Classificação de Estágios de Glaucoma Utilizando Volumes OCT e CNN3D

Matheus Levy[1], Marcos Melo Ferreira[1], Geraldo Braz Júnior[1], Marcus Vinicius[1], Gabriel Costa[1], Mario Freitas[1], Luis Rivero[1]

[1]Núcleo de Ciência Aplicada – Universidade Federal do Maranhão (UFMA)
CEP 65.080-805 – São Luís – MA – Brasil

{matheus.levy, gabriel.jesus, marcus.oliveira, mario.freitas}@discente.ufma.br, geraldo@nca.ufma.br, marcos.melo@ifma.edu.br

Abstract. *Glaucoma is one of the diseases responsible for blindness. The detection of this disease at an early stage can help patients have treatment to improve their quality of life. This study outlined the use of a method based on Deep Learning to make the classification of glaucoma stages in early, mid/advanced and without glaucoma. For this, we will use GAMMA database, composed of 100 pairs of images, color fundus photographys and three-dimensionals Optical Coherence Tomography (OCT). To classify this type of three-dimensional data, we propose the use of a 3D CNN based on the principles of DenseNet and used Transfer Learning. The best result was 84% of ACC with 2 Classes and 67% with 3 Classes. Even with limited results, the technique proves to be a good basis for future improvement work.*

Resumo. *Glaucoma é uma das doenças responsáveis pela cegueira. A detecção desta doença em estágios iniciais pode ajudar os pacientes em um tratamento para melhorar a qualidade de vida. Este estudo propõem a utilização de um método baseado em Deep Learning para fazer a classificação dos estágios do glaucoma em inicial, moderado/avançado e sem glaucoma. Para isso, utilizamos a base GAMMA, composta por 100 pares de imagens, fundos de olho e volumes OCT. Para fazer a classificação deste tipo de dado tridimensional propomos a utilização de uma CNN 3D baseada nos princípios da DenseNet e usamos Transfer Learning. O melhor resultado alcançado foi 84% de ACC com 2 Classes e 67% com 3 Classes. Mesmo com resultados limitados a técnica se mostra como uma boa base para trabalhos futuros de melhorias.*

1. Introdução

O glaucoma é apontado como sendo uma das principais causas de comprometimento da visão, e a principal causa de cegueira irreversível [Quigley and Broman 2006, Li et al. 2017, Xu et al. 2019, Fu et al. 019a]. Segundo a Organização Mundial de Saúde (OMS), estima-se que 64 milhões de pessoas em todo mundo tenham glaucoma [WHO 2019]. Deste total, estimativas de 2018 apontam que cerca de 7,7 milhões de pessoas, possuem comprometimento moderado ou severo da visão, ou cegueira resultante de formas mais graves da doença [Bourne et al. 2021]. Segundo a mesma publicação da OMS, uma projeção feita em 2017 aponta que aproximadamente 2,2 bilhões de pessoas possuem algum tipo de comprometimento da visão, e que deste total de casos, ao menos 1

bilhão ainda não foi identificado ou poderia ter sido evitado, caso houvesse o tratamento das doenças nos estágios iniciais [Fricke et al. 2018, WHO 2019].

O glaucoma é uma condição que tem como principal fator de risco uma elevada pressão intraocular (PIO), que provoca danos progressivos ao nervo óptico, levando inicialmente à cegueira bilateral, que pode evoluir para uma perda total da visão. Além da PIO elevada, outros fatores de risco como idade mais avançada, histórico familiar e raça são outros fatores que devem ser levados em consideração [SBG 2009]. Apesar de ser irreversível, a perda da visão pode ser evitada, caso procedimentos médicos sejam realizados em estágios iniciais da doença para evitar o seu progresso [WHO 2019]. No entanto, como o glaucoma é assintomático nos estágios iniciais, o diagnóstico precoce é dificultado, sendo que a doença geralmente é diagnosticada quando os efeitos do aumento da PIO já provocaram danos à visão [Sarhan et al. 2019].

Neste contexto, métodos automáticos que possam auxiliar especialistas a detectar a doença utilizando exames de imagem, podem vir a ter grande potencial para uso clínico. Métodos baseados em técnicas que utilizam aprendizagem profunda têm mostrado bons resultados em tarefas de classificação de imagens médicas [Moreira et al. 2021, Fernandes et al. 2019, Martins et al. 2021]. As redes neurais convolucionais (CNN, do inglês *Convolutional Neural Network*), principal tipo de rede profunda empregada em classificação de imagens, extrai características das imagens que podem ser consideradas irrelevantes para os seres humanos [Maetschke et al. 2019], que necessitam fazer extração de parâmetros quantitativos.

Este trabalho tem como objetivo analisar e avaliar o comportamento de uma CNN 3D na classificação de estágios de glaucoma, baseada em um volume OCT utilizando *Transfer Learning*. Como classificar glaucoma em 3 classes é uma tarefa desafiadora, é pretendido contribuir com um método para o diagnóstico de Glaucoma em OCT, sem qualquer tipo de pré-processamento, além de evoluir na utilização de pesos pré-treinados de redes 2D para redes 3D.

2. Trabalhos Relacionados

Diversos métodos tem sido propostos para detecção de glaucoma, principalmente métodos aplicados em bases de dados formadas por imagens de fundo de olho. Recentemente, métodos que utilizam outros tipos de imagem, como a OCT e a biomicroscopia ultrassônica têm sido propostos. Estas imagens são muito utilizadas, pois por meio delas é possível detectar doenças como degeneração da mácula, retinopatia diabética, e o já citado glaucoma. Neste trabalho, foram buscados artigos que descrevem trabalhos de pesquisa com métodos aplicados a bases de dados de OCTs.

Em [Muhammad et al. 2017] foi desenvolvido um método que utiliza parâmetros extraídos de imagens OCT para realização da classificação. Estes parâmetros estão relacionados à espessuras que sofrem alterações em pessoas com glaucoma, a RNFL (do inglês, *Retinal Nerve Fiber Layer Thickness*) e RGC (do inglês, *Retinal ganglion cell*). Foram realizados testes com um modelo que combina a CNN AlexNet com o classificador Random Forest. Já [Maetschke et al. 2019] desenvolveu um método para detecção de glaucoma em imagens OCT. O método não é baseado em segmentação nem no cálculo das espessuras das membranas. As imagens são utilizadas como entrada do modelo sem nenhum pré-processamento ou extração de parâmetros clínicos. Utiliza uma CNN composta por 5

camadas de convolução, e obteve melhor resultado com uso da regressão logística como classificador. Em [An et al. 2019], duas modalidades de imagem são usadas em conjunto, fundo de olho e OCT. O método consiste em obter mapas de características baseados nas espessuras citadas RNFL e RGC. Utilizou uma rede convolucional com classificador Random Forest, tendo como melhor resultado um experimento que combinou todas os mapas.

A pesquisa descrita por [Ma et al. 2021] propõe um método para classificar imagens de pacientes com glaucoma em estágio inicial, suspeitos, e saudáveis. O método utiliza ao todos 15 características, clínicas, obtidas das imagens de fundo e das imagens OCT. Foram realizados experimentos com diferentes classificadores, com o melhor resultado obtido pelo classificador SVM. O trabalho de [Xiong et al. 2021] propõe o modelo batizado de FusionNet. O modelo utilizado combina imagens OCT com gráficos obtidos de relatórios de avaliação de campo visual. Utiliza uma rede multinível, com níveis diferentes formados por camadas convolucionais, seguida de um classificador. Já [Mehta et al. 2021] utiliza um modelo multimodal, que combina redes convolucionais para extrair características de imagens de fundo e OCT. A Tabela 1 apresenta o resumo dos principais aspectos apresentados nos trabalhos relacionados.

Tabela 1. Resumo dos Trabalhos Relacionados.

Trabalho	Tipo de Dataset	Classes	Técnica
[Muhammad et al. 2017]	Volume OCT	2	Derivação de características
[Maetschke et al. 2019]	Volume OCT	2	CNN 3D e volume 3D
[An et al. 2019]	Multimodal	2	Derivação de características
[Ma et al. 2021]	Multimodal	3	Derivação de características
[Xiong et al. 2021]	Multimodal	2	CNN 2D Otimizada
[Mehta et al. 2021]	Multimodal	2	CNN 2D Otimizada

3. Metodologia

O método para realização dos experimentos consiste em 5 etapas principais: aquisição e preparação das imagens, construção do modelo CNN 3D, realização de experimentos com o modelo e estimação de parâmetros, tendo como objetivo classificar as imagens em glaucoma estágio inicial, intermediário/avançado e não-glaucoma. O fluxograma do método é apresentado na Figura 1.

3.1. Aquisição e Preparação das Imagens

A base de dados utilizada neste trabalho foi disponibilizada para os participantes do desafio GAMMA (do inglês *Glaucoma Grading from Multi-Modality Images*) [Wu et al. 2022]. A base multimodal, é composta por dois tipos de exames de imagem, fundo de olho e volumes OCT 3D, utilizadas para diagnóstico de glaucoma. A base foi provida por Sun Yat-sen Ophthamic Center, Sun Yat-sen University, Guangzhou, China. Os volumes OCT foram adquiridos utilizando uma Topcon DRI OCT Triton, estando centralizados na mácula, com um campo de visão frontal de 3 x 3 mm. O volume sendo formado por 256 imagens bidimensionais de corte transversal com resolução 992 x 512 pixeis. As imagens de fundo foram adquiridas usando uma KOWA com resolução 2000 x 2992 pixeis e uma Topcon TRC-NW 400 com resolução 1934 x 1956 pixeis.

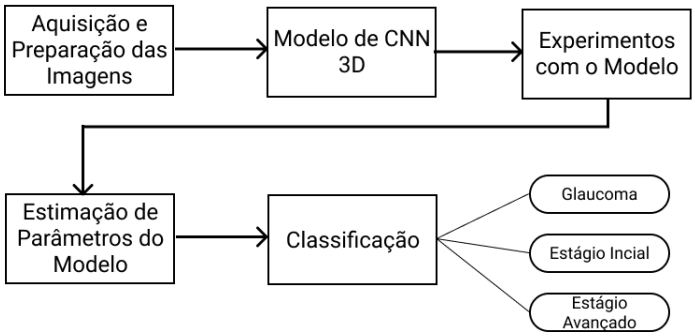

Figura 1. Fluxo de Etapas do Método Utilizado

A base é composta por pares de imagens. Cada um é composto por duas modalidades de imagem, uma fundo de olho e um volume OCT, Figura 2. Ao total, foram disponibilizados 200 pares, 100 que formam o conjunto de treino e 100 que formam o conjunto de validação. Os dois conjuntos possuem pares que pertencem a uma de três classes: sem glaucoma (classe 0 - 50 pares), estágio inicial (classe 1 - 26 pares) ou estágio moderado/avançado de glaucoma (classe 2 - 24 pares). A Figura 3 apresenta exemplos de imagens de cada uma das três classes que compõem a base. No entanto, para este estudo será utilizado somente os volumes OCT para fazer a classificação.

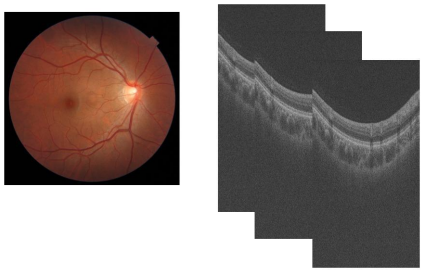

Figura 2. Par de Fundo de Olho e OCT da Base GAMMA

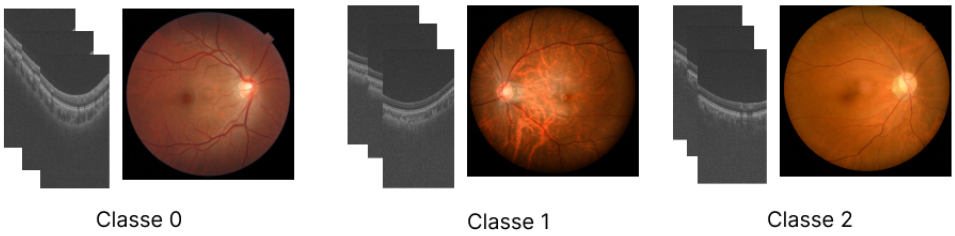

Figura 3. Divisão das Classes dos Pares na Base

Em relação às imagens OCT, foram realizados processos de aumento de dados: rotação, transformação elástica, corte aleatório das bordas, ruído gaussiano e gamma

aleatório. Também foi gerada uma segunda base a partir de fatias dos volumes recortadas na ROI (do inglês *Region of Interest*, como mostrado na Figura 4. Esta nova base foi gerada utilizando um método para identificar a ROI em uma fatia do volume. São performadas operações de dilatação, abertura e extração de arestas na imagem a fim de diminuir o número de contornos e deixa-los mais evidentes. Após são encontrados os contornos na imagem e geradas áreas retangulares que englobem estes. A maior área encontrada corresponde a ROI desejada, as demais são ruído.

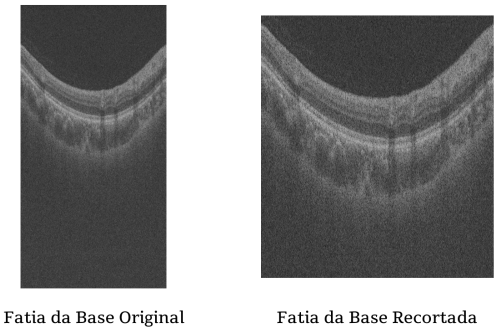

Figura 4. Exemplo de Recorte de Fatia da Base Original

3.2. Modelo de CNN 3D

Para processar um conjunto tridimensional de OCTs, propomos a utilização de uma CNN 3D, pois através de convoluções 3D é possível realizar a extração de características presentes no volume como um todo e não somente em uma fatia do volume, como ocorre em convoluções 2D.

Portanto, utilizamos uma CNN 3D, baseada nos princípios da DenseNet, especificamente escolhendo o modelo Densenet121 3D. Este modelo possui blocos densos e blocos de transição, mas com convoluções, *padding* e *pooling* 3D. Este modelo da DenseNet121 3D possui um conjunto de operações iniciais, chamada de haste, composta por camadas sequenciais: *Zero Padding 3D*, Convolução 3D, *Batch Normalization*, Ativação com relu, *Zero Padding 3D* e *Max Pooling 3D*.

Após a haste seguem 4 blocos densos intercalados por 3 blocos de transição. Um bloco denso é composto por uma quantidade de blocos de convolução sequenciais e esses por sua vez são um conjunto de operações, sendo estas: *Batch Normalization*, Ativação com relu, Convolução 3D, *Batch Normalization*, Ativação com relu e Convolução 3D. Por sua vez, os blocos de transição são compostos por *Batch Normalization*, Ativação com relu, Convolução 3D e *Average Pooling 3D*. Estes tipos de blocos podem ser visualizados na Figura 5.

Este modelo possui a haste, que foi descrita anteriormente, e a seguinte sequência: um Bloco Denso de 6 Blocos de Convolução, um Bloco de Transição, um Bloco Denso de 12 Blocos de Convolução, um Bloco de Transição, um Bloco Denso de 24 Blocos de Convolução, um Bloco de Transição, um Bloco Denso de 16 Blocos de Convolução, uma camada de *Batch Normalization*, uma Ativação com relu, uma camada de *Flatten* e o classificador com uma camada de 512 neurônios. O modelo pode ser visualizado na Figura 6.

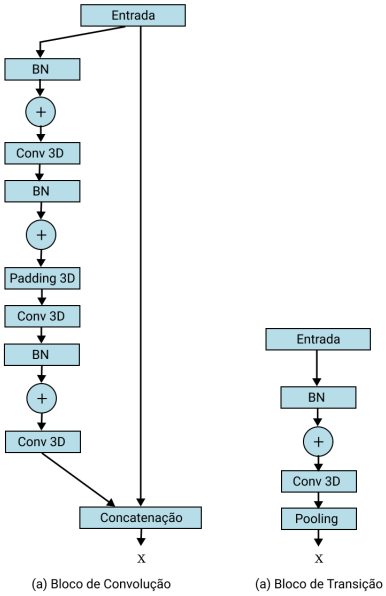

Figura 5. Bloco de Convolução e Bloco de Transição da DenseNet 3D

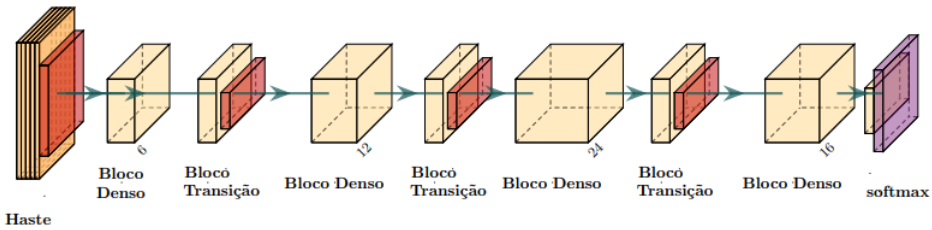

Figura 6. Arquitetura Modelo DenseNet121 3D

Como a base possui somente 100 pacientes temos uma dificuldade quanto ao volume de dados. Trabalhos relacionados que se utilizam de poucas imagens, como o de [Muhammad et al. 2017], que possui 102 pacientes, se utilizam de técnicas de derivação de características, como mapas de espessura, o que não é o nosso caso. Já [Maetschke et al. 2019] que não utiliza derivação de características possui 1110 *scans* de 624 pacientes. Para que seja possível superar essa dificuldade existem técnicas amplamente utilizadas, como a transferência de aprendizado. Esta técnica consiste de utilizar o que foi aprendido na classificação de um conjunto de dados para aumentar a capacidade de generalização em outro conjunto [Goodfellow et al. 2016]. Neste caso, o primeiro conjunto é muito maior que o segundo, como mostrado na seção anterior. Uma prática comum é a utilização dos pesos de uma CNN previamente treinada na base *ImageNet*, que possui mais de 1 milhão de imagens e 1000 classes [Araújo et al. 2017]. No caso deste estudo, utilizou-se o modelo 3D pré-treinado baseado na DenseNet121 com pesos 2D da *ImageNet* que foram expandidos para 3D [Solovyev et al. 2022][1].

[1]A implementação utilizada para os experimentos foi disponibilizada pelo ZFTurbo, disponível em `https://github.com/ZFTurbo/classification_models_3D`

3.3. Experimentos com o Modelo

Para um melhor entendimento do funcionamento de um modelo é necessário a execução de testes, com esses analisamos o comportamento do modelo. Primeiramente, o formato da entrada para o modelo é analisado, sendo que devido a limitações de memória o volume tridimensional deve ser reduzido e será observado como isso impacta o modelo. Executados alguns testes, optando-se pela utilização do formato (96, 128, 128), ou seja 96 fatias de 128 x 128 pixeis. Embora, outros formatos como (128, 64, 64) também desempenharam similarmente, optou-se por utilizar o primeiro formato, pois já havia sido testado nesta mesma rede, DenseNet121 3D, por [Solovyev et al. 2022].

Outro ponto a ser decidido são quantas camadas irão ser congeladas e quantas serão treinadas, então descongelamos a haste da rede seguida por dois blocos densos e um bloco de transição. Pois, os blocos inicias da rede extraem características mais elementares como gradientes e arestas [Albawi et al. 2017].

Para avaliar a capacidade de generalização do modelo nos experimentos decidimos utilizar o hold-out. Esta escolha se deu, principalmente, em função do custo computacional de se realizar a validação cruzada em um modelo e base 3D. O custo de memória para carregar o modelo e os volumes durante a validação cruzada estava ultrapassando a memória disponível. No entanto, um único hold-out não é confiável para analisar os resultados, pois existe uma variância a cada execução, de 5% a 15% de acurácia em relação a execução anterior. Para contornar este problema foi decidido executar 5 hold-outs e tirar a média e a variância. Com isso é esperado uma melhor análise dos resultado.

3.4. Estimação de Parâmetros do Modelo

A tarefa de classificação de imagens é por muitas vezes desafiadora. No caso da base de dados utilizada existem dificuldades no treinamento dos modelos. A base é rotulada em três classes, um desafio muito maior que uma tarefa de classificação binária. A base é pequena, sendo composta por imagens obtidas de 100 indivíduos. Outro fator que dificulta o treinamento é o desbalanceamento, uma vez que 50% das imagens são rotuladas como sem glaucoma, 26% como glaucoma em estágio inicial e 24% como glaucoma em estágio moderado ou avançado.

O próximo passo consistiu na realização de experimentos com o modelo. Utilizando a seguinte estratégia: Avaliar o desempenho do modelo para classificar os volumes OCT, em duas classes (glaucoma e não glaucoma) e em seguida avaliar o desempenho do modelo para classificar os volumes OCT, em três classes (glaucoma inicial, moderado/avançado e não glaucoma).

Como preparação para o treino o conjunto de imagens de treino sofrem um aumento de dados, como descrito na seção de Aquisição e Preparação das Imagens, e são redimensionados para (96, 128, 128, 3), isso utilizando a biblioteca Volumentations. O conjunto de teste não sofre nenhum tipo de aumento de dados, somente é redimensionado.

Em seguida, foi realizado o treinamento do modelo, utilizando a biblioteca *Keras* com *Tensorflow* como *framework*. Utilizamos o otimizador Adam e definimos o *learning rate* como 1e-5. Utilizando um *early stopping* para o valor máximo de acurácia, com um *patient* de 30 épocas. Utilizando redução de *learning rate* dinâmica também, para o valor de *loss* na validação, com fator de decréscimo de 0.1 e um valor mínimo de 1e-8. Estes

valores foram definidos empiricamente através de testes. Pesos foram utilizados para as balancear classes, utilizando a biblioteca *Sklearn*. Uma NVIDIA GeForce RTX 3060 foi usada para dar suporte a computação com um *batch-size* mínimo de 1.

O primeiro modelo foi treinado para aprender a classificar os volumes em glaucoma e não glaucoma. O segundo modelo foi treinado para classificar entre glaucoma em estágio inicial, moderado/avançado e não glaucoma. Assim podemos analisar o salto de dificuldade que o modelo encontra em classificar entre 2 classes e 3 classes, como podemos ver na Figura 7.

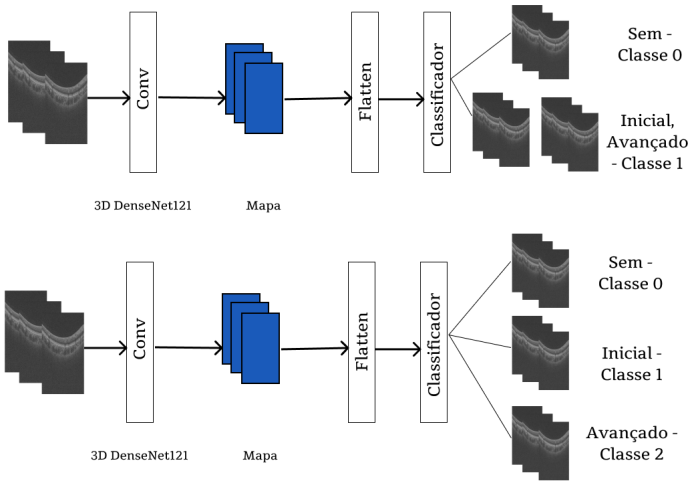

Figura 7. Modelos Utilizados na Análise de Classificação

4. Resultados

Foram realizados testes separados para cada modelo. O primeiro experimento visa classificar em duas classes, portanto as imagens de pacientes com glaucoma recebem o mesmo rótulo. A base foi dividida em treino e teste, sendo 80% das imagens agrupadas no conjunto de treinamento e as demais no conjunto de teste. Assim não havendo desbalanceamento entre as classes. No experimento para classificação em três classes, a base foi dividida entre treino e teste, sendo o conjunto de de treino formado por 80 imagens, 40 imagens da classe 0, 21 da classe 1 e 19 da classe 2, e o conjunto de teste formado por 10 imagens da classe 0, 5 da classe 1 e 5 da classe 2.

Para avaliação dos modelos, utilizou-se a mesma estratégia. Inicialmente, a base de dados foi organizada de forma a ter 2 classes, e a proporção da divisão treino/teste foi a mesma citada anteriormente. Após avaliar este modelo com 2 classes, avaliamos com 3 classes. Isto para cada uma das bases, original e recortada. Utilizamos o hold-out como descrito na seção de Experimentos com o Modelo. Os resultados estão demonstrados na Tabela 2.

Os resultados alcançados demonstram que a rede se comporta melhor com as imagens não recortadas e no problema de classificação binária, com resultado de 85% de acurácia. É provável que ao executar o recorte, mesmo que reduza o tamanho do problema, parte da coerência espacial das estruturas do volume tenha sido perdida. Outro ponto consiste numa limitação das redes profundas que consiste na grande quantidade

Tabela 2. Resultados preliminares obtidos nos testes do modelo.

Classe	Base	Acurácia	Recall			Precisão		
			Classe 0	Classe 1	Classe 2	Classe 0	Classe 1	Classe 2
2	Original	83% (± 12,8)	0,78 (± 0,24)	0,88 (± 0,09)	-	0,86 (± 0,08)	0,83 (± 0,16)	-
3	Original	67% (± 5,09)	0,86 (± 0,10)	0,64 (± 0,15)	0,32 (± 0,27)	0,79 (± 0,13)	0,55 (± 0,22)	0,48 (± 0,20)
2	Recortada	68% (± 11,6)	0,62 (± 0,23)	0,74 (± 0,18)	-	0,71 (± 0,10)	0,67 (± 0,12)	-
3	Recortada	53% (± 2,44)	0,90 (± 015)	0,12 (± 0,16)	0,20 (± 0,21)	0,52 (± 0,03)	0,30 (± 0,40)	0,47 (± 0,45)

de dados rotulados para serem utilizados para treinamento supervisionado dos modelos. Além de ser pequena, a base de dados é rotulada em três classes, um desafio mais complexo do que a classificação binária.

Continuando, agora na análise dos desvios-padrões apresentados pelas execuções do métodos. Embora tenham sido apresentados resultados promissores para um modelo inicial, verifica-se que existe uma dependência da base de treinamento que deve ser minimizada. Quando comparamos a acurácia atingida pelo nosso modelo com outros trabalhos que se utilizam de imagens OCT (Tabela 3), embora com bases diferentes, percebemos que ainda existem aspectos a serem refinados.

Tabela 3. Comparação com Trabalho Relacionados.

Modelo	Tamanho da Base	Classes	Métrica
[Muhammad et al. 2017]	102 OCTs	2	0,93 (ACC)
[Maetschke et al. 2019]	1110 OCTs	2	0,94(AUC)
[Ma et al. 2021]	301 Conjuntos de Exames	3	0,90 (AUC)
DenseNet1213D	100 OCT	2	0,83 (AC)
DenseNet1213D	100 OCT	3	0,68 (AC)

Para melhor análise do modelo podemos observar duas execuções, a primeira com acurácia de 65% e a segunda com 80%, na classificação binária. Na Figura 8 podemos ver as matrizes de confusão de cada execução. O primeiro demonstra ter uma tendência a classificar como glaucoma, o que pode ser visto pelos falsos positivos, onde 6 não glaucomas foram classificados como glaucomas. Já na classe 1 apenas uma imagem foi classificada errada, sendo um glaucoma classificado como não glaucoma. Isso é mostrado por um *recall* de 0,4 na classe 0 e 0,9 na classe 1. Já na segunda execução, o modelo mostra-se melhor em separar as duas classes, apesar de 3 falsos negativos e 1 falto positivo, o que pode ser visto por um *recall* de 0,9 na classe 0 e 0,7 na classe 1. Isto demonstra a variância do modelo de acordo com os conjuntos de treino e teste, onde existem elementos dentro da base que são mais facilmente classificados e outros que são mais complexos.

Seguindo uma estratégia semelhante a anterior iremos analisar duas execuções para 3 classes, a primeira execução atingiu 60% de acurácia e a segunda 75%. A primeira execução errou todos os glaucoma em estágio moderado/avançado, também errou 2 glaucomas em estágio inicial. A tendencia deste modelo foi classificar as imagens como não glaucoma, o que pode ser visto pelo *recall* de 0.9 na classe 0, 0,6 na classe 1 e 0,0 na classe 2. Já a segunda execução se mostrou mais precisa na classe 2, pois conseguiu classificar 3 imagens corretamente, porém ainda apresenta 2 falsos negativos, quando a classe 1 mostrou uma precisão menor, isso demonstra uma confusão esperado do modelo ao classificar entre glaucoma em estágio inicial e não glaucoma. Porém, dois falsos nega-

tivos na classe 2 sendo classificados como classe 0 não eram esperados do modelo, uma vez que se espera uma diferença maior das estruturas oculares, capturadas pela OCT, entre não glaucoma e glaucoma em estágio moderado/avançado. As matrizes de confusão podem ser visualizadas na Figura 8.

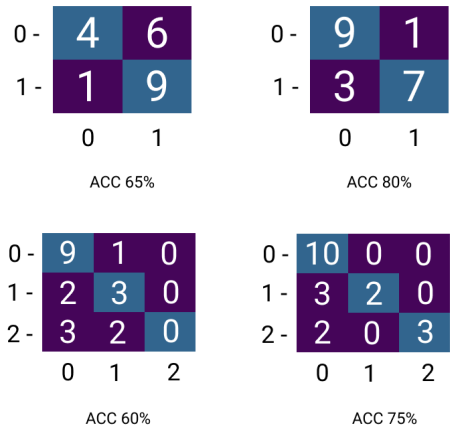

Figura 8. Matriz de Confusão dos Testes de 2 e 3 Classes

O salto de dificuldade entre a classificação binária e entre 3 classes sugere a possibilidade futura de utilizar dois classificadores. O primeiro para classificar entre glaucoma e não glaucoma e o segundo para classificar o estágio do glaucoma (inicial e moderado/avançado).

O uso somente de volumes OCT mostra a dificuldade de utilizar uma rede dependente somente das informações de um tipo de entrada e sugere a utilização de uma rede multinível com as imagens de fundo, além de ser necessário um estudo sobre aprendizado auto supervisionado (*Self Supervised Learning*, uma vez que estão disponíveis as imagens de 100 indivíduos, no entanto as mesmas não estão rotuladas. O resultado inferior com a base recortada sugere a hipótese de que recortar as fatias atrapalha a rede a formar uma imagem tridimensional e aprender as características.

Os resultados obtidos apesar de longe do desejado sugerem a possibilidade de uma melhor otimização do modelo utilizado, com uma análise mais profunda sobre a otimização da quantidade de blocos de convolução na rede. Outro ponto a ser levado em consideração para trabalhos futuros é a exploração de técnicas de regularização, pois como visto com uma base pequena e desbalanceada o modelo por vezes tende ao *overfit* se focando nas características da classe 0, pois esta tem mais elementos do que as outras duas.

5. Conclusão

Este trabalho descreveu um método baseado no modelo de uma CNN 3D, em específico na DeseNet121, utilizando-se de pesos pré-treinados na base da *ImageNet* e expandidos para a 3D, visando classificar estágios de glaucoma na base GAMMA. Mesmo com resultados limitados a técnica se mostra como uma boa base para melhorias futuras e assim gerando modelos mais refinados. Assim contribuindo para auxiliar na detecção do glaucoma em estágio inicial.

Agradecimentos

Os autores agradecem às instituições brasileiras que ajudaram no desenvolvimento desta pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) pelo financiamento Apoio.

Referências

Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). Understanding of a convolutional neural network. pages 1–6. Ieee.

An, G., Omodaka, K., Hashimoto, K., Tsuda, S., Shiga, Y., Takada, N., Kikawa, T., Yokota, H., Akiba, M., and Nakazawa, T. (2019). Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images. *Journal of healthcare engineering*, 2019.

Araújo, F., Carneiro, A., Silva, R., Medeiros, F., and Ushizima, D. (2017). Redes neurais convolucionais com tensorflow:teoria e prática. 1:382–406.

Bourne, R. R. et al. (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. *The Lancet Global Health*, 9(2):e144–e160.

Fernandes, A. G., Martins, C., Lima, A. C., Junior, G., de Almeida, J. D., and de Paiva, A. (2019). Meta aprendizagem de extração de características aplicada ao diagnóstico de glaucoma. In *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 342–347, Porto Alegre, RS, Brasil. SBC.

Fricke, T. R., Tahhan, N., Resnikoff, S., Papas, E., Burnett, A., Ho, S. M., Naduvilath, T., and Naidoo, K. S. (2018). Global prevalence of presbyopia and vision impairment from uncorrected presbyopia: systematic review, meta-analysis, and modelling. *Ophthalmology*, 125(10):1492–1499.

Fu, H., Baskaran, M., Xu, Y., Lin, S., Wong, D. W. K., Liu, J., Tun, T. A., Mahesh, M., Perera, S. A., and Aung, T. (2019a). A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. *American journal of ophthalmology*, 203:37–45.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge. `http://www.deeplearningbook.org`.

Li, J., Chen, Y., Cai, L., Davidson, I., and Ji, S. (2017). Dense transformer networks. *arXiv preprint arXiv:1705.08881*.

Ma, J., Lv, B., Li, Y., Fan, P., Zhao, X., Yuan, H., and Zhang, Y. (2021). Multimodal primary open angle glaucoma early diagnosing program based on clinical process.

Maetschke, S., Antony, B., Ishikawa, H., Wollstein, G., Schuman, J., and Garnavi, R. (2019). A feature agnostic approach for glaucoma detection in oct volumes. *PloS one*, 14(7):e0219126.

Martins, C., Santos, R. D., Almeida, J. D., Júnior, G. B., and Teixeira, J. A. (2021). Detecção de patologias oculares em imagens de reflexo vermelho utilizando descritores de cor. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 346–357, Porto Alegre, RS, Brasil. SBC.

Mehta, P., Petersen, C. A., Wen, J. C., Banitt, M. R., Chen, P. P., Bojikian, K. D., Egan, C., Lee, S.-I., Balazinska, M., Lee, A. Y., et al. (2021). Automated detection of glaucoma with interpretable machine learning using clinical data and multimodal retinal images. *American Journal of Ophthalmology*, 231:154–169.

Moreira, J. M., Almeida, J. D., Júnior, G. B., and Paiva, A. (2021). Detecção de glaucoma usando redes em cápsula. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 188–199, Porto Alegre, RS, Brasil. SBC.

Muhammad, H., Fuchs, T. J., De Cuir, N., De Moraes, C. G., Blumberg, D. M., Liebmann, J. M., Ritch, R., and Hood, D. C. (2017). Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *Journal of glaucoma*, 26(12):1086.

Quigley, H. A. and Broman, A. T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, 90(3):262–267.

Sarhan, A., Rokne, J., and Alhajj, R. (2019). Glaucoma detection using image processing techniques: A literature review. *Computerized Medical Imaging and Graphics*, page 101657.

SBG (2009). 3º consenso de glaucoma primário de Ângulo aberto. `https://www.sbglaucoma.org.br/wp-content/uploads/2020/06/consenso03-v2.pdf`. Acessado em: 12-Janeiro-2022.

Solovyev, R., Kalinin, A. A., and Gabruseva, T. (2022). 3d convolutional neural networks for stalled brain capillary detection. *Computers in Biology and Medicine*, 141:105089.

WHO (2019). World report on vision. `https://www.who.int/publications/i/item/9789241516570`. Acessado em: 12-Janeiro-2022.

Wu, J., Fang, H., Li, F., Fu, H., Lin, F., Li, J., Huang, L., Yu, Q., Song, S., Xu, X., et al. (2022). Gamma challenge: Glaucoma grading from multi-modality images. *arXiv preprint arXiv:2202.06511*.

Xiong, J., Li, F., Song, D., Tang, G., He, J., Gao, K., Zhang, H., Cheng, W., Song, Y., Lin, F., et al. (2021). Multimodal machine learning using visual fields and peripapillary circular oct scans in detection of glaucomatous optic neuropathy. *Ophthalmology*.

Xu, B. Y., Chiang, M., Chaudhary, S., Kulkarni, S., Pardeshi, A. A., and Varma, R. (2019). Deep learning classifiers for automated detection of gonioscopic angle closure based on anterior segment oct images. *American journal of ophthalmology*, 208:273–280.