Villeneve de Oliveira Soares

Análise Comparativa de Redes Neurais Convolucionais para Classificação de Retinopatia Diabética: Configurações Binárias e Multiclasse

João Pessoa - PB

Fevereiro de 2025

Villeneve de Oliveira Soares

Análise Comparativa de Redes Neurais Convolucionais para Classificação de Retinopatia Diabética: Configurações Binárias e Multiclasse

Versão apresentada à banca examinadora para o Exame de Qualificação de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal da Paraíba, como requisito parcial necessário à obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento Digital de Sinais

Carlos Danilo Miranda Regis, Prof. Dr.

Orientador

Leonardo V. Batista, Prof. Dr.

Co-orientador

João Pessoa - PB, Fevereiro de 2025

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Nilo Peçanha do IFPB, *campus* João Pessoa

S676a Soares, Villeneve de Oliveira.

Análise comparativa de redes neurais convolucionais para classificação de retinopatia diabética : configurações binárias e multiclasse / Villeneve de Oliveira Soares. – 2025.
86 f. : il.
Dissertação (Mestrado – Engenharia Elétrica) – Instituto Federal de Educação da Paraíba / Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), 2025.
Orientação : Prof. Dr. Carlos Danilo Miranda Regis.

1.Redes neurais artificiais. 2. Retinopatia diabética. 3. Aprendizado de máquina. 4. Diagnóstico assistido por computador. 5. Arquitetura de redes neurais convolucionais. I. Título.

CDU 004.032.26:617.7(043)

Bibliotecária responsável: Lucrecia Camilo de Lima – CRB 15/132

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA

PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU*

MESTRADO EM ENGENHARIA ELÉTRICA

VILLENEVE DE OLIVEIRA SOARES

ANÁLISE COMPARATIVA DE REDES NEURAIS CONVOLUCIONAIS PARA CLASSIFICAÇÃO DE RETINOPATIA DIABÉTICA: CONFIGURAÇÕES BINÁRIA E MULTICLASSE

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia Elétrica pelo Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB - Campus João Pessoa.

Aprovado em 13 de fevereiro de 2025.

Membros da Banca Examinadora:

Dr. Carlos Danilo Miranda Regis

IFPB – PPGEE

Dr. Leonardo Vidal Batista

UFPB

Dr. Patric Lacouth da Silva

IFPB - PPGEE

Dr. Yuri de Almeida Malheiros Barbosa

UFPB

João Pessoa/2025

Agradecimentos

Inicialmente, gostaria de agradecer ao meu orientador **Dr. Carlos Danilo**, por sua incomensurável paciência durante toda esta fase. Seria, para mim, impossível passar por tudo e não conseguir observar o quanto sua presença foi capaz de realizar a diferença, tanto profissional como emocional. Neste segundo ponto, posso dizer que sua fé sempre foi impressionante para mim, fé esta que, sem sombras de dúvidas, o caracterizam como um verdadeiro homem de Deus. Aqui deixo meu muito obrigado.

Como alguém que passei a conhecer após escolher qual tema de dissertação iria seguir, foi um prazer imenso conhecer **Dr. Leonardo Vidal**, meu coorientador. Como sempre me perguntaram e como eu sempre respondi: um homem doce, alegre, sorridente e portador de uma quantidade de conhecimento que foi de grande ajuda em diversos momentos desta jornada. Ao senhor, também, deixo meus agradecimentos.

Por agora, gostaria de agradecer aos amigos que fiz durante esse tempo e, para eles, tomarei um parágrafo: Agradeço a minha ilustre amiga **Dr. Brenda Nogueira**, por todas as palavras, sorrisos, lanches e almoços (os quais ganhei gratuitamente). Com toda certeza lembrarei de todos com muito carinho e sentirei falta de nossos momentos de caminhada pelos corredores; Agradeço ao meu amigo **Pedro Antonio** pelas gameplays, discussões e toda sorte de alegria e felicidade a mim causada. Destaco, assim como a Danilo, um homem de fé, caráter e moral impecáveis. Espero que nossa amizade continue assim até o fim de nossos dias; Destacados estes dois, gostaria de agradecer ao **GPDS** como um todo. Este grupo, por mais que eu não tenha admitido, me mostrou pontos de vista diferentes com os quais aprendi e refleti por muitas de minhas noites. Deixo a vocês, meu obrigado.

Em especial, e separado dos outros (não haveria de ser diferente), agradeço ao meu melhor amigo **Emanuel Tiago**. Até hoje, eu nunca compreendi como podemos pensar de maneiras tão diferentes, mas tão parecidas ao mesmo tempo. Agradeço pelas piadas, brincadeiras, discussões, tramas e, por inúmeras vezes, ter me feito esquecer de problemas que assolavam meus pensamentos como se nunca estivessem estado lá. Espero que nossa amizade se mantenha desta maneira e que eu possa contar com você assim como você sempre poderá contar comigo. Ao meu consagrado Emanuel, deixo também meu muito obrigado.

Gostaria, também, de dedicar um parágrafo para agradecer a **Ana Carolina**. Assim como todos

os citados, preciso agradecer por sua paciência e compreensão durante este período. Também sou grato por mostrar-me alguns comportamentos que, ao serem mudados, me ajudaram a chegar ao fim dessa jornada. Deixo aqui meus mais sinceros agradecimentos.

Como sempre e nunca me esquecendo, agradeço aos meus **familiares**, minhas tias **Ivanete**, **Inalda**, **Ivete** e minha tia **Irene** que, assim como meu tio e seu marido, **Coriolano**, não puderam assistir esta conquista. Tenho certeza que, se meu tio Cori estivesse aqui, me chamaria de Mestre Ville. Mas acredito, em meu coração, que ainda haveremos de nos encontrar. Também agradeço aos meus irmãos, **Vanlendell** e **Luís Alberto**, por todas as conversas e carinhos prestados.

Em separado, para mais destaque, agradeço a minha mãe **Ivanilda** e meu pai **Ivan** por terem me dado todas e, reitero, todas as oportunidades e suportes possíveis (e até impossíveis) para que eu chegasse onde cheguei. Vocês foram os dois gigantes dos quais me apoiei nos ombros, e não haveria alegria neste momento se vocês não estivessem aqui. Muito obrigado.

Por fim, mas jamais menos importante, agradeço ao meu **Deus**. O mesmo Deus de Isaque, Abraão e Jacó, Rei dos Reis e Senhor dos Senhores, por me permitir chegar até aqui. Por me trazer esperança, força e alegria onde muitos não teriam enxergado. Mas, apesar de tudo, ainda peço que mantenha minha fé tão inabalável quanto os firmamentos do mundo e que sua sabedoria me abençoe até que as estrelas caiam dos céus. Amém!

Lista de Siglas e Abreviaturas

ACC Acurácia, *Accuracy*

AUC Área sob a curva, *Area Under the Curve*

CLAHE *Contrast Limited Adaptive Histogram Equalization*

CNN redes neurais artificiais Convolucionais, *Convolutional Neural Network*

DL Aprendizado Profundo, *Deep Learning*

DR Retinopatia Diabética, *Diabetic retinopathy*

DM *Diabetes Mellitus*

DMT1 *Diabetes Mellitus* tipo 1

DMT2 *Diabetes Mellitus* tipo 2

IA Inteligência Artificial

RGB (Vermelho, Verde, Azul), *(Red, Green, Blue)*

RN redes neurais artificiais, *Neural Network*

MLP Perceptron de multicamadas, *MultiLayer Perceptron*

NLP Processamento de linguagem natural, *Natural Language Processing*

PCA Análise de componentes principais, *Principal Component Analysis*

Lista de Figuras

Lista de Tabelas

Sumário

Resumo

A retinopatia diabética é uma das principais causas de cegueira no mundo, afetando milhões de pessoas anualmente. A detecção precoce e o diagnóstico preciso são fundamentais para evitar a progressão da doença, mas a avaliação manual das imagens de retina é um processo demorado e sujeito a variações entre especialistas. Nesse contexto, a aplicação de redes neurais convolucionais surge como uma alternativa promissora para automatizar essa tarefa, tornando o diagnóstico mais rápido e acessível. Este trabalho apresenta uma análise comparativa de redes neurais convolucionais para a classificação de retinopatia diabética em diferentes níveis de gravidade. A pesquisa foca em quatro arquiteturas populares – Xception, VGG16, MobileNet e InceptionV3 – aplicadas em bases de dados amplamente utilizadas, como APTOS2019 e EyePACS2015. Métodos de pré-processamento, incluindo corte e normalização de imagens, foram empregados para maximizar a qualidade dos dados e facilitar o aprendizado das redes.

A avaliação foi conduzida utilizando métricas como acurácia, sensibilidade e precisão, em cenários de classificação binária e multiclasse, com foco em identificar os parâmetros para otimização das redes. Os resultados destacam o potencial da arquitetura Xception, que apresentou melhor desempenho em cenários de classificação com cinco níveis de gravidade. Apesar disso, as limitações computacionais e o impacto do desequilíbrio de classes evidenciaram desafios significativos para a reprodutibilidade e aplicabilidade clínica. Este estudo reforça a relevância de redes convolucionais no diagnóstico automatizado de retinopatia diabética, enquanto aponta para a necessidade de mais rigor em análises futuras, contribuindo para avanços metodológicos e maior robustez nas aplicações práticas.

Os melhores resultados obtidos foram alcançados utilizando a arquitetura Xception, que mostrou desempenho superior na classificação de retinopatia diabética em cinco classes distintas. Após ajustes nos parâmetros e pré-processamento das imagens, a rede alcançou uma acurácia de 71,70% no conjunto de dados APTOS2019. Em contrapartida, a MobileNet destacou-se por sua eficiência computacional, apresentando resultados competitivos em menor tempo de execução, enquanto a VGG16 e a InceptionV3 não atingiram o mesmo nível de desempenho esperado, sendo a última descartada nos experimentos finais. Esses resultados evidenciam o potencial da Xception como uma solução robusta para a detecção de retinopatia diabética em cenários clínicos.

Abstract

Diabetic retinopathy is one of the leading causes of blindness worldwide, affecting millions of people annually. Early detection and accurate diagnosis are essential to prevent disease progression, but manual evaluation of retinal images is a time-consuming process subject to variations among specialists. In this context, the application of convolutional neural networks emerges as a promising alternative to automate this task, making diagnosis faster and more accessible. This work presents a comparative analysis of convolutional neural networks for the classification of diabetic retinopathy at different levels of severity. The research focuses on four popular architectures – Xception, VGG16, MobileNet and InceptionV3 – applied to widely used databases, such as APTOS2019 and EyePACS2015. Preprocessing methods, including image cropping and normalization, were employed to maximize data quality and facilitate network learning.

The evaluation was conducted using metrics such as accuracy, sensitivity, and precision, in binary and multiclass classification scenarios, with a focus on identifying parameters for network optimization. The results highlight the potential of the Xception architecture, which performed best in classification scenarios with five levels of severity. Despite this, computational limitations and the impact of class imbalance highlighted significant challenges for reproducibility and clinical applicability. This study reinforces the relevance of convolutional networks in the automated diagnosis of diabetic retinopathy, while pointing to the need for more rigor in future analyses, contributing to methodological advances and greater robustness in practical applications.

The best results obtained were achieved using the Xception architecture, which showed superior performance in the classification of diabetic retinopathy into five distinct classes. After parameter adjustments and image preprocessing, the network achieved an accuracy of 71.70% on the APTOS2019 dataset. In contrast, MobileNet stood out for its computational efficiency, presenting competitive results in shorter execution time, while VGG16 and InceptionV3 did not reach the same level of expected performance, with the latter being discarded in the final experiments. These results highlight the potential of Xception as a robust solution for the detection of diabetic retinopathy in clinical scenarios.

1 Introdução

A retinopatia é um termo abrangente que se refere a um grupo de doenças que afetam a retina, a camada sensível à luz na parte posterior do olho, sendo uma das causas mais comuns de perda de visão em indivíduos adultos (KIM; WALSH; GRUESSNER, 2023). As causas para esta patologia podem ir desde hipertensão ou diabetes, ou até mesmo em decorrência de processos clínicos, como pacientes que passam por transplante de pâncreas (KIM; WALSH; GRUESSNER, 2023).

Já a doença diabetes (*Diabetes Mellitus* - DM), é um dos problemas de saúde com maior taxa de crescimento no mundo, tendo um número de brasileiros afetados de mais de 20 milhões em 2023 (BRASIL, 2023) e projetada para atingir cerca de 693 milhões de adultos em todo o mundo até 2045 (COLE; FLOREZ, 2020). Diversas pesquisas evidenciaram a presença inequívoca de uma componente genética tanto no diabetes quanto em suas complicações.

A DM, uma condição metabólica complexa, exibe uma classificação abrangente dividida em três tipos fundamentais: diabetes tipo 1 (DMT1), diabetes tipo 2 (DMT2) e diabetes gestacional. Cada uma dessas categorias se distingue por características peculiares, não apenas em relação às origens e fatores de risco, mas também em termos de abordagens terapêuticas (ZIA et al., 2022).

O DMT1 é uma condição decorrente da resposta autoimune que resulta na destruição das células β no pâncreas endócrino. Nesse cenário, o sistema imunológico ataca erroneamente as células responsáveis pela produção de insulina, levando a uma deficiência dessa substância crucial para a regulação adequada dos níveis de glicose no organismo. O DMT1 representa apenas 10% dos casos de diabetes em todo o mundo, mas ocorre com uma incidência crescente muito mais cedo na vida. Este tipo específico de diabetes demanda atenção especial devido à sua origem autoimune e à necessidade vital de intervenções terapêuticas, como a administração de insulina, para garantir um manejo eficaz e a qualidade de vida dos afetados (PASCHOU et al., 2018).

O DMT2 abrange aproximadamente 88% de todas as ocorrências de diabetes. Nessa condição, conhecida como resistência à insulina, a resposta ao hormônio insulina é reduzida, resultando na sua ineficácia. Inicialmente, para preservar a homeostase da glicose, há uma compensação com um aumento na produção de insulina. Contudo, ao longo do tempo, essa produção diminui, culminando no desenvolvimento do DMT2 (GOYAL; JIALAL, 2018). Há também a diabetes gestacional, onde a patologia se comporta como a DMT2 e ocupa 2% restantes do número de casos.

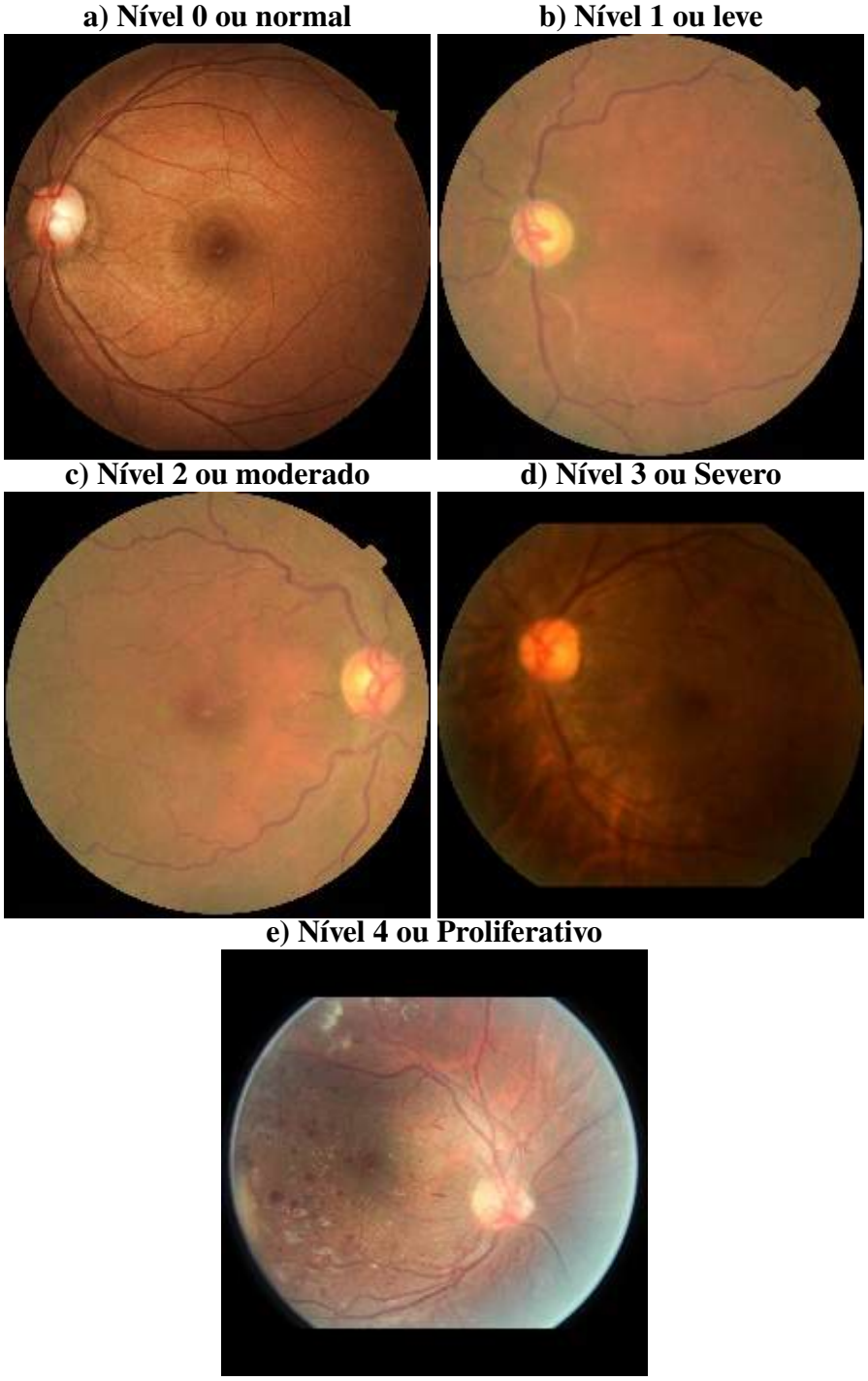
A Retinopatia Diabética (DR, do inglês *Diabetic Retinopathy*) pode gerar como uma condição oftalmológica debilitante, afetando os vasos sanguíneos delicados da retina. Esta patologia não apenas compromete a visão, mas também representa uma ameaça significativa de cegueira quando não diagnosticada precocemente. A DR é caracterizada por cinco estágios distintos ou categorias: normal, leve, moderado, severo e Retinopatia Diabética Proliferativa (DRP). Cada estágio desse espectro reflete uma progressão na gravidade dos danos causados aos vasos retinianos, delineando a complexidade e a variabilidade da condição ao longo de sua evolução (QUMMAR et al., 2019).

A DR se manifesta em dois estágios principais: a Retinopatia Diabética Não Proliferativa (DRNP) e a Retinopatia Diabética Proliferativa (DRP). Na fase não proliferativa, sendo os níveis normal (0), leve (1), moderado (2) e severo (3), as alterações nos vasos sanguíneos podem levar a microaneurismas, pequenos inchaços que representam uma resposta vascular inicial ao estresse. Conforme a condição avança para a fase proliferativa, ocorre neovascularização, um processo no qual novos vasos sanguíneos anômalos se formam, exacerbando o risco de hemorragias e potencialmente levando à perda de visão (QUMMAR et al., 2019).

Na classificação da retinopatia diabética proliferativa (4), proposta por Scanlon, Sallam e Wijngaarden (2017), há a inclusão de quatro novos níveis - leve, moderada, alto risco e avançada. Essa subdivisão permite uma análise mais refinada do grau proliferativo das características específicas de cada estágio, mas neste trabalho, foi utilizado apenas os níveis discriminatórios para a DRNP, tratando a DRP apenas como de nível 4.

A representação visual presente na Figura 1.1 ilustra os estágios que compõem a condição da retinopatia diabética.

Figura 1.1: Níveis de Retinopatia Diabética.



Fonte: Dugas et al. (2015)

1.1 Objetivo Geral

O objetivo deste trabalho foi analisar, comparar e otimizar o desempenho das diferentes arquiteturas de redes neurais convolucionais: Xception, VGG16, InceptionV3 e MobileNet, na tarefa de classificação de retinopatia diabética em cinco classes distintas, representando os graus de gravidade da patologia. Assegurou-se que os modelos desenvolvidos fossem confiáveis, reprodutíveis e eficientes, considerando as limitações e características dos dados disponíveis.

1.1.1 Objetivos Específicos

Os objetivos específicos detalham os principais aspectos investigados neste estudo, com foco em ampliar a compreensão e aprimorar o desempenho dos modelos avaliados:

- Implementar e ajustar algoritmos de processamento digital de imagens, como ajuste de contraste, nitidez e correção de cores, otimizando a qualidade das amostras clínicas para melhorar o desempenho dos modelos.

- Explorar e revisar a literatura sobre os principais sinais clínicos e biomarcadores associados à retinopatia diabética, para fundamentar as decisões metodológicas e modelagens propostas.

- Treinar, validar e comparar modelos de redes neurais convolucionais, incluindo Xception, VGG16, InceptionV3 e MobileNet, utilizando bases de dados de imagens de retina, com análise de métricas como acurácia, sensibilidade e especificidade em diferentes cenários de classificação.

- Investigar e avaliar o impacto de técnicas de pré-processamento e *data augmentation* sobre a performance dos modelos, analisando como essas etapas influenciam os resultados finais e a generalização dos modelos em condições reais.

- Otimizar os modelos selecionados, explorando ajustes nos hiperparâmetros e testes com diferentes resoluções de imagem para maximizar o desempenho e minimizar as limitações computacionais.

1.2 Trabalhos Relacionados

Um levantamento sobre o estado atual das tecnologias foi conduzido para identificar as mais recentes abordagens empregadas na classificação de imagens de fundoscopia. Esse estudo prévio permitiu compreender as técnicas mais avançadas utilizadas para enfrentar esse desafio específico, fornecendo uma base sólida para a pesquisa em questão e orientando o desenvolvimento de estratégias eficazes para a classificação precisa das imagens oftalmológicas. Esses estudos compreendem não apenas o uso de técnicas tradicionais de processamento digital de imagens, mas também exploram abordagens mais avançadas, como o aprendizado de características e o uso de classificadores.

Além disso, serão discutidos trabalhos que se dedicam à otimização de redes neurais artificiais, visando aprimorar sua eficácia e precisão na detecção da patologia oftalmológica. Essa análise oferecerá uma visão sobre o estado atual da pesquisa nesse campo e fornecerá uma base sólida para o desenvolvimento de novas estratégias e técnicas no diagnóstico da retinopatia diabética.

A reprodutibilidade de trabalhos sobre retinopatia diabética enfrenta sérias limitações devido à ausência de informações cruciais que deveriam acompanhar as pesquisas. Muitos estudos carecem de dados fundamentais, como as matrizes de confusão, que são essenciais para a compreensão completa do desempenho dos modelos de classificação. Além disso, a falta de detalhes sobre os algoritmos implementados e a escassez de uma descrição precisa das imagens utilizadas – incluindo quais foram removidas e os motivos para tal – dificulta a replicação dos resultados. A situação é agravada pela dificuldade em encontrar dados confiáveis e consistentes, o que compromete a validade das conclusões e impede que outros pesquisadores possam verificar ou construir sobre os achados apresentados.

Artigo: **Convolutional neural networks for diabetic retinopathy**, (PRATT et al., 2016)

Pratt et al. (2016) exploraram o conjunto de dados do desafio Kaggle (DUGAS et al., 2015), realizando um pré-processamento que redimensionou todas as imagens, que estão em resoluções variáveis na base, para uma resolução de 512×512 pixels. As imagens originais, após o pré-processamento, foram empregadas apenas uma vez para treinar a rede neural com uma arquitetura customizada. Em seguida, durante todo o treinamento, foi adotada uma técnica de *data augmentation* em tempo real, com o objetivo de aprimorar a capacidade da rede em localizar características relevantes nas imagens. A cada época de treinamento, cada imagem foi aleatoriamente alterada

com rotação variando de 0 a 90 graus, inversões horizontais e verticais, além de deslocamentos aleatórios tanto na horizontal quanto na vertical. Após o treinamento, a rede neural final alcançou resultados de 95% de especificidade, indicando sua habilidade em identificar corretamente os casos negativos para retinopatia diabética, 75% de precisão, indicando a proporção de casos corretamente identificados em relação ao total de casos, e 30% de sensibilidade, mostrando sua baixa capacidade de identificar corretamente os graus para retinopatia diabética. Em um cenário clínico, esta baixa sensibilidade pode levar a um número elevado de falsos negativos, onde pacientes que necessitam de tratamento não receberiam devida atenção.

Artigo: **Diagnosis of diabetic retinopathy using deep neural networks**, (GAO et al., 2018)

O estudo de Gao et al. (2018) demonstrou uma abordagem abrangente e detalhada no pré-processamento das imagens utilizadas para diagnóstico de retinopatia diabética. Além de ajustar o formato das imagens para uma forma padronizada, o grupo realizou uma normalização das cores, o que não apenas corrigiu o contraste e o brilho, mas também equilibrou a coloração das imagens, garantindo uma representação visual mais precisa dos tecidos retinianos. Para o *data augmentation* foi realizado o uso de espelhamento horizontal e vertical, juntamente com rotações aleatórias da imagem em ângulos de ± 25 graus. Além disso, a aplicação de zoom aleatório, variando de $0,85\times$ a $1,15\times$, e distorções aleatórias adicionaram ainda mais diversidade às imagens. O trabalho utilizou 5 arquiteturas de redes, uma classificação em 4 níveis e sua maior acurácia foi de $88,72\%$

Artigo: **Automated detection of diabetic retinopathy using deep learning**, (LAM et al., 2018)

No estudo conduzido por Lam et al. (2018), eles exploraram as bases de dados Messidor-1 e MildDR para realizar a classificação da presença ou ausência de DR. Utilizando a arquitetura da rede neural GoogLeNet. Um dos desafios enfrentados foi a variabilidade na forma como os dados foram adquiridos, bem como a disponibilidade limitada de amostras para treinamento do modelo. Para contornar essas questões, foi aplicado uma série de técnicas de *data augmentation*, incluindo zoom, rotação aleatória e translação aleatória. Além disso, utilizaram uma técnica avançada de pré-processamento chamada *contrast limited adaptive histogram equalization* (CLAHE) para melhorar a qualidade das imagens, tornando-as mais adequadas para análise por algoritmos de aprendizado de máquina. O resultado desse estudo foi uma acurácia de $74,5\%$ em classificação binária, demonstrando uma performance promissora na detecção da retinopatia diabética.

Artigo: **A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection**, (QUMMAR et al., 2019)

Qummar et al. (2019) optaram pelo conjunto de dados do desafio da Kaggle (DUGAS et al., 2015) para uma tarefa que envolvia a classificação de patologias em cinco classes. A abordagem desses autores incluiu a utilização de uma rede neural composta por um *ensemble* de arquiteturas convolucionais, incluindo ResNet50, Inception_v3, Xception, DenseNet121 e DenseNet169. Para aumentar a quantidade de dados, foram aplicadas operações de espelhamento, rotação aleatória e normalização. O redimensionamento dos dados foi realizado por meio de *up-sampling* com recortes de 512×512 pixels e *down-sampling* com descarte de imagens para igualar o número de imagens à classe com menor quantidade. Essa abordagem resultou em uma acurácia de $80,8\%$, demonstrando um desempenho robusto na tarefa de classificação de patologias oftalmológicas.

Artigo: **A data-driven approach to referable diabetic retinopathy detection**, (PIRES et al., 2019)

No estudo conduzido por Pires et al. (2019), foram empregados conjuntos de dados de grande relevância, incluindo o Kaggle (DUGAS et al., 2015), DR2 e Messidor-2, para examinar a presença da patologia. Eles adotaram uma abordagem utilizando redes neurais artificiais semelhantes à o_O e VGG em três resoluções distintas: 128×128, 256×256 e 512×512. Além disso, para aumentar a diversidade e a robustez dos dados, empregaram técnicas de *data augmentation*, como zoom, espelhamento, rotação aleatória, translação aleatória, esticamento e ajuste dos valores RGB. A performance do modelo foi avaliada utilizando o conjunto de dados do Kaggle (DUGAS et al., 2015), onde alcançaram uma AUC de $95,5\%$ na classificação binária. Esses resultados destacam a eficácia dessas técnicas no diagnóstico e detecção precoce de patologias oftalmológicas.

Artigo: **Deep learning approach to diabetic retinopathy detection**, (TYMCHENKO; MARCHENKO; SPODARETS, 2020)

Tymchenko, Marchenko e Spodarets (2020) realizaram tanto uma classificação binária como em 5 classes. Para treinamento e validação do modelo, utilizou versões pré-processadas das imagens originais, cortadas e redimensionadas. Devido a correlações entre o estágio da doença e características da imagem, como resolução e brilho, foram aplicadas várias técnicas de *data augmentation* para evitar o *overfitting* da CNN. Foi utilizado o conjunto de dados EyePACs de 2015 (DUGAS et al., 2015) para pré-treinamento das CNNs. Também foi inicializado o extrator de características

(*Encoder*) com pesos de uma CNN pré-treinada utilizando a base ImageNet. Este *Encoder* foi uma arquitetura de CNN já conhecida, EfficientNet-B4, EfficientNet-B5 e SE-ResNeXt50. Após o pré-treinamento, os pesos do codificador foram utilizados para os próximos estágios. O treinamento principal foi realizado na base de dados APTOS2019, IDRID e MESSIDOR combinados. Durante o treinamento, foi monitorada a separabilidade das características. O trabalho obteve acurácia máxima de $92,9\%$ utilizando a classificação em 5 níveis no conjunto de validação e $99,3\%$ utilizando classificação binária no conjunto de teste. Não há descrição de quais imagens específicas foram utilizadas de cada base.

Artigo: **Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network**, (SAEED; HUSSAIN; ABOALSAMH, 2021)

O trabalho de Saeed, Hussain e Aboalsamh (2021) apresenta um sistema automatizado para a detecção da retinopatia diabética utilizando técnicas avançadas de *deep learning* e aprendizado por transferência. O modelo utiliza uma rede convolucional (CNN) pré-treinada com ajustes específicos, incluindo a reconfiguração das camadas iniciais para captar padrões locais nas lesões e a substituição das camadas finais para melhorar a extração de características globais. Essa abordagem otimiza a adaptação do modelo às imagens fundoscópicas, destacando estruturas relevantes para o diagnóstico de DR.

Um dos diferenciais do estudo é a utilização de análise de componentes principais (PCA) nas camadas totalmente conectadas, reduzindo a dimensionalidade dos dados e melhorando a generalização do modelo. Além disso, o sistema inclui uma etapa de classificação baseada em *gradient boosting*, que potencializa a detecção de padrões complexos nas imagens. Essa combinação de técnicas garante um equilíbrio entre simplicidade e precisão, reduzindo os riscos de sobreajuste e aumentando a robustez do modelo.

A avaliação do sistema foi realizada em dois conjuntos de dados renomados, EyePACS2015 e Messidor. Os resultados obtidos indicam que o modelo supera métodos existentes, demonstrando excelente desempenho em métricas como acurácia, sensibilidade e especificidade. Essas evidências confirmam a eficácia da abordagem para a triagem automatizada de DR, facilitando a priorização de casos que requerem avaliação clínica mais detalhada. O trabalho obteve acurácia de 99,73%.

Artigo: **Diabetic Retinopathy Classification Using CNN and Hybrid Deep Convolutional Neu-**

ral Networks, (SAROBIN; PANJANATHAN, 2022)

O trabalho de Sarobin e Panjanathan (2022) aborda a detecção e classificação automática da retinopatia diabética (DR) em diferentes estágios utilizando arquiteturas avançadas de aprendizado profundo e aprendizado por transferência. A pesquisa explora três modelos principais: CNN pura, CNN híbrida com ResNet e CNN híbrida com DenseNet, aplicados a um conjunto de dados contendo 3662 imagens para treinamento. As imagens, categorizadas em cinco estágios de DR (Nenhuma, Leve, Moderada, Severa e Proliferativa), são processadas para identificar o progresso da condição com base em características extraídas das imagens fundoscópicas.

Os resultados destacam a eficácia da CNN híbrida com DenseNet, que alcançou a maior acurácia (96,22%) comparada às outras abordagens. Enquanto a CNN híbrida com ResNet obteve 93,18% e a CNN pura alcançou 75,61%, a análise comparativa sugere que a integração de DenseNet melhora a capacidade do modelo de capturar padrões complexos nas imagens. Essa abordagem é particularmente relevante para a classificação precisa dos estágios intermediários de DR, onde diferenças sutis podem ser decisivas para o diagnóstico.

Artigo: **A multilevel deep feature selection framework for diabetic retinopathy image classification**, (ZIA et al., 2022)

O trabalho realizado por Zia et al. (2022) utilizou o conjunto de dados disponibilizado pelo desafio da Kaggle (DUGAS et al., 2015) para treinar e classificar suas redes neurais artificiais. Este mesmo conjunto de dados é adotado neste estudo e será minuciosamente descrito na seção de Metodologia, fornecendo uma base consistente para a compreensão dos métodos empregados e dos resultados obtidos. Zia et al. (2022) propuseram uma abordagem que emprega uma variedade de estratégias para classificar a retinopatia diabética. Além de corrigir o desequilíbrio de dados, o estudo inclui etapas como pré-processamento, redimensionamento de imagem, extração, fusão e seleção de características. O trabalho também explorou duas arquiteturas de redes neurais artificiais, VGG19 e Inception V3, e utilizou diversos classificadores, como C-SVM, ESD, F-KNN, W-KNN, entre outros, com o objetivo de aprimorar a precisão e a eficácia da classificação das imagens de fundo de olho afetadas pela retinopatia diabética. Neste trabalho, a maior acurácia obtida foi de $96,4\%$ em classificação de 5 níveis.

Artigo: **Classification of Diabetic Retinopathy Using Image Pre-processing Techniques** , (DUV-

VURI et al., 2023)

O trabalho de Duvvuri et al. (2023) explora a classificação de imagens de retinopatia diabética (DR) utilizando técnicas de pré-processamento e aprendizado profundo para aprimorar o desempenho do modelo. Inicialmente, foram aplicadas técnicas como erosão e equalização de histograma às imagens fundoscópicas, provenientes do banco de dados EyesPACS2015. Essas etapas visaram melhorar a qualidade das imagens, realçando características relevantes para a detecção da DR e otimizando o treinamento do modelo.

A abordagem proposta baseia-se em uma rede neural convolucional (CNN) para realizar a classificação multiclasse da DR, abrangendo seus diferentes estágios. Após o pré-processamento, a CNN alcançou uma acurácia de 86,4%, demonstrando um avanço significativo em relação ao desempenho obtido sem a aplicação dessas técnicas. Este resultado evidencia a eficácia do pré-processamento na captura de padrões relevantes nas imagens de retina, especialmente para condições que afetam a precisão do diagnóstico.

Os achados de Duvvuri et al. (2023) destacam o papel do pré-processamento no contexto de redes convolucionais aplicadas à saúde ocular. O trabalho não apenas melhora a precisão do diagnóstico automatizado, mas também apresenta uma solução viável para implementação em sistemas de triagem, com potencial para impactar positivamente o manejo da DR em larga escala.

Artigo: **An Automated Detection and Multi-stage classification of Diabetic Retinopathy using Convolutional Neural Networks**, (NANDHINI et al., 2023)

O trabalho de Nandhini et al. (2023) apresenta uma abordagem inovadora para a detecção da retinopatia diabética (DR) utilizando o modelo DiaNet (DNM), explorando técnicas avançadas de pré-processamento e aprendizado profundo. O pré-processamento das imagens fundoscópicas é realizado com o uso de filtros de *Gabor*, que melhoram a visibilidade dos vasos sanguíneos e facilitam a análise de textura, reconhecimento de objetos e extração de características. Essas melhorias são essenciais para destacar as regiões afetadas pela DR.

No estágio de *data augmentation*, o modelo aplica a Análise de Componentes Principais (PCA) para reduzir as dimensões do conjunto de dados, otimizando as características relevantes para o treinamento do DNM. Essa técnica reduz a complexidade do modelo, promovendo um aprendizado mais eficiente e diminuindo o impacto de atributos menos significativos. Como resultado, a

abordagem proposta atingiu uma acurácia média de classificação de 90,02%.

Os resultados de Nandhini et al. (2023) reforçam o impacto positivo do DiaNet Model na classificação de imagens médicas. A combinação de pré-processamento eficiente, redução dimensional e aprendizado profundo demonstra como avanços tecnológicos podem aprimorar a precisão e a viabilidade de sistemas de diagnóstico, contribuindo para a melhoria da saúde ocular em larga escala.

Artigo: **Detection and Classification of Diabetic Retinopathy using Pretrained Deep Neural Networks**, (PRIYA et al., 2023)

O estudo de Priya et al. (2023) apresenta um modelo baseado em aprendizado profundo para a identificação de retinopatia diabética (DR) em seus diversos estágios. Reconhecendo que até 80% dos indivíduos com diabetes há mais de 10 anos podem desenvolver essa complicação, a pesquisa aborda uma importante lacuna, principalmente em regiões com carência de recursos e conhecimento especializado para o diagnóstico adequado. O sistema proposto utiliza redes neurais convolucionais (CNNs) pré-treinadas, como VGG-16 e MobileNet-V2, para classificar olhos normais e os diferentes graus de DR: leve, moderada, grave e proliferativa.

O modelo alcançou taxas de acurácia de 90% e 92%, respectivamente, ao classificar o estágio da DR em imagens fundoscópicas. Essas redes, reconhecidas por sua eficiência em extração de características, são empregadas para substituir métodos tradicionais baseados em extração manual de características ou identificação direta da doença, que demandam maior esforço humano e estão sujeitos a variabilidades.

Os resultados obtidos por Priya et al. (2023) destacam o potencial das CNNs pré-treinadas no avanço do diagnóstico assistido por computador, proporcionando maior acessibilidade e precisão na identificação de DR. O estudo reforça a relevância de combinar modelos robustos com estratégias de classificação que abrangem múltiplos estágios da doença, contribuindo para melhorar os cuidados com a saúde ocular globalmente.

Na Tabela 1, cada linha representa um estudo distinto, detalhando as características fundamentais de cada pesquisa. Além de listar os trabalhos e as bases de dados associadas, são fornecidas informações sobre o número de classes tratadas em cada estudo. As métricas de avaliação também são essenciais para compreender a eficácia de cada abordagem. Esses dados organizados facilitam a comparação entre os estudos e fornecem uma visão abrangente do panorama atual no campo da

detecção de retinopatia diabética.

Tabela 1: Alguns aspectos dos trabalhos relacionados.

Autor	**Base de Dados**	**Método**	**Classificação**	**Acurácia**
Pratt et al. (2016)	Kaggle.	CNN customizada.	5 níveis	ACC 75%
Gao et al. (2018)	STARE; DRIVE; DIARETDB0; DIARETDB1.	Inception@4; InceptionV3; ResNet18; ResNet101; VGG19.	4 níveis	ACC $88,72\%$
Lam et al. (2018)	Messidor-1; MildDR.	AlexNet; VGG16; GoogLeNet; ResNet; InceptionV3.	Binária	ACC 74,5%
Qummar et al. (2019)	Kaggle.	União de vários métodos.	5 níveis	ACC $80,8\%$
Tymchenko, Marchenko e Spodarets (2020)	Kaggle; IDRiD; Messidor; APTOS.	Arquitetura proposta com base na Xception.	Binária	ACC $99,3\%$
Zia et al. (2022)	Kaggle.	VGG com classificador de saída: C-SVM; ESD; F-KNN; W-KNN; ES-KNN; Q-SVM; LD.	5 níveis	ACC $96,4\%$
Saeed, Hussain e Aboalsamh (2021)	EyePACS2015, Messidor.	ResNetGB	5 níveis, 2 níveis	ACC 99,73%
Sarobin e Panjanathan (2022)	APTOS2019.	CNN customizada	5 níveis	ACC 96,22%
Duvvuri et al. (2023)	APTOS2019 GrayScale.	CNN customizada	5 níveis	ACC 73,33%
Nandhini et al. (2023)	APTOS2019, EYESPACS2015.	DenseNet169, Dianet	5 níveis	ACC 90,02%
Priya et al. (2023)	APTOS2019.	VGG16, MobileNetV2	5 níveis	ACC 92% e 74,21% de ACC calculado[1]

Fonte: Autoria própria (2024)

Os estudos revisados apresentam diversas abordagens para a classificação da retinopatia diabé-

[1]ACC calculado a partir da matriz de confusão apresentada no trabalho.

tica em imagens de fundoscopia, explorando desde arquiteturas convolucionais simples até modelos híbridos mais avançados. Técnicas de pré-processamento e *data augmentation* são amplamente utilizadas para melhorar a qualidade das imagens e aumentar a robustez dos modelos. Entre as estratégias adotadas, destacam-se normalizações de cor, ajustes de contraste e brilho, além de transformações geométricas, como rotações e espelhamentos. Modelos como VGG16, ResNet, DenseNet, InceptionV3 e EfficientNet demonstraram desempenhos variados, sendo que abordagens que combinam múltiplas arquiteturas frequentemente apresentaram melhores resultados. A acurácia variou de 74,5% a 99,73%, com modelos híbridos e técnicas de aprendizado por transferência mostrando os melhores desempenhos na classificação da patologia.

Entretanto, a reprodutibilidade dos trabalhos analisados enfrenta desafios devido à falta de informações detalhadas sobre as bases de dados utilizadas e os critérios de remoção de imagens. A ausência de matrizes de confusão e descrições completas dos algoritmos compromete a validação independente dos resultados. Além disso, observa-se uma grande variação nas métricas utilizadas para avaliar os modelos, tornando difícil uma comparação direta entre os diferentes estudos. Apesar dessas limitações, a tendência geral aponta para um avanço no uso de redes neurais profundas para a detecção da retinopatia diabética, com crescente interesse em estratégias que combinam aprendizado supervisionado e técnicas de extração de características para melhorar a precisão do diagnóstico automatizado.

2 Retinopatia Diabética

A DR é uma complicação grave do diabetes, caracterizada por danos nos vasos sanguíneos da retina, o que pode levar à perda da visão (TYMCHENKO; MARCHENKO; SPODARETS, 2020). Essa condição ocorre quando níveis elevados de açúcar no sangue danificam os pequenos vasos que fornecem sangue à retina, resultando em vazamento de fluidos e hemorragias. A visão pode ficar distorcida e, se não tratada adequadamente, pode levar à cegueira.

Além de outras doenças oculares, como catarata e glaucoma, a retinopatia diabética é uma das complicações mais comuns associadas ao diabetes, com incidência significativa relatada em países como Estados Unidos, Reino Unido e Singapura (CHA; VILLARROEL; VAHRATIAN, 2019). Esses dados destacam a importância da detecção precoce e do controle rigoroso do diabetes para prevenir ou retardar o desenvolvimento da retinopatia diabética e suas complicações.

Os dados apresentados por Rohan, Frost e Wald (1989), sugerem que uma parcela significativa, equivalente a pelo menos 56%, dos novos casos dessa doença poderia ser mitigada por meio de tratamentos adequados, oportunos e devidamente monitorados dos olhos. Essa constatação ressalta a importância da detecção precoce e do acompanhamento contínuo da saúde ocular para prevenir o agravamento da condição e seus potenciais impactos adversos na visão. Essa porcentagem expressiva destaca a necessidade de medidas proativas e eficazes no manejo dessa enfermidade, visando não apenas reduzir sua incidência, mas também melhorar a qualidade de vida dos pacientes afetados.

A progressão da perda de visão ao longo do desenvolvimento gradual da DR pode ser variada. Essa condição geralmente é dividida em duas fases principais: a DR não proliferativa (DRNP) e a DR proliferativa (DRP), esta última caracterizada por neovascularização ou hemorragia vítreo/pre-retiniana. Anualmente, até 10% dos pacientes diabéticos sem DR desenvolvem DRNP, e para aqueles com DRNP grave, o risco de evoluir para DRP em um ano chega a 75% (GAO et al., 2018). A transição do estado de normalidade, onde não há anormalidades aparentes na retina, para a DRP, geralmente leva muitos anos. Consequentemente, a DRNP é frequentemente subdividida em três estágios: leve, moderado e grave. Esses cinco estágios juntos compõem a "Escala de Gravidade da Doença Retinopatia Diabética Clínica Internacional" (GAO et al., 2018).

As melhores opções de tratamento para pacientes variam conforme os estágios da doença. Para pacientes sem DR ou com DRNP leve, apenas exames regulares são necessários e para pacientes

com DRNP moderada ou pior, as opções de tratamento variam desde o tratamento com laser disperso até a vitrectomia.

A Figura 2.1 apresenta uma comparação visual entre a visão normal de um paciente sem retinopatia diabética (Figura 2.1(a)) e a visão de uma imagem afetada pelo estágio mais avançado da patologia (Figura 2.1(b)). Na Figura 2.1(a), observa-se a imagem original, sem sinais de distorções ou perda de informações. Já na Figura 2.1(b), é possível notar as consequências visíveis da retinopatia diabética. Essa comparação visual ressalta a importância da detecção precoce e do acompanhamento regular para prevenir complicações graves associadas à retinopatia diabética.

Figura 2.1: Representação da visão patológica sendo a) a imagem sem patologia e b) uma representação da visão patológica.

Fonte: Zia et al. (2022)

Para fornecer aos pacientes o tratamento adequado, é importante primeiro classificar a gravidade da DR. Clinicamente, o diagnóstico de DR é frequentemente feito com imagens do fundo do olho, que podem ser obtidas fotografando diretamente o fundo do olho. As lesões comuns que indicam DR incluem exsudatos duros ou moles, microaneurismas e hemorragias. Todas essas lesões podem ser identificadas a partir de imagens do fundo do olho.

A fundoscopia é uma técnica fundamental para adquirir imagens utilizadas no diagnóstico de uma variedade de condições oftalmológicas, incluindo a retinopatia diabética. Apesar de ser uma ferramenta essencial, muitos estudantes e até mesmo médicos experientes enfrentam desafios ao utilizar o oftalmoscópio direto. Isso se deve, em parte, ao seu campo de visão limitado, que pode

dificultar a visualização adequada da retina, especialmente em pacientes com dificuldades de posicionamento ou cooperação (RAMSAMY; ARUNAKIRINATHAN; COOMBES, 2017).

Há outras técnicas para fazer um diagnóstico mais preciso, como a angiografia com fluoresceína. A mesma pode ser usada, pois é capaz de revelar estruturas finas dos vasos sanguíneos na retina. No entanto, os corantes de fluoresceína podem causar reações alérgicas e requerem rins funcionais para excreção, e geralmente não estão disponíveis em hospitais pequenos (GAO et al., 2018).

Esses obstáculos destacam a importância de uma abordagem cuidadosa e habilidosa durante o exame oftalmoscópico, garantindo uma avaliação completa e precisa do fundo do olho. Além disso, é fundamental que os profissionais de saúde recebam treinamento adequado e pratiquem regularmente o uso do oftalmoscópio direto para aprimorar suas habilidades e garantir um diagnóstico preciso das condições oftalmológicas (RAMSAMY; ARUNAKIRINATHAN; COOMBES, 2017).

Por meio dessa técnica, os profissionais de saúde podem avaliar diretamente a saúde da retina e identificar alterações que podem indicar condições oftalmológicas subjacentes. Entre essas condições, destacam-se as seguintes (RAMSAMY; ARUNAKIRINATHAN; COOMBES, 2017):

- Atrofia óptica,
- Retinopatia hipertensiva,
- Alterações nos vasos sanguíneos da retina,
- Retinopatia diabética,
- Descolamento de retina.

A classificação da DR envolve a ponderação de numerosas características e a localização de tais características (GROUP, 1991). Na análise das imagens de fundoscopia, os profissionais buscam por sinais que indiquem a presença da retinopatia diabética, tais como microaneurismas, hemorragias e exsudatos (NAGPAL et al., 2022). Esses indicadores fornecem pistas cruciais para o diagnóstico e tratamento eficaz da condição, destacando a importância da avaliação cuidadosa das imagens fundoscópicas para a saúde ocular dos pacientes.

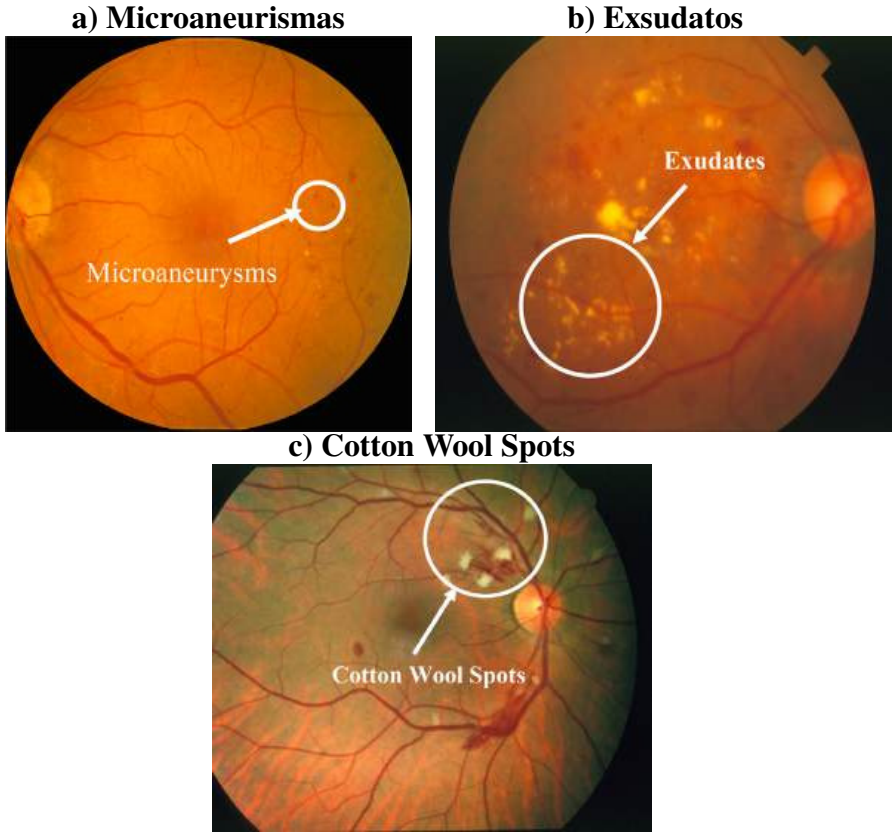
A Figura 2.2-a) mostra uma região específica onde foram detectados microaneurismas na retina, que são pequenas dilatações nos vasos sanguíneos causadas pelo enfraquecimento das paredes dos

capilares. Essas anormalidades são marcadores importantes da retinopatia diabética e podem indicar o estágio e a gravidade da condição ocular.

Na Figura 2.2-b), podemos observar a identificação dos exsudatos em um exame de fundoscopia. Os exsudatos são depósitos de material lipídico ou proteico que se acumulam na retina, muitas vezes como resultado de danos aos vasos sanguíneos devido à retinopatia diabética. Essas lesões são uma indicação visual crucial da presença e progressão da doença.

As *Cotton Wool Spots*, ou "Manchas de algodão"em português (2.2 -c)), são uma característica da patologia que os médicos oftalmologistas observam durante exames de fundoscopia. O nome "Manchas de Algodão" deriva da aparência borrada e difusa dessas áreas, que se assemelham a pequenos pedaços de algodão espalhados pela superfície da retina, porém alguns autores se referem às mesmas como exsudatos moles (JAYA; DHEEBA; SINGH, 2015).

Figura 2.2: **a)**Microaneurismas, pequenas dilatações que ocorrem nos capilares sanguíneos da retina; **b)** Exsudatos, depósitos de proteínas e lipídios na retina; **c)** *Cotton Wool Spots*, pequenas áreas brancas e fofas na retina, causadas por microinfartos da camada de fibras nervosas.



Fonte: Nagpal et al. (2022)

A formação dessas manchas está intimamente ligada à interrupção do fluxo sanguíneo na rede capilar da retina. Quando o suprimento de sangue é comprometido, as células da retina começam a sofrer de falta de oxigênio e nutrientes, resultando em danos teciduais e eventualmente na formação das manchas características. Embora não sejam exclusivas de uma condição médica específica, as manchas de algodão são frequentemente associadas a condições que afetam a microcirculação sanguínea, como a diabetes e a hipertensão (NAGPAL et al., 2022).

As hemorragias podem assumir diferentes formas na retina, sendo classificadas como hemorragias pré-retinianas, sub-retinianas e retinianas. Na Figura 2.3-a), podemos observar um exemplo de hemorragia pré-retiniana, que ocorre entre a retina e o vítreo. Esse tipo de hemorragia pode ser causado por ruptura de vasos sanguíneos na parte anterior do olho.

Na Figura 2.3-b), temos um exemplo de hemorragia sub-retiniana, onde o sangue se acumula entre a retina e a coróide. Essa condição pode ser decorrente de lesões traumáticas ou de distúrbios vasculares.

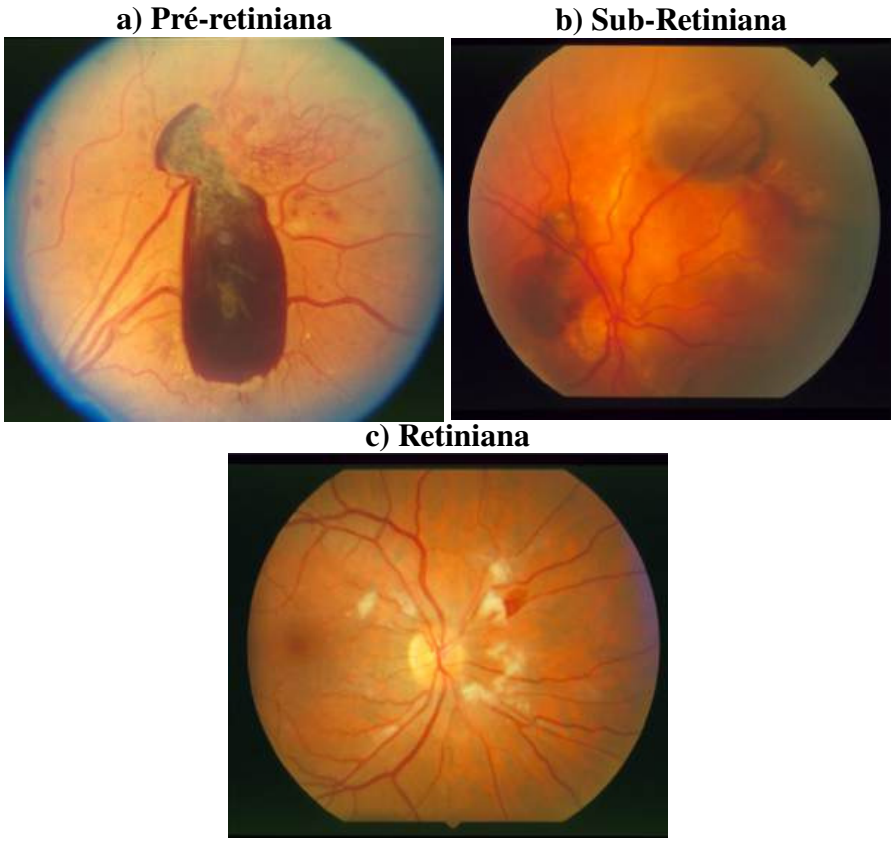
Por fim, na Figura 2.3-c), podemos ver uma hemorragia retiniana, que ocorre dentro da própria retina. Essas hemorragias são frequentemente associadas a doenças vasculares, como a retinopatia diabética, e podem ser um sinal de comprometimento da saúde ocular.

Embora as características mencionadas sejam fundamentais para o diagnóstico da retinopatia diabética, há outros sinais mais discretos que podem ser indicativos de alerta para uma detecção precoce da condição, como ilustrado na Figura 2.4. Podemos observar a ocorrência de uma reduplicação de um laço venoso. Na Figura 2.5, observa-se um fenômeno que ocorre quando um canal pré-existente se dilata ou quando surge a proliferação de um novo canal adjacente, com calibre aproximadamente igual ao do vaso original (SCANLON; SALLAM; WIJNGAARDEN, 2017).

Há sinais mais sutis que a reduplicação venosa pode indicar alertas para o diagnóstico precoce da retinopatia diabética. Um exemplo é o estrangulamento arteriovenoso, como ilustrado na Figura 2.5, onde uma artéria é comprimida por uma veia. Esses eventos, embora menos óbvios, podem ser indicativos de complicações oculares, mas são mais difíceis e complexos de serem identificados.

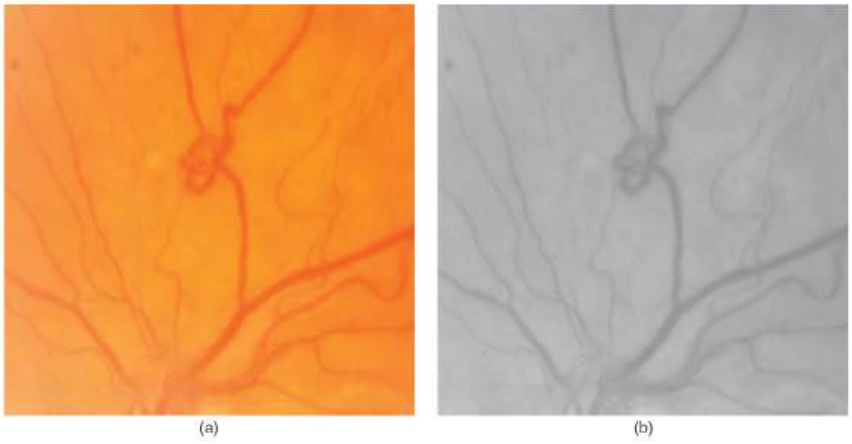
Embora as características citadas anteriormente sejam relevantes, as técnicas de diagnóstico clínico da retinopatia diabética concentram-se principalmente na identificação e avaliação de microaneurismas, hemorragias, exsudatos moles e exsudatos duros (NAGPAL et al., 2022).

Figura 2.3: Variedades de hemorragias que podem surgir em casos de retinopatia diabética.



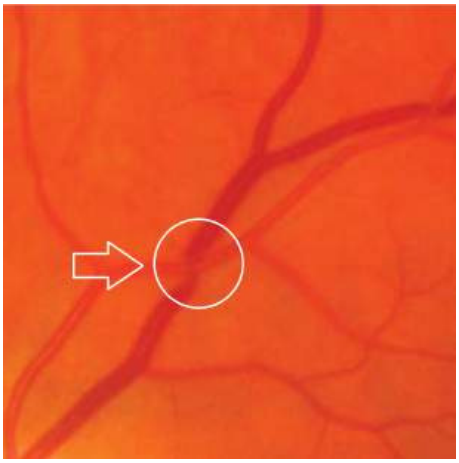
Fonte: Nagpal et al. (2022)

Figura 2.4: Reduplicação venosa: a) foto colorida e b) fotografia com vermelho zerado e em escala de cinza.



Fonte: Scanlon, Sallam e Wijngaarden (2017)

Figura 2.5: Compressão arteriovenosa onde uma artéria está comprimindo uma veia.



Adaptado de: Scanlon, Sallam e Wijngaarden (2017)

2.1 Análise quantitativa da deterioração da visão em função do tempo

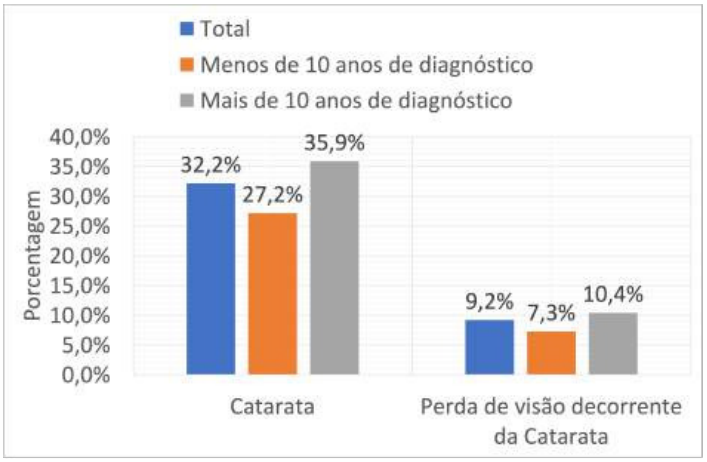
Além da retinopatia, diversas patologias oculares ocorrem em decorrência da diabetes. Segundo Cha, Villarroel e Vahratian (2019), a diabetes apenas progride negativamente com a idade. No ano de 2017, a prevalência de diabetes diagnosticada aumentou significativamente com a idade (CHA; VILLARROEL; VAHRATIAN, 2019). Enquanto era de $13,2\%$ em adultos com idades entre 45 e 64 anos, essa taxa subiu para $20,1\%$ entre indivíduos de 65 a 74 anos e $19,8\%$ naqueles com mais de 75 anos (CHA; VILLARROEL; VAHRATIAN, 2019). Essa tendência reflete o aumento do risco de desenvolver diabetes com o envelhecimento, destacando a importância de medidas preventivas e de cuidados específicos para diferentes faixas etárias.

Adicionalmente, a extensão temporal do diabetes constitui um fator de risco significativo para o avanço de complicações visuais (CHA; VILLARROEL; VAHRATIAN, 2019). Adultos diagnosticados com diabetes apresentam maior propensão a desenvolver distúrbios oculares e a sofrer perda de visão associada a esses distúrbios quando comparados àqueles sem diabetes (SOLOMON et al., 2017). Segundo aponta Cha, Villarroel e Vahratian (2019), adultos com 45 anos ou mais com diabetes diagnosticada, $32,2\%$ tinham catarata, $8,6\%$ tinham retinopatia diabética, $7,1\%$ tinham glaucoma e $4,3\%$ tinham degeneração macular.

A análise comparativa dos dados estatísticos relacionados a diversas patologias oculares com os números específicos da retinopatia diabética revela um panorama intrigante e relevante para a compreensão mais profunda das complicações oftalmológicas associadas ao diabetes (CHA; VIL-

LARROEL; VAHRATIAN, 2019). Na Figura 2.6, pode-se observar um crescimento de $8,7$ pontos percentuais de ocorrência nos casos de catarata com mais de 10 anos de diagnóstico da diabetes quando comparado com menos de 10 anos de diagnóstico.

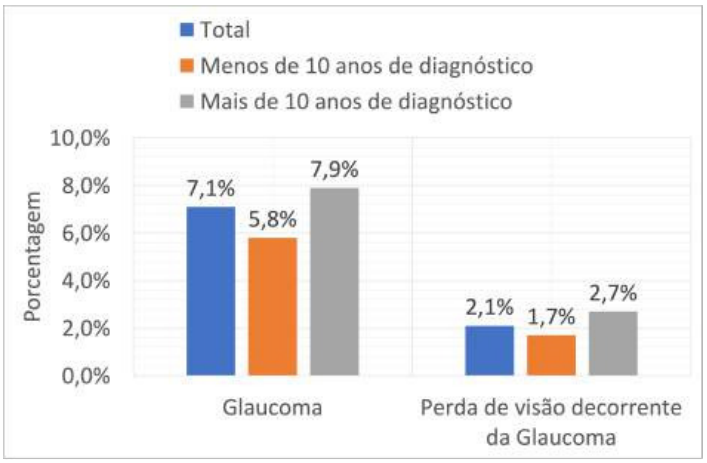
Figura 2.6: Porcentagem de adultos acima de 45 anos com diagnóstico de diabetes e que possuem catarata.



Adaptador de: Cha, Villarroel e Vahratian (2019)

Já quando analisado o panorama do glaucoma na Figura 2.7, pode-se observar um crescimento de $2,1$ pontos percentuais nos casos com mais de 10 anos de diagnóstico de diabetes quando comparado com os de menos de 10 anos de diagnóstico.

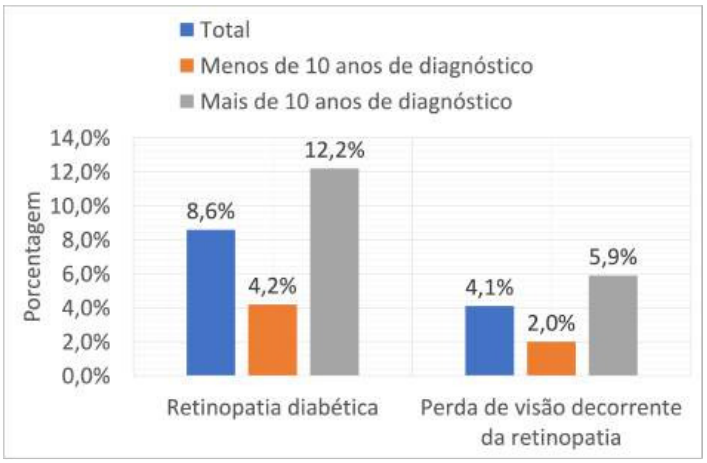
Figura 2.7: Porcentagem de adultos acima de 45 anos com diagnóstico de diabetes e que possuem glaucoma.



Adaptador de: Cha, Villarroel e Vahratian (2019)

Considerando os dados apresentados sobre o glaucoma e a catarata, podemos realizar uma análise mais aprofundada dos números relacionados à retinopatia diabética. Observa-se, na Figura 2.8, um aumento de 8 pontos percentuais no número de casos após 10 anos do diagnóstico de diabetes, em comparação com menos de 10 anos. Embora esse aumento possa parecer modesto em comparação com a catarata, por exemplo, ele representa aproximadamente três vezes mais casos após 10 anos do diagnóstico.

Figura 2.8: Porcentagem de adultos acima de 45 anos com diagnóstico de diabetes e que possuem retinopatia diabética.



Adaptador de: Cha, Villarroel e Vahratian (2019)

Essa discrepância destaca a importância de se monitorar de perto a saúde ocular em pacientes com diabetes, especialmente ao longo do tempo. A retinopatia diabética é uma complicação séria e progressiva, e a detecção precoce é fundamental para evitar danos irreversíveis à visão. Portanto, estratégias de prevenção e cuidados oftalmológicos regulares são essenciais para garantir a saúde visual dos pacientes diabéticos. Além disso, vale ressaltar que aproximadamente $5,9\%$ dos casos de retinopatia diabética resultam em perda de visão (CHA; VILLARROEL; VAHRATIAN, 2019). Esse dado posiciona essa condição entre o glaucoma e a catarata em termos de probabilidade de perda visual.

2.2 Diagnóstico Assistido por Computadores

A detecção automática da retinopatia diabética por computadores representa um avanço significativo na área da saúde ocular, pois simplifica e acelera o processo de classificação e diagnóstico.

Esse avanço é especialmente importante considerando a natureza muitas vezes tediosa e demorada da análise manual realizada por oftalmologistas. Com a detecção automática, é possível processar um maior número de exames em menos tempo, buscando uma taxa acima de 90%, aumentando assim a eficiência e consistência dos serviços oftalmológicos e permitindo um maior número de atendimentos (MOOKIAH et al., 2013).

É fundamental ressaltar que todo diagnóstico médico deve ser realizado por um profissional de saúde qualificado. A proposta deste trabalho não é substituir a avaliação clínica, mas fornecer uma ferramenta de suporte para auxiliar no diagnóstico. O uso de redes neurais na detecção de doenças, como a retinopatia diabética, busca complementar a análise médica, contribuindo para uma identificação mais rápida e precisa de casos suspeitos, mas sempre mantendo o médico como responsável final pela interpretação e decisão clínica.

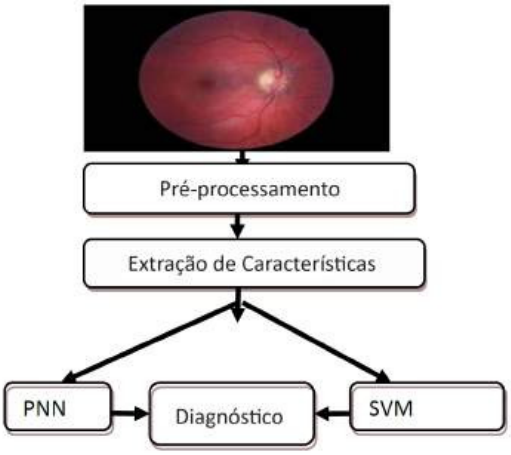
Além da agilidade no diagnóstico, a detecção automática também contribui significativamente para a prevenção da perda de visão causada pela retinopatia diabética. Ao permitir um diagnóstico mais rápido e preciso, os pacientes podem receber tratamento adequado em estágios iniciais da doença, antes que ocorram danos irreversíveis à visão. Técnicas como aprendizado de máquina com redes neurais artificiais convolucionais também permitem a detecção de sinais indicativos de outros tipos de patologias oculares (LAM et al., 2018). Isso destaca a importância da tecnologia na saúde, não apenas para melhorar a eficiência dos serviços, mas também para salvar a visão e melhorar a qualidade de vida dos pacientes com diabetes.

Durante anos, as técnicas tradicionais de classificação assistida por computadores foram fundamentais no campo da detecção da retinopatia diabética (WANG; YANG, 2018). Métodos como a Transformada de Hough, Filtros de Gabor e análises de variações de intensidade foram amplamente empregados. Após a etapa de extração de características, algoritmos de detecção de objetos, como Máquinas de Vetores de Suporte (SVM) e o K-vizinhos mais próximos (KNN), eram aplicados para identificar e localizar exsudatos e hemorragias na retina (WANG; YANG, 2018). A Figura 2.9 exemplifica uma estrutura de classificação.

As técnicas tradicionais de classificação de imagens podem ser divididas em três estágios principais, conforme descrito por Gao et al. (2018): pré-processamento de imagem, extração de características e classificação das características.

O pré-processamento de imagem é a etapa inicial e fundamental, na qual a imagem da retina

Figura 2.9: Diagrama em Blocos da comparação entre classificadores para diagnóstico de DR.



Adaptado de (PRIYA; ARUNA, 2012)

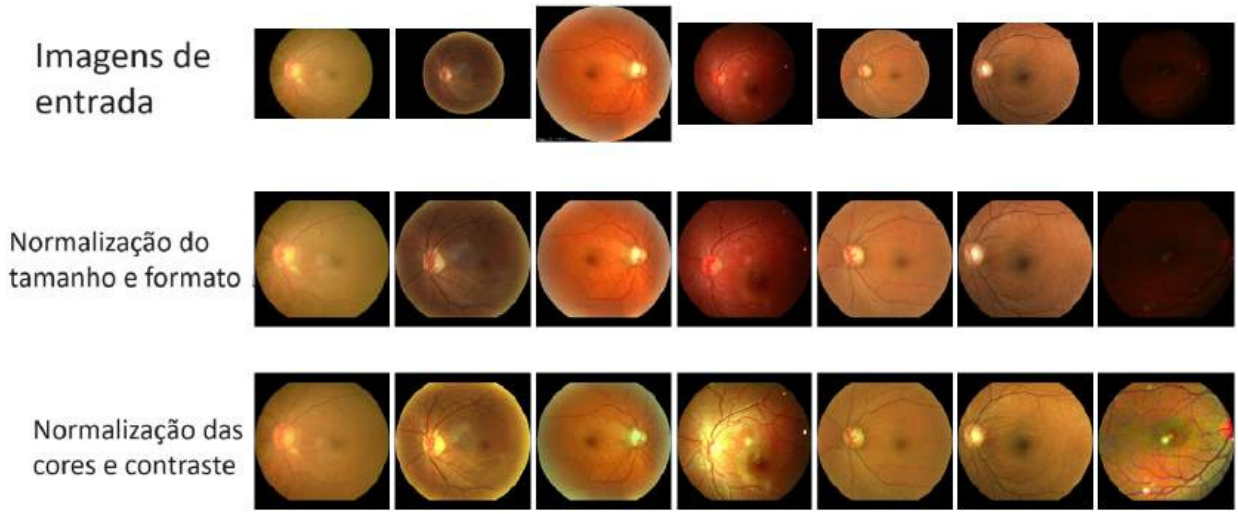
é processada para melhorar sua qualidade e facilitar a identificação de características relevantes. Isso pode incluir a correção de distorções, o aumento do contraste e a redução de ruídos (GAO et al., 2018). Em seguida, na etapa de extração de características, são identificados os aspectos mais importantes da imagem que serão utilizados para a classificação. Isso pode envolver a identificação de padrões como vasos sanguíneos, exsudatos e hemorragias, que são indicativos de retinopatia diabética (GAO et al., 2018). Por fim, na etapa de classificação das características, é aplicado um algoritmo de classificação para determinar a presença ou gravidade da retinopatia diabética com base nas características extraídas. Essa classificação pode ser binária (presente/ausente) ou multiclasse (indicando diferentes estágios da doença) (GAO et al., 2018).

Esses três estágios combinados formam a base das técnicas tradicionais de classificação de imagens para a detecção da retinopatia diabética, sendo complementadas por abordagens mais avançadas, como o uso de redes neurais artificiais convolucionais, que têm demonstrado um desempenho superior em muitos casos (TYMCHENKO; MARCHENKO; SPODARETS, 2020). Embora as abordagens tradicionais tenham um desempenho sólido, suas limitações são evidentes. A dependência de características simples e diretas torna a criação manual de novas características eficazes uma tarefa árdua (GAO et al., 2018). Esse processo pode resultar em conjuntos de características insuficientes ou pouco representativos, afetando a capacidade do modelo de generalizar para novos dados ou condições.

Os estudos mencionados, como Gao et al. (2018), Qummar et al. (2019) e Tymchenko, Mar-

chenko e Spodarets (2020), utilizaram redes neurais artificiais Convolucionais (CNNs) para classificar imagens de retinopatia diabética. Embora Gao et al. (2018) tenham alcançado uma acurácia de classificação de $88,72\%$, várias técnicas de pré-processamento foram implementadas para melhorar ainda mais a capacidade de classificação das redes neurais artificiais. A Figura 2.10 mostra um exemplo de pré-processamento.

Figura 2.10: Exemplo de pré-processamento.



Adaptado de Gao et al. (2018)

Essas técnicas de pré-processamento desempenham um papel crucial na preparação das imagens para a análise das CNNs. Elas podem incluir operações como *data augmentation*, normalização de intensidade, equalização de histograma e remoção de ruído. Essas etapas são essenciais para melhorar a qualidade das imagens e destacar características relevantes, o que pode resultar em melhorias significativas na precisão da classificação.

3 Aprendizado de máquina

O aprendizado de máquina é um campo que aprimora o desempenho de sistemas ao aprender com experiências através de métodos computacionais. Os computadores podem aprender com base nos seus erros anteriores (ZHOU, 2021). Conforme descrito por Goodfellow, Bengio e Courville (2016), o aprendizado pode ser definido como a habilidade de transformar a experiência em conhecimento.

Já o autor Géron (2019), apresenta duas citações para embasar a definição de aprendizado de máquina sendo uma mais geral, de que aprendizado de máquina é o campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados (SAMUEL, 1959), e uma definição orientada à engenharia onde um programa de computador aprende com a experiência e com relação a alguma tarefa e alguma medida de desempenho. Para avaliar o aprendizado é medido o desempenho e verificado se melhora com a experiência (MITCHELL, 1997).

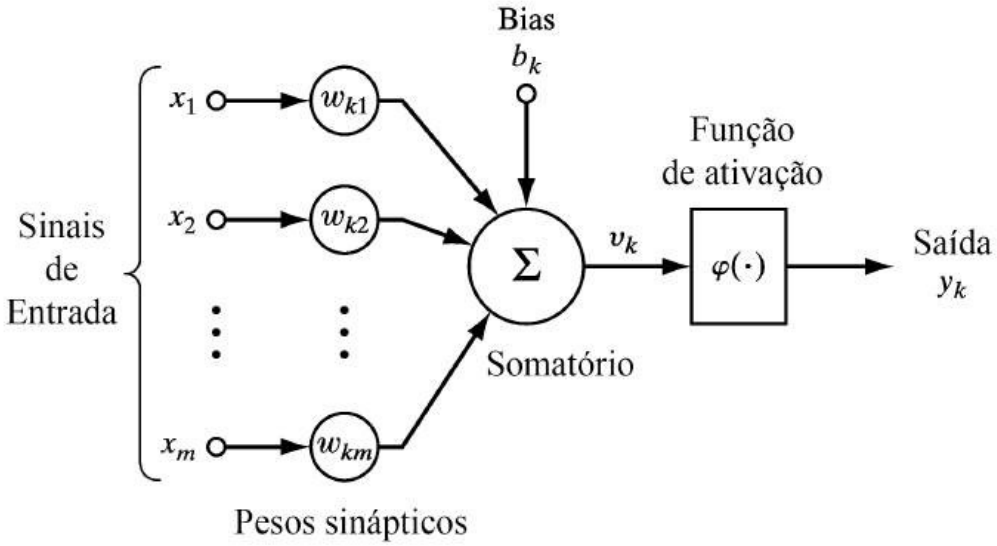
A experiência nos sistemas computacionais é representada pelos dados, e o principal objetivo do aprendizado de máquina é desenvolver algoritmos de aprendizado que construam modelos capazes de realizar aprendizado a partir destes dados. Esses modelos podem então fazer previsões sobre novas observações. Em resumo, se considerarmos a ciência da computação como o estudo de algoritmos, o aprendizado de máquina é o estudo de algoritmos de aprendizado (GéRON, 2019)

3.1 Redes Neurais artificiais

Entre as diversas abordagens de aprendizado de máquina, as redes neurais artificiais se destacam pela sua capacidade de modelar relações complexas nos dados. Essas redes foram inspiradas na estrutura biológica dos neurônios, cuja organização básica serviu como referência na construção desses modelos. O neurônio é caracterizado por possuir múltiplos dendritos, que recebem sinais de outros neurônios, e um axônio longo, responsável por transmitir sinais para outros neurônios ou células (HAYKIN, 2008). Essa organização foi crucial para o desenvolvimento dos primeiros modelos de redes neurais artificiais, que buscam replicar de forma simplificada o funcionamento do cérebro humano.

O neurônio artificial (Figura 3.1) é uma estrutura composta por diversos elementos, cada um desempenhando um papel específico e se assemelhando às partes de um neurônio biológico. Os

Figura 3.1: Diagrama em bloco de um neurônio artificial.



Adaptado de Haykin (2008)

sinais de entrada correspondem aos dendritos, que recebem informações de outros neurônios; os pesos sinápticos e o somatório desses sinais assemelham-se ao corpo celular, onde ocorre a integração e processamento das informações; a função de ativação é comparável ao axônio, responsável por transmitir o sinal processado; por fim, a saída ou saídas do neurônio artificial correspondem aos terminais sinápticos, que transmitem o sinal processado para outros neurônios ou células.

A representação diagramática da Figura 3.1 pode ser expressa matematicamente pela Equação 3.2, na qual cada elemento desempenha um papel específico. No contexto dessa equação, x_m representa o sinal da entrada m; w_{km} representa o peso sináptico da entrada m para o neurônio k; b_k é o viés de correção do neurônio k; v_k é o potencial de ativação do neurônio k; e y_k é a resposta do neurônio k após a aplicação da função de ativação. Assim, a equação que descreve essa rede neural pode ser escrita por,

$$v_k = \sum_{j=1}^{m}(x_j \cdot w_{kj}) + b_k \tag{3.1}$$

$$y_k = \varphi(\sum_{j=1}^{m}(x_j \cdot w_{kj}) + b_k) \tag{3.2}$$

As redes neurais artificiais (RNAs) são sistemas computacionais inspirados no funcionamento do cérebro humano. Elas são compostas por neurônios artificiais, unidades básicas de processamento, interconectadas em camadas. As RNs têm a capacidade de aprender com dados e identificar relações complexas entre as entradas e saídas.

Haykin (2008) definiu as redes neurais artificiais como sistemas computacionais que se baseiam em modelos biológicos de processamento neural para aprender e resolver problemas. Uma característica marcante das RNs é sua capacidade de aprendizado a partir de exemplos, o que lhes permite generalizar para novos casos de forma eficiente (WASSERMAN, 1989).

Existem dois métodos de aprendizado distintos, o supervisionado e o não supervisionado. Estes métodos são válidos para qualquer aprendizado de máquina, seja com redes neurais artificiais ou não. No aprendizado supervisionado, a rede neural é treinada com exemplos rotulados, permitindo que ela aprenda a associar entradas a saídas conhecidas (HAYKIN, 2008). Já no aprendizado não supervisionado, a rede busca identificar padrões nos dados de entrada sem depender de rótulos, explorando estruturas ou agrupamentos significativos nos dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

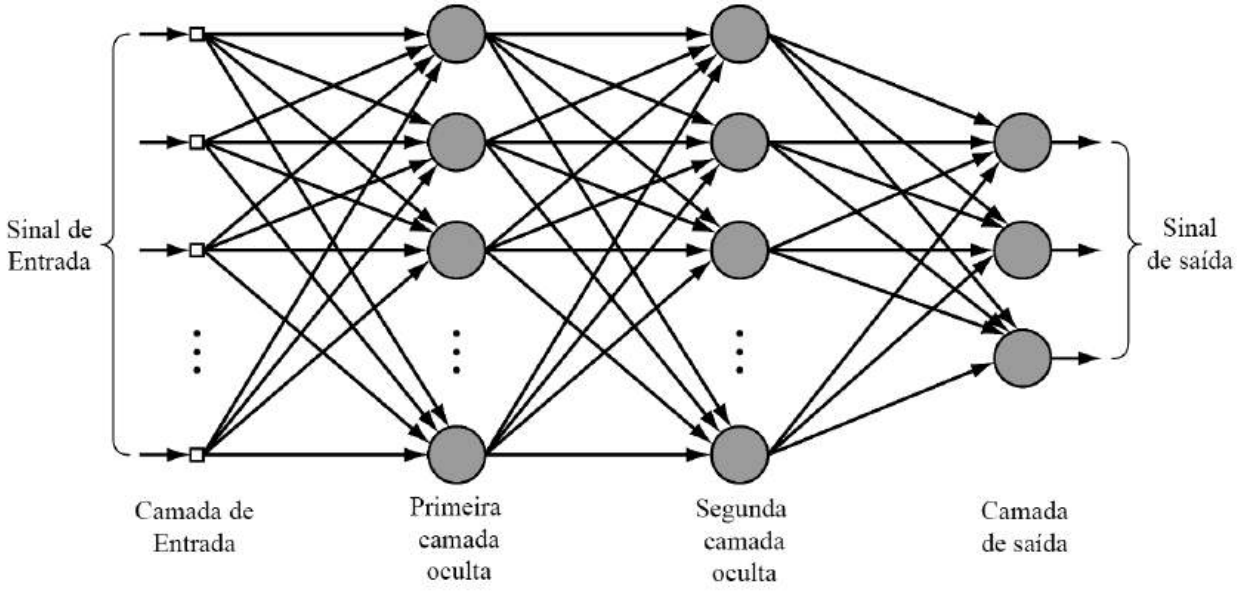
Além desses métodos, também é relevante mencionar o aprendizado por reforço, no qual a rede aprende interagindo com um ambiente, recebendo recompensas ou penalidades com base em suas ações, e ajustando suas estratégias para maximizar as recompensas ao longo do tempo (SUTTON; BARTO, 2018).

Estes diferentes paradigmas de aprendizado são fundamentais para a construção de arquiteturas de redes neurais artificiais eficientes, como a Rede Neural Perceptron Multicamada (MLP, do inglês *Multilayer Perceptron*), ilustrada na Figura 3.2. Esta estrutura típica é composta por uma camada de entrada, duas camadas ocultas e uma camada de saída. Na camada de entrada, o número de neurônios é determinado pelo número de características de entrada do sinal. Por sua vez, o número de camadas ocultas e o número de neurônios em cada camada oculta são determinados manualmente pelo programador quando o mesmo não importa arquiteturas pré-fabricadas, e têm um impacto significativo na capacidade da rede de aprender padrões complexos nos dados.

A escolha adequada desses parâmetros é crucial, pois uma rede com muitas camadas ou neurônios pode levar ao *overfitting*, fenômeno que ocorre em modelos de aprendizado de máquina quando um modelo aprende muito bem os detalhes e ruídos dos dados de treinamento, a ponto de afetar ne-

gativamente seu desempenho em novos dados, enquanto uma rede muito simples pode não capturar a complexidade dos dados. Além disso, o número de neurônios na camada de saída define a forma como a rede responde ao problema, sendo geralmente igual ao número de classes em problemas de classificação ou 1 em problemas de regressão. A arquitetura de uma MLP, portanto, é um aspecto fundamental a ser considerado no desenvolvimento de modelos de aprendizado de máquina eficazes.

Figura 3.2: Exemplo de uma MLP (MultiLayer Perceptron com 2 camadas ocultas).



Adaptado de: Haykin (2008)

As redes neurais artificiais enfrentam vários desafios que limitam sua aplicabilidade e eficácia em diferentes contextos. Um dos desafios mais significativos é a interpretabilidade, ou seja, a capacidade de compreender e explicar o funcionamento interno da rede e as decisões que ela toma (LIPTON, 2018). redes neurais artificiais de multi-camadas (MLPs), em particular, são frequentemente consideradas caixas-pretas devido à complexidade de suas operações e o nível de abstração das camadas ocultas, tornando difícil para os humanos entenderem como e por que uma decisão foi tomada.

Outro desafio importante é a necessidade de grandes conjuntos de dados para treinar redes neurais artificiais de forma eficaz. Redes mais complexas e profundas requerem quantidades substanciais de dados para aprender padrões significativos e evitar o *overfitting*. Isso pode ser um obstáculo

em áreas onde os dados são escassos ou difíceis de obter (LIPTON, 2018).

Além disso, as redes neurais artificiais também estão sujeitas ao viés algorítmico, que se refere à tendência de um algoritmo de aprendizado de máquina para favorecer certas classes ou resultados em detrimento de outros, com base nas características dos dados de treinamento (LIPTON, 2018). Isso pode resultar em discriminação injusta ou decisões tendenciosas, especialmente em sistemas de aprendizado automatizado utilizados em contextos críticos, como na saúde ou na justiça.

3.1.1 Redes Neurais Artificiais Convolucionais (CNNs)

As redes neurais artificiais Convolucionais (CNNs) têm sido amplamente adotadas em uma variedade de tarefas de visão computacional devido à sua capacidade de capturar e processar informações espaciais em imagens (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Ao contrário das redes neurais artificiais tradicionais, que exigem que os dados de entrada sejam transformados em um formato unidimensional, as CNNs preservam a estrutura espacial dos dados, o que as torna ideais para lidar com imagens e outros tipos de dados multidimensionais (CHOLLET, 2017).

Isso é alcançado por meio de camadas convolucionais, que aplicam filtros espaciais às entradas para extrair características relevantes em diferentes localizações da imagem. Essas características são então agrupadas e processadas por camadas subsequentes, como camadas de *pooling* (As camadas de pooling reduzem a dimensionalidade dos mapas de ativação, preservando características essenciais e melhorando eficiência), e camadas totalmente conectadas, para realizar tarefas específicas, como classificação, detecção de objetos e segmentação de imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

Essa capacidade de extrair características em qualquer região das imagens é fundamental (ALBAWI; MOHAMMED; AL-ZAWI, 2017), especialmente em exames como a fundoscopia, onde as características relevantes para o diagnóstico podem estar distribuídas por toda a imagem (SCANLON; SALLAM; WIJNGAARDEN, 2017). Na retinopatia diabética, por exemplo, os exsudatos podem aparecer em diferentes partes da retina, enquanto os microaneurismas podem estar localizados em regiões específicas (SCANLON; SALLAM; WIJNGAARDEN, 2017). Portanto, a capacidade da arquitetura de rede neural em detectar esses elementos de forma eficaz e precisa é crucial para um diagnóstico preciso e rápido da condição.

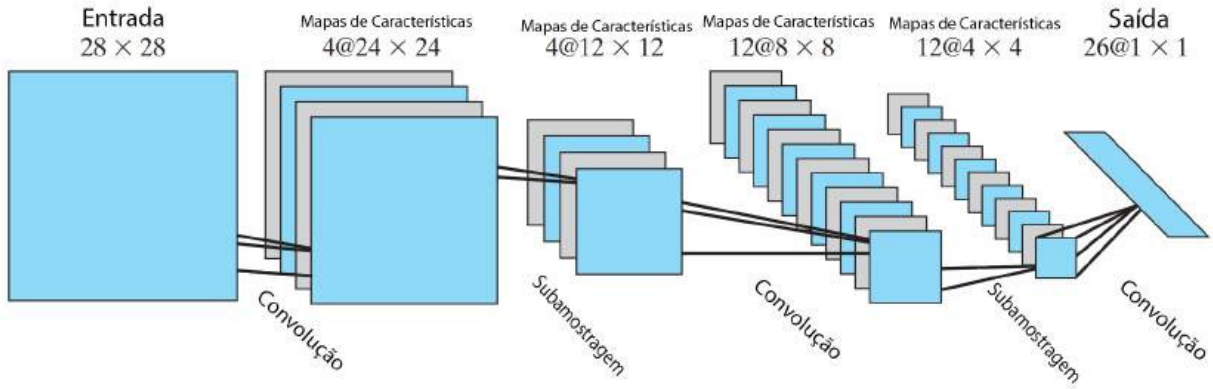
As redes neurais artificiais convolucionais possuem uma capacidade de processar vastas quanti-

dades de dados de forma eficiente, o que as torna ideais para uma variedade de aplicações, incluindo classificação de imagens e processamento de linguagem natural (JOSHI et al., 2021). Essa eficácia se deve à sua capacidade de aprender padrões complexos em dados de entrada, tornando-as especialmente adequadas para lidar com a natureza rica e variada de imagens e texto (JOSHI et al., 2021). Segundo descreve o autor Krizhevsky, Sutskever e Hinton (2012), o poder das CNNs reside na combinação de três fatores: sua capacidade de aprender representações hierárquicas, sua eficiência computacional e sua robustez a variações na entrada.

Como resultado, as CNNs têm se destacado em diversas tarefas, como reconhecimento de objetos em imagens. Essa versatilidade e capacidade de lidar com dados complexos tornam as CNN uma ferramenta poderosa e cada vez mais indispensável em áreas que demandam o processamento eficiente de informações visuais (SIMONYAN; ZISSERMAN, 2014).

A Figura 3.3 apresenta a arquitetura de uma rede convolucional composta por uma camada de entrada para imagens de 28×28 pixels, quatro camadas ocultas e uma camada de saída.

Figura 3.3: Rede neural convolucional para processamento de imagens.



Adaptado de Haykin (2008)

A primeira camada oculta executa a operação de convolução, com quatro mapas de características, cada um composto por 24×24 neurônios.

A segunda camada oculta realiza pooling utilizando a média local, com quatro mapas de características, mas agora com 12×12 neurônios em cada mapa.

A terceira camada oculta realiza uma segunda convolução, com 12 mapas de características de 8×8 neurônios cada, com possíveis conexões sinápticas de vários mapas da camada anterior.

A quarta camada oculta realiza uma segunda pooling e média local, com 12 mapas de características, cada um com 4×4 neurônios.

Por fim, a camada de saída realiza uma etapa final de convolução, com 26 neurônios, cada um associado a um dos 26 caracteres possíveis, utilizando um campo receptivo de 4×4.

Assim, ao intercalar camadas computacionais entre convolução e subamostragem, observamos um efeito "bipiramidal" (HAYKIN, 2008). Isso significa que, em cada camada de convolução ou subamostragem, há um aumento no número de mapas de características, ao passo que a resolução espacial é reduzida em relação à camada anterior correspondente. A abordagem de convolução seguida por subamostragem é fundamentada na ideia de células "simples"precedendo células "complexas" (HAYKIN, 2008).

3.1.2 Arquiteturas

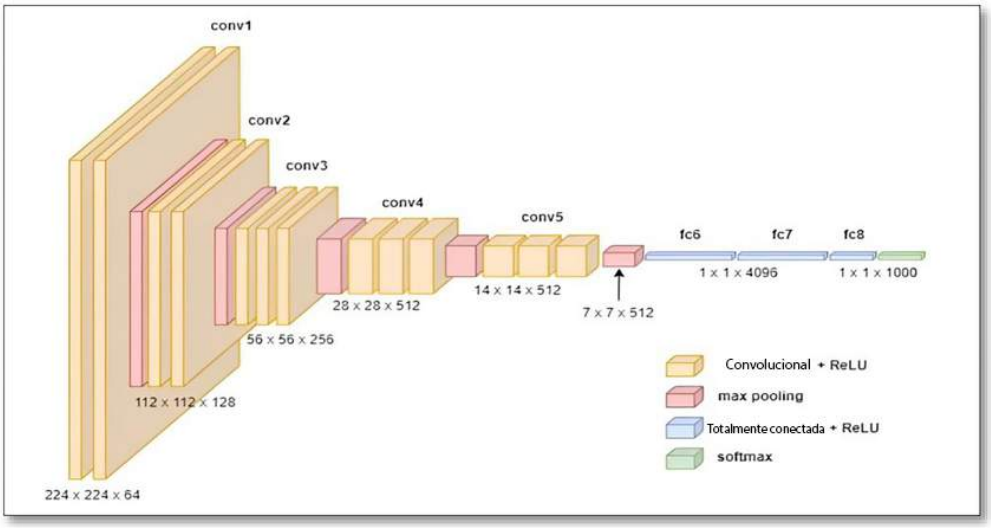
Neste trabalho, foram utilizadas algumas arquiteturas de redes neurais artificiais já conhecidas, como a VGGNet (especificamente a VGG16) (WAN; LIANG; ZHANG, 2018), a Inception_V3 (GAO et al., 2018) e a Xception (QUMMAR et al., 2019). A estrutura, as camadas e o funcionamento dessas redes serão detalhados ao longo desta secção.

Os aprimoramentos mencionados estendem significativamente a capacidade de análise e interpretação em uma vasta gama de aplicações de visão computacional, evidenciando a crescente necessidade de extrair características visuais com elevado nível de detalhamento e precisão (SIMONYAN; ZISSERMAN, 2014). A habilidade para discernir e caracterizar nuances visuais complexas se torna um diferencial crítico, impulsionando avanços em áreas que vão desde o reconhecimento facial até a análise de imagens médicas. Tais arquiteturas de redes convolucionais são rigorosamente testadas e validadas em cenários de classificação, com a base de dados *ImageNet* emergindo como um campo de prova fundamental. Esta base de dados, reconhecida por sua diversidade e volume, tem servido como um marco para o aperfeiçoamento e avaliação de várias gerações de sistemas de classificação em larga escala, estabelecendo padrões de desempenho que impulsionam a inovação contínua no campo da visão computacional (SIMONYAN; ZISSERMAN, 2014).

3.1.3 VGG16

Para introduzir o conceito da arquitetura *VGG16*, é essencial compreender a importância da profundidade e simplicidade nas redes neurais artificiais convolucionais, especialmente no que se refere ao processo de extração de características em imagens. A *VGG16*, proposta por Simonyan e Zisserman (SIMONYAN; ZISSERMAN, 2014), tornou-se um marco no desenvolvimento de redes convolucionais profundas, demonstrando que o aumento do número de camadas pode resultar em melhorias significativas no desempenho de tarefas de classificação de imagens. Sua arquitetura pode ser observada na Figura 3.4.

Figura 3.4: Arquitetura VGG16.



Adaptado de Singh, Dwivedi e Rastogi (2024)

A *VGG16* é caracterizada por sua arquitetura profunda, composta por 16 camadas de convolução e 3 camadas totalmente conectadas. A principal contribuição da *VGG16* é a utilização de pequenos filtros 3×3, que permitem a extração de características complexas ao longo de várias camadas sucessivas.

Além disso, a *VGG16* emprega camadas de *max-pooling* após blocos de camadas convolucionais para reduzir a dimensionalidade das características extraídas, mantendo as informações mais relevantes e eliminando redundâncias. Isso contribui para a robustez da rede e facilita o processo de generalização em tarefas de visão computacional. Conforme destacado por Simonyan e Zisserman (2014), essa abordagem de utilizar múltiplas camadas com filtros pequenos e camadas de pooling

resulta em uma rede que, apesar de ser profunda, é eficiente em termos de parâmetros e possui um desempenho superior em comparação a redes com camadas convolucionais maiores e menos profundas.

Em termos de aplicação, a *VGG16* foi amplamente utilizada em competições de reconhecimento de imagem, como o *ImageNet*, onde apresentou resultados notáveis para a classificação em 1.000 classes. Seu sucesso demonstrou que a profundidade da rede, combinada com o uso estratégico de pequenos filtros e pooling, pode levar a uma extração de características mais precisa, favorecendo a classificação correta de objetos em imagens complexas (SIMONYAN; ZISSERMAN, 2014). Assim, a *VGG16* consolidou-se como uma das arquiteturas de referência no campo de visão computacional, influenciando o desenvolvimento de redes subsequentes.

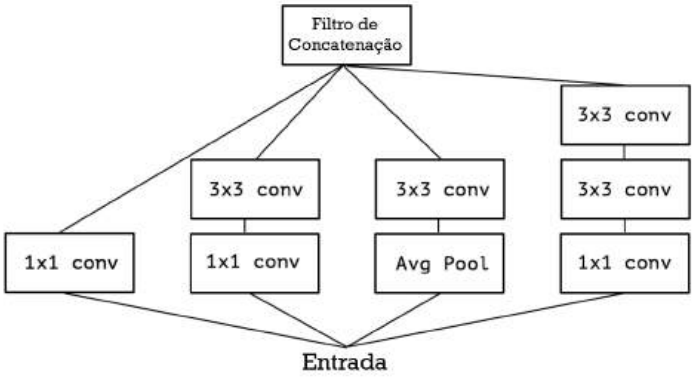
3.1.4 InceptionV3

Para introduzir o conceito de arquitetura *Inception* e sua influência na arquitetura subsequente *Xception*, é importante destacar o papel das camadas de convolução nesse contexto. Essas camadas são responsáveis por aprender filtros em um espaço tridimensional, composto por duas dimensões espaciais (largura e altura) e uma dimensão de canais de cor, geralmente seguindo o padrão RGB. Após a primeira camada, as arquiteturas tornam-se difíceis de realizar uma interpretação comum dado que as saídas dos filtros se tornam mapas de características. O conceito central da *Inception* é simplificar e otimizar o processo de aprendizado, dividindo-o explicitamente em uma sequência de operações que examinam separadamente as correlações entre canais e as correlações espaciais (CHOLLET, 2017).

Como descreve o autor Chollet (2017), de forma mais específica, o módulo *Inception* padrão começa examinando as relações entre os canais usando convoluções 1×1. Isso leva a uma redução dos dados de entrada para 3 ou 4 conjuntos menores, cada um representando um subespaço do espaço original. Em seguida, são aplicadas convoluções regulares 3×3 ou 5×5 para analisar todas as correlações dentro desses espaços 3D reduzidos. Uma ilustração desta definição pode ser vista na Figura 3.5.

A premissa fundamental subjacente ao conceito do *Inception* é que as correlações entre canais e as correlações espaciais são independentes o suficiente para que seja mais vantajoso não processá-las em conjunto. Essa abordagem baseia-se na ideia de que é mais eficaz e eficiente tratar essas duas

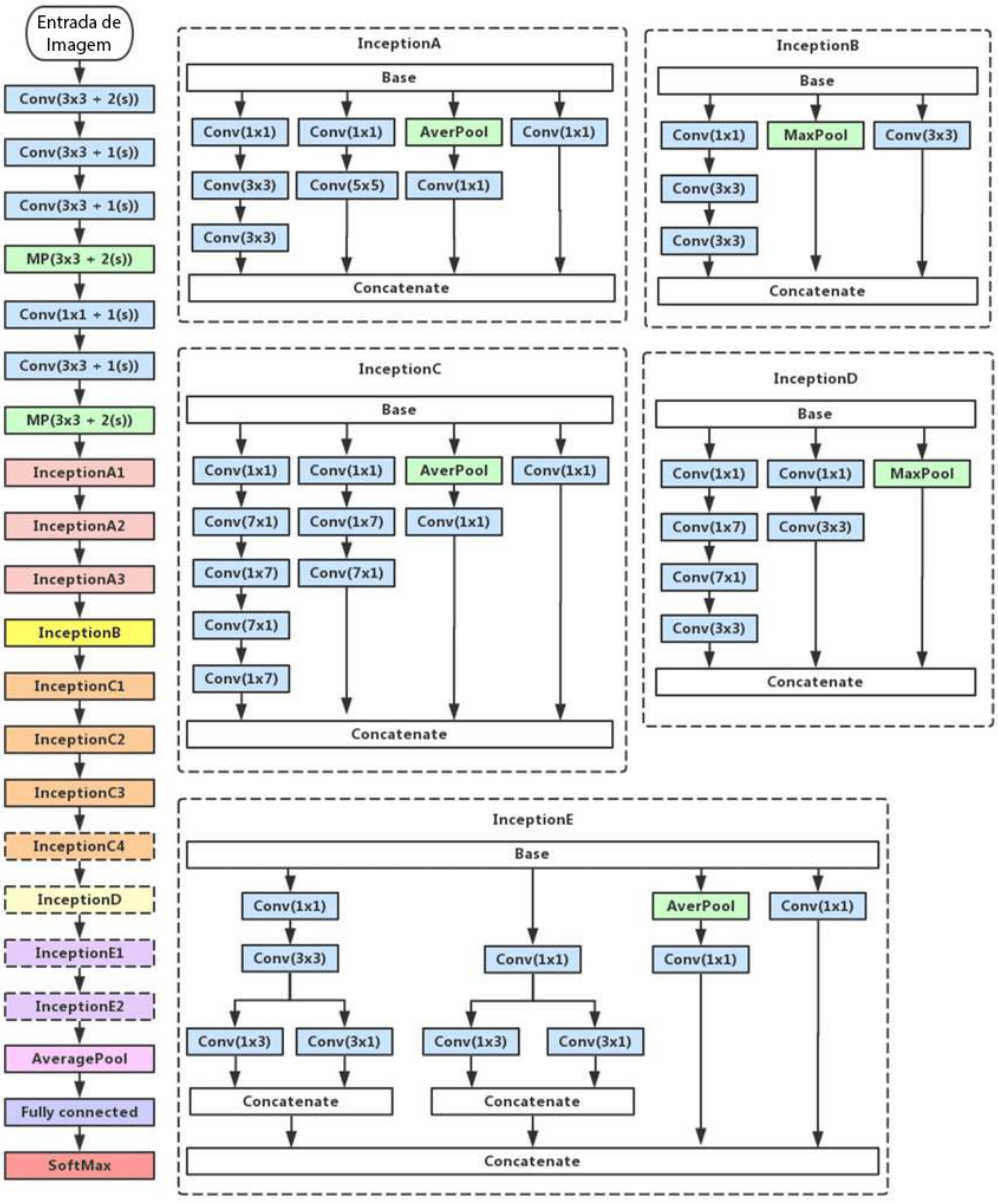
Figura 3.5: Módulo Inception Canônico.



Adaptado de Chollet (2017)

formas de correlação separadamente, permitindo que o modelo aprenda e capture informações específicas de cada tipo de correlação de forma mais precisa e eficaz (CHOLLET, 2017). A Figura 3.6 mostra a arquitetura da InceptionV3.

Figura 3.6: Arquitetura completa da InceptionV3.



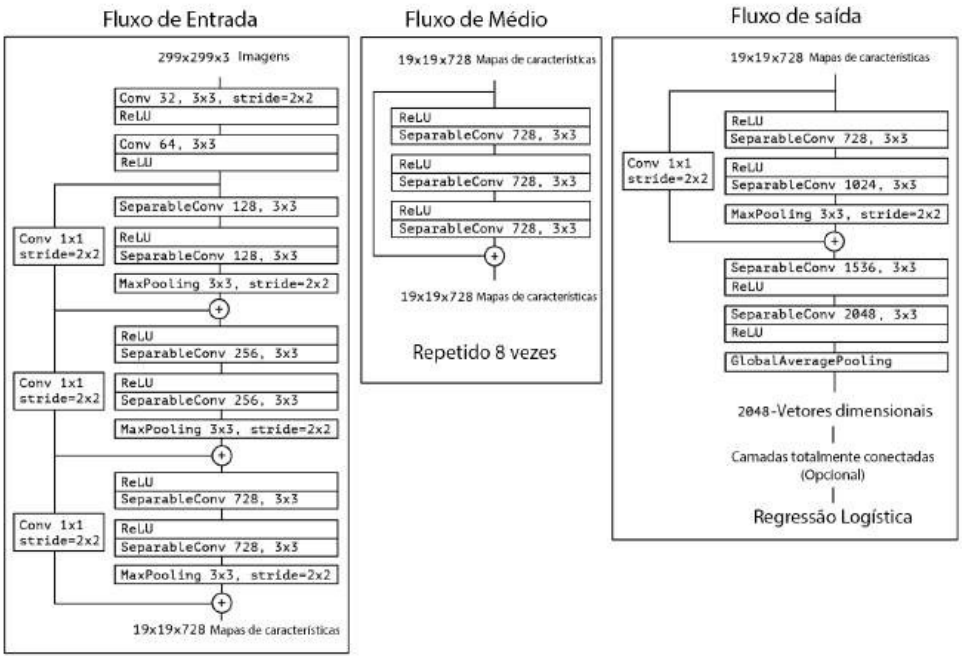
Adaptado de Ji et al. (2019)

3.1.5 Xception

Já a arquitetura *Xception* (que significa *Extreme Inception*) foi proposta por Chollet (2017) e se baseia inteiramente em camadas de convolução separáveis em profundidade. A hipótese apresentada era de que as correlações entre os canais de cor e as correlações entre as dimensões espaciais nos mapas de características das redes neurais artificiais convolucionais podem ser totalmente desacopladas (CHOLLET, 2017).

A arquitetura *Xception* possui 36 camadas convolucionais, que são a estrutura responsável pela extração das características da imagem feita pela rede. Chollet (2017) resume sua arquitetura como uma pilha linear de camadas de convolução separáveis em profundidade com conexões residuais.

Figura 3.7: A arquitetura da rede Xception proposta por Chollet (2017).



Adaptado de Chollet (2017)

3.1.6 MobileNet

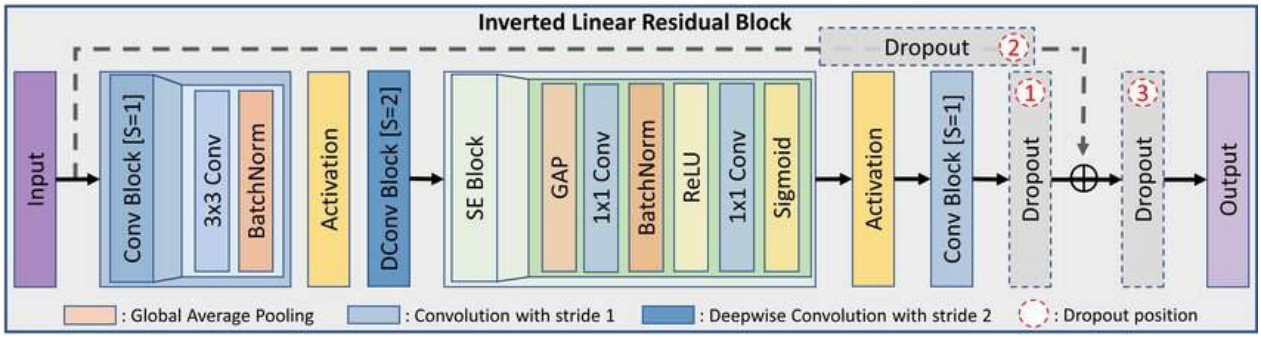
A arquitetura MobileNet foi introduzida em 2017 por Howard (2017), como uma abordagem eficiente para redes neurais profundas projetadas para dispositivos móveis e aplicações com recursos limitados. Diferentemente das redes convolucionais tradicionais, a MobileNet emprega convoluções separáveis em profundidade como uma técnica principal para reduzir significativamente a complexidade computacional, mantendo uma boa capacidade de aprendizado. Essa inovação torna a MobileNet adequada para tarefas de visão computacional em dispositivos embarcados e com baixo consumo de energia, como smartphones e câmeras IoT (HOWARD, 2017).

O núcleo da arquitetura está na separação entre a convolução espacial (que aprende padrões em posições da imagem) e a convolução por canal (que aprende as características dentro de cada canal). Isso substitui as convoluções padrão por operações mais leves e eficientes, reduzindo tanto

o número de parâmetros quanto o custo computacional. Por exemplo, em vez de aplicar um filtro tridimensional a todos os canais simultaneamente, a convolução *depthwise* processa cada canal de forma independente, enquanto a convolução *pointwise* combina esses resultados em um único canal de saída usando filtros 1×1. Essa abordagem resulta em um modelo mais compacto, ideal para inferência em tempo real em hardware limitado (HOWARD, 2017), (SANDLER et al., 2018).

A estrutura da MobileNet pode ser visualizada na Figura 3.8, que ilustra como as camadas de convolução separáveis em profundidade são organizadas em blocos modulares. Cada bloco é composto por uma convolução *depthwise*, seguida de uma convolução *pointwise*, frequentemente combinada com uma ativação ReLU e uma normalização por lotes (*BatchNorm*). Essa organização é replicada ao longo da rede, criando uma arquitetura simples, mas poderosa, capaz de capturar padrões hierárquicos em imagens com uma computação reduzida.

Figura 3.8: Arquitetura MobileNet.



Fonte: Zhu et al. (2024)

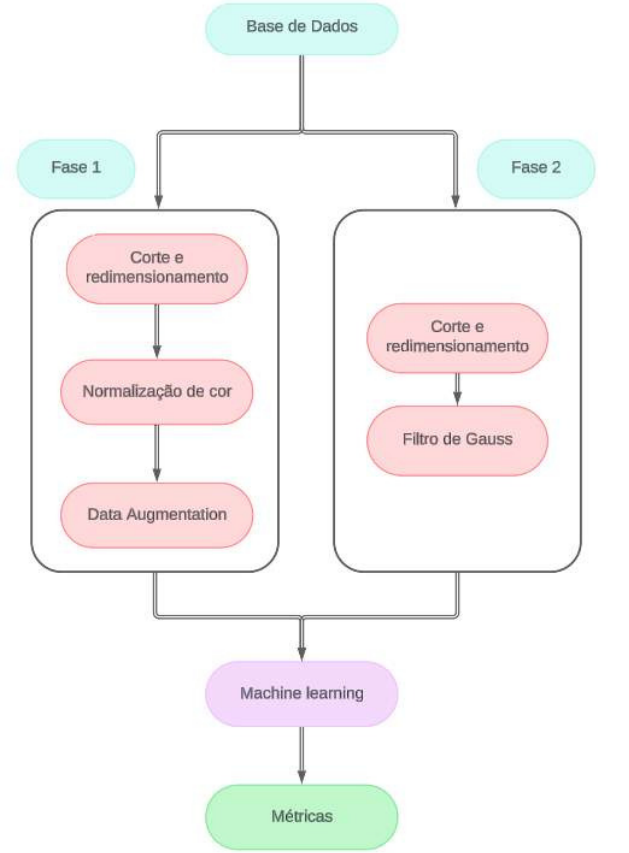
Com a publicação do artigo original, a MobileNet inspirou várias iterações, incluindo a MobileNetV2 e V3, que aprimoraram a eficiência e o desempenho introduzindo blocos invertidos residuais e otimizando o uso de ativadores não lineares. Essas melhorias foram motivadas pela necessidade de atender demandas crescentes em aplicações como reconhecimento facial, realidade aumentada e visão computacional em larga escala. Essa evolução contínua reforça a relevância das MobileNets no campo da inteligência artificial aplicada a dispositivos de baixa potência (HOWARD, 2017), (SANDLER et al., 2018).

4 Metodologia

Para o desenvolvimento deste trabalho, a metodologia foi estruturada em quatro categorias principais: base de dados, pré-processamento, aprendizado de máquina e métricas, conforme ilustrado na Figura 4.1. O processo foi realizado em duas fases, diferenciadas pelo pré-processamento aplicado às imagens.

Na primeira fase, foram utilizadas técnicas básicas de normalização, redimensionamento e *data augmentation*, enquanto na segunda fase, métodos mais avançados foram implementados, mas removendo o *data augmentation*. Essa abordagem permitiu uma análise comparativa do impacto do pré-processamento no desempenho das redes neurais avaliadas.

Figura 4.1: Diagrama de blocos usado para as arquiteturas apresentadas.



Fonte: Autoria própria

Na nomenclatura adotada neste trabalho, o sufixo "_gauss"indica que o filtro gaussiano foi aplicado às imagens durante o pré-processamento, a etapa de corte e redimensionamento não foi remo-

vida em nenhuma execução. Essa técnica foi utilizada exclusivamente na Fase 2, com o objetivo de suavizar ruídos e destacar estruturas relevantes, como vasos sanguíneos e lesões características da retinopatia diabética. O filtro foi parametrizado para garantir que as características essenciais das imagens fossem preservadas, maximizando sua utilidade para o treinamento e a validação dos modelos de redes neurais convolucionais.

4.1 Base de dados

Os conjuntos de dados utilizados foram do Desafio de Detecção de Retinopatia Diabética do Kaggle de 2015 (EyesPACS2015) (DUGAS et al., 2015) e a base APTOS2019 (SOCIETY, 2019), utilizado para treinar redes neurais artificiais convolucionais.

4.1.1 EyesPACS2015

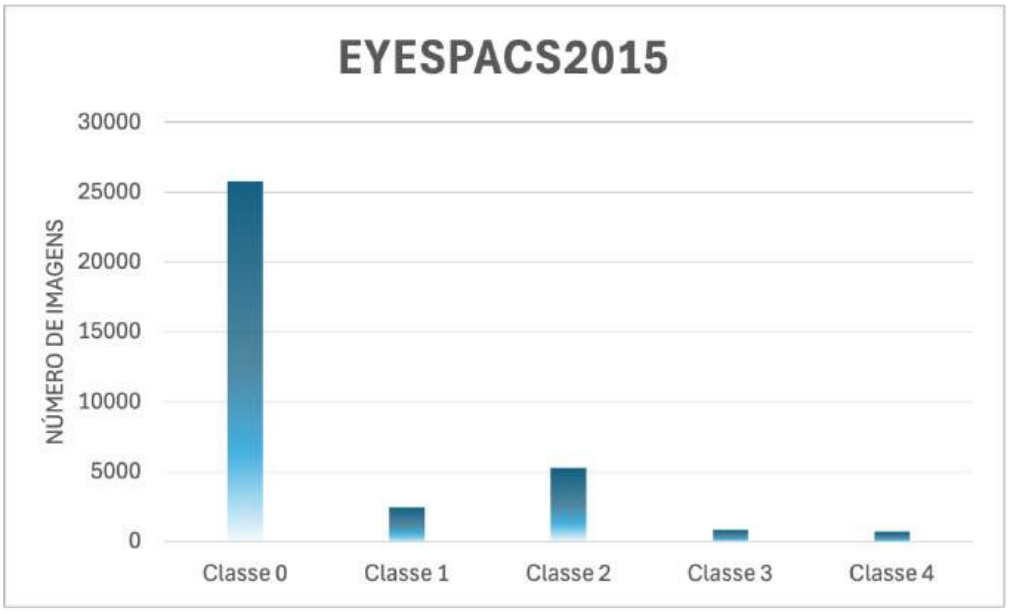
Acerca do EyesPACS2015, o conjunto de dados consiste em 35.126 imagens pré-classificadas em 5 níveis da doença. As imagens são de exames de fundoscopia dos olhos esquerdo e direito de indivíduos americanos. Os dados estão rotulados da seguinte forma:

- Nível 0: Sem retinopatia diabética;
- Nível 1: Retinopatia diabética leve;
- Nível 2: Retinopatia diabética moderada;
- Nível 3: Retinopatia diabética grave;
- Nível 4: Retinopatia diabética proliferativa.

Na Figura 4.2 é apresentada a distribuição das imagens em cada classe no conjunto de dados do desafio do Kaggle (DUGAS et al., 2015). Pode-se observar que a Classe 0 representa 73,47% de todas as imagens no conjunto de dados. Este desequilíbrio de classes é relevante, especialmente se os dados de treinamento não forem passados para a rede com os pesos apropriados calculados para cada classe.

Assim, optamos por limitar o conjunto de dados utilizado pelo número de imagens na classe com a menor quantidade de dados. A classe escolhida foi a Classe 4, que contém um total de

Figura 4.2: Distribuição das classes no conjunto de dados do Desafio Kaggle 2015.



Fonte: Autoria Própria

708 imagens. Também foi realizada uma divisão de 70% das imagens para treinamento, 15% para validação e 15% para teste. Portanto, com 5 classes, o conjunto de treinamento consistiu em 2478 imagens, e os conjuntos de validação e teste consistiram cada um em 531 imagens. O mesmo ocorreu para as combinações em 2 classes, às quais ficaram limitadas em 708 por classe, sendo um total de 1416 imagens para a combinação 0-4.

O conjunto de dados foi dividido em várias combinações, cada uma descrita pelos números de suas classes separados por hífens. Por exemplo, "0-4" representa uma classificação binária utilizando imagens das classes 0 e 4, enquanto "0-1-2-3-4"representa uma classificação em 5 categorias utilizando imagens das classes 0, 1, 2, 3 e 4.

4.1.2 APTOS2019

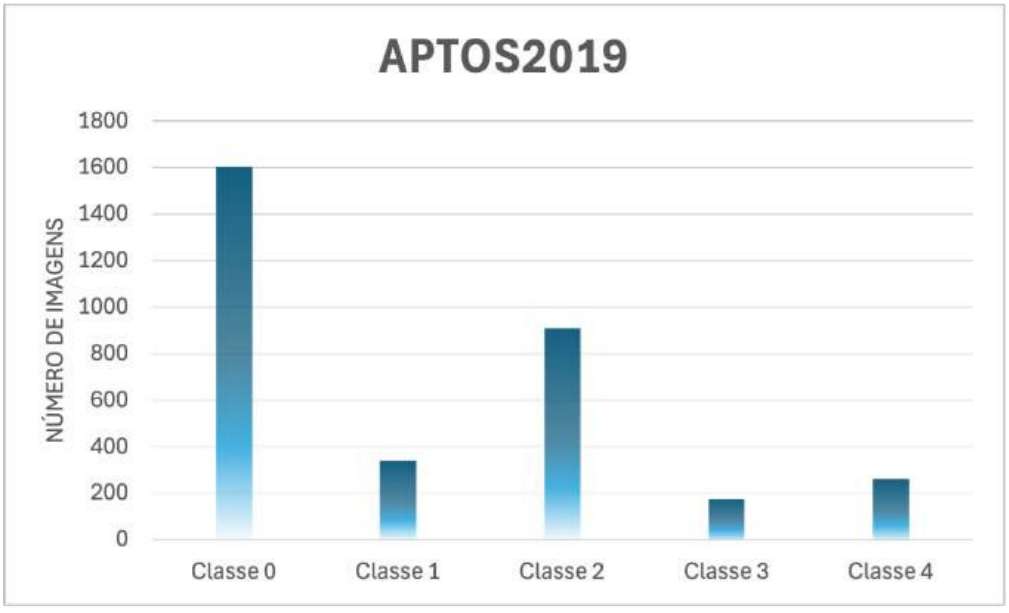
A base de dados APTOS2019 *Blindness Detection* (SOCIETY, 2019) possui um total de 3.296 imagens de retina capturadas em diferentes condições de iluminação e com equipamentos variados, refletindo a diversidade das práticas clínicas. As imagens são classificadas em cinco categorias de severidade assim como no *dataset* EyesPACS2015, facilitando estudos sobre detecção automática da patologia.

O APTOS2019, ao contrário do EyesPACS2015, não foi subdividido em classificações binárias. Todos os testes realizados foram efetuados usando a classificação em 5 classes, por possuir um número menor de imagens quando comparado ao EyesPACS2015. Dessa forma, optou-se por utilizar todas as imagens, de maneira desbalanceada, e efetuar a correção através dos pesos devidamente calculados entregues à rede.

Este banco de dados foi disponibilizado como parte de uma competição no Kaggle, organizada pela Asia Pacific Tele-Ophthalmology Society (APTOS), com o objetivo de promover o desenvolvimento de soluções baseadas em aprendizado de máquina para diagnóstico precoce da doença. As imagens apresentam resolução mínima de 640×480 pixels e são frequentemente utilizadas em pesquisas por sua alta qualidade e relevância clínica.

Além disso, as bases incluem metadados importantes, como rótulos associados a diagnósticos, permitindo não apenas a classificação da doença, mas também a análise do desempenho de diferentes modelos de redes neurais artificiais em tarefas de detecção e classificação. A Figura 4.3 mostra a distribuição do número de imagens por classe do banco de dados.

Figura 4.3: Distribuição das classes no conjunto de dados APTOS2019.



Fonte: Autoria Própria

4.2 Pré-processamento

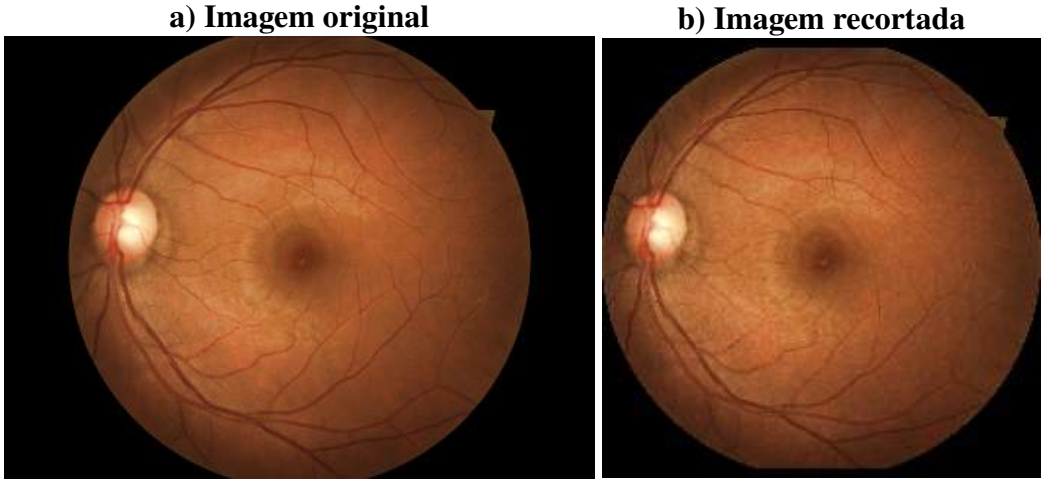
As técnicas de pré-processamento utilizadas foram selecionadas de acordo com sua aplicação nos problemas de classificação da retinopatia diabética encontrados na literatura, sendo elas de classificação em 5 classes e classificação binária.

4.2.1 Corte e redimensionamento

Na Fase 1, as imagens foram cortadas para remover as bordas pretas e centralizar a retina. Um algoritmo foi desenvolvido para centralizar a retina e eliminar as bordas pretas com base em um valor de limiar. Esse procedimento redimensionou as imagens para 224×224. Esta resolução foi observada nos trabalhos citados, resultando na menor perda possível de qualidade na fundoscopia devido ao redimensionamento. As imagens originais têm uma resolução superior a 3840×2160 pixels, o que tornaria o treinamento inviável devido às nossas limitações de hardware.

O algoritmo funcionou detectando as bordas da retina e cortando a imagem o mais próximo possível destas bordas, mantendo uma proporção de aspecto 1:1. Essa medida foi adotada para preservar o máximo possível de qualidade da parte útil da imagem (região da presença da retina no exame). O resultado deste corte pode ser visto na Figura 4.4.

Figura 4.4: **a)** Imagem original 13_left.png; **b)** Exemplo de recorte e redimensionamento com a figura 13_left.png.



Fonte: Nagpal et al. (2022)

O algoritmo calcula a soma dos valores de pixel ao longo das linhas e colunas da imagem, armazenando esses valores em dois vetores. O vetor X, com 224 posições, contém a soma dos

valores de cada coluna correspondente, enquanto cada posição das 224 do vetor Y armazena a soma dos valores de cada linha da imagem. Esses vetores foram então normalizados pelo maior valor encontrado para garantir consistência na comparação.

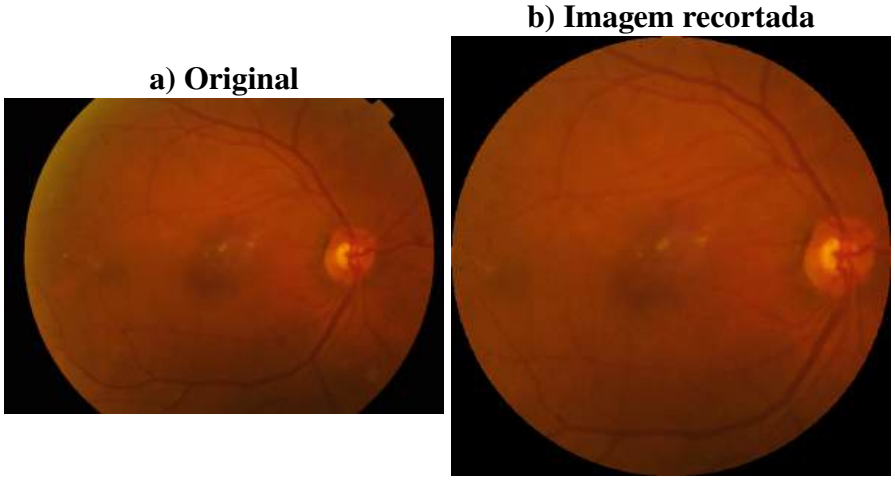
A seguir, as bordas dos vetores normalizados foram comparadas com um valor de limiar de 10%, obtido com base em testes, de maneira empírica. Esse limiar foi utilizado para determinar os pontos em que a soma dos valores de pixel se tornou igual ou superior ao limiar, indicando a transição da área preta da imagem para a área útil.

Embora o método da Fase 1 do corte e redimensionamento tenha sido eficaz para centralizar a retina e remover as bordas pretas, foi identificado que ajustes adicionais poderiam refinar ainda mais a representatividade e a qualidade das imagens resultantes. Com base nesta avaliação, um novo método foi implementado, focando em otimizar a delimitação da área útil e melhorar a preservação de detalhes relevantes na fundoscopia. Esse aprimoramento visou alinhar de forma mais precisa as imagens ao formato necessário para o treinamento, reduzindo possíveis perdas de informações críticas e maximizando o aproveitamento do conjunto de dados.

Na Fase 2 foi realizada uma operação de recorte para remover as bordas escuras e isolar a região central da retina. Esse procedimento foi otimizado com o uso de um limiar adaptativo, que garantiu uma maior precisão na identificação da área de interesse, mantendo as características centrais da imagem.

A implementação de uma abordagem circular também foi adotada para melhorar a captura da área relevante da retina, utilizando um corte centrado no meio da imagem e ajustando o raio conforme as dimensões da imagem, o que ajudou a reduzir a área de fundo e focar nos dados mais importantes (XU et al., 2024). A Figura 4.5 mostra o resultado da aplicação do processo de recorte realizado, centralizando a imagem e padronizando as bordas de forma arredondada.

Figura 4.5: **a)** Imagem original 531b39880c32.png; **b)** Imagem 531b39880c32.png após o processo de recorte.



Fonte: Autoria Própria

4.3 Normalização de cor

Uma normalização de cor também foi realizada nas imagens durante a Fase 1, uma vez que diferentes dispositivos de imagem podem introduzir variações em características como brilho, contraste e temperatura de cor (GAO et al., 2018). O processo de normalização consistiu em ajustar cada canal de cor das imagens para aproximá-los de um valor médio. Esse valor médio foi determinado com base em uma seleção aleatória de 5000 imagens do conjunto de dados. A equação utilizada pode ser vista na Eq. 4.1.

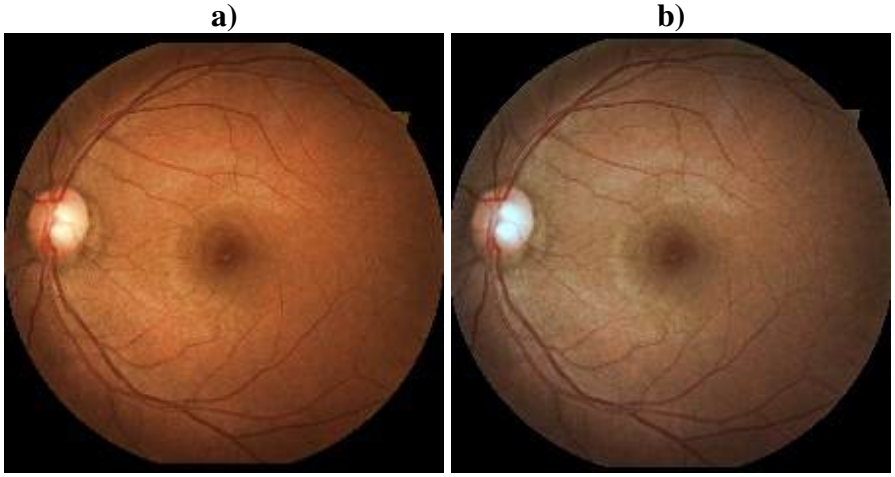
$$\begin{aligned} R_n &= min\left(\frac{R_n}{mean(R)} \cdot r^*, 255\right) \\ G_n &= min\left(\frac{G_n}{mean(G)} \cdot g^*, 255\right) \\ B_n &= min\left(\frac{B_n}{mean(B)} \cdot b^*, 255\right) \end{aligned} \tag{4.1}$$

Os valores R_n, G_n e B_n representam os componentes de cor do pixel n na imagem original. R, G e B representam os canais de cor da imagem original, dos quais é calculada a média. r^*, g^* e b^* são as médias obtidas a partir dos canais das 5000 imagens utilizadas.

Esta normalização foi realizada para melhorar a consistência na iluminação e tornar o conjunto

de dados mais uniforme. Este procedimento também pode reduzir o ruído gerado por artefatos relacionados ao instrumento de captura ou variações na luz, além de facilitar a comparação entre imagens, o que pode interferir no treinamento do modelo. A Figura 4.6 mostra um exemplo de normalização de cores da imagem 13_left.png pela média.

Figura 4.6: **a)** Imagem 13_left.png recortada e redimensionada; **b)** Normalização de cores da imagem 13_left.png.



Fonte: Autoria Própria

Embora esta normalização tenha sido um passo importante no pré-processamento das imagens, garantindo maior uniformidade e facilitando a comparação entre elas, a técnica não foi o suficiente para atingir os resultados desejados. A normalização ajudou a reduzir as variações de iluminação e minimizar alguns artefatos causados pelos dispositivos de captura, mas não resolveu completamente as limitações de representação da área útil da retina e da preservação de detalhes essenciais nas imagens.

Assim, a metodologia foi revisada e uma nova abordagem de correção de cor foi implementada na Fase 2, com foco em técnicas mais avançadas de melhoria na qualidade das imagens para o treinamento do modelo (XU et al., 2024).

Na etapa de pré-processamento, foi utilizada uma técnica para ajustar o contraste e o brilho das imagens baseada na combinação linear entre a imagem original e uma versão suavizada por um filtro gaussiano. O método combina os valores de pixels das duas imagens conforme a fórmula

$$\text{img}_{\text{output}} = \alpha \cdot \text{img}_{\text{original}} - \beta \cdot \text{GaussianBlur}(\text{img}_{\text{original}}, \sigma) + \gamma,$$

na qual α controla o contraste da imagem original, β define o contraste da versão suavizada, σ controla o nível de aplicação do ruído gaussiano, e γ adiciona um deslocamento constante. Os valores utilizados podem ser vistos na Tabela 2

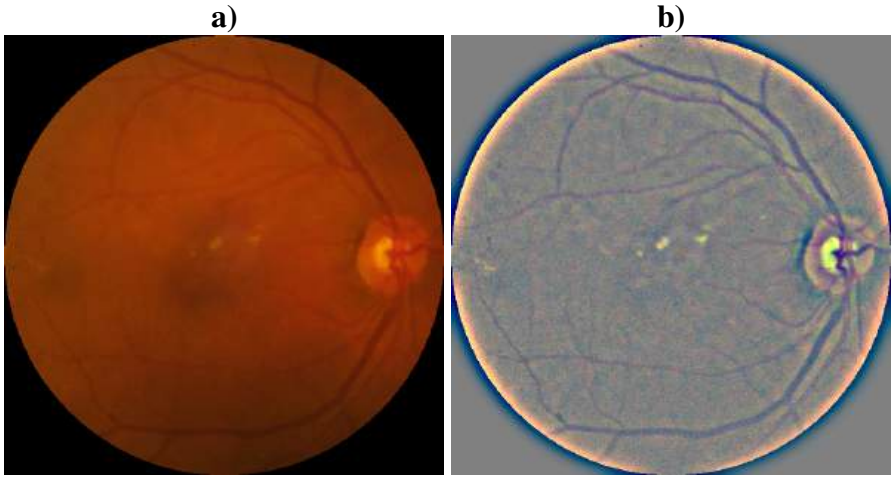
Tabela 2: Valores dos parâmetros utilizados para processamento das imagens.

Parâmetros	**Valores**
α	4
β	4
σ	30
γ	128

Fonte: Autoria Própria

O método foi configurado com um desvio padrão (σ_X) especificado para reduzir ruídos e enfatizar estruturas importantes. Esse processamento resultou em uma imagem com maior uniformidade de contraste, melhorando a clareza das características relevantes, enquanto minimizava distorções causadas por ruídos ou variações indesejadas, como pode ser visto na Figura 4.7.

Figura 4.7: **a)** Imagem recortada; **b)** Imagem após processamento.



Fonte: Autoria Própria

4.4 *Data Augmentation*

O *Data Augmentation* é um conjunto de técnicas utilizadas em aprendizado de máquina e, em especial, no treinamento de redes neurais convolucionais, com o objetivo de aumentar a quantidade de dados disponíveis para o treinamento do modelo, sem a necessidade de coletar novos dados. Isso

é feito aplicando transformações variadas aos dados originais, como rotações, translações, escalas, espelhamentos, adição de ruído, entre outras. Essas modificações geram novas amostras que são semelhantes, mas não idênticas, aos dados originais, visando o aumento e a robustez do modelo, ajudando a prevenir *overfitting* ao expor o modelo a uma maior diversidade de cenários durante o treinamento.

Cinco tipos de técnicas de *Data Augmentation* foram empregadas na Fase 1 para melhorar o conjunto de treinamento sem afetar características fundamentais da imagem, como brilho, contraste e temperatura de cor (TYMCHENKO; MARCHENKO; SPODARETS, 2020). Essas técnicas específicas incluíram: rotação de 90 graus no sentido horário, rotação de 90 graus no sentido anti-horário, rotação de 180 graus, espelhamento vertical e espelhamento horizontal. Ao aplicar essas transformações, o conjunto de dados de treinamento foi expandido significativamente, resultando em um total de 12.390 imagens para treinamento, aumentando a robustez do modelo e sua capacidade de generalização.

O processo de normalização foi crucial para garantir que o *data augmentation* se concentrasse exclusivamente em variações de orientação e posição, eliminando possíveis inconsistências introduzidas por diferenças nos dispositivos de captura das imagens. Dessa forma, o modelo pode focar no aprendizado dos padrões intrínsecos das imagens sem ser influenciado por variações indesejadas de brilho, contraste ou temperatura de cor.

É importante ressaltar que as técnicas de *data augmentation* mencionadas foram aplicadas exclusivamente ao conjunto de dados de treinamento, preservando a integridade dos conjuntos de validação e teste. Manter esses conjuntos inalterados foi essencial para assegurar a precisão e a confiabilidade na avaliação do desempenho do modelo. Essa abordagem metodológica garante que os resultados obtidos durante a validação e os testes reflitam as condições reais de uso, evitando o risco de *overfitting* a características artificiais introduzidas pelo *Data Augmentation*. Assim, a integridade dos dados de validação e teste é mantida, proporcionando uma avaliação precisa e confiável da capacidade de generalização do modelo treinado.

Na análise dos resultados da Fase 1, constatou-se que os ganhos obtidos com a aplicação de *data augmentation* permaneceram dentro do intervalo do desvio padrão das execuções. Isso indica que as melhorias observadas podem ser atribuídas a flutuações estatísticas naturais, tornando seu impacto irrelevante para a otimização do modelo. Além disso, a implementação de *data augmentation* requer

etapas adicionais de processamento e ajustes metodológicos, o que eleva o tempo necessário para o pré-processamento. Assim, optou-se por não utilizá-la na Fase 2, priorizando estratégias que apresentassem ganhos mensuráveis e justificassem o esforço adicional.

4.5 Arquiteturas utilizadas

Nesta seção, são descritas as quatro arquiteturas de redes neurais convolucionais utilizadas no trabalho: VGG16, InceptionV3, Xception e MobileNet. Em cada uma delas, os pesos pré-treinados na ImageNet foram aproveitados, as camadas de classificação originais foram removidas, e uma nova camada densa foi adicionada com cinco unidades para classificação multiclasse ou duas unidades para classificação binária.

4.5.1 VGG16

A arquitetura VGG16 possui a camada de saída original projetada para 1.000 classes, a qual foi removida e substituída por uma camada densa com cinco unidades e ativação *softmax*, adaptada para a classificação dos níveis de gravidade da retinopatia diabética. A simplicidade estrutural da VGG16, com pequenas convoluções 3×3 e profundidade consistente, facilitou sua adaptação e integração ao conjunto de dados.

4.5.2 InceptionV3

A InceptionV3, foi utilizada com uma abordagem semelhante. Sua arquitetura original, que combina convoluções de diferentes tamanhos para capturar padrões em múltiplas escalas, foi mantida até o topo da última camada convolucional. A cabeça original foi removida, e uma nova camada totalmente conectada com duas ou cinco unidades foi adicionada para a tarefa de classificação. Embora tenha sido testada no início do trabalho, o desempenho inferior em comparação com outras arquiteturas levou à sua exclusão dos experimentos finais, destacando a necessidade de redes mais adequadas ao problema específico.

4.5.3 Xception

A arquitetura Xception, que expande os conceitos da InceptionV3 utilizando convoluções separáveis em profundidade. As camadas de classificação originais foram removidas e substituídas por uma camada densa com duas ou cinco unidades e ativação *softmax*. Essa arquitetura demonstrou alto desempenho na extração de características detalhadas das imagens, devido à sua capacidade de desacoplar correlações entre canais e dimensões espaciais. Foi amplamente explorada nos experimentos devido à sua eficiência computacional e a habilidade de lidar com padrões complexos presentes nos dados.

4.5.4 MobileNet

A MobileNet foi utilizada pela sua eficiência em tarefas computacionalmente limitadas. Suas camadas finais originais foram substituídas por uma camada densa com cinco unidades, adaptando a rede ao problema de classificação. Devido ao uso de convoluções separáveis em profundidade, a MobileNet apresentou uma abordagem eficiente para a extração de características com menor custo computacional. Esse modelo foi testado especialmente em condições que exigiam eficiência, como resoluções maiores, mostrando-se uma alternativa promissora em cenários de uso restrito.

4.6 Métricas

Para uma avaliação abrangente do desempenho da rede, utilizamos as métricas de acurácia (Equação 4.2), sensibilidade (Equação 4.3) e precisão (Equação 4.4), derivadas manualmente das matrizes de confusão para fornecer uma análise detalhada do desempenho de classificação da rede. Os termos VP (Verdadeiro Positivo), VN (Verdadeiro Negativo), FP (Falso Positivo) e FN (Falso Negativo) são utilizados na avaliação de modelos de classificação.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \tag{4.2}$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \tag{4.3}$$

$$\text{Precisão} = \frac{VP}{VP + FP} \tag{4.4}$$

Para calcular a acurácia de uma matriz de confusão com 5 classes, primeiro é necessário somar todos os valores da diagonal principal, que representam as classificações corretas (verdadeiras positivas), e depois dividir essa soma pelo total de instâncias no conjunto de dados. Ou seja, a acurácia é dada pela fórmula:

$$\text{Acurácia} = \frac{\sum_{i=1}^{5} \text{VP}_i}{\sum_{i=1}^{5} \sum_{j=1}^{5} \text{Matriz Confusão}(i, j)},$$

na qual o VP_i são os valores da diagonal (verdadeiras positivas) para cada classe e Matriz Confusão(i, j) representa o número de instâncias da classe i sendo classificadas como classe j. A acurácia, portanto, é a razão entre o número total de previsões corretas e o número total de instâncias no conjunto de dados.

5 Resultados

Esta seção apresenta os resultados obtidos a partir da análise das matrizes de confusão obtidas das execuções das redes. Além de detalhar como as previsões do modelo se comparam aos valores reais das classes, são discutidas métricas derivadas, como acurácia, precisão e sensibilidade. O trabalho foi dividido em duas fases principais para facilitar a organização e análise dos resultados. Na Fase 1, o foco foi na avaliação preliminar das arquiteturas Xception, VGG16, InceptionV3 e MobileNet, analisando seu desempenho inicial com diferentes configurações de dados e técnicas de pré-processamento, incluindo data augmentation. Já na Fase 2, o objetivo foi refinar o modelo com base nos aprendizados da primeira etapa, explorando técnicas adicionais de otimização, ajustes nos parâmetros e a exclusão da InceptionV3 devido ao seu desempenho inferior.

Com essa abordagem, cada fase contribui para um entendimento progressivo e detalhado da eficácia das arquiteturas testadas. A Fase 1 estabelece uma base para identificar os modelos mais promissores e as configurações mais eficazes, enquanto a Fase 2 busca consolidar e aprimorar os resultados. Essa estrutura permite que os resultados apresentados nesta seção sejam analisados de forma clara, destacando tanto as limitações quanto os avanços obtidos em cada etapa do estudo.

5.1 Fase 1: Análise das redes em configuração binária

A análise apresentada nesta seção aborda a análise das redes neurais VGG16, InceptionV3 e Xception em uma configuração binária para a classificação de retinopatia diabética. O objetivo é avaliar o desempenho dessas arquiteturas em distinguir entre imagens classificadas como presença ou ausência da doença, considerando métricas de acurácia, sensibilidade e precisão. Pode-se destacar também que as imagens utilizadas na configuração binária se encontravam na resolução de 224×224 pixels.

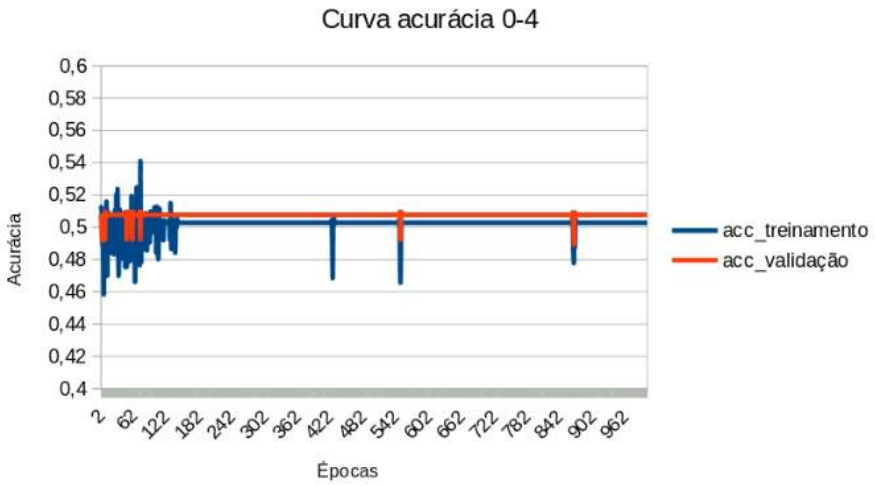
5.1.1 VGG16

Nesta subseção, são avaliados os resultados experimentais obtidos com a arquitetura VGG16 em tarefas de classificação. A análise concentra-se no comportamento da rede durante o treinamento e validação, destacando a dificuldade da VGG16 em ultrapassar 50% de acurácia em ambos os casos. A combinação 0-4 foi utilizada para ilustrar o desempenho da rede, evidenciando a limitação na

capacidade de generalização da VGG16 nas configurações testadas.

Na Figura 5.1 é apresentado o comportamento do aprendizado da arquitetura VGG16 durante o treinamento e sua comparação com a curva de validação. Pode-se observar que a rede não atingiu 100% de acurácia de treinamento, ficando estagnado nos 50% tanto em treinamento como em validação. A combinação 0-4 foi escolhida arbitrariamente por todas as combinações não expressarem aprendizado significativo, como mostra a Tabela 3.

Figura 5.1: Curva de aprendizado da rede VGG16 com a combinação 0-4.



Fonte: Autoria Própria

Tabela 3: Resultados das combinações usando rede VGG16.

Combinações	Acurácia	Sensibilidade	Precisão	*Dataset*
0-1234	79,91%	100%	79,91%	EyesPACS2015
01-234	59,63%	100%	59,63%	EyesPACS2015
012-34	40,16%	100%	40,16%	EyesPACS2015
0123-4	19,47%	100%	19,47%	EyesPACS2015
0-4	50,00%,	100%	49,23%	EyesPACS2015

Fonte: Autoria Própria

Na Tabela 3 observa-se que os resultados obtidos para a rede VGG16 em tarefas de classificação binária não foram satisfatórios. Embora a rede tenha conseguido extrair padrões das imagens durante o treinamento, a validação permaneceu estagnada no nível do número de imagens da classe selecionada pela rede sobre o número total de imagens, o que significa que a possivelmente rede

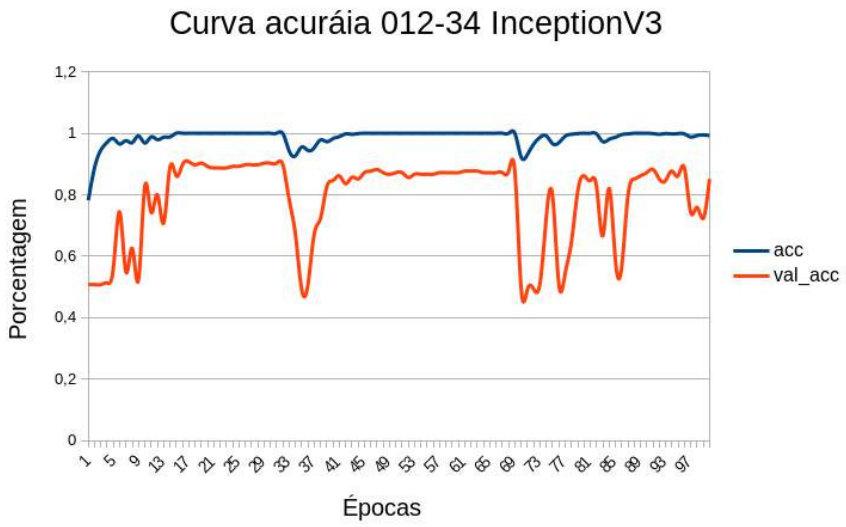
aleatorizou uma classe. Esse comportamento é evidenciado pela sensibilidade de 100% e pela precisão igual à acurácia, indicando que a arquitetura classificou todas as imagens como pertencentes à Classe 1.

Deve-se destacar a presença de desbalanceamento entre as classes durante as combinações que unem mais de 1 classe, fazendo com que a união tenha o somatório das imagens utilizadas nas classes unidas. Esse desbalanceamento é evidenciado pelo aumento da acurácia conforme as combinações aumentam a quantidade de imagens na Classe 1. Para mitigar a influência deste fator, foram utilizados pesos ajustados durante o treinamento.

5.1.2 InceptionV3

Nesta subseção, são apresentados e analisados os resultados obtidos utilizando a arquitetura InceptionV3. Esta rede, conhecida por sua eficiência na extração de características e precisão na classificação de imagens, foi testada com diversas combinações de classes para avaliar seu desempenho. A seguir, discutimos o comportamento da curva de aprendizado e os resultados quantitativos em termos de acurácia, sensibilidade e precisão.

Figura 5.2: Curva acurácia 012-34 InceptionV3.



Fonte: Autoria Própria

A Figura 5.2 mostra o comportamento do aprendizado da arquitetura InceptionV3 durante o treinamento e sua comparação com a curva de validação. Esta curva foi selecionada por ser pertencente

à combinação de melhor resultado da primeira etapa da arquitetura Xception. Pode-se observar que a rede rapidamente atingiu 100% de acurácia de treinamento, enquanto que a curva de validação ficou acima de 80% a maior parte das épocas.

A Tabela 4 apresenta os resultados obtidos utilizando a arquitetura InceptionV3 da biblioteca TensorFlow. Esses resultados demonstram uma capacidade de aprendizado superior em comparação com a arquitetura VGG16. Os melhores desempenhos foram alcançados na combinação 0-4, onde não houve necessidade de utilizar pesos, pois a base de dados estava balanceada.

Tabela 4: Resultados das combinações usando rede InceptionV3.

Combinações	Acurácia	Sensibilidade	Precisão	*Dataset*
0-1234	65,72%	67,51%	**86,64%**	EyesPACS2015
01-234	78,90%	79,59%	84,17%	EyesPACS2015
012-34	83,16%	74,24%	82,12%	EyesPACS2015
0123-4	83,77%	40,62%	62,90%	EyesPACS2015
0-4	**85,64%**,	**91,66%**	81,48%	EyesPACS2015

Fonte: Autoria Própria

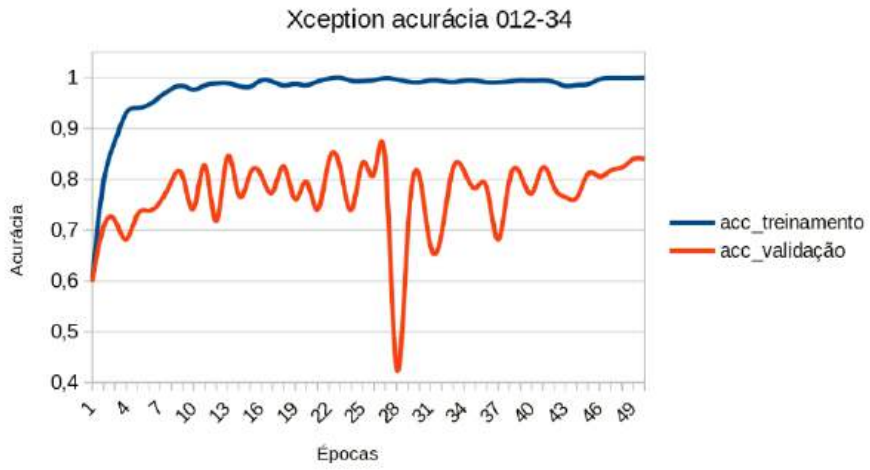
É relevante notar que a maior precisão foi obtida na combinação 0-1234. Isso indica que, mesmo com um maior número de imagens na Classe 1, os pesos ajustados durante o treinamento foram eficazes, permitindo que a rede identificasse padrões adequadamente, em vez de classificar todas as imagens como pertencentes à Classe 1.

5.1.3 Xception

Nesta subseção, são apresentados os resultados obtidos com a arquitetura Xception, uma rede avançada conhecida por sua capacidade de capturar características complexas em tarefas de classificação. A análise inclui o comportamento da rede durante o treinamento e validação, destacando a rápida convergência para uma acurácia próxima de 100% no treinamento, enquanto a validação permaneceu em torno de 89%.

A Figura 5.3 ilustra o comportamento do aprendizado da arquitetura Xception durante o treinamento e sua comparação com a curva de validação. Nota-se que a rede rapidamente atingiu quase 100% de acurácia no treinamento, enquanto a curva de validação permaneceu em torno de 80% na maioria das épocas. A combinação 012-34 foi escolhida por apresentar os melhores resultados de acurácia e sensibilidade utilizando a Xception, conforme indicado na Tabela 5.

Figura 5.3: Curva de aprendizado da rede Xception com a combinação 012-34.



Fonte: Autoria Própria

Tabela 5: Resultados das combinações usando rede Xception.

Combinações	Acurácia	Sensibilidade	Precisão	*Dataset*
0-1234	64,91%	62,94%	90,18%	EyesPACS2015
01-234	78,09%	71,42%	89,74%	EyesPACS2015
012-34	**89,74%**	**88,54%**	90,42%	EyesPACS2015
0123-4	87,02%	66,66%	66,66%	EyesPACS2015
0-4	87,18%,	78,12%	**94,94%**	EyesPACS2015

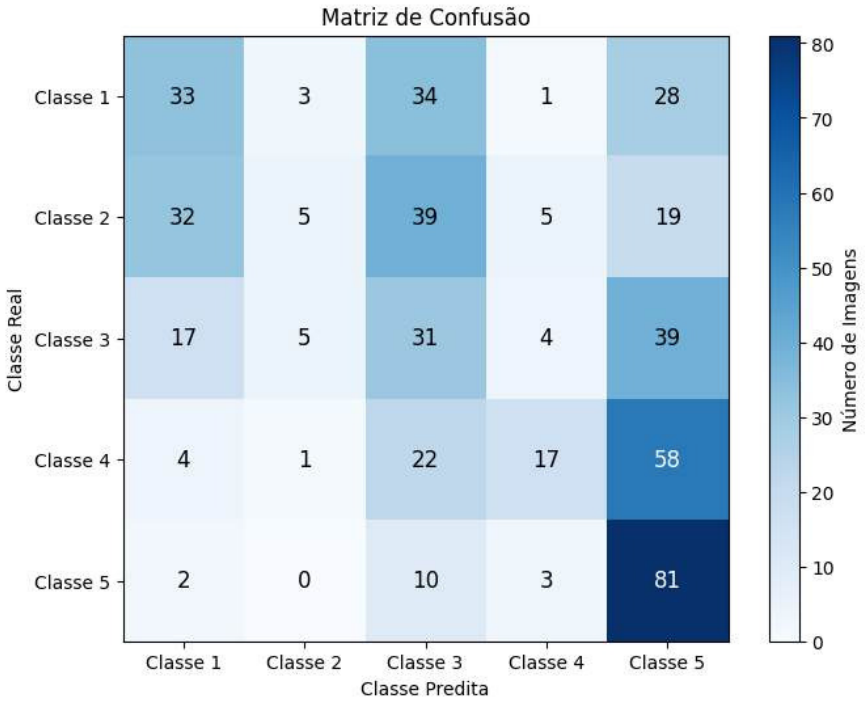
Fonte: Autoria Própria

Sendo a Xception o modelo mais complexo de classificação utilizado neste trabalho no que tange quantidade de camadas convolucionais, a matriz de confusão obtida para a validação entre 5 classes foi extraída por meio da rede e pode ser observada na Figura 5.4.

Na Fase 1, os testes com a arquitetura Xception utilizando cinco classes foram realizados com o objetivo inicial de compreender como o modelo se comportava ao lidar com a classificação detalhada das imagens de retinopatia diabética. Essa configuração permitiu avaliar a capacidade da rede em diferenciar os níveis distintos de gravidade da patologia, fornecendo informações preliminares sobre a eficácia da Xception nesse cenário mais complexo.

Esta matriz revela comportamentos da classificação. As Classe 0 e Classe 1 foram classificadas como Classe 2, o qual pode ter se dado devido a macrosemelhança entre essas imagens. Da mesma forma, mostra o número de imagens da Classe 1 que foram classificadas como Classe 0. Adicionalmente, destaca casos em que imagens da Classe 2 e Classe 3 foram classificadas erroneamente

Figura 5.4: Matriz de confusão das 5 classes individuais do desafio do Kaggle utilizando a Xception.



Fonte: Autoria Própria

como Classe 4.

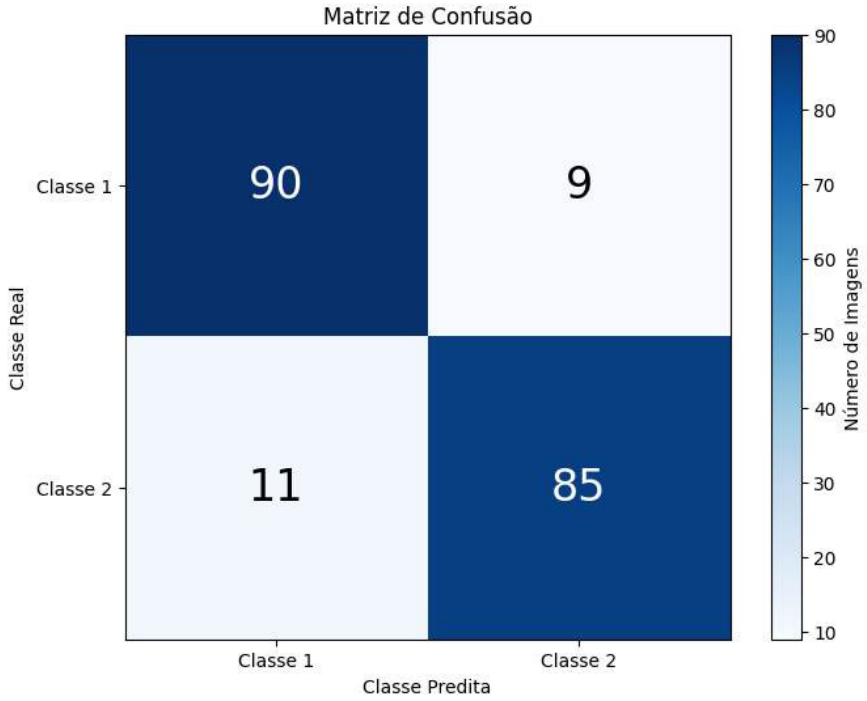
Também é possível observar que 28 imagens da Classe 0 foram erroneamente classificadas como pertencentes ao outro extremo, Classe 4. Esse erro pode ter sido introduzido devido à ruídos ou artefatos presentes nas imagens, sendo alguns deles semelhantes a exsudatos. O mesmo efeito não foi observado quando o treinamento foi realizado utilizando a combinação 0-4, o que pode indicar o problema mencionado anteriormente.

A Figura 5.5 apresenta a matriz de confusão da combinação 012-34, que se destacou como a melhor em termos de acurácia e sensibilidade. Esta matriz está incluída para permitir uma observação detalhada do comportamento do modelo em relação à classificação binária, bem como para possibilitar o cálculo de outras métricas de desempenho. As classes foram balanceadas com objetivo de testes.

Pode-se observar, ainda na Figura 5.5, a capacidade de reconhecimento de padrões feitos pela rede Xception, onde apenas algumas imagens não foram validadas corretamente. Isto pode ter se dado devido a imagens semelhantes, erroneamente classificadas ou ruídos e artefatos que ocasiona-

ram nestes erros.

Figura 5.5: Os resultados de treinamento da rede para a combinação 012-34 que foram analisados com base equilibrada.



Fonte: Autoria Própria

5.1.4 Comparação

A comparação entre as redes neurais VGG16, InceptionV3 e Xception para a classificação de retinopatia diabética revela diferenças em resultados, evidenciando a evolução das arquiteturas de redes convolucionais (CNNs) em tarefas de visão computacional. A VGG16 apresentou uma acurácia de 40,16%. Apesar de atingir 100% de sensibilidade, indicando que todos os casos positivos foram identificados, sua precisão de 40,16% demonstra um elevado número de falsos positivos, o que compromete sua aplicabilidade prática. A Tabela 6 mostra a comparação direta das 3 arquiteturas na combinação do melhor resultado da melhor rede.

Por outro lado, a InceptionV3 e a Xception, arquiteturas mais modernas, demonstraram resultados superiores à VGG16. A InceptionV3 alcançou uma acurácia de 83,16%, com uma sensibilidade de 74,24% e precisão de 82,12%, equilibrando melhor a identificação de casos positivos e a redução de falsos positivos. No entanto, a Xception superou as outras redes, obtendo uma acurácia de

Tabela 6: Comparação dos resultados para combinação 012-34.

Arquiteturas	Acurácia	Sensibilidade	Precisão
VGG16	40,16%	100%	40,16%
InceptionV3	83,16%	74,24%	82,12%
Xception	**89,74%**	**88,54%**	90,42%

Fonte: Autoria Própria

89,74%, sensibilidade de 88,54% e precisão de 90,42%. Esses valores destacam sua capacidade de extrair características relevantes de maneira mais eficiente, resultado da integração de convoluções separáveis em profundidade e uma maior complexidade arquitetural. Assim, a Xception se mostra como a escolha mais robusta para classificação de retinopatia diabética entre as três redes analisadas na classificação binária.

5.2 Fase 2: Otimização dos resultados para 5 classes

Nesta seção, são apresentados os resultados obtidos com diferentes arquiteturas e metodologias para a classificação de retinopatia diabética em cinco classes.

Na Fase 2, a metodologia foi ajustada em relação à Fase 1 para explorar estratégias que pudessem superar as limitações observadas anteriormente. Além de continuar os testes com as arquiteturas selecionadas, foram realizadas mudanças significativas, incluindo o aumento da resolução das imagens de entrada. Enquanto a Fase 1 utilizou majoritariamente imagens com resolução de 224×224 pixels, nesta etapa foram testadas resoluções maiores, como 512×512 e 600×600, com o objetivo de avaliar o impacto dessa mudança na captura de detalhes importantes para a classificação.

Em vez da InceptionV3, que não obteve um bom desempenho na seção anterior, a MobileNet foi adotada devido à sua velocidade e eficiência em tarefas de classificação, sem comprometer significativamente a precisão. Além disso, a VGG16 e a Xception continuam sendo avaliadas, mas com ajustes que buscam otimizar a acurácia nas cinco classes de severidade da doença.

Na Fase 2, a escolha de utilizar a arquitetura VGG16, mesmo após seus resultados insatisfatórios na Fase 1, foi motivada pela necessidade de compreender melhor o comportamento da rede ao lidar com a classificação em cinco classes. Apesar de ter apresentado dificuldades significativas em aprender as características dos dados na etapa inicial, a VGG16 teoricamente deveria ser capaz de realizar essa tarefa devido à sua robustez comprovada em outras aplicações de classificação.

Os experimentos nesta etapa foram conduzidos mantendo a mesma base utilizada na configuração binária, com ajustes na camada de saída para refletir as cinco classes. Isso permitiu avaliar se o desempenho inferior da rede observado anteriormente estava relacionado a limitações intrínsecas da arquitetura ou a possíveis configurações inadequadas durante o treinamento. Essa investigação foi essencial para assegurar uma análise abrangente das arquiteturas e para validar se os resultados insatisfatórios persistiriam mesmo sob condições otimizadas.

A comparação entre as redes revela as vantagens e limitações de cada uma no contexto da classificação multiclasse. A VGG16, embora simples, apresentou uma boa sensibilidade, enquanto a Xception se destacou pela robustez em termos de acurácia e precisão. Por sua vez, a MobileNet mostrou-se eficiente, alcançando resultados promissores em termos de tempo de processamento, o que a torna uma opção interessante para aplicações em tempo real, como diagnóstico assistido por IA em clínicas e hospitais.

5.2.1 VGG16

Durante os experimentos utilizando a arquitetura VGG16 na classificação em 5 classes com resoluções diferentes, os resultados obtidos foram insatisfatórios para o objetivo de classificação de retinopatia diabética.

Tanto o desempenho do treinamento quanto da validação não ultrapassaram o nível de acerto esperado por um chute aleatório, indicando dificuldade do modelo em aprender as características relevantes dos dados. Mesmo com a adição de uma camada de *Batch Normalization*, observou-se uma melhora limitada apenas no treinamento, sem impacto significativo no desempenho final durante a avaliação no conjunto de teste, que continuou no nível de aleatoriedade.

Foram realizadas diversas tentativas de ajuste, incluindo alterações no tamanho do lote (*batch size*) e outros parâmetros de treinamento, mas nenhuma dessas intervenções conseguiu melhorar os resultados de maneira substancial. Esses resultados evidenciam que a VGG16, no formato testado, não foi eficaz para a tarefa proposta, demandando a exploração de outras arquiteturas ou abordagens mais adequadas ao problema. Os resultados podem ser observados na Tabela 7. Todas as execuções apresentadas na tabela foram realizadas na base de dados EyesPACS2015.

O comportamento previamente observado de "chutar" uma classe específica de maneira aparentemente aleatória se manteve. Foi possível observar uma melhora ao aumentar a resolução das

Tabela 7: Resultados das combinações usando rede VGG16 com 5 classes.

Combinações	Acurácia	Resolução	Batch Size	Épocas	*Dataset*
0-1-2-3-4	19,47%	224x224	50	50	EyesPACS2015
0-1-2-3-4	19,47%	224x224	50	100	EyesPACS2015
0-1-2-3-4	26,00%	600x600	10	100	EyesPACS2015
0-1-2-3-4_gauss	35,28%	600x600	10	50	EyesPACS2015

Fonte: Autoria Própria

imagens de entrada, e uma melhora adicional ao aplicar o preprocessamento de contraste e cor com o filtro gaussiano. Apesar disto, os resultados foram inferiores aos obtidos com outras arquiteturas, como Xception.

5.2.2 MobileNet

Nesta subseção, são apresentados os resultados obtidos com a arquitetura MobileNet, amplamente reconhecida por sua eficiência em aplicações de visão computacional em dispositivos com recursos limitados. Durante o treinamento, a MobileNet apresentou uma boa convergência inicial, beneficiando-se de sua estrutura compacta e convoluções separáveis em profundidade para otimizar o aprendizado de padrões hierárquicos.

A MobileNet obteve a maior acurácia de validação para classificação em 5 classes utilizando imagens 224x224 no *dataset* APTOS2019. Esses resultados podem ser vistos na Tabela 8. A inclusão apenas da métrica de acurácia para a MobileNet no trabalho deve-se à priorização de uma avaliação inicial da arquitetura, dada a natureza exploratória dessa etapa. O foco foi compreender seu comportamento geral na classificação em cinco classes, utilizando o dataset APTOS2019. Embora métricas como sensibilidade e precisão sejam importantes para uma análise mais abrangente, a decisão de restringir o relatório inicial à acurácia reflete uma abordagem simplificada frente ao desempenho inferior observado na fase exploratória.

Ademais, a escolha de não calcular outras métricas detalhadas decorreu da percepção de que os resultados da MobileNet estavam aquém do esperado. Dessa forma, os esforços foram redirecionados para outras arquiteturas e ajustes metodológicos que apresentaram maior potencial de contribuir para os objetivos do estudo. Assim, a acurácia foi considerada suficiente para identificar padrões iniciais de desempenho e viabilizar comparações preliminares.

Tabela 8: Resultados obtidos utilizando a arquitetura MobileNet.

Combinações	Acurácia	Resolução	*BatchSize*	Épocas	Dataset
0-1-2-3-4	57,27%	224x224	50	50	APTOS2019
0-1-2-3-4	38,13%	224x224	50	50	EyesPACS2015
0-1-2-3-4	46,67%	600x600	10	50	EyesPACS2015
0-1-2-3-4_gauss	**73,63%**	600x600	10	50	APTOS2019
0-1-2-3-4_gauss	36,04%	600x600	10	50	EyesPACS2015

Fonte: Autoria Própria

Os resultados apresentados na Tabela 8 destacam o desempenho da MobileNet em diferentes condições de treinamento e conjuntos de dados. O melhor desempenho foi alcançado no conjunto APTOS2019, com uma acurácia de 57,27% utilizando resolução de entrada 224x224, tamanho de *batch* de 50 e treinamento por 50 épocas.

Por outro lado, no conjunto EyesPACS2015, o desempenho foi significativamente inferior, com acurácias variando entre 36,04% e 46,67%, dependendo da resolução e do pré-processamento aplicado. Isso reflete a sensibilidade da arquitetura à qualidade dos dados e à variabilidade do conjunto.

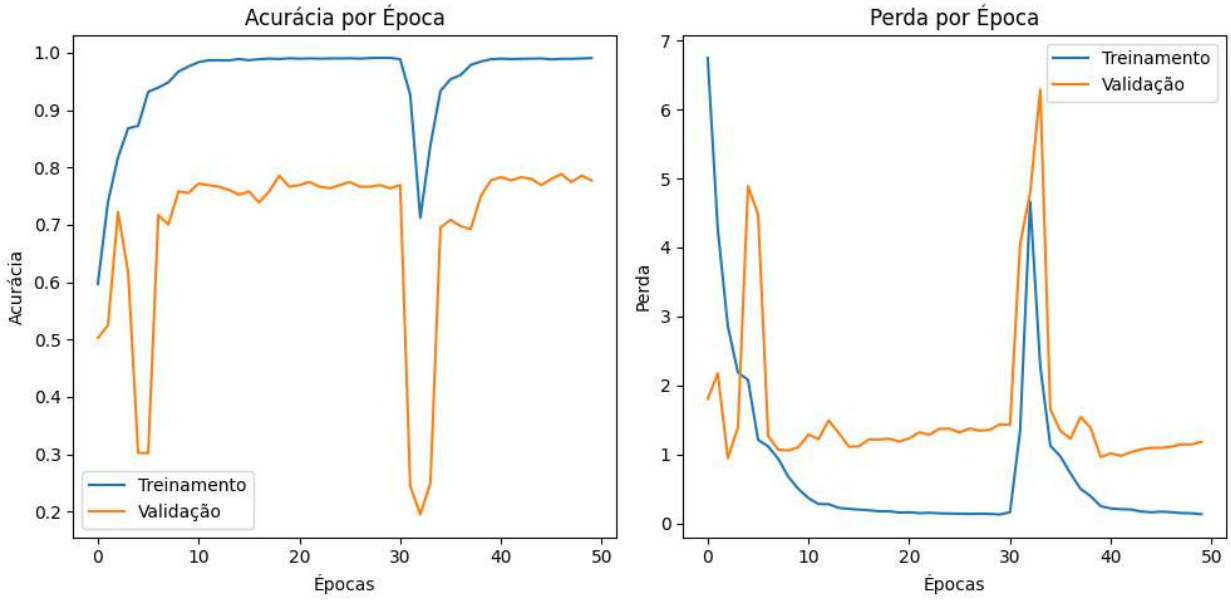
O impacto da resolução e do *batch size* também é evidente. Enquanto a resolução 224x224 foi suficiente para alcançar bons resultados no APTOS2019, a resolução maior de 600x600 não trouxe melhorias significativas para o EyesPACS2015 em comparação com o APTOS2019 em 224x224.

Esses achados mostraram que a MobileNet apresentou dificuldades para o conjunto de dados EYESPACS2015. Isto evidencia a importância de otimizar o pré-processamento e ajustar parâmetros em função do conjunto de dados utilizado.

A Figura 5.6 ilustra as curvas de aprendizado da MobileNet para o cenário de melhor resultado observado. Uma análise cuidadosa revela que o valor final da acurácia de validação representado na curva não corresponde exatamente ao valor registrado na Tabela 8. Essa discrepância ocorre porque os valores apresentados na tabela foram obtidos utilizando o método *evaluate* da biblioteca *TensorFlow*, que aplica um cálculo final sobre o conjunto de validação completo, sem as dinâmicas presentes durante o treinamento.

Durante o treinamento, as métricas de validação são computadas em *batches*, muitas vezes com dados embaralhados (*shuffle*), o que pode suavizar ou otimizar temporariamente os valores apresentados nas curvas. Já o método *evaluate* opera sobre o conjunto de validação com os pesos finais do modelo fixados, calculando as métricas de forma determinística, sem o impacto de ajustes em tempo

Figura 5.6: Curvas de acurácia e erro durante o treinamento do melhor caso da MobileNet.



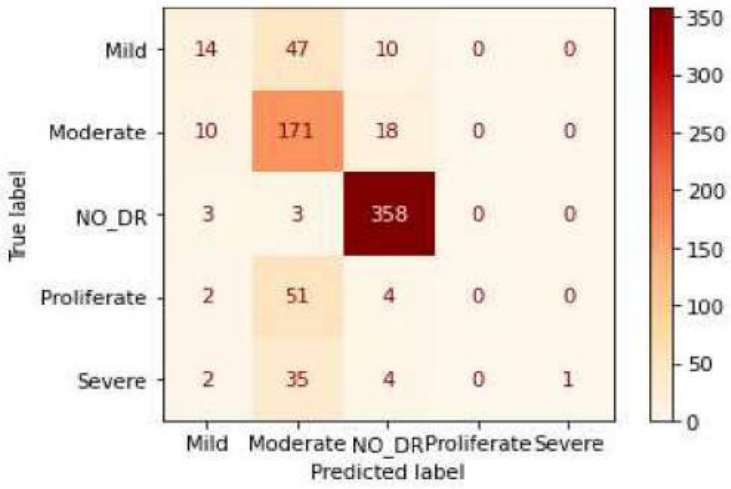
Fonte: Autoria Própria

real. Essa abordagem pode destacar fraquezas no modelo que foram mascaradas pelas dinâmicas do treinamento, como dependências específicas do *batch size* ou da ordem dos dados.

Esse efeito foi observado em trabalhos citados no estado da arte, os quais disponibilizaram suas matrizes de confusão, permitindo o cálculo manual da acurácia. Um exemplo claro disso foi o trabalho de Priya et al. (2023), o qual expõe como resultado final 90% de acurácia com a VGG16 e 92% de acurácia com a MobileNetV2. Porém, como mostra a Figura 5.7, a acurácia calculada a partir dessa matriz de confusão é de 74,21%.

Portanto, a diferença entre as métricas de validação observadas nas curvas de aprendizado e os resultados finais é um reflexo direto das condições distintas sob as quais essas métricas são calculadas. Enquanto as curvas fornecem uma visão do progresso durante o ajuste do modelo, os valores do *evaluate* representam uma análise mais fiel à capacidade de generalização do modelo. Isso reforça a importância de usar ambas as abordagens para uma avaliação mais completa e realista da performance do modelo.

Figura 5.7: Matriz de confusão para VGG16 apresentada por Priya et al. (2023).



Fonte: Priya et al. (2023)

5.2.3 Xception

Após a análise detalhada dos resultados da Xception utilizando combinações binárias de classes, novos experimentos foram conduzidos explorando sua capacidade de classificação em um cenário mais desafiador: a utilização de cinco classes distintas.

Os resultados obtidos com essas configurações ajustadas podem ser observados na Tabela 9 incluindo a curva de aprendizado do melhor caso que pode ser visto na Figura 5.8. Essa abordagem permite avaliar se as técnicas de pré-processamento e a capacidade de generalização da Xception são suficientes para lidar com classes visualmente semelhantes, destacando tanto os avanços quanto as limitações da rede nesse novo contexto.

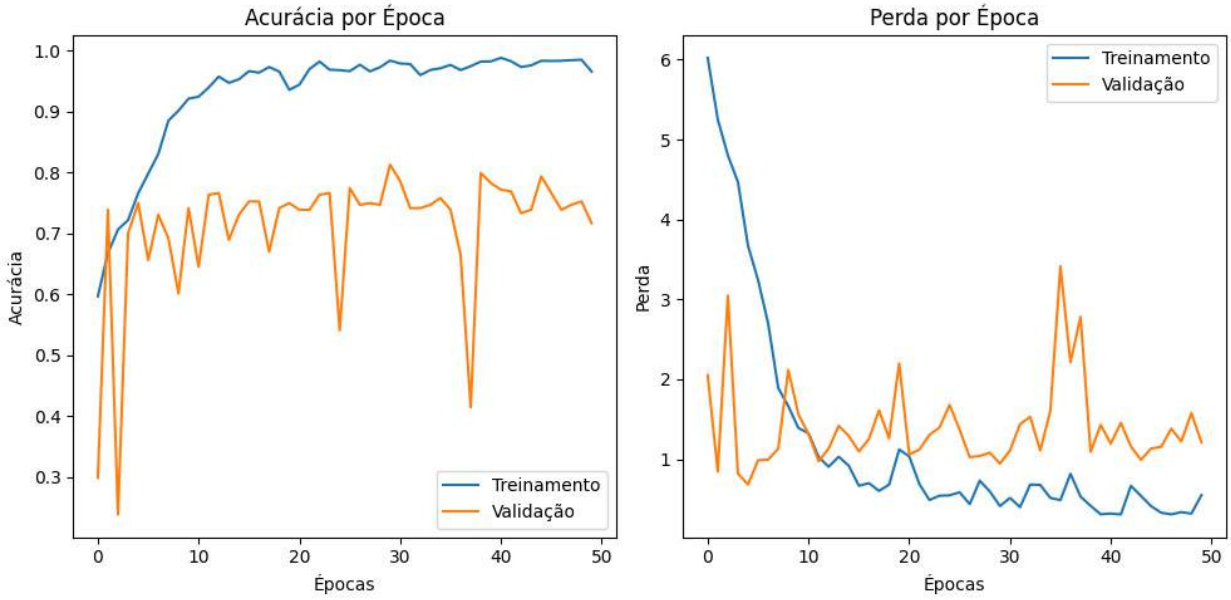
Tabela 9: Resultados obtidos utilizando a arquitetura Xception.

Combinações	Acurácia	Resolução	*BatchSize*	Épocas	Dataset
0-1-2-3-4_gauss	61,82%	224x224	50	50	APTOS2019
0-1-2-3-4	36,51%	224x224	50	50	EyesPACS2015
0-1-2-3-4_gauss	**71,70%**	512x512	10	50	APTOS2019
0-1-2-3-4	50,00%	600x600	10	100	EyesPACS2015
0-1-2-3-4_gauss	41,13%	600x600	10	50	EyesPACS2015

Fonte: Autoria Própria

A Tabela 9 apresenta os resultados obtidos utilizando a arquitetura Xception em diferentes con-

Figura 5.8: Curvas de acurácia e erro durante o treinamento do melhor caso da Xception.



Fonte: Autoria Própria

figurações de dados, resoluções e datasets. Observa-se que o melhor desempenho foi alcançado com o pré-processamento *gaussiano*, obtendo 71,70% de acurácia na resolução de 512×512, utilizando o dataset APTOS2019. Este resultado indica uma relação entre a acurácia e a resolução das imagens de entrada, entretanto, as imagens com resolução 600×600 não obtiveram resultados melhores, pode-se atribuir esse efeito ao número reduzido de épocas nesta execução.

Por outro lado, os experimentos realizados com o dataset EyesPACS2015 apresentaram desempenhos inferiores, mesmo com variações na resolução e nas combinações de classes. A melhor acurácia nesse dataset foi de 50,00%, alcançada na resolução de 600x600 e com 100 épocas. Isso sugere que a qualidade intrínseca do dataset e a adequação da rede aos padrões das imagens desempenham papéis cruciais na obtenção de melhores resultados, indicando possíveis desafios para generalização da Xception em cenários mais adversos.

6 Considerações Finais

Neste trabalho, foram realizadas uma série de abordagens experimentais com o intuito de aprimorar o desempenho de modelos de redes neurais convolucionais na análise de imagens de fundo de olho. Diversas técnicas, como o *data augmentation* e o redimensionamento de imagens, foram aplicadas com o objetivo de melhorar as métricas de classificação. Embora os resultados obtidos com essas abordagens não tenham demonstrado ganhos expressivos em termos de acurácia, foi possível identificar padrões importantes relacionados ao pré-processamento das imagens e à escolha das resoluções. Além disso, a análise das diferentes arquiteturas e conjuntos de dados evidenciou a importância de se adaptar as estratégias de processamento às características específicas das imagens, considerando as limitações e as particularidades de cada base de dados para otimizar os resultados.

Os resultados foram obtidos utilizando técnicas de *data augmentation* durante a Fase 1, que contribuíram para a melhoria das métricas de desempenho. No entanto, os ganhos proporcionados por essa abordagem não foram significativamente expressivos. O aumento da acurácia permaneceu dentro de um desvio padrão de aproximadamente 1% para as arquiteturas que demonstraram capacidade de aprendizado.

O redimensionamento das imagens foi realizado de forma a centralizar a fundoscopia, garantindo que ela ocupasse a maior área útil possível sem introduzir distorções. Durante a Fase 1, as imagens foram ajustadas para uma resolução de 224×224 pixels, o que permitiu uma maior utilização de pixels e a preservação de detalhes importantes para a análise. No entanto, vale destacar que essa resolução foi adotada especificamente nessa fase, sendo um padrão amplamente utilizado na literatura. Nessa fase, foram feitas exclusivamente avaliações de combinações binárias, onde a Xception obteve o melhor resultado.

Em fases subsequentes, exploraram-se técnicas que possibilitaram o uso de imagens de maior resolução nas redes convolucionais, com o objetivo de capturar mais detalhes e potencialmente melhorar o desempenho do modelo.

Diante das referências apresentadas, observa-se que o presente trabalho não supera as métricas do estado da arte, exceto no caso do estudo de Lam et al. (2018), onde a classificação binária foi inferior e, também, no trabalho de Pratt et al. (2016) onde resultados muito semelhantes foram obtidos na classificação em cinco níveis. No entanto, esses resultados indicam um avanço nas técnicas

empregadas, demonstrando seu potencial e relevância para futuras pesquisas e aprimoramentos na área.

A análise dos dados apresentou variações significativas entre os diferentes conjuntos e resoluções utilizados. No caso do dataset APTOS2019, a arquitetura Xception alcançou uma acurácia de 71,70% ao utilizar imagens de 512×512 pixels com técnica de suavização (*gaussian blur*). Isso demonstra que a utilização de resoluções maiores, associada a técnicas de pré-processamento, pode aprimorar o desempenho da rede ao preservar mais detalhes relevantes. Em contrapartida, para o conjunto EyesPACS2015, mesmo com ajustes semelhantes, o desempenho máximo foi de 50%, o que indica diferenças inerentes à qualidade e variabilidade dos dados dos datasets.

Os experimentos evidenciam que a resolução da imagem tem impacto direto nos resultados. Enquanto imagens de 224×224 pixels são amplamente utilizadas, experimentos com 512×512 mostraram superioridade para o APTOS2019, mas sem os mesmos benefícios para o EyesPACS2015. Além disso, a combinação de dados com suavização gaussiana apresentou desempenho inferior em situações de menor resolução, sugerindo que este pré-processamento pode ser mais adequado a contextos específicos. Assim, estratégias futuras podem explorar ajustes finos nos parâmetros de pré-processamento para maximizar os ganhos em diferentes condições de dados e arquitetura.

Por fim, o contraste entre os resultados dos datasets destaca a importância de se considerar as características específicas de cada conjunto de dados. O APTOS2019, composto por imagens de maior qualidade e consistência, mostrou-se mais responsivo a melhorias no pré-processamento e nas resoluções. Já o EyesPACS2015 apresentou limitações claras, possivelmente devido à variabilidade entre as imagens e à presença de artefatos que dificultam a classificação. Este cenário ressalta a necessidade de maior uniformização e padronização nos dados, bem como o desenvolvimento de abordagens que sejam mais robustas a essas variações.

Como perspectiva para trabalhos futuros, sugere-se a exploração de abordagens que não foram testadas neste estudo, mas que apresentam potencial para melhorar o desempenho dos modelos. Entre elas, destaca-se a utilização de resoluções de imagem ainda maiores, como 800×800 ou superiores, que podem capturar detalhes mais sutis relevantes para a classificação. Além disso, seria interessante investigar outras técnicas de *data augmentation*, como transformações geométricas mais avançadas, ajustes de iluminação variados e métodos baseados em aprendizado adversarial, que podem enriquecer o conjunto de treinamento. Também vale considerar combinações de con-

figurações que não foram exploradas neste trabalho, como o uso da MobileNet com resoluções de 600×600 pixels, o que pode oferecer um equilíbrio entre eficiência computacional e qualidade de classificação. Essas propostas podem contribuir significativamente para o avanço das técnicas de detecção automatizada de retinopatia diabética, permitindo análises ainda mais robustas e aplicáveis a diferentes cenários clínicos.

6.1 Artigo Publicado

SOARES, V. de O.; SILVA, P. A. de A.; SILVA, E. T. de A.; REGIS, C. D. M.; BATISTA, L. V. Classification of Diabetic Retinopathy with Different Combinations of Levels Using Xception Architecture. In: Congresso Brasileiro de Engenharia Biomédica (CBEB 2024), 2024, Ribeirão Preto - SP.

Referências

ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. *2017 international conference on engineering and technology (ICET)*. [S.l.], 2017. p. 1–6.

BRASIL, M. da S. *Vigitel Brasil 2023: Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico*. Brasília, Brasil, 2023. Acesso em: 04 jun. 2024. Disponível em: <http://www.saude.gov.br/vigitel>.

CHA, A. E.; VILLARROEL, M. A.; VAHRATIAN, A. Eye disorders and vision loss among us adults aged 45 and over with diagnosed diabetes, 2016–2017. 2019.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1251–1258.

COLE, J. B.; FLOREZ, J. C. Genetics of diabetes mellitus and diabetes complications. *Nature reviews nephrology*, Nature Publishing Group UK London, v. 16, n. 7, p. 377–390, 2020.

DUGAS, E. et al. *Diabetic Retinopathy Detection*. Kaggle, 2015. Disponível em: <https://kaggle.com/competitions/diabetic-retinopathy-detection>.

DUVVURI, K. et al. Classification of diabetic retinopathy using image pre-processing techniques. In: IEEE. *2023 3rd International Conference on Intelligent Technologies (CONIT)*. [S.l.], 2023. p. 1–6.

GAO, Z. et al. Diagnosis of diabetic retinopathy using deep neural networks. *IEEE Access*, IEEE, v. 7, p. 3360–3370, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT press, 2016.

GOYAL, R.; JIALAL, I. Diabetes mellitus type 2. 2018.

GROUP, E. T. D. R. S. R. Grading diabetic retinopathy from stereoscopic color fundus photographs–an extension of the modified airlie house classification. etdrs report number 10. early treatment diabetic retinopathy study research group. *Ophthalmology*, v. 98, n. 5 Suppl, p. 786–806, May 1991.

GéRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2nd. ed. Sebastopol, CA: O'Reilly Media, 2019. Revision History for the Early Release: 2018-11-05: First Release, 2019-01-24: Second Release, 2019-03-07: Third Release, 2019-03-29: Fourth Release.

HAYKIN, S. *Neural Networks and Learning Machines*. 3rd. ed. Upper Saddle River, New Jersey: Pearson Prentice Hall, 2008. Rev. ed of: Neural networks. 2nd ed., 1999. ISBN 978-0-13-147139-9.

HOWARD, A. G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

JAYA, T.; DHEEBA, J.; SINGH, N. A. Detection of hard exudates in colour fundus images using fuzzy support vector machine-based expert system. *Journal of Digital Imaging*, Springer, v. 28, p. 761–768, 2015.

JI, Q. et al. Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images. *Algorithms*, v. 12, p. 51, 02 2019.

JOSHI, S. et al. Analysis of preprocessing techniques, keras tuner, and transfer learning on cloud street image data. In: IEEE. *2021 IEEE International Conference on Big Data (Big Data)*. [S.l.], 2021. p. 4165–4168.

KIM, Y. J.; WALSH, A. W.; GRUESSNER, R. W. G. retinopathy. *Transplantation of the Pancreas*, Springer, p. 845–857, 2023.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

LAM, C. et al. Automated detection of diabetic retinopathy using deep learning. *AMIA summits on translational science proceedings*, American Medical Informatics Association, v. 2018, p. 147, 2018.

LIPTON, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, v. 16, n. 3, p. 31–57, 2018.

MITCHELL, T. M. *Machine learning*. [S.l.]: McGraw-hill, 1997.

MOOKIAH, M. R. K. et al. Computer-aided diagnosis of diabetic retinopathy: A review. *Computers in biology and medicine*, Elsevier, v. 43, n. 12, p. 2136–2155, 2013.

NAGPAL, D. et al. A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends. *Journal of King Saud University-Computer and Information Sciences*, Elsevier, v. 34, n. 9, p. 7138–7152, 2022.

NANDHINI, S. et al. An automated detection and multi-stage classification of diabetic retinopathy using convolutional neural networks. In: IEEE. *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*. [S.l.], 2023. p. 1–5.

PASCHOU, S. A. et al. On type 1 diabetes mellitus pathogenesis. *Endocrine connections*, Bioscientifica Ltd, v. 7, n. 1, p. R38–R46, 2018.

PIRES, R. et al. A data-driven approach to referable diabetic retinopathy detection. *Artificial intelligence in medicine*, Elsevier, v. 96, p. 93–106, 2019.

PRATT, H. et al. Convolutional neural networks for diabetic retinopathy. *Procedia computer science*, Elsevier, v. 90, p. 200–205, 2016.

PRIYA, R.; ARUNA, P. Svm and neural network based diagnosis of diabetic retinopathy. *International Journal of Computer Applications*, Citeseer, v. 41, n. 1, 2012.

PRIYA, S. S. S. et al. Detection and classification of diabetic retinopathy using pretrained deep neural networks. In: IEEE. *2023 International Conference on Innovations in Engineering and Technology (ICIET)*. [S.l.], 2023. p. 1–7.

QUMMAR, S. et al. A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access*, v. 7, p. 150530–150539, 2019.

RAMSAMY, G.; ARUNAKIRINATHAN, M.; COOMBES, A. The essentials of fundoscopy. *British Journal of Hospital Medicine*, MA Healthcare London, v. 78, n. 2, p. C28–C32, 2017.

ROHAN, T. E.; FROST, C. D.; WALD, N. J. Prevention of blindness by screening for diabetic retinopathy: a quantitative assessment. *BMJ: British Medical Journal*, BMJ Publishing Group, v. 299, n. 6709, p. 1198, 1989.

SAEED, F.; HUSSAIN, M.; ABOALSAMH, H. A. Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network. *IEEE Access*, IEEE, v. 9, p. 41344–41359, 2021.

SAMUEL, A. L. Machine learning. *The Technology Review*, v. 62, n. 1, p. 42–45, 1959.

SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2018.

SAROBIN, V. R.; PANJANATHAN, R. Diabetic retinopathy classification using cnn and hybrid deep convolutional neural networks. *Symmetry*, MDPI, v. 14, n. 9, p. 1932, 2022.

SCANLON, P. H.; SALLAM, A.; WIJNGAARDEN, P. V. *A practical manual of diabetic retinopathy management*. [S.l.]: John Wiley & Sons, 2017.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SINGH, A.; DWIVEDI, R. K.; RASTOGI, R. Machine learning based framework for lung cancer detection and image feature extraction using vgg16 with pca on ct-scan images. *SN Computer Science*, v. 5, 11 2024.

SOCIETY, A. A. P. T.-O. *APTOS 2019 Blindness Detection Dataset*. 2019. <https://www.kaggle.com/competitions/aptos2019-blindness-detection>. Accessed: 2024-12-16.

SOLOMON, S. et al. Diabetic retinopathy: A position statement by the american diabetes association. *Diabetes Care*, American Diabetes Association, v. 40, n. 9, p. 412–418, 2017.

SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. 2. ed. Cambridge, MA: MIT Press, 2018.

TYMCHENKO, B.; MARCHENKO, P.; SPODARETS, D. Deep learning approach to diabetic retinopathy detection. *arXiv preprint arXiv:2003.02261*, 2020.

WAN, S.; LIANG, Y.; ZHANG, Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Computers & Electrical Engineering*, Elsevier, v. 72, p. 274–282, 2018.

WANG, Z.; YANG, J. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. In: *Workshops at the thirty-second AAAI conference on artificial intelligence*. [S.l.: s.n.], 2018.

WASSERMAN, P. D. *Neural Computing: Theory and Practice*. New York: Van Nostrand Reinhold, 1989.

XU, H. et al. A hybrid neural network approach for classifying diabetic retinopathy subtypes. *Frontiers in Medicine*, v. 10, 01 2024.

ZHOU, Z.-H. *Machine learning*. [S.l.]: Springer Nature, 2021.

ZHU, W. et al. Beyond mobilenet: An improved mobilenet for retinal diseases. In: . [S.l.: s.n.], 2024. p. 56–65. ISBN 978-3-031-54856-7.

ZIA, F. et al. A multilevel deep feature selection framework for diabetic retinopathy image classification. *Comput. Mater. Contin*, v. 70, p. 2261–2276, 2022.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
Campus João Pessoa - Código INEP: 25096850
Av. Primeiro de Maio, 720, Jaguaribe, CEP 58015-435, João Pessoa (PB)
CNPJ: 10.783.898/0002-56 - Telefone: (83) 3612.1200

Documento Digitalizado Ostensivo (Público)

Dissertação de Mestrado

Assunto:	Dissertação de Mestrado
Assinado por:	Villeneve Oliveira
Tipo do Documento:	Dissertação
Situação:	Finalizado
Nível de Acesso:	Ostensivo (Público)
Tipo do Conferência:	Cópia Simples

Documento assinado eletronicamente por:

- **Villenêve de Oliveira Soares, DISCENTE (20222630003) DE MESTRADO ACADÊMICO EM ENGENHARIA ELÉTRICA - JOÃO PESSOA**, em 24/03/2025 11:34:40.

Este documento foi armazenado no SUAP em 24/03/2025. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse https://suap.ifpb.edu.br/verificar-documento-externo/ e forneça os dados abaixo:

Código Verificador: 1431412
Código de Autenticação: 9986d56b5a