisões. Essa abordagem,

Figura 2.11 – Compartilhamento de parâmetros suaves para aprendizado multitarefa em redes neurais profundas.



Fonte: (Produzida pelo próprio autor).

permite uma avaliação mais robusta do modelo, pois envolve a repetição do processo de treinamento e avaliação em diferentes combinações de subconjuntos de treinamento e teste, garantindo uma melhor estimativa do desempenho geral do modelo (Leite, 2020).

O *k-fold* é um método estatístico que desempenha um papel crucial na comparação e seleção de modelos. Sua abordagem é direta e de fácil compreensão. Para ilustrar seu funcionamento, consideremos um exemplo de validação cruzada com cinco iterações, isto é,  $k = 5$ . Nesse cenário, o conjunto de dados é dividido em cinco partes ou dobras. Durante a primeira iteração, o modelo utiliza a primeira dobra como conjunto de teste, enquanto as demais dobras são utilizadas como conjunto de treinamento. Na segunda iteração, a segunda dobra é utilizada como conjunto de teste, enquanto as demais continuam a ser utilizadas para treinamento. Esse processo é repetido até que cada uma das cinco dobras tenha sido utilizada como conjunto de teste. Esse procedimento permite uma avaliação mais robusta do modelo, pois todas as amostras do conjunto de dados têm a oportunidade de serem usadas tanto para treinamento quanto para teste, resultando em uma estimativa mais precisa do desempenho geral do modelo (Team, 2020). A Figura 2.12 ilustra o processo de *k-fold*.

### 2.1.9 Métricas de Avaliação de Modelos

As métricas de avaliação utilizadas em estudos de aprendizado de máquina desempenham um papel fundamental na análise e compreensão dos resultados obtidos. As métricas mais

Figura 2.12 – Exemplo da técnica *k-fold* com cinco subdivisões, ou seja, com *k* igual a 5.

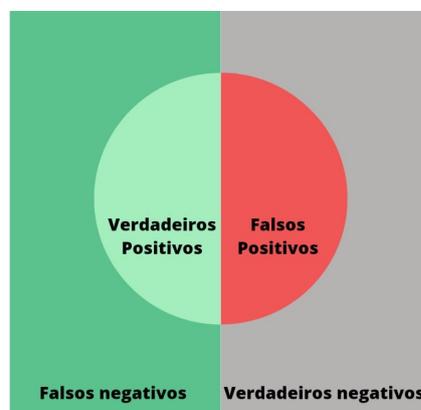


Fonte: (Shaikh, 2018).

utilizadas na literatura são acurácia, precisão, revocação e *F-score*.

Na Figura 2.13, são apresentadas quatro regiões distintas que representam as possíveis predições de um modelo. A região de cor verde escura representa os falsos negativos (FN), ou seja, o modelo faz uma previsão incorreta ao indicar que uma determinada imagem não pertence a uma classe, quando na verdade ela pertence. A região de cor verde clara indica os verdadeiros positivos (VP), onde o modelo prevê corretamente que uma imagem pertence a uma determinada classe. A região de cor vermelha representa os falsos positivos (FP), em que o modelo erroneamente prevê que uma imagem pertence a uma classe, quando na verdade não pertence. Por fim, a região de cor cinza representa os verdadeiros negativos (VN), em que o modelo acerta ao prever que uma imagem não pertence a uma determinada classe.

Figura 2.13 – Exemplo ilustrativo para melhor exemplificar as métricas de avaliação.



Fonte: (Produzida pelo próprio autor).

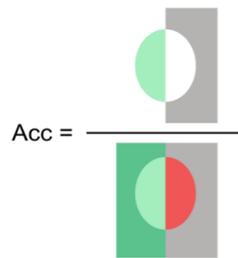
A acurácia é uma métrica amplamente conhecida e utilizada para avaliar o desempenho geral de um modelo de aprendizado de máquina. Ela representa a proporção de predições corretas

em relação ao número total de amostras avaliadas, exemplificando com base na Figura 2.13, seria as predições presentes na região de coloração verde escura e verde clara dividido pelo total de amostras. A acurácia pode ser definida por:

$$Acc = \frac{VP + VN}{VP + FP + FN + VN}, \quad (2.1)$$

onde, a acurácia é expressa como uma porcentagem e varia de 0 a 100%. Uma acurácia de 100% indica que todas as amostras foram corretamente classificadas pelo modelo (Mariano, 2021). Na Figura 2.14 pode-se analisar graficamente como a fórmula da acurácia é obtida.

Figura 2.14 – Exemplo da obtenção da fórmula da acurácia.



Fonte: (Produzida pelo próprio autor).

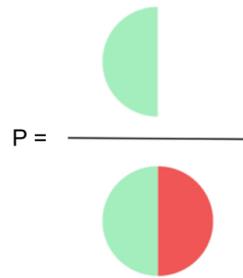
A precisão é uma métrica que fornece informações sobre a confiabilidade das predições positivas feitas pelo modelo. Ela representa a proporção de predições positivas corretamente classificadas em relação ao número total de predições positivas, exemplificando com base na Figura 2.13, seria as predições localizadas na região de coloração verde clara dividido pelas predições da região verde clara e vermelha. A precisão pode ser definida por:

$$P = \frac{VP}{VP + FP}. \quad (2.2)$$

Esta representação auxilia na avaliação da capacidade do modelo em evitar classificar erroneamente amostras negativas como positivas. Uma alta precisão indica que o modelo tem uma taxa baixa de falsos positivos (Mariano, 2021). A precisão é baseada para um problema binário (duas classes), em um contexto multi-classe, essa métrica é calculada para cada classe. Na Figura 2.15 pode-se analisar como a fórmula da precisão é obtida.

A revocação, também conhecida como sensibilidade, é uma métrica que mede a capacidade do modelo em encontrar corretamente todas as amostras positivas. Ela representa a proporção de amostras positivas corretamente identificadas em relação ao número total de amostras positivas no conjunto de dados, exemplificando com base na Figura 2.13, seria as predições presentes na região de coloração verde clara dividido pelas predições localizadas na região verde clara e verde escura. A revocação pode ser definida como:

Figura 2.15 – Exemplo da obtenção da fórmula da precisão.

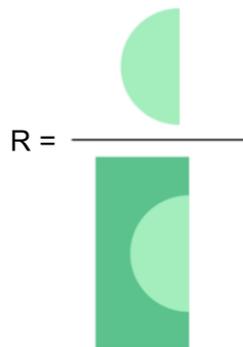


Fonte: (Produzida pelo próprio autor).

$$R = \frac{VP}{VP + FN}, \quad (2.3)$$

e é particularmente útil quando é fundamental evitar falsos negativos, ou seja, quando é crucial identificar corretamente todas as amostras positivas. Uma revocação alta indica que o modelo tem uma taxa baixa de falsos negativos (Mariano, 2021). Semelhante a precisão, a revocação é uma métrica medida por classe. Na Figura 2.16 pode-se analisar como a fórmula da revocação é obtida.

Figura 2.16 – Exemplo da obtenção da fórmula da revocação.



Fonte: (Produzida pelo próprio autor).

Por fim, o *F-score* ( $F1$ ) é uma métrica que combina a precisão e a revocação em uma única medida, fornecendo uma visão mais abrangente do desempenho do modelo. O *F-score* é calculado pela média harmônica da precisão e da revocação e é definido por:

$$F1 = 2 \times \frac{\text{precisão} \times \text{revocação}}{\text{precisão} + \text{revocação}}. \quad (2.4)$$

Essa métrica é útil quando é importante equilibrar a taxa de falsos positivos e falsos negativos. O valor do *F-score* varia de 0 a 1, onde 1 representa um modelo com alta precisão e alta revocação (Mariano, 2021).

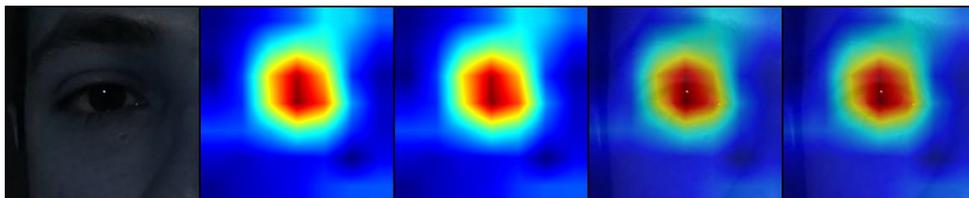
### 2.1.10 Interpretabilidade

A interpretabilidade é um conceito que vem ganhando atenção no campo de IA. Ele se refere à capacidade de compreender e explicar o funcionamento e as decisões de um modelo de *machine learning* de maneira clara e intuitiva (MEDICINA, 2023). Em outras palavras, trata-se da capacidade de um sistema de IA de ser compreendido e justificado por seres humanos, especialmente por especialistas não técnicos, como profissionais de diversas áreas que podem não ter conhecimento avançado em ciência de dados ou aprendizado de máquina.

Diversas técnicas são empregadas para alcançar a interpretabilidade, incluindo a visualização de dados, a simplificação de modelos complexos, a geração de explicações compreensíveis e a identificação de padrões relevantes nos processos de tomada de decisão (Mesameki *et al.*, 2023). Essas técnicas são fundamentais para promover a confiança e a transparência na aplicação de sistemas de IA, principalmente de reconhecimento biométrico, facilitando a detecção de possíveis vieses, erros ou comportamentos indesejados.

Inicialmente, para a interpretação da Figura 2.17, é imperativo reconhecer a representação em forma de mapa de calor. Neste contexto, cores mais frias, exemplificadas pelo azul, denotam regiões nas quais o algoritmo não incorporou significativamente em sua decisão. Em contrapartida, tonalidades mais quentes, como vermelho, alaranjado e amarelo, indicam áreas que exerceram maior influência na determinação do algoritmo, respectivamente.

Figura 2.17 – Demonstração de uma saída da técnica de interpretabilidade usando o Grad-CAM para um problema de reconhecimento periocular.



Fonte: (Produzida pelo próprio autor).

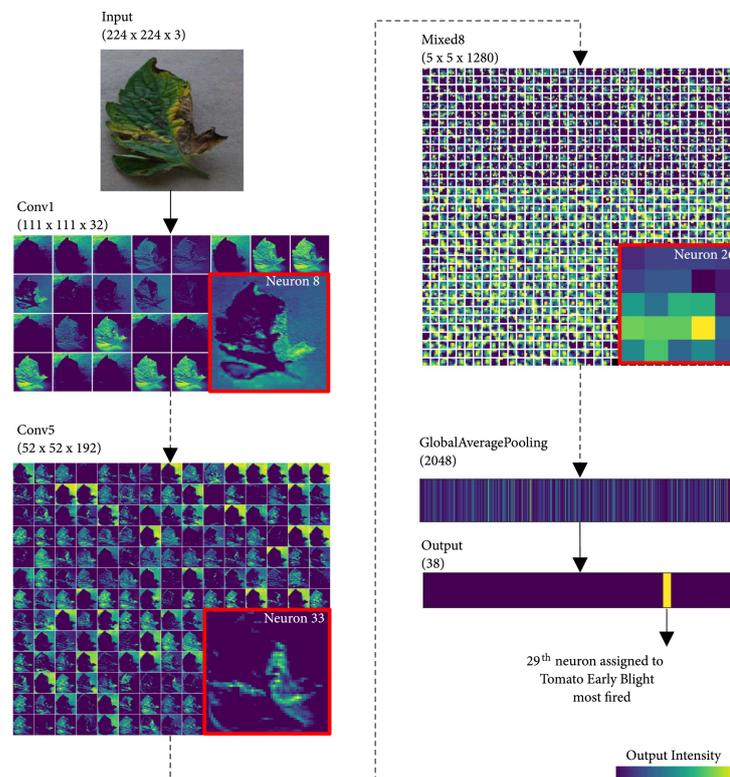
A compreensão da interpretabilidade de modelos de aprendizado de máquina abrange uma variedade de abordagens, sendo notáveis os métodos que se baseiam na análise das ativações em redes neurais convolucionais predefinidas. Esses métodos buscam examinar o comportamento das camadas convolucionais, mapeando as projeções dos níveis de ativação resultantes (Souza, 2023). É pertinente ressaltar que as camadas de *pooling*, por produzirem representações menos minuciosas das entradas, frequentemente são excluídas durante as análises fundamentadas em ativação (Souza, 2023).

Nesse contexto, destacam-se dois métodos notáveis de avaliação embasados em ativação: Ativações Intermediárias (Souza, 2023) e Maximização da Ativação (Souza, 2023). O método de Ativações Intermediárias quantifica os níveis de ativação de um conjunto específico de neurônios em uma rede neural. Isso é obtido pela análise dos níveis de ativação em relação a uma entrada particular, permitindo a identificação dos neurônios mais ativos durante o processamento desta

entrada (Souza, 2023). Durante a implementação desse método, procede-se à modificação da rede neural, direcionando os neurônios a serem analisados para se tornarem os neurônios de saída. Posteriormente, ao selecionar uma entrada, a rede é processada, e a saída resultante representa os níveis de ativação do segmento analisado (Souza, 2023).

A Figura 2.18 ilustra os níveis de ativação nas camadas intermediárias de um modelo de *Convolutional Neural Network* (CNN), especialmente treinado para o diagnóstico de doenças em plantas. A figura apresenta os níveis de ativação derivados de uma amostra de uma folha de tomate infectada com ferrugem precoce.

Figura 2.18 – Visualização das Saídas Intermediárias Produzidas por uma CNN Treinada. Uma imagem de uma folha de tomate infectada com ferrugem precoce (label 29) foi empregada como entrada para a rede, enquanto as saídas intermediárias das camadas são exibidas para visualização. Os módulos *Layer* e *Inception*, bem como o tamanho de suas saídas, são especificados abaixo de cada saída intermediária.



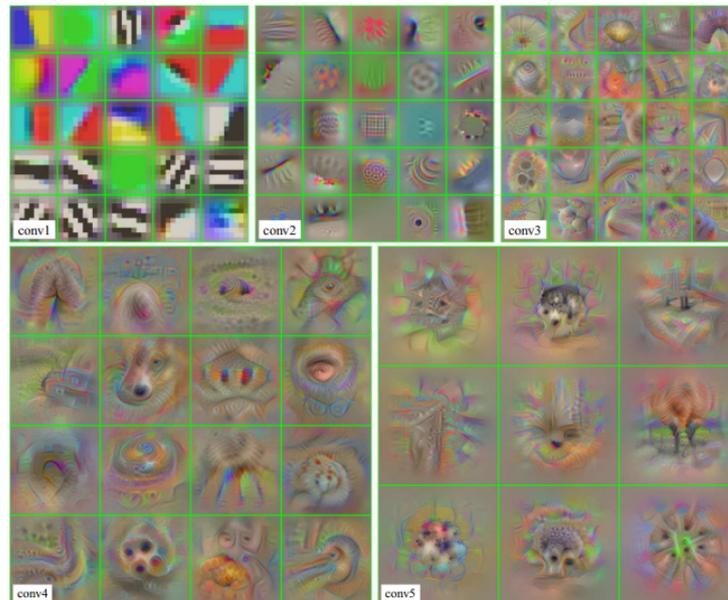
Fonte: (Toda; Okura, 2019).

Por sua vez, o método de Maximização da Ativação, proposto por Erhan e Courville (2009), tem como propósito identificar um padrão de entrada que maximize os níveis de ativação de uma unidade específica nas redes neurais. Esse processo é modelado como um problema de otimização, em que os neurônios de entrada da rede são tratados como variáveis de decisão e os níveis de ativação como a função objetivo (Souza, 2023).

A Figura 2.19 apresenta um exemplo de ativações máximas em uma CNN. Neste caso, os autores (Mahendran; Vedaldi, 2015) empregam o método de Maximização da Ativação na

VGG-M. A imagem retrata as ativações máximas dos primeiros filtros de cada uma das camadas da referida rede.

Figura 2.19 – Exploração de Ativações nos Primeiros Filtros de Cada Camada Convolutiva da VGG-M.



Fonte: (Mahendran; Vedaldi, 2015).

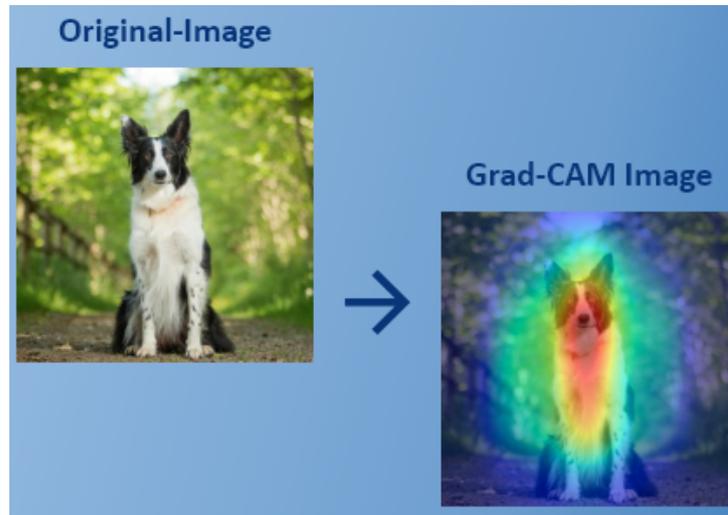
### 2.1.10.1 GradCAM

O Mapeamento de Ativação de Classe Ponderado pelo Gradiente, ou *Gradient-weighted Class Activation Mapping* (GradCAM), é uma técnica empregada em aprendizado profundo e visão computacional que facilita a visualização das regiões de uma imagem nas quais uma rede neural convolutiva (CNN) concentra sua atenção ao efetuar previsões. Constitui-se como uma ferramenta robusta para elucidar quais segmentos de uma imagem são preponderantes no processo decisório da rede neural (Trade, 2023).

Conforme ilustrado na Figura 2.20, apresenta-se inicialmente a representação original da imagem, a qual é submetida ao modelo para análise e subsequente tomada de decisões. Em contrapartida, a figura adjacente revela a mesma imagem, porém enriquecida com a sobreposição de um mapa de calor. O mapa de calor evidencia as regiões da imagem que exerceram influência na decisão tomada pelo modelo, neste contexto, atribuindo relevância à representação do cachorro para sua classificação como cão.

O Grad-CAM gera um mapa de calor correspondente a cada categoria de classificação possível. Inicialmente, são determinados os gradientes associados a cada canal da camada convolutiva final em relação à categoria alvo. Por meio da avaliação desses gradientes em relação aos mapas de características da derradeira camada convolutiva da CNN, o Grad-CAM estabelece a pontuação pertinente à classe em questão. Os mapas de características, então, são ponderados conforme tais gradientes, resultando em um mapa de calor que evidencia as regiões

Figura 2.20 – Ilustração comparativa entre imagens pré e pós-aplicação do GradCAM, destacando, por meio de um mapa de calor, as regiões determinantes nas decisões do modelo.



Fonte: (Trade, 2023).

de destaque da imagem original. Para visualizar as regiões mais relevantes para a CNN, esse mapa de calor pode ser integrado à imagem de origem (Dakshina, 2023).

Conclusivamente, o Grad-CAM tem evidenciado sua aplicabilidade multifacetada em diversos domínios, englobando a classificação de imagens, o reconhecimento de objetos e a segmentação. Este método se estabelece como uma ferramenta robusta e eficiente na visualização dos padrões de ativação inerentes a modelos de aprendizagem profunda (Dakshina, 2023). Adicionalmente, destaca-se sua capacidade de aprimorar a interpretabilidade e a transparência dos modelos, bem como de facilitar processos de depuração e refinamento em contextos de aprendizagem profunda. Por meio do Grad-CAM, é viabilizada uma compreensão visual elucidativa acerca do mecanismo decisório subjacente às redes neurais (Dakshina, 2023).

## 2.2 Trabalhos Relacionados

Kaur, Singh e Kumar (2014) conduziram um levantamento da literatura acerca do reconhecimento biométrico, por meio de diversas modalidades, tais como: reconhecimento facial, reconhecimento de íris, reconhecimento da palma da mão, reconhecimento por voz, reconhecimento por impressão digital, e o sistema de reconhecimento baseado no sinal do eletrocardiograma (ECG). A pesquisa abarcou uma revisão sistemática da literatura científica pertinente à época, e expôs tanto as vantagens quanto as desvantagens de cada método de reconhecimento biométrico em questão. Cada um desses métodos é acompanhado de diferentes técnicas para a extração de recursos de reconhecimento. Ademais, esses métodos apresentam características distintas no que tange à identificação de indivíduos. Tais características são determinantes para a escolha do método mais adequado em diferentes contextos, tais como iluminação, qualidade da imagem, presença de objetos que possam interferir na detecção, dentre outros fatores. Por conseguinte,

foi possível concluir o desempenho de diferentes modalidades de biometria, como ilustrado na Tabela 2.1.

Tabela 2.1 – Comparativo entre diferentes biometrias.

Biometria	Singularidade	Atuação	Mensurabilidade
Reconhecimento facial	Alto	Médio	Alto
Reconhecimento de íris	Alto	Alto	Médio
Reconhecimento de ECG	Alto	Alto	Médio
Reconhecimento de voz	Baixo	Baixo	Médio
Reconhecimento de palma	Médio	Médio	Alto

Fonte: (Kaur; Singh; Kumar, 2014) - Traduzida pelo autor.

Barcellos e Gonzaga (2021) propuseram uma abordagem alternativa para a extração de características em uma Rede Neural Convolutacional (CNN). Em vez de utilizar as ativações da CNN como um vetor de características, a CNN foi utilizada para extrair características diretamente dos mapas de características. Em seguida, um descritor de textura baseado no padrão de pontos binários locais uniformes (LBP) foi aplicado a esses mapas de características. Para avaliar a eficácia do método proposto pelos autores, foi realizado um experimento de transferência de aprendizado (*transfer learning*) usando a rede AlexNet, pré-treinada no conjunto de dados do *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)* (Russakovsky *et al.*, 2014), como extrator de recursos. O conjunto de dados *Visible Light Mobile Ocular Biometric (VISOB)* (Rattani *et al.*, 2016) foi utilizado para testar o desempenho do método proposto em imagens perioculares coletadas em diferentes *smartphones* e condições de iluminação.

Os resultados obtidos pelos autores foram apresentados na Tabela 2.2, que mostra a menor taxa de erro de igualdade (EER) obtida e a camada correspondente da CNN para cada variação do método proposto em diferentes *smartphones* e condições de iluminação do conjunto de dados VISOB. Com base nos resultados obtidos, os autores concluíram que a abordagem proposta, que utiliza um descritor de textura na CNN, supera as abordagens artesanais de visão computacional e até mesmo a extração padrão de características de CNN em termos de desempenho.

O estudo realizado por Dias e Jr. (2019) tem como objetivo avaliar o desempenho da rede neural artificial profunda no reconhecimento da região periocular. Para tal, o modelo escolhido para a avaliação foi a CNN ResNet-34, que foi pré-treinada na base de imagens de região periocular Ubiris.v2 (Proenca *et al.*, 2010). O modelo em questão foi treinado durante um período de duas semanas, abrangendo 245 épocas. O resultado mais satisfatório obtido foi de cerca de 92% de precisão na base de dados Ubiris.v2. Diante disso, os autores concluíram que, embora a transferência de conhecimento apresente perspectivas promissoras, não é suficiente para garantir a precisão adequada de um sistema de reconhecimento de biometria periocular.

Alahmadi, Hussain e Aboalsamh (2022) propuseram um sistema de reconhecimento biométrico sem contato, utilizando a região periocular como região de interesse. Os autores utilizaram um modelo CNN pré-treinado como base e introduziram um modelo de reconhecimento

Tabela 2.2 – Avaliação de desempenho de cada método comparado nos diferentes *smartphones* e condições de iluminação da base de dados VISOB (Rattani *et al.*, 2016).

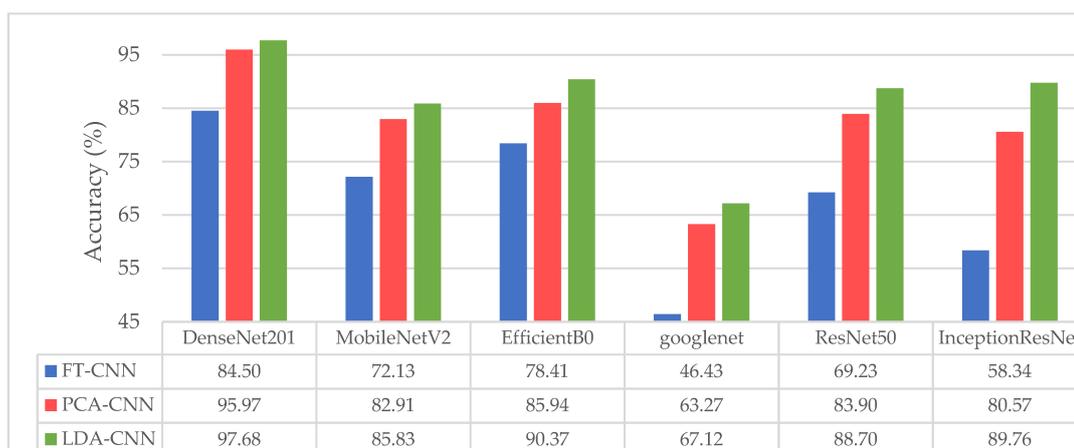
Light	EER (%)																							
	Day Light						Dim Light						Office Light											
	iPhone		Oppo		Samsung		iPhone		Oppo		Samsung		iPhone		Oppo		Samsung							
Phone	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right		
Eye Side																								
uLBP	30,65	29,08	50,73	42,91	19,85	29,33	53,22	48,91	25,8	27,32	27,45	27,74	41,66	35,54	32,15	30,19	26,72	30,44	30,61	40,34	27,22	27,16		
HOG	30,02	29,63	47,43	43,97	27,27	33,08	44,06	44,04	31,67	30,69	31,97	36,62	30,89	33,7	33,18	32,11	30,61	40,34	30,61	40,34	27,22	27,16		
SURF	18,09	18,52	19,4	21,29	20,68	20,6	17,01	16,66	19,34	17,71	22,31	21,62	26,45	25,9	27,72	27,84	27,16	27,22	27,16	27,22	27,22	27,16		
SIFT	7,79	8,5	9,26	10,78	8,62	7,22	7,93	7,68	7,73	7,35	8,32	8,94	8,31	9,86	14,37	15,69	12,41	10,08	12,41	10,08	10,08	10,08		
FE-raw <sup>d</sup>	5,84	6,47	10,03	9,91	9,46	7,87	5,54	5,59	4,57	5,13	6,99	5,54	7,78	8,93	9,29	11,91	16,71	13,31	16,71	13,31	13,31	13,31		
FE-LBP-FM-Conc <sup>a</sup>	7,76	8,22	11,56	12,3	11,53	11,32	10,35	12,23	7,38	8,62	8,15	6,73	9,77	11,35	11,35	13,36	13,37	12,88	13,37	12,88	12,88	12,88		
FE-LBP-FM-Norm-Conc <sup>a</sup>	3,37	3,96	9,77	9,34	7,09	5,59	5,94	5,98	4,09	4	4,07	2,71	5,78	6,38	7,63	9,55	7,17	5,61	7,17	5,61	5,61	5,61		

Fonte: (Barcellos; Gonzaga, 2021) - Traduzida pelo autor.

periocular profundo chamado *Linear Discriminant Analysis CNN* (LDA-CNN), no qual a camada LDA foi incorporada após a última convolução do modelo base. O modelo proposto foi avaliado em quatro conjuntos de dados distintos: UFPR-Periocular (Zanlorensi *et al.*, 2022), UBIRIS.v2 (Proenca *et al.*, 2010), VISOB (Rattani *et al.*, 2016) e UBIPr (Proenca *et al.*, 2012). Os melhores resultados alcançados pelo modelo foram de 97.68% utilizando a base de dados UFPR-Periocular, 99.82% utilizando a base de dados UBIRIS.v2 e 99.17% utilizando a base de dados UBIPr. A base de dados VISOB foi utilizada para experimentos com diferentes *smartphones*, resultando em uma acurácia máxima de 98.49%, e também para experimentos de iluminação cruzada, obtendo uma acurácia de 98.52%.

A Figura 2.21 apresenta as variações de acurácia do modelo proposto por Alahmadi, Hussain e Aboalsamh (2022) (LDA-CNN) ao utilizar diferentes modelos base em conjunto com a base de dados UFPR-Periocular. É possível observar que o modelo base DenseNet-201 obteve a melhor acurácia nessa base de dados, alcançando um valor de 97,68%.

Figura 2.21 – Avaliação comparativa do desempenho do modelo proposto LDA-CNN em relação a outros modelos, utilizando diferentes modelos base, no conjunto de dados UFPR-Periocular.



Fonte: (Alahmadi; Hussain; Aboalsamh, 2022).

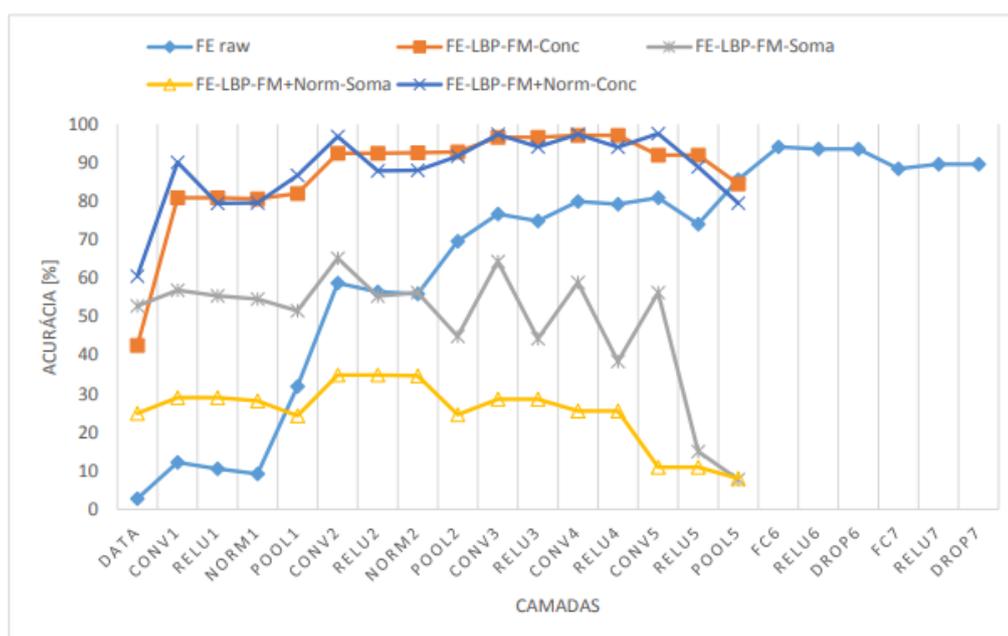
Em (Barcellos *et al.*, 2019), foi realizada uma avaliação do desempenho de técnicas de aprendizagem por transferência (em inglês, *Transfer Learning*) em redes neurais convolucionais para a aplicação em biometria periocular. As técnicas avaliadas no aprendizado por transferência (aprendizado de máquina que se concentra em armazenar o conhecimento adquirido ao resolver um problema e aplicá-lo a um problema diferente, mas relacionado) foram a Extração de Características e o *Fine Tuning*. Para essa avaliação, foram testadas duas arquiteturas de redes neurais convolucionais, a AlexNet e a VGG-16, em dois bancos de imagens da região ocular comumente utilizadas para o reconhecimento de íris, a base LAVI DB2 (LAVI, 2019) e a base ND-IRIS-0405 *Iris Image Dataset* (LG2200) (Bowyer, 2016). As bases de imagens utilizadas apresentam diversas características distintas.

O estudo e a avaliação do desempenho das técnicas de aprendizagem por transferência foi

dividido em três etapas. Na primeira etapa, a precisão da Extração de Características foi medida. Na segunda etapa, avaliou-se o desempenho do *Fine Tuning*. Na terceira etapa, utilizou-se a arquitetura AlexNet para o ajuste fino em um banco de dados. Com base nos resultados obtidos, os autores concluíram que a qualidade dos dados, o desequilíbrio de classe e o método de seleção do treinamento e teste influenciam diretamente na precisão da rede neural convolucional. Além disso, constatou-se que a técnica de Extração de Características apresenta uma menor precisão do que o *Fine Tuning*. Por fim, o ajuste fino de uma rede neural convolucional com imagens perioculares de um banco de dados não aumenta a precisão desta rede no modo de Extração de Características para outra base de dados de imagens perioculares, sendo a precisão bastante semelhante à obtida pela rede pré-treinada original.

Barcellos (2022) propôs o uso do descritor LBP uniforme nas ativações das camadas profundas de uma rede neural convolucional (CNN). Esse método foi aplicado aos vetores de características extraídos de imagens da região periocular, incluindo o olho, para fins de reconhecimento biométrico. A abordagem proposta foi avaliada na base de imagens UBIPr (Proenca *et al.*, 2012), e os resultados de acurácia obtidos nas diversas camadas da CNN AlexNet foram apresentados na Figura 2.22. Os resultados mostraram que a aplicação do descritor LBP nas ativações das camadas da CNN AlexNet resultou em desempenho superior em relação às metodologias tradicionais, como uLBP, HOG, SURF e SIFT. A análise dos resultados permitiu concluir que as imagens intermediárias geradas pelos filtros previamente treinados por uma CNN *off-the-shelf* (pronta para uso) possuem características texturais distintas. A extração dessas características pode, portanto, melhorar o reconhecimento periocular com a inclusão do olho, em comparação com outros métodos tradicionais de visão computacional.

Figura 2.22 – Resultados de acurácias obtidas pela abordagem proposta em (Barcellos, 2022).



Fonte: (Barcellos, 2022).

Em (Luz *et al.*, 2018) é apresentada uma técnica para o reconhecimento de indivíduos com base na região periocular, que é a área ao redor dos olhos. Utiliza-se uma CNN pré-treinada como extratora de recursos para a região periocular. A arquitetura da CNN usada foi a Inception-v3, que é uma rede profunda popularmente utilizada para tarefas de classificação de imagens, mas o que difere essa abordagem das demais é que após a extração de recursos da região periocular usando a CNN, os autores utilizaram uma técnica de classificação baseada em SVM para identificar indivíduos. SVM é um algoritmo de aprendizado de máquina supervisionado amplamente utilizado para classificação de dados. Seus resultados para a modalidade da região periocular do rosto foram de até 95% de taxa de reconhecimento.

Por fim, Zanlorensi *et al.* (2022) propõem um novo conjunto de dados periorculares, o UFPR-Periocular. Esses dados foram coletados por meio de dispositivos móveis, abrangendo cenários irrestritos. Tal conjunto de dados consiste em amostras de 1.122 indivíduos, coletadas em 3 sessões distintas.

As imagens presentes na base de dados UFPR-Periocular (Zanlorensi *et al.*, 2022) foram capturadas pelos próprios indivíduos em ambientes não restritos, ou seja, eles tiveram a liberdade de escolher locais diversos para a captura das imagens. A única instrução fornecida aos participantes foi que posicionassem seus olhos em uma região de interesse. Os autores conduziram uma extensa avaliação comparativa, utilizando a base de dados proposta pelos mesmos e diversos modelos de CNNs, aplicados em abordagens de última geração, como classificação multi-classe, aprendizado multitarefa, rede de filtros emparelhados e rede siamesa. Para o protocolo de mundo fechado, foram utilizados modelos de CNN como VGG-16, ResNet-50-V2, MobileNet-V2 e um modelo Multitarefa, com resultados alcançados de 50,56%, 63,18%, 77,9% e 84,32%, respectivamente.

Os métodos utilizados no protocolo de mundo fechado, juntamente com seus resultados correspondentes, podem ser visualizados na Tabela 2.3.

Tabela 2.3 – Resultados de *benchmark* no protocolo de mundo fechado para a tarefa de identificação.

Modelo	Identificação (1:N)	
	Classificação 1 (%)	Classificação 5 (%)
VGG-16	50,56 ± 3,30	68,73 ± 3,01
VGG16-Face	56,29 ± 1,62	73,84 ± 1,48
Xception	57,43 ± 1,43	75,88 ± 1,52
ResNet50V2	63,18 ± 2,14	77,79 ± 1,81
InceptionResNet	65,16 ± 2,45	81,53 ± 1,99
ResNet50	71,06 ± 1,14	85,22 ± 0,82
ResNet50-Face	74,76 ± 1,43	86,86 ± 1,02
DenseNet121	75,54 ± 1,36	88,53 ± 0,97
MobileNetV2	77,98 ± 1,08	90,19 ± 0,79
Multitarefa	84,32 ± 0,71	94,55 ± 0,58

Fonte: (Zanlorensi *et al.*, 2022) - Traduzida pelo autor.

Na Tabela 2.4, são descritas as técnicas empregadas pelos autores referenciados neste estudo em suas respectivas bases de dados.

Tabela 2.4 – Técnicas utilizadas pelos autores citados e as respectivas bases de dados utilizadas.

Autores	Ano	Técnica Utilizada	Base de Dados Utilizada
Vizoni e Marana (2019)	2019	VGG-Face	Face Recognition Grand Challenge (FRGC)
Barcellos e Gonzaga (2021)	2021	AlexNet	ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
Dias e Jr. (2019)	2019	ResNet-34	Uribis.v2
Alahmadi, Hussain e Aboalsamh (2022)	2022	Linear Discriminant Analysis CNN (LDA-CNN)	UFPR-Periocular, UBIRIS.v2, VISOB e UBIPr
Barcellos <i>et al.</i> (2019)	2019	AlexNet e VGG-16	LAVI DB2 e ND-IRIS-0405 Iris Image Dataset (LG2200)
Barcellos (2022)	2022	AlexNet	UBIPr
Luz <i>et al.</i> (2018)	2018	Inception-v3	UBIRIS.v2 e MobBio
(Zanlorensi <i>et al.</i> , 2022)	2022	VGG16, VGG16-Face, Xception, ResNet50V2, InceptionResNet, ResNet50, ResNet50-Face, DenseNet121 e MobileNetV2	UFPR-Periocular

Fonte: (Produzida pelo próprio autor).

No presente contexto, mediante a avaliação da Tabela 2.4, é possível inferir a relevância contemporânea do reconhecimento biométrico focado na região periocular, dado o emprego de múltiplas técnicas visando sua potencial resolução. Notavelmente, as arquiteturas AlexNet e VGG16 emergem como as mais frequentemente adotadas neste domínio. Adicionalmente, é pertinente destacar a variedade de conjuntos de dados empregados nos estudos pertinentes.

## 3 Desenvolvimento

Neste capítulo, serão apresentadas a base de dados usada na [Seção 3.1](#) e metodologia na [Seção 3.2](#).

### 3.1 Descrição da Base de Dados

A base de dados empregada neste trabalho foi a UFPR-Periocular *Dataset* (Zanlorensi *et al.*, 2022), que abrange uma compilação de imagens específicas da região periocular, destinadas à finalidade de biometria. O conjunto de dados é composto por um total de 33.660 imagens correspondentes a olhos individuais, ou seja, 16.830 imagens em que os dois olhos foram capturados simultaneamente. Essas imagens foram adquiridas a partir de uma amostra de 1.122 indivíduos distintos, resultando em um conjunto de dados composto por 2.224 classes únicas.

As imagens foram adquiridas mediante a colaboração dos participantes utilizando seus *smartphones* pessoais, por meio de um aplicativo móvel especificamente desenvolvido pelos responsáveis pela base de dados. Cada indivíduo contribuiu com um total de 15 imagens dos seus olhos, obtidas em três cenários distintos, os quais foram realizados com um intervalo mínimo de oito horas entre eles. Sendo assim, cada cenário contém cinco imagens de cada indivíduo, e as imagens pertencentes aos respectivos cenários possuem uma diferença temporal. É importante ressaltar que a resolução das imagens apresenta variação, abrangendo uma faixa que varia de  $360 \times 160$  *pixels* a  $1862 \times 1008$  *pixels*, sendo influenciada pelos diferentes modelos de *smartphones* utilizados para a captura das imagens. Nesse contexto, foram empregados 196 dispositivos móveis diferentes.

Além disso, a base de dados engloba dois protocolos distintos, conhecidos como mundo fechado e mundo aberto, cada um desses protocolos são compostos por três cenários. No contexto do mundo fechado, o conjunto de dados é particionado de forma a incluir amostras provenientes do mesmo sujeito, ou seja, tanto os conjuntos de treinamento quanto de teste contêm amostras do mesmo sujeito. Por outro lado, no protocolo de mundo aberto, há sujeitos diferentes presentes nos conjuntos de treinamento e de teste. Nesse contexto, a tarefa de identificação é executada sob o protocolo de mundo fechado, onde a capacidade de reconhecimento se estende a sujeitos não presentes no conjunto de treinamento.

Devido às imagens terem sido capturadas pelos próprios indivíduos em diferentes sessões, a base de dados apresenta uma variedade de condições desafiadoras para os algoritmos de identificação periocular. Essas condições incluem variações na iluminação, presença de oclusões, reflexões especulares, desfoque de movimento, uso de óculos, diferentes ângulos de visão, fixação do olhar, maquiagem e expressões faciais.

No processo de inclusão das imagens na base de dados, os criadores realizaram anotações manuais, marcando quatro pontos de referência na região do olho em cada imagem. Essas anotações foram utilizadas para normalizar a região periocular em termos de escala e rotação. Além disso, as imagens foram cortadas para manter apenas a região de interesse, como pode ser observado na Figura 3.1, a fim de reduzir a influência de informações irrelevantes e focar especificamente na área periocular.

Figura 3.1 – Amostra de imagens da base de dados UFPR-Periocular, onde é possível observar uma grande diversidade nas imagens.



Fonte: (Zanlorensi *et al.*, 2022).

O processo de divisão da base de dados UFPR-Periocular funciona como um *k-fold* com  $k = 3$ , em cada um dos cenários disponibilizados, uma parte diferente da base de dados é selecionada como conjunto de treinamento, conjunto de validação e conjunto de teste.

### 3.1.1 Divisão dos Dados

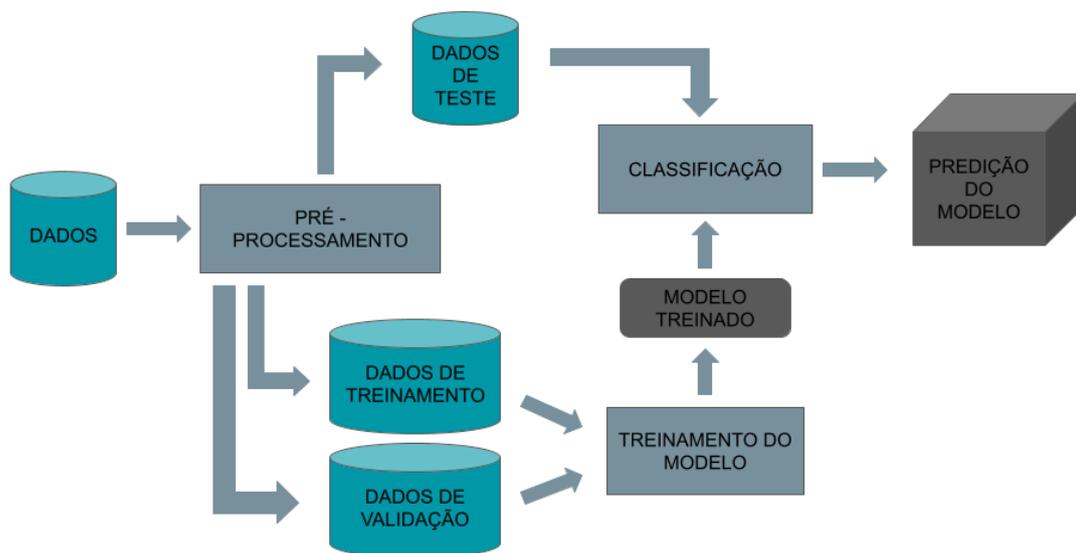
Os dados utilizados para a condução deste trabalho foram obtidos por meio do protocolo de mundo fechado. A base de dados é composta por subconjuntos distintos, incluindo um conjunto de treinamento, um conjunto de validação e um conjunto de teste. O subconjunto de treinamento, que consiste em 13.464 imagens, é empregado para instruir o algoritmo durante a etapa de treinamento. Por sua vez, o subconjunto de validação, composto por 8.976 imagens, é utilizado para avaliar o desempenho do modelo, isto é, a capacidade de generalizar o conhecimento adquirido a partir do subconjunto de treinamento. Por fim, o subconjunto de teste, contendo 11.220 imagens, desempenha o papel de avaliar a eficácia final do modelo. Assim, as imagens dos respectivos subconjuntos estão distribuídas em um total de 2.244 classes distintas. A divisão dos

dados em treino, validação e teste foi realizada através da técnica *k-fold* ( $k = 3$ ). Vale ressaltar que foi seguida a divisão proposta pelos próprios autores da base (Zanlorensi *et al.*, 2022).

## 3.2 Metodologia Proposta

Na Figura 3.2, apresenta-se uma representação visual do fluxo de treinamento, adotado neste trabalho. Inicialmente, os dados são adquiridos, ou seja, as imagens são obtidas a partir da base de dados UFPR-Periocular. Em seguida, é realizado o pré-processamento das imagens, envolvendo a aplicação de uma série de procedimentos com o propósito de prepará-las de forma adequada e informativa para o processo de reconhecimento periocular. Posteriormente, as imagens são particionadas em três conjuntos distintos: teste, treinamento e validação. O conjunto de treinamento e o conjunto de validação são empregados para o treinamento do modelo proposto. Após a adaptação do modelo com base nesses conjuntos de imagens, tem-se um modelo treinado, o qual é utilizado para realizar a classificação das imagens presentes no conjunto de teste. Dessa forma, são efetuadas as previsões do modelo, proporcionando resultados de classificação para as imagens em questão.

Figura 3.2 – Fluxograma do modelo descrevendo os estágios sequenciais, iniciando no processo de aquisição de dados e culminando nas previsões geradas pela abordagem proposta.



Fonte: (Produzida pelo próprio autor).

A seguir, são descritos cada passo presentes no fluxograma da Figura 3.2.

### 3.2.1 Descrição do Pré-processamento dos Dados

Durante o pré-processamento das imagens, uma série de etapas são realizadas para prepará-las para o processamento subsequente. Uma dessas etapas é a normalização, na qual as imagens são divididas por 255. Essa normalização tem como objetivo ajustar os valores dos

*pixels* no intervalo de 0 a 1, tornando-os mais adequados para o processamento do treinamento (Carneiro, 2020).

Além disso, as dimensões das imagens são reordenadas durante o pré-processamento. Mais especificamente, o canal de cores das imagens é colocado na primeira posição da matriz, enquanto as outras dimensões (como altura e largura) são reorganizadas de acordo, ficando (*canais, altura, largura*). Essa reordenação é realizada para que as imagens estejam no formato adequado para o modelo de rede neural convolucional, que espera um tensor nesse formato.

Por fim, uma etapa de redimensionamento é aplicada às imagens lidas. Essa etapa tem como objetivo ajustar o tamanho das imagens para uma dimensão específica, nesse caso,  $(224 \times 224)$ . O redimensionamento é realizado utilizando o método de interpolação por área, que ajusta o tamanho da imagem preservando as informações visuais importantes por meio de um cálculo de média ponderada. Dessa forma, as imagens são padronizadas para terem o mesmo tamanho, o que é essencial para garantir a consistência e a compatibilidade durante o processamento posterior.

### 3.2.2 Periocular EfficientNet-B0 (PEN)

A arquitetura de modelo base selecionada é a EfficientNet-B0, a qual é uma das arquiteturas pertencentes à família de modelos da EfficientNet. A família de modelos da EfficientNet foi inicialmente proposto por Tan e Le (2019), estabelecendo um paradigma para a ampliação de modelos de redes neurais. Este enfoque abrange o aprimoramento das dimensões de profundidade, largura e resolução, proporcionando uma notável melhoria na eficácia. O EfficientNet é um método que se insere no domínio do dimensionamento e concepção de CNNs, sendo caracterizado pela aplicação de um coeficiente composto que atua de maneira uniforme sobre todas as dimensões supracitadas (Shamila Ebenezer *et al.*, 2022).

Para aproveitar os benefícios dos pesos pré-treinados disponíveis na biblioteca, a arquitetura EfficientNet-B0 é carregada utilizando pesos pré-treinados à partir da ImageNet como ponto de partida.

No entanto, tratando-se do modelo focalizado exclusivamente em uma única tarefa, algumas modificações são realizadas na arquitetura antes do modelo final ser criado. A primeira modificação consiste na remoção da camada de classificação final. Essa camada é eliminada para permitir a adição de camadas personalizadas que atendam ao problema específico em questão.

As camadas personalizadas incorporadas ao modelo incluem a função *Flatten()*, empregada para transformar o tensor de entrada em um vetor unidimensional, seguida por uma camada densamente conectada. A dimensão da saída desta camada é definida em conformidade com o número de classes, que, neste contexto, corresponde a 2.244. Para obter uma distribuição de probabilidade relativa às classes, as saídas do modelo são submetidas à função de ativação *Softmax*.

Portanto, o modelo final é composto pela arquitetura EfficientNet-B0, na qual foram

adicionadas camadas personalizadas para se adequar à tarefa de identificação. Essas alterações permitem que o modelo aprenda representações complexas dos dados e realize a classificação adequada. Essa nova arquitetura criada é nominada no presente trabalho como Periocular EfficientNet-B0, ou PEN.

### 3.2.3 Multitarefa Periocular EfficientNet-B0 (MPEN)

Para a construção do modelo utilizando a técnica de multitarefas, utilizou-se um compartilhamento de parâmetros rígidos. Essa nova arquitetura denominada MPEN é delineada por um *backbone*, que é a EfficientNet-B0, seguido por três camadas totalmente conectadas independentes: uma destinada à classificação do olho, outra à classificação do *smartphone* utilizado e uma terceira para a classificação biométrica, na qual todos os dados são submetidos à análise conjunta. Cada camada totalmente conectada é sucessivamente seguida por uma função de ativação *Softmax*.

Dessa maneira, os modelos designados para as tarefas específicas apresentam as seguintes estruturas de camadas após o emprego do modelo base. Inicialmente, é adotada uma camada de achatamento (*Flatten*), seguida por uma camada linear, cuja saída é dimensionada de acordo com o número de classes a serem classificadas por cada modelo. Nesse contexto, são consideradas 2 classes para a tarefa de reconhecimento dos olhos, sendo esquerdo ou direito, 196 classes para a tarefa associada ao *smartphone* utilizado e 2.244 classes para a classificação biométrica.

A função de perda empregada para as três saídas distintas foi a *CrosEntropyLoss*. A expressão matemática para o cálculo do erro (*CELoss*) para uma única saída é dada por:

$$CELoss = - \sum_i y_i \cdot \log(\hat{y}_i), \quad (3.1)$$

onde,  $y_i$  representa a verdadeira distribuição de probabilidade (rótulos) e  $\hat{y}_i$  representa a distribuição de probabilidade predita pelo modelo para a classe  $i$ . O índice  $i$  percorre todas as classes.

Em seguida, cada saída específica (qual olho, qual *smartphone* e biometria) produz uma perda individual ( $CELoss_{olho}$ ,  $CELoss_{smart}$ ,  $CELoss_{bio}$  respectivamente) com base na *CrosEntropyLoss*. A *Loss* final ( $Loss_{Final}$ ) é determinada pela soma dessas perdas individuais:

$$Loss_{Final} = CELoss_{olho} + CELoss_{smart} + CELoss_{bio}. \quad (3.2)$$

Dessa forma, todas as funções de perda têm o mesmo peso no treinamento do modelo.

### 3.2.4 Treinamento dos Modelos

No processo de treinamento de ambos os modelos PEN e MPEN, os dados de treinamento foram divididos em lotes e passados pelo modelo. Para cada lote, as previsões foram calculadas e comparadas com as respectivas etiquetas de classe usando as funções de perda. Em seguida, o otimizador realizou a retropropagação do gradiente para atualizar os parâmetros do modelo.

Durante o treinamento, também foram monitoradas a perda média e a acurácia do modelo nos dados de validação. Após cada época, o modelo foi avaliado utilizando o conjunto de validação para medir sua capacidade de generalização e fazer ajustes necessários.

Após o treinamento, o modelo foi utilizado para realizar previsões nos dados de teste. Para cada entrada do conjunto de teste, as previsões foram calculadas e comparadas com as classes reais para medir o desempenho do modelo.

Na avaliação do modelo são utilizadas as métricas de perda (*loss*) e de acurácia (*accuracy*). A perda é calculada usando a função de perda *CrossEntropyLoss*, e a acurácia é calculada comparando as previsões do modelo com as classes reais.

Para avaliar o treinamento dos modelos, utilizou-se as métricas de acurácia, precisão, revocação e F1.

### 3.2.5 Interpretação do Aprendizado

Além das métricas tradicionais de avaliação do modelo treinado, outra técnica de avaliação empregada consistiu na análise da interpretabilidade. O propósito dessa técnica é conferir clareza e acessibilidade às informações, simplificando a análise e compreensão dos resultados. A abordagem implementada para esta finalidade envolve a geração de mapas de ativação visualmente interpretáveis, conhecidos como *Gradient-weighted Class Activation Mapping* (GradCAM).

Primeiramente, é realizado o carregamento do modelo previamente treinado, sendo este adaptado para o modo de avaliação, com a remoção de sua camada de classificação final. Subsequentemente, são gerados mapas de ativação para cada imagem fornecida, mediante a utilização da técnica GradCAM. Por fim, os resultados obtidos são armazenados em formato de imagem, visando análises visuais posteriores.

Este procedimento é essencial para a avaliação da interpretabilidade do modelo, permitindo a identificação das áreas de maior influência nas imagens e promovendo uma compreensão mais profunda do processo decisório da referida modelagem. A visualização dos mapas de ativação proporciona *insights* sobre as regiões que mais influenciam as previsões, contribuindo para a interpretabilidade e confiança nas decisões do modelo.

## 4 Experimentos e Resultados

Neste capítulo, serão apresentados o *Setup* empregado na condução dos experimentos e os resultados correspondentes alcançados em cada um deles.

### 4.1 *Setup* de Experimentos

Para os experimentos descritos a seguir, o modelo passou por um processo de treinamento que abrangeu 15 épocas, com a divisão dos dados em lotes, cada um contendo 8 imagens. Durante esse período, foram observadas reduções na taxa de erro e aumentos na acurácia, indicando um progresso favorável no treinamento do modelo. O otimizador Adam foi utilizado, com um *learning rate* de 0.0001 para ajustar os parâmetros do modelo.

Os experimentos foram conduzidos em um sistema equipado com uma máquina Intel i9-10900, que possui 10 núcleos físicos e 20 *threads*, operando a uma frequência de 2,80GHz, além de uma capacidade de memória RAM DDR4 de 128GB. A máquina também é equipada com uma unidade de processamento gráfico (GPU) RTX 3090 com uma memória dedicada de 24GB GDDR6X.

### 4.2 Resultados Obtidos

#### 4.2.1 Treinamento do Modelo PEN

Os resultados alcançados nas avaliações conduzidas em cada cenário, mediante a utilização do modelo PEN, são apresentados na Tabela 4.1.

Tabela 4.1 – Resultados dos experimentos realizados com o modelo PEN em três cenários distintos.

Modelo	Cenário	Erro (%)	Acurácia (%)
PEN	1	0.0224	99.45
	2	0.0249	99.47
	3	0.0278	99.31
	1-3	$0.0250 \pm 0.0027$	$99.41 \pm 0.0711$

Fonte: (Produzida pelo próprio autor).

Com base nos resultados reportados, pode-se concluir que o modelo teve um melhor desempenho no cenário 2, enquanto encontrou um pouco mais de dificuldade na classificação do cenário 3. Dessa forma, como os três cenários foram submetidos às mesmas condições, indica-se que as amostras presentes no cenário 3 apresentaram características que dificultaram sua correta classificação. Essas características podem incluir oclusão da região periocular, baixa

qualidade da imagem, problemas de iluminação, ou outros fatores que impactaram negativamente no desempenho do modelo.

É importante ressaltar que a acurácia de teste do modelo nos três cenários foram próximas, indicando uma capacidade de generalização. No entanto, no cenário 3, a taxa de erro foi um pouco mais alta, enquanto a acurácia foi um pouco menor no teste em comparação com os outros dois cenários.

Por fim, conforme apresentado na Tabela 4.2, são delineados os resultados alcançados pelo modelo PEN, avaliados por meio de diversas métricas. Tais dados reiteram a eficácia e a confiabilidade do referido modelo em distintos critérios de avaliação. É possível observar que o modelo é equilibrado em relação a precisão e revocação, apresentando um alto F-score.

Tabela 4.2 – Análise dos resultados experimentais conduzidos com o modelo PEN em três configurações específicas, empregando as métricas de avaliação: Precisão, Revocação e F-Score.

Modelo	Cenário	Precisão (%)	Revocação (%)	F-Score (%)
PEN	1	99.39	99.27	99.25
	2	99.41	99.28	99.27
	3	99.44	99.32	99.31
	1-3	$99.41 \pm 0.0205$	$99.29 \pm 0.0216$	$99.27 \pm 0.0249$

Fonte: (Produzida pelo próprio autor).

#### 4.2.2 Treinamento do Modelo MPEN

Na Tabela 4.3, é possível realizar uma análise dos resultados obtidos pelo modelo MPEN.

Tabela 4.3 – Resultados dos experimentos realizados com o modelo MPEN Multitarefa em três cenários distintos.

Modelo	Cenário	Erro (%)	Acurácia (%)
MPEN	1	0.0981	98.86
	2	0.0989	98.97
	3	0.0738	99.10
	1-3	$0.0903 \pm 0.0143$	$98.97 \pm 0.1201$

Fonte: (Produzida pelo próprio autor).

Nesse contexto, mediante a análise dos resultados obtidos pelo modelo multitarefa, observa-se que o desempenho do mesmo é semelhante ao modelo PEN (sem multitarefa). O modelo em questão apresentou acurácia significativa em todos os diferentes cenários em que foi submetido, se saindo melhor no terceiro cenário de teste e tendo um pouco mais de dificuldade no primeiro cenário.

Finalmente, os resultados obtidos pelo modelo MPEN são apresentados na Tabela 4.4, sendo avaliados por meio de diversas métricas. Os dados apresentados corroboram a confiabilidade

e eficácia do modelo MPEN. Observa-se que o referido modelo demonstra equilíbrio entre precisão e revocação, manifestando um elevado valor de F-score.

Tabela 4.4 – Análise dos resultados experimentais conduzidos com o modelo MPEN em três configurações específicas, empregando as métricas de avaliação: Precisão, Revocação e F-Score.

Modelo	Cenário	Precisão (%)	Revocação (%)	F-Score (%)
MPEN	1	99.14	98.92	98.90
	2	99.12	98.91	98.88
	3	99.24	99.07	99.05
	1-3	99.16 ± 0.0643	98.96 ± 0.0896	98.94 ± 0.0929

Fonte: (Produzida pelo próprio autor).

Comparando os resultados obtidos com o modelo PEN e o MPEN através das diversas técnicas de avaliação utilizadas, é possível observar que a adição da técnica de multitarefa introduziu uma complexidade adicional ao modelo. Nesse caso, esta complexidade suplementar não resultou em aprimoramentos em seu desempenho, destacando uma leve diminuição na eficácia do modelo em comparação com a arquitetura convencional.

### 4.2.3 Comparação com a Literatura

Diante dos resultados obtidos pelos modelos propostos neste estudo (PEN e MPEN), é pertinente realizar uma comparação com os resultados alcançados no trabalho de (Zanlorensi *et al.*, 2022), que propôs a base de dados utilizada neste trabalho.

Na Tabela 4.5, são apresentados os resultados alcançados neste trabalho, utilizando as arquiteturas PEN (adaptação da arquitetura EfficientNet-B0 para o contexto de reconhecimento periocular) e MPEN. A comparação inclui, igualmente, o desempenho mais destacado obtido por Zanlorensi *et al.* (2022) no contexto da tarefa de identificação, seguindo o protocolo de mundo fechado, utilizando uma arquitetura Multitarefa (melhor resultado apresentado pelos autores). Adicionalmente, são contemplados os resultados mais promissores alcançados por Alahmadi, Hussain e Aboalsamh (2022), atual estado-da-arte, que faz uso da arquitetura DenseNet201 como modelo base.

Ao comparar os resultados apresentados na Tabela 4.5 com os obtidos pelo PEN e MPEN, é observada uma vantagem significativa deste trabalho em relação ao de Zanlorensi *et al.* (2022), evidenciando uma acurácia média superior de 15.09%. Além disso, quando comparado com a abordagem estado-da-arte (Alahmadi; Hussain; Aboalsamh, 2022), percebe-se uma vantagem considerável, evidenciada por uma média de acurácia superior em 1.73%. Este cenário evidencia a robustez do PEN/MPEN em lidar com as diversas variações presentes na base de dados utilizada.

Tabela 4.5 – Apresentação dos resultados mais destacados no âmbito deste estudo, abrangendo as performances do modelo PEN, do MPEN, conforme os achados de [Alahmadi, Hussain e Aboalsamh \(2022\)](#), e os resultados delineados por [Zanlorensi \*et al.\* \(2022\)](#). Quadro comparativo dos melhores resultados obtidos na base de dados UFPR-Periocular.

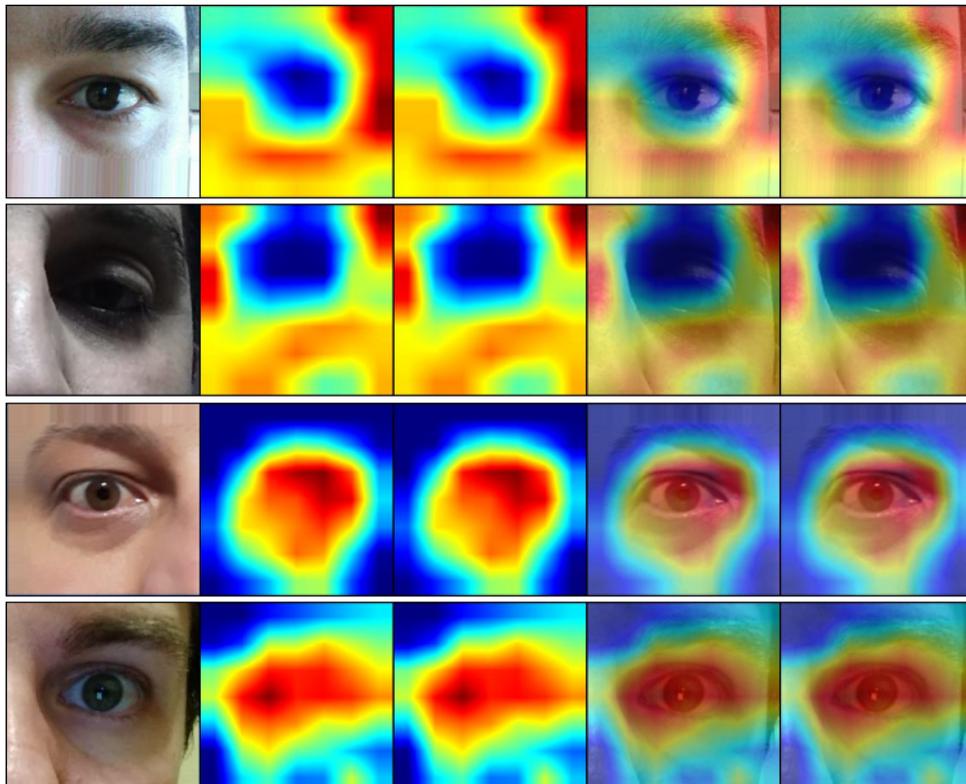
Trabalho	Acurácia (%)
PEN (Trabalho Proposto)	99.41 ± 0.07
MPEN (Trabalho Proposto)	98.97 ± 0.12
<a href="#">Alahmadi, Hussain e Aboalsamh (2022)</a>	97.68 ± 0.31
<a href="#">Zanlorensi <i>et al.</i> (2022)</a>	84.32 ± 0.71

Fonte: (Produzida pelo próprio autor).

### 4.3 Análise da Interpretabilidade do Modelo PEN

Na Figura 4.1, são apresentados os resultados da aplicação de técnicas de interpretabilidade baseadas em Grad-CAM para gerar mapas de ativação que destacam áreas decisivas para tomada de decisão do modelo. A adoção desta abordagem específica no mencionado modelo decorreu da observação de seu desempenho superior. Por meio desta análise, busca-se compreender as características intrínsecas consideradas pelo modelo em suas decisões.

Figura 4.1 – Imagens geradas por meio de técnicas de interpretabilidade baseadas em Grad-CAM, onde a primeira coluna exibe a imagem inalterada e as colunas subsequentes exibem a imagem interpretativa. Nestas últimas colunas, tem-se uma redução na intensidade cromática, visando facilitar a visualização das características faciais.



Fonte: (Produzida pelo próprio autor).

Na primeira linha da Figura 4.1, é possível discernir a presença de uma imagem ocular com uma discrepância na intensidade luminosa da região ocular. Tal variação sugere que o algoritmo, em sua tomada de decisão, não atribuiu relevância a essa região específica, direcionando sua atenção para a região ao redor do olho. Analogamente, na segunda linha, devido à significativa variação de luminosidade, o algoritmo também desconsiderou a região ocular em sua análise.

Na terceira e quarta linhas da Figura 4.1, é evidente que o algoritmo atribuiu maior relevância à região ocular, destacando a importância dos olhos em seu processo decisório. Nesse contexto, a nitidez da região ocular assume um papel crucial, pois a presença de uma resolução adequada nas imagens em questão facilita a extração e análise de características pertinentes pelo algoritmo.

Por fim, na Figura 4.2, são apresentados exemplos representativos das imagens utilizadas no experimento destinado a avaliar a influência da íris nas decisões do modelo. Semelhantemente à metodologia anteriormente empregada, a presente abordagem foi selecionada para este modelo em virtude de seu desempenho superior. Sendo assim, para realizar essa avaliação, foi adicionado um bloco de cor branca sobre a íris em todas as imagens de teste. Essa inclusão foi facilitada pelo fato de que a região central de todas as imagens da base de dados coincide, assim, foi suficiente obter a posição da íris em uma imagem (coincidente com as demais) e zerar os canais de cores dessa região, tornando-a branca. Posteriormente, o modelo foi submetido a uma bateria de testes.

Adiante, na Tabela 4.6, são apresentados os resultados alcançados pelo modelo nos diferentes cenários, considerando as imagens nas quais a íris se encontra oculta.

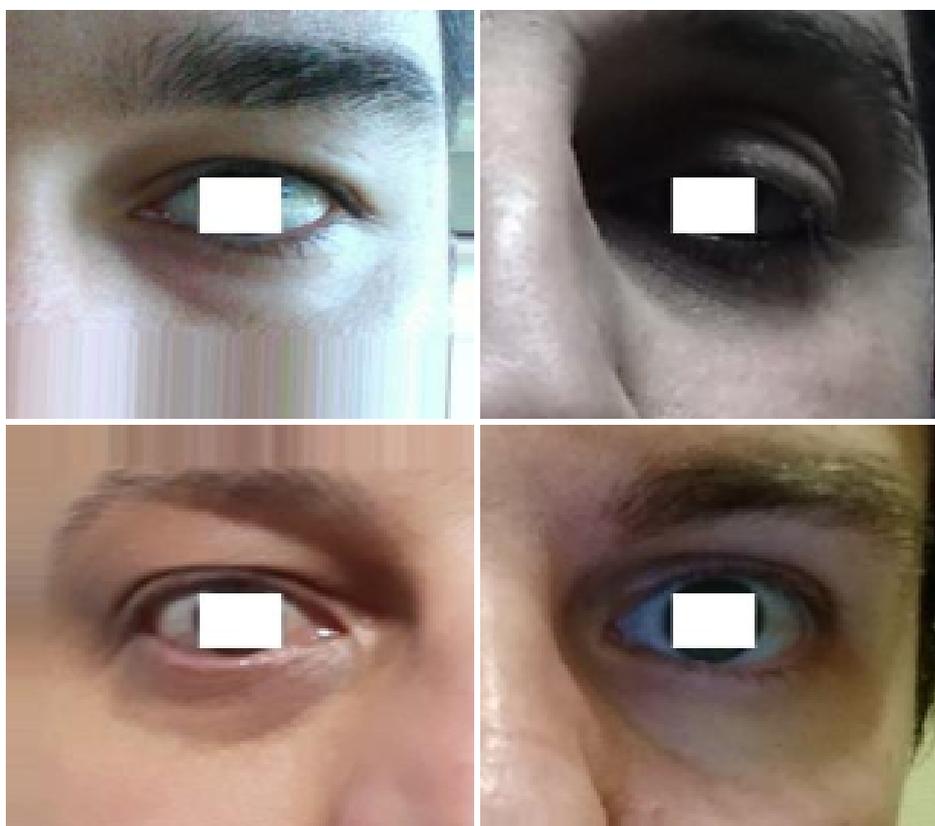
Tabela 4.6 – Resultados obtidos a partir dos experimentos conduzidos com o modelo PEN nos três cenários, considerando a ocultação da íris em todas as imagens analisadas.

Modelo	Cenário	Erro (%)	Acurácia (%)
PEN	1	0.1124	97.47
	2	0.0975	97.83
	3	0.1188	97.34
	1-3	$0.1095 \pm 0.0109$	$97.54 \pm 0.2538$

Fonte: (Produzida pelo próprio autor).

Com base nos dados apresentados na Tabela 4.6, observa-se que o modelo tende a fazer uso limitado da íris em seu processo decisório. Isso é evidenciado pelo desempenho significativo do modelo mesmo na ausência visual da íris, resultando em métricas comparáveis às aquelas registradas nas imagens sem qualquer oclusão.

Figura 4.2 – Imagens nas quais a íris foi ocultada por um bloco branco para avaliar a influência da íris nas decisões do modelo.



Fonte: (Produzida pelo próprio autor).

## 5 Considerações Finais

No contexto deste estudo, é apresentada uma abordagem que utiliza redes neurais para o reconhecimento periocular em um cenário de galeria fechada. O modelo EfficientNet-B0 é adotado como a base inicial dos modelos propostos, e em seguida são adicionadas camadas personalizadas para aprimorar o desempenho dos modelos denominados como Periocular EfficientNet-B0 (PEN) e Multitarefa Periocular EfficientNet-B0 (MPEN).

Sendo assim, o uso de *Convolutional Neural Network* (CNN) mostrou-se capaz de construir um sistema de reconhecimento periocular obtendo bons resultados, uma vez que os experimentos conduzidos com o modelo PEN revelaram acurácias 1.73% acima do estado-da-arte para a base de dados UFPR-Periocular. Essa base de dados apresenta desafios adicionais devido às características específicas das imagens, tais como o uso de óculos, presença de sombras, oclusão parcial do rosto, entre outros fatores. Notavelmente, verificou-se que o modelo é robusto, pelo fato de que manteve uma acurácia alta nos três cenários distintos de teste, ou seja, independente das variações contidas nas imagens.

Ademais, constatou-se que o aprendizado multitarefa não evidenciou superioridade em relação à abordagem de aprendizado de tarefa única para o problema em análise. Esta constatação emergiu dada a comparação dos resultados alcançados pelos modelos PEN e MPEN, visto que ambos exibiram resultados equiparáveis com os resultados de MPEN ligeiramente inferiores.

A arquitetura do modelo PEN demonstrou um desempenho superior em comparação com várias outras arquiteturas avaliadas no estudo conduzido por [Zanlorensi et al. \(2022\)](#), inclusive superando aquela que obteve o melhor resultado, a *MobileNet-V2*.

Considerando os resultados alcançados neste estudo, uma direção promissora para investigações futuras envolve a implementação e otimização do modelo *Visual Transformer* (ViT) para a análise dos conjuntos de dados. O ViT pode proporcionar uma melhoria substancial na representação e interpretação dos dados, abrindo portas para descobertas mais profundas e precisas. Ademais, a integração do método de *conformal prediction* surge como uma estratégia complementar valiosa, visando quantificar e avaliar a incerteza associada às previsões geradas pelo modelo. Ao incorporar essa técnica, espera-se não apenas melhorar a confiabilidade das estimativas, mas também fornecer uma estrutura robusta para a interpretação e validação externa dos resultados obtidos. Por fim, proceder à implementação da técnica *Out-of-Distribution* (OOD), cujo propósito reside na identificação e manejo de amostras que se encontram fora da distribuição das utilizadas durante o processo de treinamento do modelo. Portanto, a combinação estratégica dessas abordagens inovadoras promete potencializar a qualidade e a aplicabilidade das análises desenvolvidas neste estudo, abrindo novas perspectivas para avanços significativos no campo em questão.

# Referências

- ALAHMADI, A.; HUSSAIN, M.; ABOALSAMH, H. Lda-cnn: Linear discriminant analysis convolution neural network for periocular recognition in the wild. **Mathematics**, v. 10, n. 23, 2022. ISSN 2227-7390. Acessado em 21/07/2023. Disponível em: <<https://www.mdpi.com/2227-7390/10/23/4604>>.
- ALECRIM, E. **Redes Neurais Artificiais**. INFOWESTER, 2004. Acessado em 19/05/2023. Disponível em: <<https://www.infowester.com/redesneurais.php>>.
- ALEKSANDER, I.; MORTON, H. **An introduction to neural computing**. [S.l.]: Chapman & Hall London, 1990. v. 3. Acessado em 09/05/2023.
- AMAZON. **O que é uma rede neural?** 2023. Acessado em 09/05/2023. Disponível em: <<https://aws.amazon.com/pt/what-is/neural-network/>>.
- BARCELLOS, W. **Reconhecimento biométrico usando a região periocular e o olho por meio de extração de características texturais de camadas profundas de uma CNN**. Tese (Doutorado) — Universidade de São Paulo, 2022. Acessado em 11/05/2023. Disponível em: <<https://doi.org/10.11606/D.18.2022.tde-16082022-103558>>.
- BARCELLOS, W.; GONZAGA, A. **Periocular Authentication in smartphones applying ULBP descriptor on CNN Feature Maps**. SBC, 2021. Acessado em 13/06/2023. Disponível em: <<https://sol.sbc.org.br/index.php/wvc/article/view/18890>>.
- BARCELLOS, W. *et al.* **Evaluation of fine tuning and feature extraction methods in biometric periocular recognition**. SBC, 2019. Acessado em 13/06/2023. Disponível em: <<https://sol.sbc.org.br/index.php/wvc/article/view/7626>>.
- BARROS, L. **Biometria Facial: o Que É Essa Tecnologia Que Revolucionou a Marcação do Ponto**. tangerino blog, 2023. Acessado em 11/05/2023. Disponível em: <<https://tangerino.com.br/blog/biometria-facial/>>.
- BOWYER, P. J. F. K. W. **The ND-IRIS-0405 Iris Image Dataset**. 2016. Acessado 6 Abril, 2023. Disponível em: <<https://arxiv.org/abs/1606.04853>>.
- BRANCO, H. **Overfitting e underfitting em machine learning**. ABRACD - Associação Brasileira de Ciência de Dados, 2020. Acessado em 09/05/2023. Disponível em: <<https://abracd.org/overfitting-e-underfitting-em-machine-learning/#>>.
- BRAUNER, D. **A semântica de dados é um futuro possível para o aprendizado de máquina?** medium, 2020. Acessado em 26/12/2023. Disponível em: <<https://medium.com/@danibrauner/a-sem%C3%A2ntica-de-dados-%C3%A9-um-futuro-poss%C3%ADvel-para-o-aprendizado-de-m%C3%A1quina-9344661b5db7>>.
- BRUNIALTI, L. *et al.* **Aprendizado de máquina em sistemas de recomendação baseados em conteúdo textual: uma revisão sistemática**. **Anais do XI Simpósio Brasileiro de Sistemas de Informação**, SBC, p. 203–210, 2015. Acessado em 10/05/2023. Disponível em: <<https://sol.sbc.org.br/index.php/sbsi/article/view/5818/5716>>.

CAMILO, C. O.; SILVA, J. C. d. **Mineração de dados: Conceitos, tarefas, métodos e ferramentas.** **Universidade Federal de Goiás (UFG)**, v. 1, n. 1, p. 1–29, 2009. Acessado em 10/05/2023. Disponível em: <<https://ww2.inf.ufg.br/node/355>>.

CARNEIRO, A. L. C. **Técnicas de processamento digital de imagens no Python com matemática.** Data Hackers, 2020. Acessado em 13/06/2023. Disponível em: <<https://medium.com/data-hackers/técnicas-de-processamento-digital-de-imagens-no-python-com-matemática-154eae204447>>.

CARVALHO, A. P. d. L. F. d. 2019. Acessado em 16/05/2023. Disponível em: <<https://sites.icmc.usp.br/andre/research/neural/>>.

CECCON, D. **Funções de ativação: definição, características, e quando usar cada uma.** Expert Academy, 2020. Acessado em 16/05/2023. Disponível em: <<https://iaexpert.academy/2020/05/25/funcoes-de-ativacao-definicao-caracteristicas-e-quando-usar-cada-uma/>>.

CECCON, D. **A importância do bias nas Redes Neurais.** IA Expert Academy, 2020. Acessado em 19/05/2023. Disponível em: <[https://iaexpert.academy/2020/09/28/importancia-do-bias-nas-redes-neurais/?doing\\_wp\\_cron=1683401843.6309609413146972656250](https://iaexpert.academy/2020/09/28/importancia-do-bias-nas-redes-neurais/?doing_wp_cron=1683401843.6309609413146972656250)>.

COSTA, E. S. **Meta aprendizado para detecção de anomalias em imagens.** Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2022. Acessado em 13/12/2023.

COSTA, M. C. S. F.; WALTER, M. E. M. T.; STADLER, P. F. **Aprendizado Por Transferência Para predição de RNAs não-codificadores.** Universidade de Brasília, 2022. Acessado em 21/07/2023. Disponível em: <<https://wpos.unb.br/images/Artigos22/st4/6.pdf>>.

COSTA, R. M. d. **Uma nova abordagem para reconhecimento biométrico baseado em características dinâmicas da íris humana.** Tese (Doutorado) — Universidade de São Paulo, 2009. Acessado em 11/05/2023. Disponível em: <<https://www.teses.usp.br/teses/disponiveis/18/18152/tde-25032010-092709/en.php>>.

DAKSHINA, Y. **GRADCAM (Gradient-weighted Class Activation Mapping) funcionando, aplicativos e sua implementação em Python.** DL, 2023. Acessado em 26/12/2023. Disponível em: <[https://deeplearningofpython.blogspot.com/2023/05/Gradcam-working-example-python.html?source=post\\_page-----28c161463231-----#point3](https://deeplearningofpython.blogspot.com/2023/05/Gradcam-working-example-python.html?source=post_page-----28c161463231-----#point3)>.

DIAS, G. M.; JR., A. d. S. B. **Verificação de Região Periocular Utilizando Aprendizagem Profunda - UEPG.** 2019. Acessado em 13/06/2023. Disponível em: <[https://siseve.apps.uepg.br/storage/EAIC2019/8\\_Glaucia\\_Maria\\_Belo\\_Dias-156868331454341.pdf](https://siseve.apps.uepg.br/storage/EAIC2019/8_Glaucia_Maria_Belo_Dias-156868331454341.pdf)>.

DICIO. **Inteligência Artificial.** Dicio, Dicionário Online de Português, 2023. Acessado em 15/05/2023. Disponível em: <<https://www.dicio.com.br/inteligencia-artificial/>>.

DUARTE, R. *et al.* **Aplicação dos Sistemas biométricos de reconhecimento facial na segurança pública.** Brazilian Journal of Forensic Sciences Medical Law and Bioethics, 2021. Acessado em 11/05/2023. Disponível em: <<https://www.bjfs.org/bjfs/index.php/bjfs/article/view/848>>.

ERHAN, Y. B. D.; COURVILLE, A. **Visualizing Higher-Layer Features of a Deep Network.** 2009. Acessado em 24/10/2023. Disponível em: <[https://www.researchgate.net/publication/265022827\\_Visualizing\\_Higher-Layer\\_Features\\_of\\_a\\_Deep\\_Network](https://www.researchgate.net/publication/265022827_Visualizing_Higher-Layer_Features_of_a_Deep_Network)>.

FIA. **Inteligencia artificial: O que É, Como Funciona e exemplos**. 2021. Acessado em 15/05/2023. Disponível em: <<https://fia.com.br/blog/inteligencia-artificial/>>.

FRANKIEWICZ, M. **Os desafios da implementação da autenticação biométrica: um guia abrangente**. TS2, 2023. Acessado em 11/05/2023. Disponível em: <<https://ts2.space/pt/os-desafios-da-implementacao-da-autenticacao-biometrica-um-guia-abrangente/>>.

FREEP!K. **Conceito de colagem de reconhecimento facial**. FREEP!K, 2024. Acessado em 09/02/2024. Disponível em: <[https://br.freepik.com/fotos-gratis/conceito-de-colagem-de-reconhecimento-facial\\_36295547.htm](https://br.freepik.com/fotos-gratis/conceito-de-colagem-de-reconhecimento-facial_36295547.htm)>.

GAEA. **Afinal, O que É deep learning?** 2021. Acessado em 09/05/2023. Disponível em: <<https://gaea.com.br/afinal-o-que-e-deep-learning/>>.

GAYHARDT, L. *et al.* **Aprendizado profundo x Aprendizado de máquina em Azure Machine Learning versus Machine Learning**. Microsoft, 2022. Acessado em 21/07/2023. Disponível em: <<https://learn.microsoft.com/pt-br/azure/machine-learning/concept-deep-learning-vs-machine-learning>>.

JOSE, I. **Redes Neurais roots - parte 01**. IMASTERS - we are developers, 2019. Acessado em 16/05/2023. Disponível em: <<https://imasters.com.br/desenvolvimento/redes-neurais-roots-parte-01>>.

KAUR, G.; SINGH, G.; KUMAR, V. A review on biometric recognition. **International Journal of Bio-Science and Bio-Technology**, v. 6, n. 4, p. 69–76, 2014. Acessado em 21/05/2023. Disponível em: <[https://gvpress.com/journals/IJBSBT/vol6\\_no4/7.pdf](https://gvpress.com/journals/IJBSBT/vol6_no4/7.pdf)>.

LAVI. **LAVI Iris Database - DB2**. 2019. Acessado 6 Abril, 2023. Disponível em: <<http://imagem.sel.eesc.usp.br/base/iris/index.html>>.

LEITE, R. **Introdução a Validação-Cruzada: K-Fold**. Data Hackers, 2020. Acessado em 13/06/2023. Disponível em: <<https://drigols.medium.com/introdução-a-validação-cruzada-k-fold-2a6bcd32a90>>.

LUZ, E. *et al.* Deep periocular representation aiming video surveillance. **Pattern Recognition Letters**, v. 114, p. 2–12, 2018. ISSN 0167-8655. Acessado em 16/05/2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865517304476>>.

MAGALHAES, P. S. T.; SANTOS, H. D. d. Biometria e autenticação. Associação Portuguesa de Sistemas de Informação (APSI), 2003. Acessado em 26/12/2023.

MAHENDRAN, A.; VEDALDI, A. **Visualizing Deep Convolutional Neural Networks Using Natural Pre-Images**. Arxiv, 2015. Acessado em 24/10/2023. Disponível em: <<https://arxiv.org/abs/1512.02017>>.

MARIANO, D. **Métricas de avaliação em machine learning**. diegomariano, 2021. Acessado em 28/06/2023. Disponível em: <<https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>>.

MCCULLOCH, W. S.; PITTS, W. Um cálculo lógico das ideias imanentes na atividade nervosa. Springer Link, Dec 1943. Acessado em 16/05/2023. Disponível em: <<https://link.springer.com/article/10.1007/bf02478259>>.

MEAZZINI, L. **Tutorial - Transfer Learning**. Ensina.AI, 2019. Acessado em 21/07/2023. Disponível em: <<https://medium.com/ensina-ai/tutorial-transfer-learning-3972cac5e9b5>>.

MEDICINA. **Entenda o que é interpretabilidade de exames e qual a importância**. 2023. Acessado em 18/10/2023. Disponível em: <<https://www.neuralmed.ai/blog/interpretabilidade-de-exames>>.

MESAMEKI *et al.* **Interpretabilidade do modelo**. 2023. Acessado em 18/10/2023. Disponível em: <<https://learn.microsoft.com/pt-br/azure/machine-learning/how-to-machine-learning-interpretability?view=azureml-api-2>>.

NEOWAY. **inteligência artificial: o que é, como funciona e exemplos**. Neoway, 2020. Acessado em 09/05/2023. Disponível em: <<https://blog.neoway.com.br/inteligencia-artificial/>>.

NISHIMOTO, B. E. **Principais conceitos por trás do Machine Learning**. medium, 2018. Acessado em 28/06/2023. Disponível em: <<https://medium.com/neuronio-br/principais-conceitos-por-tras-do-machine-learning-a4b942d5d309>>.

ORACLE. **O que é Deep Learning?** OCI, 2023. Acessado em 22/12/2023. Disponível em: <<https://www.oracle.com/br/artificial-intelligence/machine-learning/what-is-deep-learning/>>.

ORVALHO, V. Reconhecimento facial. **Revista de Ciência Elementar**, Casa das Ciências, v. 7, n. 4, 2019. Acessado em 11/05/2023. Disponível em: <<https://rce.casadasciencias.org/rceapp/art/2019/073/>>.

PAOLI, A.; DAHIA, G.; AMPARO, L. Uso de impressao digital para biometria. sep 2016. Acessado em 11/05/2023. Disponível em: <[https://d1wqtxts1xzle7.cloudfront.net/50060314/uso-de-impressao-libre.pdf?1478108267=&response-content-disposition=inline%3B+filename%3DUso\\_de\\_impressao\\_digital\\_para\\_biometria.pdf&Expires=1691693384&Signature=F6UvKB4tzPy4eIUlcEwhNR0kjfTJkThJNE16J7N5KVkGqJiGC~yrR~Sznw1id6WuDVECXueHUim~oIOYmSG-IURyC674ct8RMRxlRgleMpQZF6H7XMJ8DZo2AuN3Y6tIE06Sk2DVKPyYa1O9aB0qYrtFNaMSKU0rhjLK7YLPB~HySzb8mCCbeCO6YkhSBvjKYaykcjRjLgPiRENUdtOmodfktUlob1vBtRbmq0oiqeNDi5oyYkrbBIsr9QG4K9QAANzi8E1rL4qY5HQLewlRdWqJ7gskCiJKy-pZjSAwP~z~DuN6fj48n-Y2APDnrCICWCg34tR81ULjl6e48BCZwA\\_\\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](https://d1wqtxts1xzle7.cloudfront.net/50060314/uso-de-impressao-libre.pdf?1478108267=&response-content-disposition=inline%3B+filename%3DUso_de_impressao_digital_para_biometria.pdf&Expires=1691693384&Signature=F6UvKB4tzPy4eIUlcEwhNR0kjfTJkThJNE16J7N5KVkGqJiGC~yrR~Sznw1id6WuDVECXueHUim~oIOYmSG-IURyC674ct8RMRxlRgleMpQZF6H7XMJ8DZo2AuN3Y6tIE06Sk2DVKPyYa1O9aB0qYrtFNaMSKU0rhjLK7YLPB~HySzb8mCCbeCO6YkhSBvjKYaykcjRjLgPiRENUdtOmodfktUlob1vBtRbmq0oiqeNDi5oyYkrbBIsr9QG4K9QAANzi8E1rL4qY5HQLewlRdWqJ7gskCiJKy-pZjSAwP~z~DuN6fj48n-Y2APDnrCICWCg34tR81ULjl6e48BCZwA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)>.

PASSOS, B. T. **O mundo do ponto de vista das Redes Neurais Convolucionais (RNCs)**. blog, 2021. Acessado em 16/05/2023. Disponível em: <<https://ateliware.com/blog/redes-neurais-convolucionais>>.

PROENCA, H. *et al.* The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance. **IEEE Trans. PAMI**, v. 32, n. 8, p. 1529–1535, August 2010. Acessado em 13/06/2023. Disponível em: <<http://iris.di.ubi.pt/ubiris2.html>>.

PROENCA, H. *et al.* Ubipr. April 2012. Acessado em 13/06/2023. Disponível em: <<http://iris.di.ubi.pt/ubipr.html>>.

RATTANI, A. *et al.* Icip 2016 competition on mobile ocular biometric recognition. p. 320–324, 2016. Acessado em 13/06/2023. Disponível em: <<https://ieeexplore.ieee.org/document/7532371>>.

RAWAT, W.; WANG, Z. **Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review**. PubMed, 2017. Acessado em 25/05/2023. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/28599112/>>.

ROSENBLATT, F. **The perceptron: A probabilistic model for information storage and organization in the brain.** *Psychological Review*. 1958. Acessado em 09/05/2023. Disponível em: <<https://doi.org/10.1037/h0042519>>.

RUSSAKOVSKY, O. *et al.* Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, v. 115, 09 2014. Acessado em 13/06/2023. Disponível em: <[https://www.researchgate.net/publication/265295439\\_ImageNet\\_Large\\_Scale\\_Visual\\_Recognition\\_Challenge](https://www.researchgate.net/publication/265295439_ImageNet_Large_Scale_Visual_Recognition_Challenge)>.

SALES, S. A.; SALES, N. M. R.; SILVA, M. A. L. R. Biometria: Reconhecimento de voz. **Revista de Sistemas e Computação-RSC**, v. 11, n. 1, 2021. Acessado em 11/05/2023. Disponível em: <<https://revistas.unifacs.br/index.php/rsc/article/view/7031>>.

SEIDOR, V. **Métricas de performance e funções de perda para Machine Learning.** Viceri Seidor, 2020. Acessado em 28/06/2023. Disponível em: <<https://viceri.com.br/insights/metricas-de-performance-e-funcoes-de-perda-para-machine-learning/>>.

SHAIKH, R. **Cross Validation Explained: Evaluating estimator performance.** 2018. Acessado em 13/06/2023. Disponível em: <<https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85>>.

Shamila Ebenezer, A. *et al.* Effect of image transformation on efficientnet model for covid-19 ct image classification. **Materials Today: Proceedings**, v. 51, p. 2512–2519, 2022. ISSN 2214-7853. International Conference on Advances in Materials Science. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2214785321078469>>.

SILVA, A.; SIEBERT, C. Densenet-dc: Optimizing densenet parameters through feature map generation control. **Revista de Informática Teórica e Aplicada**, v. 27, p. 25–39, 06 2020. Acessado em 28/06/2023. Disponível em: <[https://seer.ufrgs.br/rita/article/view/Vol27\\_nr3\\_25](https://seer.ufrgs.br/rita/article/view/Vol27_nr3_25)>.

SILVA, M. A. d. **Inteligência Artificial.** UOL, 2008. Acessado em 09/05/2023. Disponível em: <<https://brasilecola.uol.com.br/informatica/inteligencia-artificial.htm>>.

SILVA, P. H. L. **Seleção de features em Representações Profundas para a íris e a região periocular Como modalidades biométricas.** RIUFOP, 2018. Acessado em 15/05/2023. Disponível em: <<http://www.repositorio.ufop.br/jspui/handle/123456789/10013>>.

SOUZA, V. S. S. **Introdução à interpretabilidade de redes neurais convolucionais.** UFOP, 2023. Acessado em 24/10/2023. Disponível em: <<https://www.monografias.ufop.br/handle/35400000/5358>>.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International conference on machine learning.** 2019. p. 6105–6114. Acessado em 13/06/2023. Disponível em: <<http://proceedings.mlr.press/v97/tan19a/tan19a.pdf>>.

TEAM, D. S. **Validação cruzada K-Fold.** Data Science, 2020. Acessado em 13/06/2023. Disponível em: <<https://datascience.eu/pt/aprendizado-de-maquina/validacao-cruzada-k-fold/>>.

TODA, Y.; OKURA, F. **How Convolutional Neural Networks Diagnose Plant Disease.** SPI, 2019. Acessado em 24/10/2023. Disponível em: <<https://spj.science.org/doi/10.34133/2019/9237136>>.

TRADE, C. **Understand the Grad-CAM: A Complete Guide With Example**. CodeTrade.io, 2023. Acessado em 26/12/2023. Disponível em: <<https://www.codetrade.io/blog/grad-cam-a-complete-guide-with-example/>>.

VARELLA, M. **Neurônio**. Portal Drauzio Varella, 2014. Acessado em 16/05/2023. Disponível em: <<https://drauziovarella.uol.com.br/corpo-humano/neuronio/>>.

VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. **Proceedings of the xxix conference on graphics, patterns and images**. [S.l.], 2016. v. 1, n. 4. Acessado em 19/05/2023.

VIZONI, M. V.; MARANA, A. N. **Person Authentication Based on the Difference of Deep Features Extracted from the Ocular and FaceRegions**. SBC, 2019. Acessado em 21/05/2023. Disponível em: <<https://sol.sbc.org.br/index.php/wvc/article/view/7628/7505>>.

Zanlorensi, L. A. *et al.* A new periocular dataset collected by mobile devices in unconstrained scenarios. **Scientific Reports**, v. 12, p. 17989, 2022. ISSN 2045-2322. Acessado em 21/05/2023. Disponível em: <<https://web.inf.ufpr.br/vri/wp-content/uploads/sites/7/2022/10/zanlorensi2022new-published.pdf>>.

ZHANG, Y.; YANG, Q. An overview of multi-task learning. **National Science Review**, v. 5, n. 1, p. 30–43, 09 2017. ISSN 2095-5138. Acessado em 23/08/2023. Disponível em: <<https://doi.org/10.1093/nsr/nwx105>>.