

Detecção Automática de Doenças Oculares em Imagens de Tomografia de Coerência Óptica

Caio H. R. Carvalho^{1,4}, Antônio M. S. Pinheiro³, Rodrigo M. S. Veras²,
Romuere R. V. Silva^{1,3}

¹ Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal do Piauí (UFPI) – Teresina, PI – Brasil

² Departamento de Computação,
Universidade Federal do Piauí (UFPI) – Teresina, PI – Brasil

³ Sistemas de Informação,
Universidade Federal do Piauí (UFPI) – Picos, PI – Brasil

⁴ Departamento de Desenvolvimento Educacional,
Instituto Federal Baiano (IFBaiano) – Bom Jesus da Lapa, BA – Brasil

caio.carvalho@ifbaiano.edu.br, romuere@ufpi.edu.br

Abstract. *Eye diseases are problems caused by different reasons that often lead to blindness. Optical Coherence Tomography (OCT) is a non-invasive test that evaluates possible changes in the retina. This work uses Computer Vision techniques to develop classification models (binary and multiclass) of ocular anomalies on an OCT image dataset. Binary classification results reach accuracy between 97-100% and kappa 93-100%. Multiclass experiments achieved 91-92% accuracy and kappa between 86-90%. The models included in this study showed promise in classifying diseases on OCT images.*

Resumo. *Doenças oculares são problemas oftalmológicos provocados por inúmeros motivos que acarretam, muitas vezes, na cegueira. A Tomografia de Coerência Óptica (OCT) é um exame não-invasivo que permite avaliar possíveis alterações na retina. Este trabalho tem como objetivo a utilização de técnicas de Visão Computacional para o desenvolvimento de modelos de classificação (binária e multiclasse) de anomalias oculares em uma base de imagens de OCT. Os resultados de classificação binária atingiram acurácia entre 97-100% e kappa 93-100%. Os experimentos multiclasse alcançaram acurácia entre 91-92% e kappa entre 86-90%. Os modelos desenvolvidos neste estudo foram promissores na classificação de doenças em imagens de OCT.*

1. Introdução

O Buraco Macular (MH) é uma doença da retina que causa perda de visão central e/ou distorção visual, que defeitos como espaços intra-retinianos [Tanaka et al. 2020]. A Retinopatia Serosa Central (CSR) é uma doença comum da meia-idade, que causa distorções e diminuição da acuidade visual [Gholami et al. 2020], sendo caracterizada por descolamentos focais da retina e da camada Epitélio Pigmentar da Retina.

Retinopatia Diabética (DR) é uma das principais causas de cegueira evitável [Santos 2020], afetando cerca de 35 a 40% dos pacientes com diabetes no Brasil [Alam et al. 2014], tornando fundamental a classificação da evolução natural da doença para tomada de decisões clínicas. Degeneração Macular relacionada à Idade (AMD) é um fator que afeta a segurança da mácula e é o quarto distúrbio ocular mais prevalente que causa cegueira [Vos et al. 2015], sendo que o rastreamento adequado pode ajudar a minimizar a cegueira em estágio inicial.

A tomografia de coerência óptica (OCT) é uma técnica de imagem oftalmológica não-invasiva. A imagem ocular baseada em OCT é amplamente acessível a uma grande população devido ao seu baixo custo e ausência de efeitos colaterais [Thomas et al. 2021]. Portabilidade, baixo custo de instrumentação e custo operacional são outras vantagens deste mecanismo. Os modelos de análise automática de imagens oculares são opções que visam reduzir a carga de trabalho dos médicos, e fornecer um método prático e com bom custo-benefício [Oliveira et al. 2011, Tufail et al. 2017].

Neste trabalho, são propostos métodos de classificação binária e multiclasse compostos pela utilização de seis Redes Neurais Convolucionais (CNN) pré-treinadas, técnicas de transferência de aprendizado, ajuste fino e aumento de dados para diagnóstico automático de doenças oculares em imagens de OCT. Os métodos foram avaliados em 572 imagens pertencentes a uma base de dados pública (OCTID *dataset*). Os resultados obtidos são comparados com trabalhos existentes no estado da arte.

O restante do artigo está organizado da seguinte forma. Trabalhos relacionados são apresentados na Seção 2. Os passos da metodologia proposta e a base de dados utilizada nos experimentos são apresentados na Seção 3. Na Seção 4 são apresentados os resultados alcançados, assim como a discussão entorno da performance obtida. Por fim, na Seção 5, apresentamos as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

Nesta seção, são apresentados alguns dos principais trabalhos relacionados que utilizam técnicas de Visão Computacional em imagens de OCT para classificação de doenças. São relatadas as técnicas utilizadas e resultados obtidos em cada trabalho.

Gholami e Peyman (2018) usaram AlexNet, GoogLeNet, ResNet pré-treinadas para diferenciar OCT saudáveis de DR na base de imagens OCTID, e obtiveram no melhor experimento (ResNet) uma acurácia de 97,55%, uma precisão de 94,49% e uma *recall* de 94,33% .

Thomas et al. (2021) propõem uma arquitetura de CNN multiescala para classificação binária de AMD ou imagens normais. A rede neural proposta foi testada na base OCTID, e alcançou uma acurácia de 98,08%.

Hassan et al. (2021) utilizaram um método para detecção totalmente automatizada de CSR usando CNN. No framework desenvolvido, a fase de pré-processamento melhorou a qualidade da imagem e eliminou o ruído. Os três classificadores pré-treinados (AlexNet, ResNet-18 e GoogleNet) foram usados para classificação. Os resultados indicaram que o AlexNet alcançou a maior acurácia de 99,64%, precisão de 98,91%, *recall* de 100% e f1-score de 99,45% no conjunto OCTID.

Mishra et al. (2022) desenvolveram o modelo MacularNet, para classificação mul-

ticlasse de doenças oculares, e, também fizeram experimentos com a mesma base de imagens que utilizamos neste estudo OCTID [Gholami et al. 2020]. Para realização dos experimentos foi utilizada a técnica de validação cruzada K-Fold, com K igual a 5. O modelo proposto obteve 93,12% de acurácia, 92,08% de precisão, 91,09% de *recall*, e AUC de 98,51%.

Similar à proposta apresentada pelos autores supracitados, contudo, não se atendo à apenas um único experimento, o presente trabalho propõe modelos de Visão Computacional para classificação automática de imagens de OCT. Para atingir esse objetivo, foram realizados diversos experimentos com todas as classes de OCTID. A proposta aplica técnicas de classificação binária e multiclasse, utilizando-se de seis (6) arquiteturas de CNNs distintas, e buscou avaliar os resultados dos testes com várias métricas de validação.

3. Metodologia

O presente trabalho propõe a criação de uma metodologia de classificação de imagens, realizando dois tipos de experimentos: classificação binária e multiclasse. No primeiro, será informado se as imagens possuem, ou não, a doença; no segundo, as imagens serão classificadas dentre todas as classes existentes na base. Para executar as duas tarefas, a metodologia proposta foi dividida em 3 etapas: 1) aquisição das imagens; 2) classificação; 3) análise dos resultados.

3.1. Aquisição das Imagens

Para o desenvolvimento deste trabalho foi utilizada a base de imagens OCTID (*Optical Coherence Tomography Image Database*), criada no Hospital do Olho Sankara Nethralaya, em Chennai, Índia [Gholami et al. 2020]. O conjunto de dados consiste em 572 imagens de OCT, divididas em cinco classes (Figura 1): Normal (206 imagens), MH (102), AMD (55), CSR (102), e DR (107). As imagens do hospital foram diagnosticadas e rotuladas com base no diagnóstico de especialistas em retina.

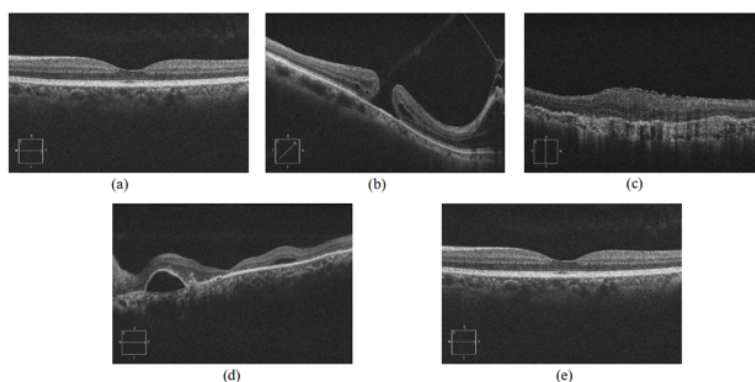


Figura 1. Exemplo de imagens de cada classe na base: (a) Normal. (b) Buraco Macular. (c) Retinopatia Diabética. (d) Retinopatia Serosa Central. (e) Degeneração Macular Relacionada à Idade.

3.2. Classificação

Após a aquisição do conjunto OCTID, realizou-se a classificação automática das imagens. Essa etapa recebe o conjunto de dados e realiza a extração de características e

classificação das imagens, seguindo os seguintes passos: 1) utilização de técnicas de pré-processamento; 2) aplicação de aumento de dados aleatório para maior generalização do modelo de classificação; 3) treinamento de modelos de Redes Neurais Convolucionais (CNN) para extração de características e classificação das imagens. Na Figura 2 são exibidas essas etapas.

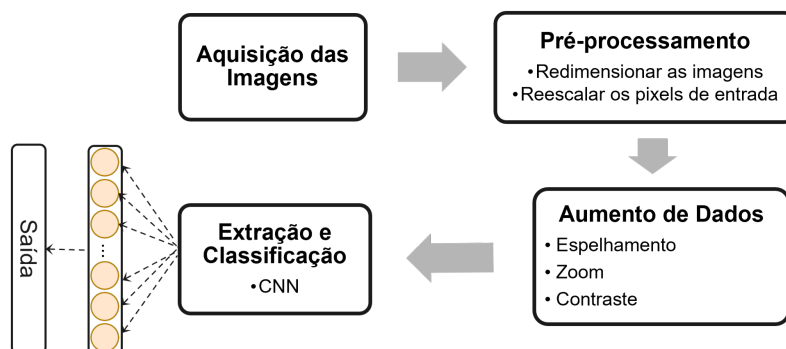


Figura 2. Visão geral da metodologia de classificação.

3.2.1. Pré-processamento

Esta etapa é executada em todas as imagens do conjunto de dados, buscando aprimorá-las e classificá-las de forma correta. A base utilizada possui imagens com diferentes resoluções. Por esse motivo, nesta fase, cada imagem foi redimensionada e normalizada para o respectivo padrão de entrada de cada CNN.

Foi utilizada a técnica de interpolação linear para a realização do redimensionamento do tamanho das imagens para a entrada padrão. As CNNs aplicadas neste estudo têm entradas de dimensões quadradas 224×224 , 299×299 e 331×331 . A interpolação linear é frequentemente usada em muitas aplicações de processamento de imagem porque fornece um compromisso entre eficiência computacional e qualidade de imagem [Gonzalez and Woods 2002].

Cada modelo de CNN espera, por padrão, que os valores dos *pixels* das imagens estejam dentro do intervalo entre -1 e 1 ou entre 0 e 1. Entretanto, neste ponto, os valores dos *pixels* das imagens estão dentro do intervalo entre 0 e 255. Portanto, antes de serem inseridas no modelo, foi realizado um pré-processamento para redimensionar os valores dos *pixels* de entrada.

3.2.2. Aumento de Dados

Uma etapa de aumento de dados (*Data Augmentation*) foi incluída com o intuito de aumentar o número de amostras da base e melhorar os resultados obtidos [Taylor and Nitschke 2018], sendo inserida ao modelo uma camada de aumento de dados aleatório provida pela API Keras. Nesta etapa, após análise empírica, as imagens que compõem a base foram submetidas a técnicas como: 1) espelhamento aleatório da imagem, no eixo horizontal; 2) zoom aleatório variando de 0% a 10%, pois valores mais altos

podem causar a perda de partes importantes da imagem; 3) contraste da imagem ajustado por um valor de fator de 20%. Além disso, ocorre combinação aleatória entre as mesmas.

3.2.3. Extração de Atributos e Classificação das Imagens

Foram realizados experimentos de classificação binária e multiclasse (Tabela 1). A classificação binária foi subdividida em 5 experimentos: todas as classes de doença rotuladas em uma única classe (Anormal) e treinadas com a classe Normal (ausência de doença); cada classe de doença treinada, separadamente, com a classe Normal. No experimento Multiclasse, todas as classes do base de imagens foram avaliadas em conjunto.

Tabela 1 - Divisão dos Experimentos de Classificação.

Classificação	Classes
Binária	Anormal × Normal
	AMD × Normal
	CSR × Normal
	DR × Normal
	MH × Normal
Multiclasse	AMD × CSR × DR × MH × Normal

Para a extração das características e classificação das imagens, foram realizados experimentos com diferentes CNNs. Para isso, utilizou-se a técnica de transferência de aprendizado com as redes pré-treinadas na ImageNet [Russakovsky et al. 2015].

As arquiteturas de CNNs foram escolhidas mediante análise dos seus desempenhos no desafio da base ImageNet (Tabela 2). São elas: DenseNet201 (DN201) [Huang et al. 2017]; InceptionResNetV2 (IRNV2) [Szegedy et al. 2017]; InceptionV3 (IV3) [Szegedy et al. 2016]; NASNetLarge (NASNL) [Zoph et al. 2018]; ResNet152V2 (RN152V2) [He et al. 2016] e Xception (XC) [Chollet 2016].

Tabela 2 - Características das arquiteturas de CNN.

Modelo	Acurácia	Entrada	Parâmetros
DenseNet201	0,923	224×224	8,1 milhões
InceptionResNetV2	0,953	299×299	55,9 milhões
InceptionV3	0,937	299×299	23,9 milhões
NASNetLarge	0,960	331×331	88,9 milhões
ResNet152V2	0,942	224×224	60,4 milhões
Xception	0,945	299×299	22,9 milhões

Todas as camadas pré-treinadas das CNNs foram congeladas para a realização da Transferência de Aprendizado. Em seguida, foi adicionada uma camada agrupamento médio global (*global average pooling*) para geração de um único vetor de características. Uma vantagem da utilização desta camada em vez de camadas totalmente conectadas é que ela é mais nativa da estrutura de convolução, estabelecendo correspondências entre mapas de características e categorias [Lin et al. 2013]. Agregada a esta camada, uma operação de *dropout* de 20% é realizada com objetivo de possibilitar uma classificação mais generalizada e evitar em grande parte o *overfitting* [Krizhevsky et al. 2017]. O conjunto de dados foi treinado durante 20 épocas, aplicando-se uma taxa de aprendizado inicial de 0,0001.

Ao final da etapa de classificação, é realizado o ajuste fino no modelo pré-treinado, descongelando-o e retreinando-o com o conjunto de dados do presente trabalho. O ajuste fino foi aplicado durante 10 épocas, com uma taxa de aprendizado inicial de 0,00001. Após análise empírica do treinamento, notou-se que os hiper-parâmetros selecionados geraram um modelo robusto e capaz de classificar os dados. A Tabela 3 apresenta esses hiper-parâmetros.

Tabela 3 - Hiper-parâmetros utilizados na classificação.

Tipo	Hiper-parâmetro
Arquitetura	DenseNet121, InceptionResNetV2, InceptionV3, NASNetLarge, ResNet152V2, Xception
Camada de classificação	<i>Global Average Pooling</i>
Dropout	0,2
Função de ativação na camada de saída	<i>Sigmoid</i> (binária), <i>Softmax</i> (multiclasse)
Função de ativação nas camadas ocultas	ReLU
Otimizador	Adam
Taxa de aprendizagem base	0,0001
Taxa de aprendizado do ajuste fino	0,00001
Taxa de aprendizagem mínima	0,0000001
Fator de decaimento da taxa de aprendizagem	0,1
Paciência para decaimento da taxa de aprendizagem	5 épocas
Paciência para parar aprendizado	10 épocas
Função de custo do treinamento	<i>Binary Crossentropy</i> (binária) <i>Categorical Crossentropy</i> (multiclasse)
Tamanho do lote	8
Épocas iniciais de treinamento	20
Épocas de ajuste fino	10

Paciência: quantidade de épocas sem diminuição da *loss* de validação.

Foi utilizado o algoritmo de otimização Adam para o treinamento dos modelos, pois tem sido amplamente empregado para melhorar o desempenho de redes neurais profundas em problemas de classificação de imagens [Kingma and Ba 2014].

Quando o modelo apresenta diminuição na perda de validação durante 5 épocas consecutivas, a taxa de aprendizado é reduzida, seguindo um fator de decaimento. Adicionalmente, se não houver redução da perda de validação por 10 épocas consecutivas, o modelo interrompe o aprendizado. Destaca-se que o modelo com a menor perda no conjunto de validação é salvo em disco para que possa ser avaliado no conjunto de testes.

Para classificar as imagens e avaliar o desempenho do método proposto, foi utilizada a validação cruzada, técnica de avaliação da capacidade de generalização de modelos. O método de validação cruzada *k-fold* consiste em dividir o conjunto total em *k* subconjuntos de mesmo tamanho. Um subconjunto é usado para teste e os outros subconjuntos *k-1* para treinamento. Esse processo é repetido *k* vezes, alterando-se circularmente o subconjunto de testes [Grimm et al. 2017].

Para os experimentos, utilizou-se 5 *folds*. Foi utilizada nesta divisão uma proporção de 80% para treinamento, 6% para validação e 14% para testes. Escolha

feita tendo em vista a pequena quantidade de imagens na base utilizada. Dessa maneira, permitiu-se que o modelo fosse treinado e validado em diferentes combinações dessas partições. A Figura 3 ilustra a divisão do conjunto de dados.

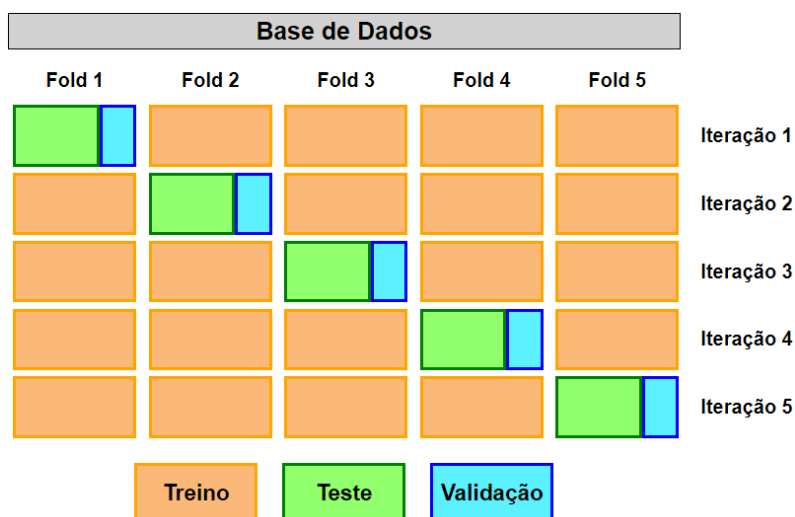


Figura 3. Divisão dos dados em *folds* para os experimentos de classificação.

Em cada iteração nos *folds*, é realizada a transferência de aprendizado e ajuste fino do modelo com base em cada arquitetura de CNN pré-treinada. São utilizados os hiper-parâmetros definidos na Tabela 3 para o treinamento de todos os modelos. Após, são calculados a média e desvio padrão dos resultados referentes a cada CNN.

Para avaliar a metodologia proposta, foram selecionadas métricas frequentemente utilizadas na literatura: Kappa, Acurácia, Precisão, *Recall*, Especificidade, F1-Score e AUC [Goodfellow et al. 2016].

4. Resultados

Esta seção apresenta a avaliação dos resultados obtidos nos dados de teste. Na primeira subseção são informados os resultados de classificação binária. Na segunda, os resultados referentes à classificação multiclasse. Os valores de cada métrica representam a média e o desvio padrão das iterações executadas em todos os *folds*.

4.1. Classificação Binária

A Tabela 4 mostra os resultados dos experimentos de detecção binária.

A metodologia proposta alcançou os melhores resultados em classificação binária com o modelo NASNetLarge, atingindo as melhores métricas em praticamente todos os experimentos, exceto na classificação CSR x Normal (InceptionV3 e DenseNet201). Além disso, todas as demais CNNs obtiveram valores de Kappa acima de 0.93, Acurácia acima de 0.97, *Recall* acima de 0.99, Precisão acima de 0.93, F1-Score acima de 0.96, Especificidade acima de 0.96 e AUC acima de 0.98, demonstrando a robustez do método proposto.

Tabela 4 - Resultados obtidos nos experimentos de classificação Binária.

Modelo	Kappa	Acurácia	Recall	Precisão	F1-score	Especific.	AUC
Anormal × Normal							
DN201	0,961±0,031	0,983±0,014	1,000±0,000	0,951±0,038	0,975±0,020	0,974±0,021	0,987±0,011
IRNV2	0,939±0,023	0,973±0,011	0,993±0,017	0,930±0,033	0,960±0,015	0,963±0,019	0,978±0,008
IV3	0,973±0,033	0,988±0,015	1,000±0,000	0,965±0,041	0,982±0,021	0,981±0,023	0,991±0,011
NASNL	0,978±0,023	0,990±0,010	1,000±0,000	0,971±0,029	0,985±0,015	0,985±0,015	0,993±0,008
RN152V2	0,955±0,015	0,980±0,007	1,000±0,000	0,943±0,018	0,970±0,010	0,970±0,010	0,985±0,005
XC	0,967±0,023	0,985±0,010	1,000±0,000	0,957±0,029	0,978±0,015	0,978±0,016	0,989±0,008
AMD × Normal							
DN201	0,983±0,037	0,995±0,012	1,000±0,000	0,993±0,015	0,997±0,008	0,975±0,056	0,988±0,028
IRNV2	0,983±0,037	0,995±0,012	1,000±0,000	0,993±0,015	0,997±0,008	0,975±0,056	0,988±0,028
IV3	0,967±0,046	0,989±0,015	1,000±0,000	0,987±0,018	0,993±0,009	0,950±0,068	0,975±0,034
NASNL	0,983±0,037	0,995±0,012	1,000±0,000	0,993±0,015	0,997±0,008	0,975±0,056	0,988±0,028
RN152V2	0,967±0,046	0,989±0,015	1,000±0,000	0,987±0,018	0,993±0,009	0,950±0,068	0,975±0,034
XC	0,948±0,078	0,984±0,024	1,000±0,000	0,980±0,029	0,990±0,015	0,925±0,112	0,963±0,056
CSR × Normal							
DN201	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,996±0,008	0,987±0,030	0,993±0,015
IRNV2	0,979±0,028	0,991±0,013	1,000±0,000	0,986±0,019	0,993±0,009	0,973±0,037	0,987±0,018
IV3	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,996±0,008	0,987±0,030	0,993±0,015
NASNL	0,979±0,028	0,991±0,013	1,000±0,000	0,986±0,019	0,993±0,009	0,973±0,037	0,987±0,018
RN152V2	0,969±0,028	0,986±0,013	0,993±0,015	0,986±0,019	0,990±0,010	0,973±0,037	0,983±0,017
XC	0,948±0,065	0,977±0,028	0,993±0,015	0,974±0,041	0,983±0,020	0,947±0,087	0,970±0,041
DR × Normal							
DN201	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,997±0,008	0,987±0,030	0,993±0,015
IRNV2	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,997±0,008	0,987±0,030	0,993±0,015
IV3	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000
NASNL	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000	1,000±0,000
RN152V2	0,990±0,023	0,996±0,010	1,000±0,000	0,994±0,014	0,997±0,007	0,987±0,030	0,993±0,015
XC	0,980±0,028	0,991±0,012	1,000±0,000	0,987±0,018	0,993±0,009	0,973±0,037	0,987±0,018
MH × Normal							
DN201	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,996±0,008	0,987±0,030	0,993±0,015
IRNV2	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,996±0,008	0,987±0,030	0,993±0,015
IV3	0,979±0,028	0,991±0,013	1,000±0,000	0,986±0,019	0,993±0,009	0,973±0,037	0,987±0,018
NASNL	0,990±0,023	0,995±0,010	1,000±0,000	0,993±0,015	0,996±0,008	0,987±0,030	0,993±0,015
RN152V2	0,979±0,028	0,991±0,013	1,000±0,000	0,986±0,019	0,993±0,010	0,973±0,037	0,987±0,018
XC	0,979±0,028	0,991±0,013	1,000±0,000	0,986±0,019	0,993±0,009	0,973±0,037	0,987±0,018

Em negrito estão os melhores resultados.

4.2. Classificação Multiclasse

A Tabela 5 apresenta os resultados das métricas obtidas no experimento de classificação multiclasse dos modelos de CNN aplicados no conjunto de imagens.

A metodologia proposta alcançou os melhores resultados em classificação multiclasse com o modelo ResNet152V2. Todas as demais CNNs obtiveram valores de Kappa acima de 0.86, Acurácia acima de 0.90, Recall acima de 0.90, Precisão acima de 0.90, F1-Score acima de 0.90, Especificidade acima de 0.97 e AUC acima de 0.91, apresentando resultados satisfatórios.

Vale destacar que a aplicação do ajuste fino após a técnica de transferência de aprendizado nos modelos de CNN influenciou positivamente o desempenho da fase de treinamento, fazendo com que o melhor modelo de cada interação (*fold*) fosse escolhido para a fase de teste (resultados exibidos nas Tabelas 4-5). É significativa a variação das

Tabela 5 - Resultados obtidos na classificação Multiclasse (5 classes).

Modelo	Kappa	Acurácia	Recall	Precisão	F1-score	Especific.	AUC
DN201	0,904±0,028	0,925±0,021	0,907±0,028	0,930±0,019	0,925±0,022	0,981±0,007	0,944±0,017
IRNV2	0,884±0,029	0,910±0,022	0,883±0,037	0,917±0,019	0,908±0,024	0,976±0,011	0,930±0,022
IV3	0,874±0,050	0,903±0,037	0,869±0,058	0,914±0,028	0,899±0,045	0,977±0,013	0,922±0,034
NASNL	0,884±0,024	0,910±0,019	0,887±0,013	0,922±0,013	0,910±0,017	0,978±0,005	0,932±0,008
RN152V2	0,904±0,044	0,925±0,034	0,908±0,045	0,932±0,029	0,925±0,034	0,981±0,012	0,945±0,027
XC	0,865±0,022	0,895±0,017	0,865±0,020	0,905±0,019	0,895±0,017	0,976±0,004	0,919±0,012

Em negrito estão os melhores resultados.

métricas de avaliação ao aplicar o ajuste fino (10 épocas) após o treinamento inicial do modelo (20 épocas), conforme exemplo mostrado na Figura 3.

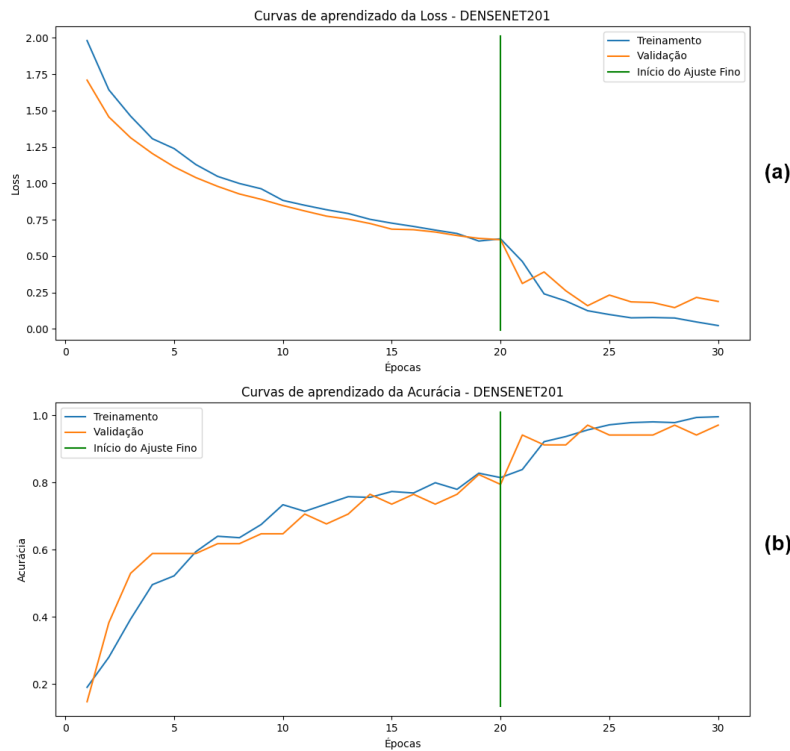


Figura 4. Curvas de aprendizado do modelo DenseNet201 no experimento multi-classe. (a) decaimento da Loss do modelo. (b) crescimento da Acurácia.

4.3. Comparação dos resultados de classificação com trabalhos do estado da arte

A Tabela 6 compara os resultados do melhor método de classificação AMDxNormal proposto com o de Thomas et al. (2021). O método proposto o superou em praticamente todos os cenários, exceção da métrica AUC. Há alguns pontos diferentes quando comparado o estudo de Thomas et al. (2021) com o presente trabalho: 1) ele tem um tamanho de imagem de entrada menor 200x200. 2) não foi aplicada a técnica de ajuste fino 3) não utilizaram a validação cruzada de classificação para apresentar o resultado das métricas, uma técnica essencial, principalmente considerando conjuntos de dados com poucas imagens.

A Tabela 7 compara a classificação CSRxNormal proposta com a de Hassan et al. (2021). O método proposto tem resultados bastantes similares em praticamente todas as métricas. Há alguns pontos que divergem as metodologias: 1) utilizaram técnicas de

Tabela 6 - AMD \times Normal

Método	Kappa	Acurácia	Recall	Precisão	F1-score	Especific.	AUC
[Thomas et al. 2021]	-	0,980	0,980	0,981	0,980	-	0,997
Método Proposto	0,983	0,995	1,000	0,993	0,997	0,975	0,988

preprocessamento como alteração na saturação e remoção de ruído nas imagens. 2) não aplicaram o ajuste fino ao final do treinamento; 3) não utilizaram a validação cruzada de classificação para apresentar o resultado das métricas. 4) Validaram o modelo com menos métricas avaliação.

Tabela 7 - CSR \times Normal

Método	Kappa	Acurácia	Recall	Precisão	F1-score	Especific.	AUC
[Hassan et al. 2021]	-	0,996	1,000	0,989	-	0,994	-
Método proposto	0,990	0,995	1,000	0,993	0,996	0,987	0,993

A Tabela 8 apresenta a comparação dos resultados do melhor método de classificação DR \times Normal proposto com aquele apresentado nos trabalhos relacionados. O método proposto o superou em praticamente todos os cenários, exceção da métrica AUC. Diferentes percebidas do estudo de Gholami e Peyman (2018) com o presente trabalho: 1) eles testaram a base apenas com 3 arquiteturas de CNN; 2) menos métricas foram avaliadas nos seus resultados; 3) não utilizaram a validação cruzada de classificação para apresentar o resultado das métricas, uma técnica essencial, principalmente considerando conjuntos de dados com poucas imagens.

Tabela 8 - DR \times Normal

Método	Kappa	Acurácia	Recall	Precisão	F1-score	Especific.	AUC
[Gholami, Peyman 2018]	-	0,975	0,943	-	0,945	-	-
Método proposto	1,000	1,000	1,000	1,000	1,000	1,000	1,000

A Tabela 9 apresenta o comparativo entre os resultados do melhor método de classificação multiclasse proposto neste estudo com aquele o de Mishra et al (2022) apresentado nos trabalhos relacionados. Ambos os trabalhos utilizaram a validação cruzada na base OCTID. E os resultados alcançados na fase de testes foram muito próximos.

Tabela 9 - Multiclasse (5 classes)

Modelo	Kappa	Acurácia	Recall	Precisão	F1-score	Especific.	AUC
[Mishra et al. 2022]	-	0,931	0,910	0,928	-	-	0,985
Método Proposto	0,904	0,925	0,908	0,932	0,925	0,981	0,945

Não foram encontrados estudos que unificando as 4 classes de doenças da OCTID em único rótulo, tampouco, realizando classificação binária de MH e imagens saudáveis.

5. Conclusões

Neste trabalho, foi desenvolvido um método de detecção de doenças oculares em imagens de OCT, utilizando Redes Neurais Convolucionais tradicionais.

Um grande diferencial deste estudo foi aplicação de diversos experimentos de classificação, tanto binária quanto multiclasse na base de imagens OCTID, inclusive, experimentos não encontrados no estado da arte (classificação binária de MH).

Após análise dos modelos, e com a comparação aos trabalhos relacionados, pode-se observar que o modelo proposto conseguiu atingir o objetivo esperado e os resultados promissores, mesmo não utilizando de técnicas de pré-processamento mais robustas. Destaca-se que a detecção correta destas doenças pode agilizar o diagnóstico do paciente, evitando a cegueira.

Espera-se que em trabalhos futuros, seja possível testar essa metodologia em: outras arquiteturas de CNN, novos conjuntos de imagens de OCT. Por fim, pretende-se incluir mais técnicas de pré-processamento à metodologia, como por exemplo: *CLAHE Histogram Equalization*; Remoção de ruídos; *Zero-padding*, buscando assim, melhorar os resultados de classificação, sobretudo, nos modelos multiclasse.

Referências

- Alam, U., Asghar, O., Azmi, S., and Malik, R. A. (2014). Chapter 15 - general aspects of diabetes mellitus. In Zochodne, D. W. and Malik, R. A., editors, *Diabetes and the Nervous System*, volume 126 of *Handbook of Clinical Neurology*, pages 211–222. Elsevier.
- Chollet, F. (2016). Deep learning with separable convolutions. *arXiv Prepr. arXiv1610*, 2357.
- Gholami, P., Roy, P., Parthasarathy, M. K., and Lakshminarayanan, V. (2020). Ocid: Optical coherence tomography image database. *Computers & Electrical Engineering*, 81:106532.
- Gholami, Peyman (2018). Developing algorithms for the analysis of retinal optical coherence tomography images. Master's thesis.
- Gonzalez, R. C. and Woods, R. E. (2002). Digital image processing. upper saddle river. J.: *Prentice Hall*.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Grimm, K. J., Mazza, G. L., and Davoudzadeh, P. (2017). Model selection in finite mixture models: A k-fold cross-validation approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(2):246–256.
- Hassan, S. A. E., Akbar, S., Gull, S., Rehman, A., and Alaska, H. (2021). Deep learning-based automatic detection of central serous retinopathy using optical coherence tomographic images. In *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, pages 206–211. IEEE.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.
- Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.
- Mishra, S. S., Mandal, B., and Puhan, N. B. (2022). Macularnet: towards fully automated attention-based deep cnn for macular disease classification. *SN Computer Science*, 3(2):142.
- Oliveira, C. M., Cristovao, L. M., Ribeiro, M. L., and Abreu, J. R. F. (2011). Improved automated screening of diabetic retinopathy. *Ophthalmologica*, 226(4):191–197.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252.
- Santos, L. R. (2020). Identificação da retinopatia diabética por aprendizado de máquina.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.
- Tanaka, S., Inoue, M., Inoue, T., Yamakawa, T., Uchio, E., Grewal, D. S., Mahmoud, T. H., and Kadonosono, K. (2020). Autologous retinal transplantation as a primary treatment for large chronic macular holes. *Retina*, 40(10):1938–1945.
- Taylor, L. and Nitschke, G. (2018). Improving deep learning with generic data augmentation. In *2018 IEEE symposium series on computational intelligence (SSCI)*, pages 1542–1547. IEEE.
- Thomas, A., Harikrishnan, P., Krishna, A. K., Palanisamy, P., and Gopi, V. P. (2021). A novel multiscale convolutional neural network based age-related macular degeneration detection using oct images. *Biomedical Signal Processing and Control*, 67:102538.
- Tufail, A., Rudisill, C., Egan, C., Kapetanakis, V. V., Salas-Vega, S., Owen, C. G., Lee, A., Louw, V., Anderson, J., Liew, G., et al. (2017). Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*, 124(3):343–351.
- Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013. *The lancet*, 386(9995):743–800.
- Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8697–8710.