GABRIEL OZEAS DE OLIVEIRA

**GERAÇÃO DE MAPAS DE ESPESSURAS
SINTÉTICOS POR MEIO DE REDES
GENERATIVAS ADVERSÁRIAS PARA O
TREINAMENTO DE REDES NEURAIS
PROFUNDAS PARA AUXÍLIO AO DIAGNÓSTICO
DE GLAUCOMA**

São Paulo
2021

GABRIEL OZEAS DE OLIVEIRA

**GERAÇÃO DE MAPAS DE ESPESSURAS
SINTÉTICOS POR MEIO DE REDES
GENERATIVAS ADVERSÁRIAS PARA O
TREINAMENTO DE REDES NEURAIS
PROFUNDAS PARA AUXÍLIO AO DIAGNÓSTICO
DE GLAUCOMA**

Versão Corrigida

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Ciências.

São Paulo
2021

GABRIEL OZEAS DE OLIVEIRA

**GERAÇÃO DE MAPAS DE ESPESSURAS
SINTÉTICOS POR MEIO DE REDES
GENERATIVAS ADVERSÁRIAS PARA O
TREINAMENTO DE REDES NEURAIS
PROFUNDAS PARA AUXÍLIO AO DIAGNÓSTICO
DE GLAUCOMA**

Versão Corrigida

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Ciências.

Área de Concentração:

Engenharia da Computação

Orientador:

Edson Satoshi Gomi

São Paulo
2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 29 de janeiro de 2021

Assinatura do autor: Gabriel Ozeas de Oliveira

Assinatura do orientador: [Assinatura]

Catálogo-na-publicação

Oliveira, Gabriel Ozeas

Geração de mapas de espessuras sintéticos por meio de redes generativas adversárias para o treinamento de redes neurais profundas para auxílio ao diagnóstico de glaucoma / G. O. Oliveira -- versão corr. -- São Paulo, 2021.

83 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Inteligência Artificial 2.Aprendizagem de Máquina 3.Glaucoma 4.Diagnóstico Por Computador 5.Redes Neurais I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

AGRADECIMENTOS

Agradecimento especial ao meu orientador Prof. Dr. Edson Satoshi Gomi, por me acolher na Escola Politécnica, pelos ensinamentos e estímulos transmitidos durante todo o trabalho.

Aos Prof. Dr. Vital Paulino Costa e Leonardo Seidi Shigueoka por tornar este trabalho possível.

Ao Prof. Dr. Marcelo Dias pelos ensinamentos e inspiração.

Aos meus colegas do laboratório KNOMA, pela amizade, conselhos e momento de descontração.

Aos meu pais Sergio e Lina e meu irmão Daniel por torcerem por mim e tolerarem minha ausência. A minha namorada Letícia pelo amor e apoio incondicional até aqui.

Aos meus empregadores e colegas de trabalho pelo tempo cedido para realização deste trabalho.

RESUMO

O glaucoma é o segundo causador de cegueira no mundo, atingindo estimadamente 76 milhões de pessoas em 2020. Devido a seu caráter irreversível, vários trabalhos de pesquisa na literatura apresentam métodos para o diagnóstico de glaucoma a fim de auxiliar na tomada de decisão do especialista. Alguns destes estudos empregam métodos de aprendizagem de máquina, apesar do lento e custoso processo de coleta de dados presente em várias áreas da medicina. Este trabalho apresentou uma análise sobre uso de redes generativas adversárias na geração de mapas de espessura RNFL (Retinal Nerve Fiber Layer) sintéticos para serem empregados no treinamento de classificadores para o diagnóstico de glaucoma. Especificamente, esta investigação tratou dos classificadores que utilizam redes neurais profundas e que necessitam de grande quantidade de dados para treinamento. Foram gerados mapas de espessura de RNFL sintéticos utilizando uma rede generativa adversária, criando 11 conjuntos de dados mistos com dados reais e dados sintéticos em diferentes proporções. Esses conjuntos de dados foram utilizados no treinamento de 8 redes neurais profundas para o auxílio ao diagnóstico de glaucoma. O desempenho dos classificadores foi medido através dos valores de AROC. Os melhores valores de AROC foram obtidos quando o conjunto de dados foi aumentado em 100 vezes, onde 50% dos dados eram sintéticos e 50% eram reais. O melhor valor de AROC foi 0,865, obtido pela rede VGG16. Tais resultados indicam que o uso de mapas de espessura de RNFL sintéticos gerados por redes generativas adversárias contribui para melhorar o desempenho de redes neurais profundas para o diagnóstico de glaucoma.

Palavras-Chave - Glaucoma, Aprendizagem de Máquina, Redes Neurais Convolucionais, Rede Generativas Adversárias

ABSTRACT

Glaucoma is the second leading cause of blindness in the world. Around 76 million people were affected by this disease in 2020. Due to the irreversible glaucoma nature, several studies in the literature presented early diagnosis methods to assist in decision-making specialists. Some of these studies employ machine learning methods despite the slow process of collecting data present in many medical fields. This work presents an analysis of generative adversarial networks in the generation of synthetic RNFL (retinal nerve fiber layer) thickness maps for use in the training of classifiers for the diagnosis of glaucoma. Mainly classifiers that use deep neural networks and require a large amount of data for training. Synthetic RNFL thickness maps were created using a generative adversarial network, creating 11 mixed datasets with real data and synthetic data in different proportions. These datasets were used in the training of 8 deep neural networks to aid in glaucoma diagnosis. Classifiers' performance was measured using AROC values. Increasing the dataset by one hundred times gave the best AROC values, where 50 % of the data were synthetic, and 50 % were real. The best AROC value was 0.865, obtained by the VGG16 network. Such results indicate that the use of synthetic RNFL thickness maps generated by generative adversarial networks contributes to improving the performance of deep neural networks for the diagnosis of glaucoma.

Keywords - Glaucoma, Machine Learning, Convolutional Neural Networks, Generative Adversarial Networks

LISTA DE FIGURAS

1	Diagrama da anatomia do olho humano. Adaptado de (HUMAN..., 2012)	17
2	Fluxo de secreção e drenagem do humor aquoso. Adaptado de (GOEL RENATA G PICCIANI; BHATTACHARYA, 2010)	18
3	Imagem do fundo do olho. Adaptado de Bourne (2006)	20
4	Imagem do fundo do olho (BOURNE, 2006)	21
5	Esquema de funcionamento do OCT	23
6	Comparação entre TD-OCT (lado esquerdo) e SD-OCT (lado direito) do mapeamento do RNFL. Adaptado de Liu et al. (2014)	24
7	Gráficos gerados pelo OCT. (a) Gráfico da média das espessuras divididas em 12 horas. (b) Média das espessuras dividido em 4 quadrantes. (c) Média de espessuras totais. (d) Gráfico de espessura de 256 pontos.	25
8	Perceptron	28
9	Rede Neural com duas camadas intermediárias	30
10	Filtro (F) aplicado a imagem (I), gerando um <i>Feature Map</i>	33
11	Max Pooling sendo aplicado em um Feature Map	34
12	Bloco da rede <i>Inception</i>	36
13	Conexão de atalho de identidade	37
14	Matriz de confusão para um classificador binário	38
15	Lado esquerdo: Pontos de execuções com diferentes configurados para um mesmo classificador. Lado direito: Interpolação de pontos de um classificador	40
16	Rede generativa adversária	41

17	Rede generativa adversária condicional	45
18	A: Escala de cores representando espessura em microns. B: Imagem de mapa de espessuras de RNFL gerado do OCT	52
19	Fluxo de treinamento da GAN e do Classificador	53
20	Arquitetura da rede geradora. PS é abreviação para <i>Padding=Same</i> . BN é abreviação de <i>Batch Normalization</i> . FC é abreviação de <i>Fully Connected</i> (camada totalmente conectada).	55
21	Arquitetura da rede discriminadora. PS é abreviação para <i>Padding=Same</i> . BN é abreviação de <i>Batch Normalization</i>	58
22	Funções de erro das redes geradora e discriminadora depois de 5 mil épocas.	64
23	Espessuras de RNFL produzidas pela rede geradora e convertidas em ima- gens	64
24	Gráfico do banco de dados normativo para faixa de 60 à 69 anos	64
25	Extração dos valores das espessuras de RNFL peripapilares das espessura sintéticas geradas	65
26	Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 10%, 50% e 100%. .	67
27	Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 2x, 10x, 20x e 50x. .	68
28	Valores de AROC dos classificadores para conjunto de dados de treina- mento (E1) e para diferentes quantidade de dados reais aumentados + dados sintéticos: 2x, 10x, 20x e 50x.	70
29	Conversão do imagem de espessura de RNFL para uma escala construída .	82
30	Imagem a partir da média das imagens de pacientes normal do conjunto de dados de treinamento	83

LISTA DE TABELAS

1	Funções de ativações	30
2	Mapas de espessuras do conjunto de dados de treinamento + dados sintéticos para os experimentos E1, E2, E3 e E4. Conjunto de dados de treinamento é composto por 275 mapas de espessura de RNFL.	66
3	Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 10%, 50% e 100%. .	66
4	Mapas de espessuras do conjunto de dados de treinamento + dados sintéticos para os experimentos E5, E6, E7 e E8. Conjunto de dados de treinamento é composto por 275 mapas de espessura de RNFL.	67
5	Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 2x, 10x, 20x e 50x. .	68
6	Mapas de espessuras do conjunto de dados de treinamento aumentado + dados sintéticos para os experimentos E9, E10 e E11. Conjunto de dados de treinamento é composto por 275 mapas de espessura de RNFL.	69
7	Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados reais aumentados + dados sintéticos: 2x, 10x, 20x e 50x.	70
8	Desempenhos dos classificadores no experimento E11. VN=Verdadeiros Negativos. FP=Falsos Positivos. FN=Falsos Negativos. VP=Verdadeiros Positivos. Acc=Acurácia. P= <i>Precision</i> . R= <i>Recall</i>	71

LISTA DE ABREVIATURAS

AUC Area Under the Roc Curve

CDR Cup to Disc Ratio

CGAN Conditional Generative Adversarial Network,

BN Batch Normalization

FC Fully Connected

FNN Feedforward Neural Network

GAN Generative Adversarial Networks

GPU Graphic Processing Unit

ILSVRC Large Scale Visual Recognition Challenge

IOP Intraocular Pressure

PACG Primary Angle Closure Glaucoma

POAG Primary Angle Open Glaucoma

OCT Optical Coherence Tomography

OMS Organização Mundial da Saúde

ROC Receiver Operating Characteristic

RNFL Retinal Neural Fiber Layer

SAP Standard Automated Perimetry

SVM Support Vector Machines

SD-OCT Spectral-Domain Optical Coherence Tomography

TD-OCT Time-Domain Optical Coherence Tomography

VCDR Vertical Cup to Disc Ratio

SUMÁRIO

1	Introdução	13
1.1	Definição do problema	14
1.2	Hipótese a ser testada	15
1.3	Objetivo da pesquisa	15
1.4	Metodologia de Pesquisa	16
1.5	Estrutura da Dissertação	16
2	Glaucoma e seu diagnóstico	17
2.1	Métodos de diagnóstico	19
2.1.1	Exames estruturais	21
3	Modelos de Aprendizagem de Máquina	26
3.1	Redes Neurais	27
3.1.1	Backpropagation	29
3.2	Redes Neurais Convolucionais	32
3.3	Redes Neurais Profundas	34
3.3.1	VGG	35
3.3.2	Inception	35
3.3.3	Resnet	36
3.4	Métodos de avaliação para classificadores	37

3.4.1	Curva ROC	39
3.5	Redes Generativas Adversárias	40
3.6	Redes Generativas Adversárias Convolucionais Profunda	43
3.7	Redes Generativas Adversárias Condicionais	44
3.8	Redes Generativas Adversárias Condicionais Auxiliares	44
4	Estado da arte do diagnóstico do glaucoma com métodos de aprendizagem de máquina	46
4.1	Diagnóstico a partir de dados do SAP	46
4.2	Diagnóstico a partir de dados do OCT	47
4.3	Diagnóstico através de Redes Neurais Convolucionais	48
5	Parte Experimental	51
5.1	Geração dos mapas de espessuras de RNFL sintéticos	53
5.1.1	Rede Geradora	54
5.1.2	Rede Discriminadora	57
5.1.3	Processo de treinamento da <i>GAN</i>	59
5.2	Treinamento dos classificadores para diagnóstico de glaucoma	61
6	Avaliação dos Resultados	63
7	Considerações Finais	72
7.1	Contribuições do Trabalho	74
7.2	Trabalhos Futuros	74
	Referências	76

A.1	Preprocessamento das imagens dos conjuntos de dados	81
-----	---	----

1 INTRODUÇÃO

Segundo Bourne et al. (2017), em 2015 mais de 216 milhões de pessoas no mundo tinham algum tipo de deficiência visual moderada ou severa e 36 milhões de pessoas eram afetadas pela perda total do campo de visão. O glaucoma é a segunda causa de cegueira no mundo, afetando parcialmente ou totalmente a visão de cerca de 76 milhões de pessoas em 2020, segundo a estimativa mais recente (THAM et al., 2014). A identificação de pacientes que sofrem de glaucoma não é trivial, sendo que mesmo em países desenvolvidos apenas 50% dos pacientes são diagnosticados com glaucoma. Em países em desenvolvimento e com baixo acesso a recursos de saúde, a porcentagem é ainda menor, restringindo-se a 10% (SAEEDI; RAMULU; FRIEDMAN, 2014). A identificação em estágios iniciais pode adiar ou retardar o avanço da doença (WEINREB; KHAW, 2004). Cerca de 86% dos pacientes com perda total de visão estão acima dos 50 anos, sendo a idade um fator importante de crescimento do número de casos (BOURNE et al., 2017). Conforme ocorre o envelhecimento da população, mais casos de glaucoma são detectados.

Identificar as causas do glaucoma é fundamental para detectar a doença precocemente e minimizar os riscos de perda total de visão. Alguns indicadores já são conhecidos na literatura, porém isoladamente não levam sempre ao diagnóstico. A elevação da pressão intraocular (IOP, Intraocular Pressure), a diminuição da espessura da camada de fibras nervosas, e perda da sensibilidade no campo visual são alguns desses indicadores (FOSTER et al., 2002).

Existem esforços no campo da computação para desenvolver métodos que antecipem o diagnóstico a fim de iniciar o tratamento adequado antes que perdas significativas no

campo visual sejam detectadas. Trabalhos como Vidotti et al. (2012) e Shigueoka et al. (2016) modelam a tarefa do diagnóstico do glaucoma como um problema de reconhecimento de padrões e aplicam técnicas de aprendizagem de máquina a fim de construir classificadores que permitam auxiliar o especialista na tomada de decisão. Outros trabalhos como Chen et al. (2015) e Asaoka et al. (2016) propõem a utilização de algumas técnicas da área de *Deep Learning*, uma sub-área da aprendizagem de máquina, no treinamento de classificadores para o diagnóstico do glaucoma. Com o aumento da quantidade de dados disponíveis (GOOGLE, 2006) e a utilização de GPUs para o treinamento de classificadores com várias camadas de abstrações (SCHMIDHUBER, 2015), a área de *Deep Learning* vem se destacando no estado da arte de problemas que envolvem o reconhecimento de imagens, voz e processamento de linguagem natural.

Por outro lado, nem todas as áreas contam com abundância de dados disponíveis. Custo ou tempo envolvidos tornam a coleta de dados inviável em alguns cenários. Em áreas, como por exemplo a física médica (ROGERS, 2006) é comum a utilização de técnicas, denominadas de métodos de Monte Carlo, que permitem a geração de dados artificiais na execução.

1.1 Definição do problema

Na oftalmologia, o diagnóstico e avaliação da progressão do glaucoma muitas vezes exigem um acompanhamento prolongado do paciente e tomadas de decisão subjetivas. Entretanto a literatura apresenta técnicas de Aprendizagem de Máquina com o intuito de fornecer ferramentas auxiliares para que especialistas possam identificar precocemente e minimizar os danos que podem surgir no campo visual decorrentes do glaucoma (KHALIL; KHALID; SYED, 2014). Em contrapartida, nos últimos anos a utilização de técnicas de *Deep Learning*, em especial as redes neurais convolucionais, vem obtendo resultados interessantes nas áreas de reconhecimento visual, reconhecimento de voz e processamento de linguagem natural, áreas onde o conjunto de dados relacionados possuem maior disponibilidade. Diferentemente na área da saúde, em cenários onde são realizados exames

complexos, a aquisição de dados pode ser mais lenta sendo frequente a publicação de trabalhos onde são utilizados conjuntos de dados na ordem de centenas de pacientes. Tais técnicas de aprendizagem de máquina, quando aplicadas a conjuntos de dados pequenos, podem originar classificadores com problemas de *overfitting* (GU et al., 2018).

1.2 Hipótese a ser testada

No contexto do diagnóstico de glaucoma, as características das redes neurais convolucionais podem ser úteis na geração de classificadores treinados através de conjuntos de dados com dados estruturais como a espessura da RNFL, obtendo valores de sensibilidade e especificidade superiores aos dos classificadores que não possuem estas características e que estão presentes na literatura. Conjuntos de dados pequenos podem acarretar problemas de *overfitting* em arquiteturas com muitas camadas, e o uso de conjunto de dados de pacientes suficientemente grandes gerados utilizando modelos generativos pode diminuir o problema de *overfitting*. Para validar esta hipótese os modelos gerados com tais técnicas devem ser comparados com modelos treinados a partir de dados de pacientes reais, avaliando seu desempenho com relação a sensibilidade e especificidade.

1.3 Objetivo da pesquisa

O objetivo deste trabalho foi analisar se o uso de dados sintéticos gerados com redes generativas adversárias melhoram o desempenho dos classificadores baseados em redes neurais profundas, para o diagnóstico de glaucoma. Foram analisadas diferentes configurações de conjunto de dados compostos por dados reais e dados sintéticos no treinamento desses classificadores. Dessa forma, contribuindo na redução dos custos e tempo envolvidos na aquisição de dados usados na geração de classificadores para o diagnóstico de glaucoma.

1.4 Metodologia de Pesquisa

A metodologia proposta para atender ao objetivo da pesquisa foi estudar o emprego de Redes Generativas Adversárias (*GAN*) na geração de conjuntos de dados sintéticos. O uso de *GANs* tem sido empregado na geração de imagens sintéticas como em Karras et al. (2017). De forma similar, nesta pesquisa é proposto a utilização de um tipo de *GAN* que permita condicionar o tipo de dado produzido por ela, ajudando a criar dados específicos para pacientes com glaucoma e para pacientes normais.

- Encontrar uma configuração de *GAN* que atenda as necessidades para a geração de dados sintéticos de pacientes com glaucoma e pacientes normais.
- Treinar classificadores baseados em redes neurais profundas com dados de pacientes reais.
- Analisar a qualidade dos dados produzidos pela *GAN* medindo o desempenho dos classificadores treinados a partir de conjunto de dados compostos por dados reais e dados sintéticos em diferentes proporções.

1.5 Estrutura da Dissertação

O capítulo 2 aborda o glaucoma e alguns dos métodos de diagnóstico utilizados na identificação da doença. O capítulo 3 é dedicado a uma breve introdução à aprendizagem de máquina, abordando os princípios das redes neurais artificiais e o funcionamento das redes neurais convolucionais. Além disso, são apresentados as redes generativas adversárias, além de algumas de suas variações que introduziram melhorias à arquitetura proposta originalmente. O capítulo 4 aborda trabalhos presentes na literatura para o diagnóstico de glaucoma através do uso de aprendizagem de máquina. Os experimentos realizados no trabalho são descritos no capítulo 5 enquanto a avaliação dos resultados é apresentada no capítulo 6. As considerações finais encerram o texto no capítulo 7.

2 GLAUCOMA E SEU DIAGNÓSTICO

Segundo Foster et al. (2002), o glaucoma é definido como uma neuropatia óptica que possui como característica danos estruturais ao nervo óptico, associado a disfunção visual, ou seja, a medida que as células ganglionares da região da cabeça do nervo óptico são perdidas, concomitantemente, acarretará perdas no campo visual.

Um parâmetro importante para a definição do Glaucoma é a pressão intraocular. Ela é regulada pela secreção e drenagem de uma substância incolor chamada humor aquoso, que preenche as câmaras anterior e posterior, fornecendo nutrientes para a íris, o cristalino e a córnea. Essas estruturas que compõem o olho são apresentadas na figura 1. O fluxo do humor aquoso inicia no corpo ciliar e termina sendo drenado pela malha trabecular (caminho principal) conforme é apresentado na figura 2.

Figura 1: Diagrama da anatomia do olho humano. Adaptado de (HUMAN..., 2012)

Este fluxo gerado entre a secreção e a drenagem cria uma pressão de aproximadamente $15,5 \text{ mmHg} \pm 2,6 \text{ mmHg}$ (média $\pm$ desvio padrão e mmHg, milímetro de mercúrio) no globo ocular em pacientes saudáveis. Alterações no processo de secreção ou drenagem fazem com que a pressão intraocular também se altere (GOEL RENATA G PICCIANI; BHATTACHARYA, 2010).

Figura 2: Fluxo de secreção e drenagem do humor aquoso. Adaptado de (GOEL RENATA G PICCIANI; BHATTACHARYA, 2010)

Existem dois tipos de glaucoma predominantes; o glaucoma primário de ângulo aberto (em inglês, Primary Angle Open Glaucoma - POAG) e o glaucoma primário de ângulo fechado (em inglês, Primary Angle Closure Glaucoma - PACG).

O glaucoma primário de ângulo aberto é caracterizado pelo vazamento do humor aquoso. Já o glaucoma primário de ângulo fechado é caracterizado pela obstrução do sistema de drenagem. Ambos os tipos podem apresentar perda da visão progressiva.

Ainda que o POAG seja mais comum que o PACG, os números de perda total de visão são aproximadamente iguais para os dois tipos, apesar de os fatores de risco não serem completamente entendidos (WEINREB; KHAW, 2004). Alguns fatores podem ser

relacionados com o surgimento e a progressão da doença, como por exemplo, fatores de risco demográficos que envolvem etnia, idade, gênero e histórico familiar e, fatores oculares que envolvem aumento da pressão intra-ocular (em inglês, Intraocular Pressure - IOP), desbaste da camada de fibras nervosas, miopia, hiperopia, entre outras patologias (SAEEDI; RAMULU; FRIEDMAN, 2014).

O glaucoma pode ser classificado como primário ou secundário, essa classificação varia de acordo com o conhecimento da causa da elevação da IOP. Se a causa for desconhecida o glaucoma é primário, se for conhecida é classificado como secundário (TAN; KAUFMAN, 2014). Mesmo que o aumento de IOP seja frequentemente um fator de risco elevado para o glaucoma, uma quantidade substancial de indivíduos que apresentam glaucoma não demonstram níveis altos de IOP, não sendo um critério absoluto para o diagnóstico da doença (MAREK; HARRIS; SIESKY, 2014).

Em quase todos os pacientes, os danos no disco óptico ocorrem antes do surgimento de dados ao campo visual (SOMMER et al., 1991). Mesmo nestes casos o diagnóstico pode ser feito antecipadamente e é denominado Glaucoma Pré-perimétrico (PPG, Preperimetric Glaucoma, em inglês)(OPHIR, 2010).

2.1 Métodos de diagnóstico

A priori, o diagnóstico se dá a partir de revisões do passado médico do paciente assim como o histórico e exames oculares funcionais e estruturais. O histórico médico pode destacar a progressão de alguns sintomas relacionados a algum tipo de glaucoma. Também podem ser identificadas drogas que estejam associadas a elevação do IOP, entre eles, os corticosteróides. O histórico ocular pode revelar fatores como miopia e hiperopia, cirurgias anteriores e traumas oculares, além da identificação de resistência a medicações para glaucoma. O histórico familiar também é considerado fator de risco para desenvolvimento do glaucoma, possibilitando as chances de serem encontrados descendentes com valor elevado de IOP (ALI; RAMOS-ESTEBAN; KATZ, 2014).

Existem inúmeros métodos que facilitam ao especialista verificar a saúde estrutural e funcional da visão e na identificação de doenças como o glaucoma. Alguns tipos de exame por imagem permitem o mapeamento e análise da estrutura ocular. Um exame comumente utilizado é a imagem do fundo do olho ou fundoscopia. Um exemplo de imagem gerada pela fundoscopia é apresentada na figura 3 e descreve alguns pontos da estrutura ocular de um paciente normal, como o disco óptico, estrutura verticalmente oval de cerca de 1,5 mm de diâmetro, constituído por uma depressão central chamada de escavação e a rima neural que a circula (WEINREB; KHAW, 2004).

Figura 3: Imagem do fundo do olho. Adaptado de Bourne (2006)

O aumento da pressão gerado por alterações no processo de drenagem e escoamento do humor aquoso levam ao aumento do tamanho da escavação e a diminuição da espessura da rima neural, e por consequência diminui a relação entre os dois, chamada de CDR (Cup to Disc Ratio). Uma variação dessa medida chamada de VCDR (Vertical Cup to Risc Ratio) calcula a razão entre os polos do disco óptico e é um bom indicador da saúde da rima neural (FOSTER et al., 2002). A evolução de um paciente com glaucoma ao decorrer de

sete anos é apresentada na figura 4. É possível observar alterações na área da escavação e o aumento na relação escavação/disco óptico principalmente na parte superior. Pacientes com CDR maior que 0,65 podem indicar glaucoma severo (MURTHI; MADHESWARAN, 2012).

Figura 4: Imagem do fundo do olho (BOURNE, 2006)

2.1.1 Exames estruturais

Pesquisas na literatura como Bowd et al. (2000) apontam para a diminuição das espessuras das fibras da cabeça do nervo óptico como um dos indicadores para o diagnóstico do glaucoma. Um exame capaz de avaliar essa estrutura é o OCT. Ele é um exame não invasivo e de tempo real que produz imagens com resoluções na escala de microns (WOLLSTEIN; SCHUMAN, 2014) e pode ser aplicado no diagnóstico de uma variedade de doenças oftalmológicas, permitindo gerar imagens das seções anterior e posterior do globo ocular (SCHUMAN et al., 2004) com facilidade.

O OCT foi primeiramente demonstrado por Huang et al. (1991) e inicialmente comercializado em 1996, sendo fabricado por empresas como Carl Zeiss Meditec (Cirrus HD-OCT) e Heidelberg Engineering GmbH (Spectralis OCT). Dois modelos estão disponíveis no mercado, TD-OCT (*Time Domain Optical Coherence Tomography*) e SD-OCT (*Spectral Domain Optical Coherence Tomography*).

É possível fazer uma analogia entre o TD-OCT e os exames por ultrassom B. No ultrassom as reflexões das ondas decorrentes do contato com o tecido ajudam a formar as imagens. Algo parecido ocorre no TD-OCT, porém ele utiliza um feixe de luz ao invés do som como meio (FUJIMOTO et al., 2000). Como a luz é muito mais rápida que o som,

não é possível ser capturada e analisada diretamente por dispositivos eletrônicos, então a forma como o TD-OCT capta as reflexões difere do ultrassom.

Para capturar e medir a interferência no feixe de luz provocada pelos tecidos, o TD-OCT conta com a ajuda da interferometria de baixa coerência. Esta técnica permite medir o atraso no tempo de retorno e a intensidade do feixe de luz que foi refletido pelo tecido sendo analisado, comparando-os com as características de um feixe de luz que percorre um caminho de tamanho e tempo conhecidos (SCHUMAN, 2008).

O funcionamento do TD-OCT, apresentado na 5, é caracterizado por um feixe de luz saindo de uma determinada fonte, e sendo dividido em dois por um divisor de feixe de luz. Enquanto parte do feixe (feixe de referência) segue até um espelho refletor, a outra metade (feixe de medição) é direcionada ao tecido sendo analisado. Ao refletirem são combinados novamente em um único feixe. Este feixe é direcionado a um fotodetector eletrônico capaz de extrair informações relacionadas ao feixe de medição. O TD-OCT executa varreduras axiais individuais, chamadas de A-scan, linearmente pelo tecido, analisando sua profundidade e gerando imagens de cortes transversais conhecidas como B-scan (SCHUMAN, 2008).

Figura 5: Esquema de funcionamento do OCT

Inicialmente comercializado em 2006, o SD-OCT realiza a análise da frequência do feixe de luz, gerando imagens com resoluções axiais superior ao TD-OCT. As distâncias do feixe analisadas pelo SD-OCT são codificadas em transformadas de Fourier das frequências da luz refletida pelo tecido, permitindo que mais dados sejam captados em um menor espaço de tempo, produzindo, assim, imagens com maior resolução. No TD-OCT o espelho de reflexão é móvel fazendo com que o feixe de referência tenha diferentes distâncias, enquanto isso, o SD-OCT possui um espelho estacionário, agilizando a captação dos dados, não dependendo da locomoção do espelho. Para armazenar as variações do espectro do feixe de luz, o SD-OCT substitui o fotodetector por um espectrômetro (YAQOOB; WU; YANG, 2005). As diferenças de resoluções do TD-OCT e do SD-OCT são apresentadas na figura 6.

Figura 6: Comparação entre TD-OCT (lado esquerdo) e SD-OCT (lado direito) do mapeamento do RNFL. Adaptado de Liu et al. (2014)

O OCT possui um banco de dados normativo que permite o especialista comparar as camada de fibras nervosas da retina do paciente com valores de RNFL para a população da faixa etária correspondente, apresentado na figura 7. O gráfico *RNFL Thickness* na figura exibe um comparativo da espessura da RNFL entre o indivíduo sendo avaliado, 1%, 5% e 95% da população correspondente a sua faixa de idade, assim é possível verificar se uma determinada região da RNFL possui uma espessura inferior à base de referência. Além disso, também é realizada uma divisão em 12 seções e apresentada uma média de cada seção em comparação do seu valor normativo (figura 7a). As 12 seções são novamente agrupadas produzindo médias por quadrantes Temporal, Nasal, Inferior e Superior na figura 7b. Uma média total da espessura da RNFL é exibida na figura 7c.

A tecnologia empregada no diagnóstico de glaucoma evoluiu permitindo análises cada vez mais detalhadas das camadas do tecido de fibras nervosas da retina. A seguir será apresentada uma introdução aos modelos de aprendizagem de máquina para posteriormente analisar seu uso para o diagnóstico de glaucoma principalmente quando aplicado a esse tipo de tecido.

Figura 7: Gráficos gerados pelo OCT. (a) Gráfico da média das espessuras divididas em 12 horas. (b) Média das espessuras dividido em 4 quadrantes. (c) Média de espessuras totais. (d) Gráfico de espessura de 256 pontos.

3 MODELOS DE APRENDIZAGEM DE MÁQUINA

Uma tarefa que o ser humano exerce com facilidade em seu cotidiano é identificar formas e padrões na natureza. A ciência tenta copiar e melhorar esse tipo de tarefa através de métodos computacionais, para que sejam aplicadas nas mais diversas áreas, seja nas previsões do mercado financeiro, ou na tomada de decisão na área da saúde. O conjunto dessas técnicas e métodos desenvolvidos pode ser chamado de aprendizagem de máquina (MURPHY, 2012). Uma definição clássica para aprendizagem de máquina vem de Mitchell (1997):

Um programa de computador **aprende** de uma experiência E com respeito a alguma classe de tarefas T e uma medida de desempenho P , se seu desempenho nas tarefas em T , medido através de P , melhora com a experiência E .

Ele exemplifica a aprendizagem no contexto de um jogo de damas, onde a tarefa T é jogar damas, a medida de desempenho P é a porcentagem de jogos vencidos contra um oponente e a experiência E é o ato de praticar vários jogos contra si mesmos.

Pode-se fazer a mesma associação no problema do diagnóstico do glaucoma, onde a tarefa T é identificar a presença ou ausência do glaucoma, a medida de desempenho P é a porcentagem de pacientes classificados corretamente e a experiência E é um conjunto de dados com as características dos pacientes e uma dada classificação para cada uma das instâncias desse conjunto.

Quando a resposta do programa de computador para uma determinada instância é

discreta como no caso da identificação do glaucoma (possui ou não possui), então este programa é chamado de **classificador**.

Existem algumas formas de medir o desempenho de um classificador, a fim de compará-lo com outros classificadores, ou até mesmo durante o processo de treinamento escolher quais parâmetros devem ser atualizados para que o classificador tenha um desempenho melhor. Acurácia, Ganho de Informação, Precisão e Recall são alguns destes métodos e são abordados na seção 3.4.

Encontrar o classificador que tem o melhor desempenho dentre uma longa lista de combinação de configurações para os parâmetros do classificador pode ser um problema devido ao alto número de possibilidades. Classes de algoritmos como *Greedy Search*, *Gradient Descent* e *Linear Programming* definem estratégias que permitem buscar em um espaço de hipóteses os classificadores que tenham os maiores desempenhos.

A aprendizagem de máquina é frequentemente dividida em duas classes principais. Na abordagem preditiva ou supervisionada, o objetivo é aprender a mapear uma entrada x em uma saída y , dado um conjunto de treinamento $D = \{(x_i, y_i)\}_{i=1}^N$, onde N é o número de exemplos do conjunto de treinamento. Cada x_i é um vetor K -dimensional que representa um conjunto de características, por exemplo, peso, altura e idade de uma pessoa ou pixels em uma imagem. O valor de y pode ser um valor real escalar, neste caso o problema em questão é conhecido como regressão. y também pode ser uma variável categórica de um conjunto finito de classes (por exemplo, masculino ou feminino), neste caso o problema é visto como problema de classificação. O segundo tipo de aprendizagem é denominada aprendizagem descritiva ou não supervisionada. Neste caso são somente dadas entradas $D = \{x_i\}_{i=1}^N$, a fim de localizar padrões de interesse (MURPHY, 2012).

3.1 Redes Neurais

Um dos tipos de representação que possuem grande importância para a aprendizagem de máquina e grande aplicabilidade na área da medicina diagnóstica (AMATO et

al., 2013) são as redes neurais artificiais. Conhecidas como *Feedforward Neural Network* ou *Multilayer Perceptions*, são inspiradas na forma como os neurônios dos mamíferos se organizam, processam e transferem informações.

Uma rede neural é uma estrutura constituída por unidades (também chamado de neurônio artificial) conectadas entre si e separadas por camadas. Normalmente uma rede neural é formada por **1 camada de entrada**, n **camadas intermediárias** e **1 camada de saída**. A quantidade de unidades nas camadas define a largura da rede, enquanto a quantidade de camadas define sua profundidade.

Apresentado por Rosenblatt (1957), o tipo mais simples de rede neural é denominado *perceptron* (ou também *Single Layer Perceptron*), permite aprender modelos lineares. Como pode ser observado na figura 8, possui somente uma camada intermediária com uma unidade, mapeando um vetor de entrada X (neste caso de duas dimensões) em uma saída binária y . É realizado o cálculo do produto escalar entre x e um vetor de pesos w ($w \cdot x$), conforme equação 3.2, onde M é o número de características do vetor de entrada x . O termo b é chamado de *bias* e permite que a função não fique presa na origem do sistema de coordenadas mesmo se o resultado do produto escalar for 0.

Figura 8: Perceptron

O valor do produto escalar então é passado para uma **função de ativação** (também chamada de **função de transferência**), que no perceptron, é utilizada uma função degrau que pode ser visualizada na equação 3.1.

$$f_w(x) = \begin{cases} 1, & \text{se } w \cdot x + b > 0 \\ 0, & \text{senão} \end{cases} \quad (3.1)$$

$$w \cdot x = \sum_{i=1}^M w_i x_i \quad (3.2)$$

O objetivo do algoritmo de aprendizagem do *perceptron* é encontrar os valores para o vetor de peso w de forma a minimizar uma função de erro $E(w)$, que neste caso é quantidade de classificações erradas feitas utilizando o vetor w , podendo ser calculada através da equação 3.3, onde f_w é a função binária que separa um conjunto A ($f_w(x) = 1$) de um conjunto B ($f_w(x) = 0$) (ROJAS, 2013).

$$E(w) = \sum_{x \in A} (1 - f_w(x)) + \sum_{x \in B} f_w(x) \quad (3.3)$$

O algoritmo de aprendizagem do *perceptron* sempre converge para um conjunto de dados que são **linearmente separáveis**. Ser linearmente separáveis significa que dois conjuntos de pontos A e B pertencentes ao mesmo espaço n-dimensional são separáveis se $\forall a \in A, \sum_{i=1}^M w_i a_i > 0$, e $\forall b \in B, \sum_{i=1}^M w_i b_i < 0$, (ROJAS, 2013).

Não conseguir separar dados que não são linearmente separáveis é uma das limitações do *perceptron*. Outra rede que permite tratar este problema é denominada *Feedforward Neural Network* (FNN), frequentemente é composta por uma ou duas camadas intermediárias, e um neurônio de uma camada é ligado aos neurônios da camada anterior através de conexões. O nome *Feedforward* é designado pois nesta arquitetura as informações fluem de uma ponta a outra da rede, começando pela entrada e cada conexão é somente utilizada uma vez, como é possível visualizar na figura 9. Diferente do *perceptron*, funções de ativação não lineares são comumente utilizadas ao invés de funções degrau. Na tabela 1 são descritas algumas destas funções.

3.1.1 Backpropagation

Da mesma forma que o *perceptron*, para que as *feedforward neural networks* o processo de treinamento da rede consiste em encontrar os valores da matriz w que minimizam a função de erro $E(w)$.

Figura 9: Rede Neural com duas camadas intermediárias

Tabela 1: Funções de ativações

	Função	Derivada	Imagem
Sigmoid	$f(z) = \frac{1}{1 + e^{-z}}$	$\frac{df(z)}{dx} = f(z)(1 - f(z))$	$[0, 1]$
Tangente Hiperbólica	$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$\frac{df(z)}{dx} = 1 - (f(z))^2$	$[-1, 1]$
Rectifier	$f(z) = \max(0, z)$		$[0, \infty]$

Então é preciso definir uma função de erro para a rede, como por exemplo, a equação 3.4 listada em (NG et al., 2013). Dada um conjunto de instâncias $\mathbf{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ podemos calcular o erro gerado através da equação 3.4, onde no primeiro termo, chamado de termo do erro da média dos mínimos quadrados, o w é o vetor de pesos da rede, b é o termo *bias*, $h_{w,b}(x^i)$ é a estimativa da rede para a instância i . O segundo termo é chamado de termo de regularização e minimiza o problema de *overfitting* que a rede possa apresentar, onde n_l é o número de camadas e s_l é o número de unidades da camada l . O termo λ permite indicar quão influente o termo de regularização é no erro.

$$E(w, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{w,b}(x^i) - y^i\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{j=1}^{s_l} \sum_{k=1}^{s_{l+1}} (w_{jk}^l)^2 \quad (3.4)$$

Segundo (NG et al., 2013), pode-se então atualizar os valores dos pesos de w , através

do método de gradiente descendente, onde cada um dos pesos w_{jk}^l é atualizado com relação a derivada parcial $\frac{\partial E(w)}{\partial w_{jk}^l}$, como é descrito na equação 3.5.

$$w_{jk}^l := w_{jk}^l - \alpha \frac{\partial E(w)}{\partial w_{jk}^l} \quad (3.5)$$

Rumelhart, Hinton e Williams (1988) sugeriu um método denominado *backpropagation* que permite calcular as derivadas parciais de forma eficiente através da regra da cadeia. O método calcula as derivadas parciais dos nós de saída um pouco diferente dos nós intermediários por já poder medir diretamente a diferença entre o que a rede produziu $a_i^{n_l}$ e o esperado y_i . Então para cada um dos nós de saída é calculado o $\delta_i^{n_l}$, de acordo com a equação 3.6.

$$\delta_i^{n_l} = \frac{\partial}{\partial z_i^{n_l}} \frac{1}{2} \|h_{w,b}(x) - y\|^2 = -(y_i - a_i^{n_l}) \cdot f'(z_i^{n_l}) \quad (3.6)$$

E para cada uma das unidade i de cada uma das camadas intermediárias l , o δ_i^l calculado é dado na equação 3.7, onde δ_j^{l+1} são as diferenças dos erros da camada da frente.

$$\delta_i^l = \left(\sum_{j=1}^{s_{l+1}} w_{ij}^l \delta_j^{l+1} \right) f'(z_i^{n_l}) \quad (3.7)$$

Após calcular os δ_j^{l+1} , é possível calcular as derivadas parciais, através da equação 3.8, permitindo assim reduzir a função de erro $E(w,b)$ repetindo os passos do gradiente descendente.

$$\frac{\partial}{\partial w_{ij}^l} E(w, b; x, y) = a_j^l \delta_i^{l+1} \quad (3.8)$$

3.2 Redes Neurais Convolucionais

A área de *Deep Learning* é a responsável por grande impacto no estado da arte do reconhecimento de padrões, principalmente na área de reconhecimento de imagens (SZEGEDY et al., 2016) e voz (HINTON et al., 2012). Uma das características fundamentais desse conjunto de técnicas é permitir que modelos consigam aprender representações (também conhecido como Aprendizagem de Representação (*Representation Learning*, em inglês) através de vários níveis de abstrações (LECUN; BENGIO; HINTON, 2015).

Uma rede neural é denominada convolucional quando pelo menos uma de suas camadas é formada por operações de convolução ao invés da multiplicação de matrizes que normalmente acontece (GOODFELLOW; BENGIO; COURVILLE, 2016). Aplicar convoluções as redes neurais não é algo realmente atual e foram inspirada por trabalhos como Hubel e Wiesel (1968), que estudaram a forma como os campos receptivos são reconhecidos no córtex cerebral de macacos. Um campo receptivo, quando relacionado a visão, pode ser definido como um espaço visual onde um estímulo de luz pode provocar respostas elétricas em células ganglionares da retina (HARTLINE, 1969). Diferentemente de uma rede neural que possui todas as entradas conectadas a todos os neurônios da primeira camadas e assim sucessivamente, o estudo de Hartline sugeriu que diferentes agrupamento de células reagem somente a regiões localizadas da imagem.

Apesar das rede neurais convolucionais já serem utilizadas mesmo quando as redes neurais não tinham muita evidência na área de reconhecimento de padrões, um dos trabalhos que ressaltaram a aplicação de redes neurais convolucionais foi em LeCun et al. (1998) através do emprego em reconhecimento de escrita manual, sendo inclusive aplicado comercialmente.

Uma convolução é uma operação matemática similar a operação de correlação. Pode-se pensar na correlação como uma operação realizada, por exemplo, em uma imagem I , onde para cada um dos itens I_{jk} e seus vizinhos é aplicado um filtro F , de tamanho $N \times N$, gerando uma nova matriz Y de tamanho $N \times N$. É realizada a soma dos itens da matriz

Y , e o valor escalar é o novo valor para o item X'_{jk} na nova imagem, que é chamada de *Feature Map*, como pode ser observado na figura 10 (JACOBS, 2005). A equação 3.9 define a operação de correlação. Segundo (JACOBS, 2005), uma das grandes utilidades da correlação é poder ser aplicada para identificar localidades na imagem que coincidam um *template*.

$$F \circ I(x, y) = \sum_{i=-N}^N F(i, y) I(x + i, y + j) \quad (3.9)$$

A convolução é basicamente uma correlação, porém o filtro é invertido (no caso do espaço R^2 é invertido horizontalmente e verticalmente), a fim de se ganhar a propriedade de associatividade ($F * (G * H) = (F * G) * H$). A convolução normalmente é definida com o símbolo $*$, como pode ser visto na equação 3.10.

$$F * I(x, y) = \sum_{i=-N}^N F(i, y) I(x - i, y - j) \quad (3.10)$$

Figura 10: Filtro (F) aplicado a imagem (I), gerando um *Feature Map*

Segundo Goodfellow, Bengio e Courville (2016), uma rede neural convolucional consiste de pelo menos 3 estágios. No primeiro estágio as camadas realizam várias convoluções, gerando um conjunto de ativações lineares que são passadas a funções não lineares como a função de ativação ReLU (segundo estágio). No terceiro estágio é empregada uma técnica

que diminui a dimensão dos parâmetros de saída, chamada de *pooling*. Existem diversas técnicas de *pooling*, que analisam a vizinhança de um determinado ponto e a partir de algum critério como o média ou o máximo, criar um novo mapa, reduzindo a quantidade de parâmetros através da sumarização. A função de *Max Pooling* seleciona o maior valor existente na vizinhança de um ponto do *Feature Map*, conforme apresentado na figura 11. Uma das características importantes do *pooling* é permitir que as representações geradas sejam invariantes a pequenas alterações de translação na entrada, uma propriedade bastante útil, principalmente em cenários onde saber se um determinado objeto existe em uma imagem é mais importante do que onde exatamente está localizado na imagem.

Figura 11: Max Pooling sendo aplicado em um Feature Map

3.3 Redes Neurais Profundas

Com o aumento do poder computacional, agora é possível treinar redes neurais com dezenas e centenas de camadas, chamadas de redes neurais profundas. O estado da arte para classificação de base de dados como *ImageNet* vem sendo repetidamente atualizado nesses últimos anos por este tipo de rede. A seguir são descritos alguns tipos de arquiteturas de rede neurais profundas utilizadas neste trabalho.

3.3.1 VGG

A rede VGG (SIMONYAN; ZISSERMAN, 2014) é um rede convolucional (ConvNet) que chamou bastante atenção devido aos resultados de acurácia nos testes de classificação e localização do *ILSVRC* (*Large Scale Visual Recognition Challenge*, em inglês). A rede possui uma variedade de configurações, mas as mais conhecidas pelas configurações VGG16 e VGG19. Sua arquitetura base é composta por uma pilha de camadas convolucionais, utilizando filtros pequenos de 3x3. Cada pilha de camada convolucional é entreposta com uma camada de max-pooling com janelas de 2x2 e stride de 2. A diferença entre a VGG16 e a VGG19 é quantidade de camadas de convolução entre as pilhas, sendo 16 e 19 camadas, respectivamente. Depois das pilhas de convolução e camadas de max-pooling, são adicionadas 3 camadas totalmente conectadas contendo 4096 nós, 4096 nós e 1000 nós, respectivamente. Esta última camada é relacionada ao número de classes do conjunto de dados do *ILSVRC*. Então esta camada é ligada a uma camada de soft-max.

3.3.2 Inception

Uma das tarefas necessárias nas redes convolucionais é escolher as configurações das camadas de convolução como, por exemplo, o tamanho de cada filtro. Filtros grandes são melhores para capturar informações globais enquanto filtros menores são melhores para capturar informações locais. A rede *Inception* (Szegedy et al., 2015), também conhecida como *GoogLeNet*, é uma rede convolucional que introduziu o conceito de blocos de *Inception*. Cada bloco é composto por um conjunto de filtros de vários tamanhos diferentes. Eles funcionam em paralelo e suas saídas são concatenadas umas as outras formando um novo volume como pode ser visto na figura 12. Assim não é necessário escolher os tamanhos dos filtros. Com essa configuração a rede *Inception* também é responsável por obter bons resultados de acurácia nos testes do *ILSVRC*.

Esta rede ainda possui mais duas versões com melhorias. A *Inception V2* (IOFFE; SZEGEDY, 2015) inclui o conceito de normalização de *batch*. Já a *Inception V3* (SZE-

Figura 12: Bloco da rede *Inception*

GEDY et al., 2016) inclui o conceito de fatoração de convolução com o intuito de diminuir o número de parâmetros da rede sem perder sua eficiência. Ambas levaram à melhora tanto da acurácia de classificação como do custo computacional associado.

3.3.3 Resnet

Apesar de parecer contradição, nem sempre a adição de camadas nas *ConvNets* tradicionais resulta em melhor desempenho. Quando este tipo de rede aumentam consideravelmente o número de camadas, ocorre problemas associados ao algoritmo de gradiente descendente. Um desses problemas é o desaparecimento do gradiente. Ocorre quando os valores do gradiente ficam tão pequenos que a rede não consegue mais se atualizar. Outro problema conhecido é a explosão do gradiente onde valores muito grandes desestabilizam o treinamento da rede.

A rede *ResNet* (He et al., 2016) aborda estes problemas, sugerindo a introdução de blocos residuais. A ideia principal é ligar camadas anteriores diretamente a camadas posteriores, criando uma espécie de atalho. Tanto que o nome dado a essa técnica é conexão de atalho de identidade (*identity shortcut connection*, em inglês) como pode ser visto na figura 13.

Figura 13: Conexão de atalho de identidade

3.4 Métodos de avaliação para classificadores

A avaliação é um componente fundamental para a comparação de classificadores, e a literatura apresenta alguns métodos para medir o desempenho de um determinado classificador ou realizar a comparação entre um grupo de classificadores.

Como exemplo inicial, considera-se um problema de classificação binário, onde um determinado modelo precisa classificar um paciente entre as classes **{Glaucomatoso, Normal}**. Há quatro cenários possíveis listados na figura 14 para a resposta que o modelo possa dar e a situação real do paciente. Esta tabela é chamada de matriz de confusão. O modelo pode classificar um paciente que realmente possui glaucoma em Glaucomatoso (**Verdadeiro Positivo**), ou Normal (**Falso Negativo**). No caso em que o paciente seja normal, o modelo pode classificá-lo como Normal (**Verdadeiro Negativo**) e Glaucomatoso (**Falso Positivo**).

Segundo Bradley (1997), apesar da matriz de confusão possuir todos os dados referentes ao desempenho do classificador, é possível extrair medidas mais significativas utilizando os valores obtidos que podem ser calculadas da seguinte forma:

$$\text{Acurácia (1 - Erro)} = \frac{\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}}{\text{Número Total}} \quad (3.11)$$

		Situação Real do Paciente		
		Glaucomatoso	Normal	
Classificação realizada pelo modelo	Glaucomatoso	Verdadeiros Positivos	Falsos Positivos	Total de positivos
	Normal	Falsos Negativos	Verdadeiros Negativos	Total de negativos
		Total de Verdadeiros	Total de Falsos	Número Total de Pacientes

Figura 14: Matriz de confusão para um classificador binário

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Total de Positivos}} \quad (3.12)$$

$$\text{Sensibilidade} = \text{Recall} = \text{TP Rate} = \frac{\text{Verdadeiros Positivos}}{\text{Total de Verdadeiros}} \quad (3.13)$$

$$\text{FP Rate} = \frac{\text{Falsos Positivos}}{\text{Total de Falsos}} \quad (3.14)$$

$$\text{Especificidade} = 1 - \text{FP Rate} = \frac{\text{Verdadeiros Negativos}}{\text{Total de Falsos}} \quad (3.15)$$

A acurácia, que é a razão da taxa de acertos pelo total de instâncias de um determinado algoritmo, é uma das medidas de desempenho que podem ser aplicadas e que, aparentemente por simplicidade, se apresenta consistente entre modelos.

Porém em um cenário onde a distribuição das classes no conjunto de dados que se deseja classificar é desigual, a acurácia não se mostra uma medida de desempenho boa. Por exemplo, dado um *dataset* com 99 instâncias da classe *A* e uma instância da classe *B*,

um classificador que sempre aponta qualquer instância como sendo da classe A resultaria em uma acurácia de 99%, uma medida enviesada pela distribuição do conjunto de dados. Do mesmo modo, outras medidas de desempenho como Recall, Precisão e Fator F podem sofrer de algum viés (POWERS, 2011).

3.4.1 Curva ROC

Segundo Fawcett (2006), os gráficos de ROC (*Receiver Operating Characteristics*) permitem visualizar classificadores baseados em seus valores de especificidade e sensibilidade. Apesar de inicialmente ser empregada na análise de processamento de sinais, hoje possui uma vasta literatura na área da saúde (ZOU, 2007).

Gráficos ROC exibem o *tradeoff* entre obter uma determinada taxa de verdadeiros positivos (*TP Rate*, em inglês) mas como consequência obter um determinado valor para a taxa de falsos positivos (*FP Rate*, em inglês). No eixo Y, ficam localizados valores de *TP Rate*, enquanto no eixo X, fica localizado valores de *FP Rate* (1 - especificidade).

Cada ponto no gráfico representa o desempenho de um determinado classificador mediante a execução em um *dataset*. O ponto (0,0) representa o desempenho onde o classificador é o mais restritivo possível, onde a taxa de falsos positivos é 0. Nesse cenário o modelo classifica os casos verdadeiros como negativos também, tornando a taxa positivos verdadeiros 0. Já no outro extremo do gráfico fica o cenário menos restritivo possível, (1,1), onde o modelo tem uma alta taxa de verdadeiros positivos. O ponto ótimo de gráfico é em (0,1), onde a taxa de falsos positivos é 0, e a taxa de verdadeiros positivos é 1. Um ponto é melhor do que outro no gráfico se ele está a mais noroeste (FAWCETT, 2006). A linha pontilhada no meio do gráfico representa um desempenho de aleatoriedade. Pontos abaixo desta linha possuem desempenho inferior a classificar o *dataset* aleatoriamente.

Apesar disso, os pontos isolados de variação de parâmetros de um classificador no gráfico ainda não exibem um valor escalar unitário que possa ser usado para realizar comparação entre classificadores.

Ao interpolar os pontos de um classificador no gráfico, é gerada uma curva que vai do ponto (0,0) ponto (1, 1), conforme observado no gráfico do lado direito na figura 15. O cálculo da área abaixo desta curva, AUC (AUC, Area Under the Roc Curve), proposto por Bradley (1997), frequentemente é utilizado como valor escalar unitário de desempenho de um classificador na comparação com outros.

Figura 15: Lado esquerdo: Pontos de execuções com diferentes configurados para um mesmo classificador. Lado direito: Interpolação de pontos de um classificador

3.5 Redes Generativas Adversárias

Atualmente existe um grande interesse em estudar modelos generativos devido sua capacidade de aprender a distribuição de conjuntos de dados e conseguir medir ou extrair amostrar dessas distribuições, mesmo estas tendo um número grande de dimensões. Esta característica dos modelos generativos permite que sejam aplicados em diversas áreas. No computação gráfica é possível encontrar aplicações de modelo generativos para gerar imagens em alta resolução baseado em imagens de baixa resolução como em Ledig et al. (2017). Um desses modelos que ganhou bastante importância para o campo dos aprendizagem de máquina são as Redes Generativas Adversárias (*Generative Adversarial Networks*, GAN, em inglês).

Uma rede generativa adversária é composta por duas funções diferenciáveis. Normal-

Figura 16: Rede generativa adversária

mente uma rede neural é usada em ambas as funções. A rede discriminadora é treinada usando técnicas de aprendizagem supervisionada e é responsável por identificar se um dado apresentado a ela pertence ao conjunto de dados (real) ou não (falso). Já a função da rede geradora G é criar amostras que com características estruturais que seguem a distribuição dos dados reais a fim de enganar a rede discriminadora D . A relação da rede geradora com a rede discriminador durante o treinamento é apresentada na figura 16.

Cada uma das redes possui um conjunto de parâmetros a serem otimizados de acordo com as seguintes funções de perda, $J^D(\Theta^D, \Theta^G)$ para a rede discriminadora, e $J^G(\Theta^D, \Theta^G)$ para a rede geradora. Porém as redes devem somente controlar seus respectivos parâmetros. A função de perda da rede discriminadora pode ser dada pela equação 3.16.

$$J^D(\Theta^D, \Theta^G) = -\frac{1}{2}E_{x \sim p_{data}} \log D(x) - \frac{1}{2}E_z \log(1 - D(G(z))) \quad (3.16)$$

Quanto mais a rede geradora conseguir produzir dados que enganam a rede discriminadora, mais acertos ela estará fazendo. Sendo assim a função de perda da rede geradora pode ser dada da seguinte forma:

$$J^G = -J^D \quad (3.17)$$

Este cenário pode ser melhor descrito como um jogo de soma zero onde os dois jogadores dependem dos parâmetros um do outro, mas só podem controlar o seus próprios (GOODFELLOW, 2017). Sendo assim o jogo pode ser descrito usando uma função de valor para um problema de otimização:

$$V(\Theta^{(D)}, \Theta^{(G)}) = -J^{(D)}(\Theta^{(D)}, \Theta^{(G)}) \quad (3.18)$$

Como a solução envolve uma minimização seguida de uma maximização, também é chamado de jogo *minimax* e pode ser descrito usando a seguinte equação para a função geradora.

$$\Theta^{(G)*} = \arg \min_{\Theta^{(G)}} \max_{\Theta^{(D)}} V(\Theta^{(D)}, \Theta^{(G)}) \quad (3.19)$$

Expandindo a função de minimax, que também pode ser chamadas de problema de minimax adversário temos a equação 3.20.

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (3.20)$$

No processo de treinamento da *GAN*, tanto a rede geradora quando a rede discriminadora são treinadas utilizando normalmente técnicas de gradiente descendente para achar o menor valor na função de perda. O algoritmo 1 apresenta lógica de treinamento para uma *GAN*. Normalmente os pesos das redes são atualizadas separadamente, onde os pesos da rede discriminadora são atualizadas primeiro utilizando amostras do conjunto de dados reais e amostras do conjunto de dados gerado pela rede geradora. Num próximo passo é realizado a atualização dos pesos da rede geradora utilizando as funções de perda discutidas anteriormente.

Algorithm 1 Treinamento de G e D através de *backpropagation*

```

1: for numero de epocas do
2:   for  $x$  em Minibatches( $X$ ) do
3:      $m \leftarrow \text{tamanho}(x)$ 
4:      $z_1 \leftarrow m$  amostras de  $p_g \sim \mathcal{N}(0, 1)$ 
5:      $\theta_D \leftarrow \Delta_{\theta_d} \frac{1}{m} \sum_1^m [\log D(x^i) + \log 1 - D(G(z_1^i))]$ 
6:      $z_2 \leftarrow m$  samples from  $p_g \sim \mathcal{N}(0, 1)$ 
7:      $\theta_G \leftarrow \Delta_{\theta_g} \frac{1}{m} \sum_1^m \log(1 - D(G(z_2^i)))$ 
8:   end for
9: end for

```

Apesar dos vários tipos de arquiteturas de GANs serem surpreendentes, existem algumas dificuldades referente ao seu processo de treinamento. De acordo com Goodfellow (2017), um dos maiores problemas com o treinamento das *GANs* é a não convergência. Devido o problema de otimização envolver dois jogadores, um cenário ótimo seria quando um jogador melhora seus parâmetros e melhorando o equilíbrio com o outro jogador. Porém na pratica quando um jogador melhora desfazendo o progresso do outro jogador causando oscilações na GAN.

Um tipo bastante conhecido de não convergência é chamado de *mode collapse* e se refere à situação em que a rede geradora começa a mapear diferentes valores da variável ruído z para um mesmo valor de saída. Esse situação faz com que a rede geradora sempre gere a mesma imagem para um subconjunto de valores de z (parcial) ou para qualquer valor de z (total). Enquanto o *mode collapse* total é raro, o parcial acontece frequentemente.

3.6 Redes Generativas Adversárias Convolucionais Profunda

Para tentar tratar os problemas citados anteriormente especialmente no aprendizado não supervisionado de imagens, Radford, Metz e Chintala (2015) introduziu uma arquitetura com algumas diferenças em relação à *GAN* original. A arquitetura proposta baseia-se quase que exclusivamente no uso de camadas convolucionais com *stride* em vez de camadas totalmente conectadas, unindo assim o conceito de *CNN* e *GANs*. Outra

alteração adicionada foi a normalização de *batch* tanto na rede geradora como na rede discriminadora melhorando a estabilidade do aprendizado. Por fim o uso da função de ativação *ReLU* na maioria das camadas da rede geradora, além da função *LeakyReLU* nas camadas da rede discriminadora mostraram-se melhor do que o uso da função de ativação *maxout* em Goodfellow et al. (2014).

3.7 Redes Generativas Adversárias Condicionais

Apesar das redes generativas adversárias conseguirem produzir dados similares à distribuição dos dados original, não é possível informar a rede geradora qual classe ele deve produzir. As redes generativas adversárias condicionais (*Conditional Generative Adversarial Network*, *cGAN*, em inglês) descritas por Mirza e Osindero (2014) tentam resolver este problema fornecendo mais informações para a rede geradora além da variável ruído Z , permitindo condicionar a rede geradora em informações extras y . Assim é possível induzir a rede geradora a produzir um certo tipo de dado baseado nas características fornecidas por y . A estrutura de uma rede generativa adversária condicional é apresentada na figura 17. Já a equação de *minimax*, equação 3.21 é alterada para condicionar levar em consideração a informações condicionais em y .

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log (1 - D(G(z|y)))] \quad (3.21)$$

3.8 Redes Generativas Adversárias Condicionais Auxiliares

Uma outra variação de *GAN* é chamada de rede generativas adversárias condicionais auxiliares (Auxiliary Classifier GAN, AC-GAN, em inglês). Nesse tipo de *GAN* a rede discriminadora além de fornecer uma distribuição de probabilidade que permite avaliar se um dado é real ou falso, também gera uma distribuição de probabilidade sobre as classes alvo. Isso permite que as classes do conjunto de dados real sejam utilizadas como auxílio

Figura 17: Rede generativa adversária condicional

no aprendizado da rede discriminadora. Assim como na *cGAN*, para este tipo de *GAN*, a rede geradora recebe como entrada uma variável ruído Z e uma variável condicional $c \sim p_c$ que influencia a rede geradora a produzir um determinado tipo de classe (ODENA; OLAH; SHLENS, 2017). A função objetivo para AC-GAN também tem duas partes, L_S (equação 3.22) para a classificação entre verdadeiro ou falso e L_C (equação 3.23) para as classes (por exemplo, se é glaucoma ou normal).

$$L_S = E[\log P(S = \text{real}|X_{\text{real}})] + E[\log P(S = \text{fake}|X_{\text{fake}})] \quad (3.22)$$

$$L_C = E[\log P(C = c|X_{\text{real}})] + E[\log P(C = c|X_{\text{fake}})] \quad (3.23)$$

O objetivo da rede discriminadora é maximizar $L_S + L_C$ enquanto que o trabalho da rede geradora é maximizar $L_C - L_S$.

4 ESTADO DA ARTE DO DIAGNÓSTICO DO GLAUCOMA COM MÉTODOS DE APRENDIZAGEM DE MÁQUINA

Alguns trabalhos presentes na literatura apresentam métodos computacionais, que incluem técnicas de Aprendizagem Computacional, a fim de permitir uma maior probabilidade de diagnóstico precoce, além de fornecer modelos para a avaliação da progressão ou estabilização de danos estruturais e funcionais. A seguir são listados alguns deles.

4.1 Diagnóstico a partir de dados do SAP

Bizios, Heijl e Bengtsson (2007) compara o desempenho de redes neurais treinadas utilizando dados de testes do campo visual gerados a partir do SAP (*SITA Standard 30-2*) de um grupo de 100 pacientes com glaucoma leve e moderado e 116 pacientes considerados normais. A AUC obtida foi de 0,984 com sensibilidade de 93% e especificidade de 94%. A estrutura da rede é descrita em Bengtsson, Bizios e Heijl (2005), sendo 74 unidades de entrada referentes ao padrão 30-2, 2 camadas intermediárias, sendo a primeira de 25 unidades e a segunda de 5 unidades e uma unidade de saída com uma função logística. A ANN (rede neural artificial) foi comparada com outros 5 algoritmos convencionais de avaliação de valores de GHT e PSD e apesar de apresentar acurácia maior, não foi estatisticamente significativa.

4.2 Diagnóstico a partir de dados do OCT

Burgansky-Eliash et al. (2005) e Barella et al. (2013) analisaram a utilização de dados do SD-OCT no treinamento de classificadores na tarefa de identificação de pacientes com glaucoma. No primeiro trabalho são analisados os desempenhos de alguns algoritmos de classificação, como Linear Discriminant Analysis, Generalized Linear Model, Generalized Additive Model, Support Vector Machine e Recursive Partitioning and Regression Tree. Dados de 47 pacientes com glaucoma e 42 sem glaucoma foram utilizados e 38 parâmetros gerados a partir da macula, cabeça do nervo óptico e camada de fibras nervosas peripapilar utilizando o SD-OCT (Stratus OCT). A maior AUC com 0,981 foi obtida utilizando Support Vector Machine e um subconjunto de 8 parâmetros que envolvem a área da rima neural e VCDR, entre outros. Enquanto no segundo trabalho, são apresentados resultados para um conjunto de dados de 103 pacientes, onde são testados 10 algoritmos de classificação, utilizando diferentes atributos fornecidos pelo SD-OCT. Os atributos que obtiveram maiores valores de AUC são a razão da área da relação escavação/disco com $AUC = 0,846$ e a média da relação escavação/disco com $AUC = 0,843$, sendo que o classificador gerado utilizando Random Forest obteve a maior AUC com 0,877.

Silva et al. (2013) analisou a eficiência de algoritmos de aprendizagem de máquina aplicados a dados do SD-OCT e SAP em conjunto, medida também através de AUC. O algoritmos selecionados foram *Bagging*, *Naive-Bayes*, *Multilayer Perceptron*, *Radial Basis Function*, *Random Forest*, *Ensemble Selection*, *Classification Tree*, *Ada Boost M1*, *Support Vector Machine Linear* e *Support Vector Machine Gaussiano* implementados no ambiente Weka. Os atributos selecionados do SD-OCT foram espessura média, 4 quadrantes e as medidas de 12 horas. Enquanto do SAP foram utilizados os atributos MD, PSD e GHT. Os dados foram coletados de 110 olhos de pacientes, deste 62 com glaucoma e 48 sem glaucoma. Uma seleção de atributos foi realizada, dentre os 20 atributos iniciais de modo a obter a maior AUC e o melhor resultado obtido utilizando OCT+SAP foi através do algoritmo Random Forest com 4 atributos resultado em uma AUC igual a 0,946, não

significativamente diferente do resultado obtido com somente o melhor parâmetro do SAP, AUC igual a 0,917 ($p=0,37$).

Devido a dificuldade na coleta de dados para treinamento dos classificadores, como pode ser notado no trabalho de Khalil, Khalid e Syed (2014), Dias (2009) recorreu ao uso de simulação a fim de gerar dados de treinamento representativos de populações com e sem glaucoma, podendo ser utilizado para o treinamento de novos classificadores. O software desenvolvido, denominado *Glaucoma Oracle* (GLOR) recebe como entrada parâmetros de distribuições estatísticas de variáveis funcionais e estruturais, além de dados epidemiológicos e clínicos e através de um método de Monte Carlo desenvolvido, gera uma quantidade também parametrizável de instâncias amostrais de pacientes normais e glaucomatosos acompanhando a distribuição estatística configurada. Equipamentos como OCT e o SAP possuem bancos de dados normativos, e para permitir o GLOR gerar dados semelhantes destes aparelhos, os respectivos bancos de dados foram reconstruídos a partir de exames de exemplo. Nos experimentos realizados foram testados alguns classificadores utilizando dados gerados (cerca de 5000 pacientes produzidos pelo GLOR), entre eles, Naive Bayes, Multilayer Perceptron e Decision Tree, Decision Tree with Bagging e os comparou com classificadores gerados com dados reais utilizando como método de avaliação gráficos AUC. Foram medidos 4 critérios: acurácia total, sensibilidade, especificidade e AUC. O melhor classificador avaliado utilizando os dados artificiais foi Multilayer Perceptron com AUC igual a 0,941.

4.3 Diagnóstico através de Redes Neurais Convolucionais

O trabalho de Chen et al. (2015) apresenta uma arquitetura de redes neurais convolucionais profundas para o diagnóstico da patologia utilizando como dados de treinamento imagens de fundo do olho disponíveis nas bases de dados ORIGA (168 imagens com glaucoma e 482 imagens normais) e SCES (46 imagens com glaucoma e 1676 imagens normais). A estrutura da rede consiste em 4 camadas convolucionais e 2 camadas total-

mente conectadas. A fim de diminuir overfitting no classificador, foram adotadas camadas de normalização de resposta e camadas de overlapping-pooling, além disso, para melhorar o desempenho da rede foram adotadas estratégias de regularização como "dropout" e data argumentation. A área do disco óptico foi automaticamente extraída para formar uma imagem menor que serve como entrada para a rede. Em um experimento 99 imagens aleatórias foram escolhidas da base de dados ORIGA para compor o conjunto de treinamento, enquanto 551 imagens foram utilizadas para teste. Em um outro experimento, a base de dados ORIGA foi utilizada com conjunto de treinamento e a base SCES foi utilizada como conjunto de teste. O desempenho de cada experimento foi medido através de AUC, igual a 0,831 e 0,887, respectivamente.

A pesquisa de Asaoka et al. (2016) apresenta um método de Aprendizagem Profunda para diagnóstico de Glaucoma Pré-perimétrico mesmo em campos visuais aparentemente normais, sugerindo que breves alterações no campo visual podem ser utilizadas no reconhecimento de casos de glaucoma. Foram utilizados 171 campos visuais de 51 pacientes selecionados entre 589 pacientes suspeitos e confirmados de OAG. No treinamento do classificador foi utilizada uma rede neural profunda com uma arquitetura de 2 camadas intermediárias com ativações através de função sigmoide em conjunto com pilhas de Autoencoders. Como entrada da rede são utilizados os 52 valores de campo visual, além dos valores de PSD e MD. O classificador gerado obteve $AUC = 0,926$ com intervalo de confiança de 95% para a rede neural profunda. Enquanto isso, classificadores como Random Forest, Gradient Boosting, SVM e Neural Network aplicados aos mesmos dados obtiveram, respectivamente, 0,79, 0,776, 0,712, 0,667, todos com intervalo de confiança de 95%.

Bisneto, de Carvalho Filho e Magalhães (2020) utiliza redes generativas adversárias condicionais para a segmentação do disco óptico em uma das etapas para o diagnóstico de glaucoma. Foi utilizado como base a arquitetura da *cGAN Pix2Pix* (Isola et al., 2017). Para o processo de classificação, foram testados diferentes classificadores como *MLP*, *SMO* e *Random Forest*. O melhor valor para a AUC foi dado pelo Random Forest (0.77).

A maioria das pesquisas relacionadas ao emprego de aprendizagem de máquina para o diagnóstico de glaucoma utilizam base de dados de imagens de fundo de olho. A seguir será discutido a parte experimental onde o conjunto de dados utilizado é composto por dados de exames mais complexos como o *OCT* e que podem oferecer maiores detalhes sobre a camada de fibras nervosas da retina para os classificadores de aprendizagem de máquina.

5 PARTE EXPERIMENTAL

A parte experimental consistiu em gerar mapas de espessura de RNFL sintéticos, utilizando uma *GAN*, para o treinamento de redes neurais profundas, com o objetivo melhorar o desempenho dos classificadores para o diagnóstico de glaucoma.

Para realizar os experimentos, foram disponibilizados dois conjunto de dados que consistem em imagens de mapas de espessura de RNFL. O conjunto de dados de treinamento contém 137 imagens de mapas de espessuras de RNFL de pacientes com glaucoma e 138 imagens de mapas de espessuras de RNFL de pacientes normais. O conjunto de dados de validação contém 56 imagens de mapas de espessuras de RNFL de pacientes com glaucoma e 46 imagens de mapas de espessuras de RNFL de pacientes normais. O conjunto de dados de treinamento possui imagens de espessuras de RNFL dos dois olhos do paciente enquanto o conjunto de dados de validação possui imagens de mapas de somente um olho por paciente. Um exemplo de tipo de imagem de mapa de espessura de RNFL presente nos dois conjuntos de dados é apresentado na figura 18B. No exame do OCT os mapas de espessura de RNFL são acompanhados de uma escala de cor, apresentada na figura 18A, para auxílio na interpretação do mapa. Essa escala representa os valores das espessuras em microns, com intervalo entre 0 e 350 microns. Para ler os valores da escala com exatidão na imagem, foi realizado um pré-processamento tanto da escala quanto das imagens dos conjuntos de dados. Esse pré-processamento é descrito no apêndice A.1.

O fluxo de execução dos experimentos é apresentado na figura 19. A primeira fase, descrita na figura como *Treinamento GAN*, consiste no treinamento de uma *GAN* para a geração de imagens sintéticas de mapas de espessura de RNFL, utilizando o conjunto de

Figura 18: A: Escala de cores representando espessura em microns. B: Imagem de mapa de espessuras de RNFL gerado do OCT

dados de treinamento. Os detalhes de arquitetura da *GAN* e seu processo de treinamento são descritos na seção 5.1.

Na segunda fase, descrita na figura 19 como *Treinamento Classificador*, foram geradas imagens de mapas de espessura de RNFL utilizando a rede geradora da *GAN*. Essas imagens foram empregadas no treinamento de 8 tipos diferentes de arquitetura de redes neurais profundas. Durante o treinamento, foram utilizadas imagens de mapas de espessura de RNFL do conjunto de dados de treinamento e imagens de mapas de espessura de RNFL sintéticos gerados pela *GAN* em diferentes proporções. Posteriormente, os classificadores foram testados com o conjunto de dados de validação. As arquiteturas escolhidas para os classificadores e os processos de treinamento e teste são descritos na seção 5.2.

Todos os experimentos foram executados utilizando um computador com processador Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz hexa-core, 32GB de memória e com gpu Nvidia GeForce RTX 2080 Ti. Foram utilizados a biblioteca CUDA versão 10.1 e os frameworks Keras 2.2.5 e Tensorflow 1.13.1.

Figura 19: Fluxo de treinamento da GAN e do Classificador

5.1 Geração dos mapas de espessuras de RNFL sintéticos

Neste trabalho é investigado o uso de GANs para a geração de mapas de espessura de RNFL sintéticos empregadas no treinamento de classificadores para o diagnóstico de glaucoma. Uma das primeiras etapas necessárias foi a escolha da arquitetura da *GAN* a ser utilizada.

A arquitetura da *GAN* escolhida levou em consideração a necessidade de induzir, através de algum parâmetro, a rede geradora a produzir mapas de espessura de RNFL de pacientes com glaucoma ou de pacientes normais, já que o objetivo é aumentar o conjunto

de dados de treinamento de forma balanceada entre as classes glaucoma e normal. Um outro ponto importante levado em consideração foi a presença de imagens já previamente rotuladas como glaucoma ou normal nos dois conjuntos de dados. Essas informações poderiam ser utilizadas como *feedback* no treinamento da rede discriminadora da *GAN*. Por esses pontos levantados, a arquitetura da *GAN* utilizada nos experimentos foi baseada na arquitetura da *AC-GAN* (ODENA; OLAH; SHLENS, 2017). Assim foi possível condicionar a rede geradora para produzir mapas de espessura de RNFL de pacientes com glaucoma e pacientes normais e também utilizar a classificação presente nos conjunto de dados como auxílio a rede discriminadora durante o treinamento. A seguir são descritas as arquiteturas da rede geradora e discriminadora e os passos executados durante o treinamento da *GAN*.

5.1.1 Rede Geradora

A estrutura da rede geradora da *GAN* utilizada nos experimentos é apresentada na figura 20. Como entrada, a rede geradora recebe informação de duas variáveis: a variável latente Z e a variável condicional c . A variável latente Z é um vetor de 100 posições em que seus valores seguem a distribuição normal com média=0 e desvio padrão=1. A variável condicional c é um vetor de 2 posições que condiciona a rede para gerar mapas de espessura de RNFL de pacientes com glaucoma ($[0, 1]$) ou para gerar mapas de espessuras de RNFL de pacientes normais ($[1, 0]$). Os valores dessas duas variáveis então são concatenados e passados para uma camada totalmente conectada de 4096 nós com a função de ativação ReLU. A saída dessa camada é então reorganizada em um volume de dimensões 1024x2x2.

Logo em seguida são utilizadas 6 camadas de convolução transposta, também chamada de *Fractionally Strided Convolution* (DUMOULIN; VISIN, 2016), descritas na figura 20 como *ConvTrans*. Esse tipo de camada de convolução permite gerar *feature-maps* maiores do que os *feature-maps* de entrada, mantendo os padrões de conectividade da convolução. Todas as camadas de convolução transposta possuem dimensão 4x4 (*kernel size*) e utilizam *stride*=2 com *padding*=same. A configuração de *padding*=same faz com que

Figura 20: Arquitetura da rede geradora. PS é abreviação para *Padding=Same*. BN é abreviação de *Batch Normalization*. FC é abreviação de *Fully Connected* (camada totalmente conectada).

as dimensões dos *feature-maps* de saída sejam iguais às dimensões dos *feature-maps* de entrada, assumindo $stride=1$. Neste caso, como foi utilizado $stride=2$, as dimensões dos *feature-maps* diminuem pela metade. Dessa maneira, a quantidade de filtros vai diminuindo pela metade a cada nova camada de convolução transposta. A quantidade de filtros em cada uma das camadas de convolução transposta é baseada em Radford, Metz e Chintala (2015) e começa com 512 filtros, até a penúltima camada, com 32 filtros. A última camada de convolução transposta tem somente 1 filtro, pois a saída que desejamos para a rede geradora é um mapa de dimensões 128x128 que representa a espessura de RNFL. Todas as camadas de convolução transposta tiveram seus pesos inicializados com valores seguindo uma distribuição normal com média=0 e desvio padrão=0,02.

Como é mostrado na figura 20, com exceção da última camada de convolução transposta, todas as outras camadas de convolução transposta na rede geradora são seguidas de uma camada de normalização de *batch* (IOFFE; SZEGEDY, 2015), uma camada de função de ativação *LeakyReLU* e uma camada de *Dropout* (SRIVASTAVA et al., 2014). A adição das camadas de *Dropout* e normalização de *batch* entre as camadas de convolução transposta contribuiu para uma maior estabilidade da rede geradora durante o treinamento, assim como acontece em Radford, Metz e Chintala (2015). A taxa de *dropout* utilizada na camada de *Dropout* foi de 0,5. Houve uma melhora na qualidade das imagens utilizando a função de ativação *LeakyReLU* ao invés da função de ativação *ReLU*. Para a função de ativação *LeakyReLU* foi utilizado o coeficiente de inclinação negativa igual a 0,2.

Devido o uso da função de ativação tangente hiperbólica na última camada de convolução transposta, os valores da saída da rede geradora estão no intervalo $[-1, 1]$. Esse intervalo de valores também é utilizado na entrada da rede discriminadora. Porém para visualizar a imagem de mapa de espessura de RNFL a partir da saída da rede geradora, os valores produzidos por ela são mapeados para o domínio de $[0, 137]$ (escala utilizada no pré-processamento das imagens descrito no apêndice A.1).

5.1.2 Rede Discriminadora

A estrutura da rede discriminadora da *GAN* utilizada nos experimentos é apresentada na figura 21. A rede discriminadora recebe como entrada um mapa de espessuras de RNFL com dimensão 128x128 e com valores no intervalo $[-1, 1]$. O mapa de espessuras de RNFL passa então por uma camada que adiciona ruídos em seus valores com a finalidade de melhorar a estabilidade e diminuir a chance de ocorrer o problema de *mode collapse* durante o treinamento da *GAN*. Essa camada de ruído soma valores, que seguem uma distribuição gaussiana com média=0 e desvio padrão=0,10, ao mapa de espessura de RNFL dado como entrada da rede discriminadora.

Depois da camada de ruído, foram adicionadas 5 camadas de convolução, mostradas na figura 21 como 'Conv'. Todas essas camadas de convolução foram configuradas para terem o tamanho de filtro (*kernel size*) com dimensão 4x4 e *stride*=2. A primeira camada possui 16 filtros e a cada camada de convolução seguinte, essa quantidade de filtros é dobrado chegando até a ultima camada de convolução com 256 filtros. A configuração de *padding* utilizada aqui foi '*same*'. Todas as camadas de convolução da rede discriminadora são seguidas das seguintes camadas: camada de normalização de *batch*, função de ativação *LeakyReLU* e camada de *Dropout*. Para a função de ativação *LeakyReLU* foi utilizado o coeficiente de inclinação negativa igual a 0,2. A taxa de *dropout* utilizada na camada de *Dropout* foi de 0,4. Os motivos para a adição das camadas de normalização e de *Dropout* são as mesmas discutidas na seção 5.1.1.

Por último, foram adicionadas duas funções de ativação em paralelo, formando as saídas da rede discriminadora, descritas na figura 21 como *S1* e *S2*. A saída *S1* é formada pela função de ativação *sigmoid*, descrita pela equação 5.1 e é destinada para indicar se o mapa de espessura de RNFL analisado pela rede discriminadora é real ou produzido pela rede geradora. A saída *S2* é formada pela função de *softmax* e é destinada para indicar se o mapa de espessura de RNFL é de um paciente com glaucoma ou de um paciente normal. A função de ativação *softmax* é descrita pela equação 5.2, onde *K* é o

Figura 21: Arquitetura da rede discriminadora. PS é abreviação para *Padding=Same*. BN é abreviação de *Batch Normalization*.

número de classes do conjunto de dados, que neste caso são duas, glaucoma e normal. Na arquitetura apresentada, a saída dessa função é um vetor binário que segue a representação *hot-encoding*, onde os valores $[0, 1]$ indicam que o mapa de espessura de RNFL é de um paciente com glaucoma e os valores $[1, 0]$ indicam que o mapa de espessura de RNFL é de um paciente normal.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5.1)$$

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \quad (5.2)$$

5.1.3 Processo de treinamento da *GAN*

O treinamento da *GAN* consistiu em atualizar os pesos da rede geradora e da rede discriminadora utilizando o algoritmo de *backpropagation* durante 5.000 épocas. Em cada uma das épocas foram feitas atualizações de pesos da rede discriminadora e da rede geradora.

O conjunto de dados de treinamento, primeiramente foi dividido na proporção de 80% e 20%, onde 80% foram utilizados para treinamento da *GAN* e os 20% foram utilizados para validação durante o processo de treinamento. Estes 80% de mapas de espessura de RNFL foram dividido em *batches*. Para cada um dos *batches* de mapa de espessuras de RNFL do conjunto de dados de treinamento era também gerado um *batch* de mapas de espessura de RNFL sintéticos, utilizando a rede geradora. Esse *batch* sintético, produzido pela rede geradora, continha uma quantidade de mapas de espessura de RNFL de pacientes com glaucoma e mapas de espessura de RNFL de pacientes normais que seguiam as proporções de uma distribuição uniforme.

Para cada um dos *batches* de mapa de espessuras de RNFL do conjunto de dados de treinamento foi realizado os seguintes passos. Primeiro, o *batch* de mapas de espessura de RNFL do conjunto de dados de treinamento era usado para atualizar os pesos da rede

discriminadora. Logo em seguida era utilizado o *batch* de mapas de espessura de RNFL sintéticos para atualizar novamente os pesos da rede discriminadora. Evitando assim a mistura de dados reais e dados sintéticos em um mesmo *batch*. Depois dessas duas atualizações na rede discriminadora, os pesos da rede discriminadora eram congelados para que ocorresse a atualização dos pesos da rede geradora assim como descrito no algoritmo 1 do capítulo 3.

O treinamento da *GAN* utilizou *batches* de 128 amostras tanto para o conjunto de dados de treinamento quanto para os mapas de espessura de RNFL sintéticos criados pela rede geradora. *Batches* com tamanhos menores do que 128 faziam com que a *GAN* entrasse em *mode collapse* produzindo pouca variedade de mapa de espessura de RNFL para pacientes com glaucoma e para pacientes normais.

O algoritmo de otimização escolhido tanto para o treinamento da rede discriminadora quanto para o treinamento da rede geradora foi o *Adam* (KINGMA; BA, 2015), convergindo mais rapidamente que os algoritmos *SGD* e *RMSprop* testados nos experimentos. Os parâmetros utilizados no algoritmo de otimização *Adam* foram *taxa de aprendizado*=0,0001, *beta 1*=0,5, *beta 2*=0,999 e taxa de decaimento=0,0001. Tanto taxas de aprendizado como taxas de decaimento com ordens de grandeza maiores que esses valores tornavam a treinamento da *GAN* instável nos experimentos realizados.

Como a rede discriminadora possui duas saídas, exibidas na figura 21 como *S1* e *S2*, foi necessário escolher duas funções de erro, uma para cada saída. A função de erro escolhida para ambas as saídas foi a entropia cruzada, dada pela equação 5.3, onde *N* é a quantidade de amostras no *batch*. Essa função foi usada para as saídas *S1* e *S2* tanto durante o treinamento da rede discriminadora quanto durante o treinamento da rede geradora.

$$J(w) = \frac{1}{N} \sum_{n=1}^N [y_n \log(\hat{y}_n) + (1 - y_n) \log(1 - \hat{y}_n)] \quad (5.3)$$

A saída *S2* da rede discriminadora, que retorna se um mapa de espessura de RNFL

é de um paciente com glaucoma ou um paciente normal, ajudou a rede discriminadora a ter um *feedback* a mais durante o processo de treinamento. Porém, a saída S1, que indica se o mapa de espessura de RNFL é real ou falso, é mais importante como *feedback* para a *GAN* do que o auxílio da saída S2. Dessa maneira, foi adicionado um peso a cada uma das funções de erro. A função de erro da saída S1 recebeu peso 10 enquanto a saída S2 recebeu peso 1. A adição desses pesos melhorou a estabilidade das redes durante o treinamento.

Outro ponto que contribuiu para a redução do problema de *mode collapse* durante o treinamento da *GAN* foi a substituição dos rótulos dos mapas de espessura de RNFL, fixados em 1 para 'glaucoma' e 0 para 'normal', por valores aproximados. Por exemplo, o rótulo 1 foi substituído por valores no intervalo $[0,7; 1,2]$. A escolha desses valores, para cada um dos mapas de espessura de RNFL, seguiu uma distribuição uniforme. Para o rótulo 'normal' (0), ocorreu o mesmo processo no intervalo $[0; 0,3]$. Esse processo é chamado de *soft label* e foi aplicado para todos os mapas de espessura de RNFL passados para a rede discriminadora sejam eles produzidos pela rede geradora ou pertencente ao conjunto de dados de treinamento.

5.2 Treinamento dos classificadores para diagnóstico de glaucoma

Para comparar quantitativamente os mapas de espessura de RNFL produzidos pela *GAN*, foram escolhidos 8 tipos diferentes de arquitetura de redes neurais profundas. O objetivo dos experimentos foi testar o desempenho dos classificadores criados com diferentes proporções de mapas de espessura de RNFL reais e mapas de espessura de RNFL sintéticos produzidos pela *GAN*. Para avaliar o desempenho dos classificadores, foi utilizado o conjunto de dados de validação e como métrica de comparação entre os experimentos foi utilizado o valor de *AROC*.

As arquiteturas de rede neurais profundas escolhidas para os experimentos foram: *Den-*

seNet 201 (Huang et al., 2017), *Inception ResNet V2* (SZEGEDY et al., 2017), *Inception V3* (SZEGEDY et al., 2016), *ResNet 152 V2* e *ResNet 50* (He et al., 2016), *VGG 16* e *VGG 19* (SIMONYAN; ZISSERMAN, 2014) e *Xception* (CHOLLET, 2017).

Por se tratarem de redes com muitas camadas, optou-se por utilizar modelos pré-treinados com o conjunto de dados *ImageNet* e, a partir daí, atualizar os pesos dos classificadores através do treinamento com os mapas de espessura de RNFL. Por se tratarem de modelos pré-treinados com o conjunto de dados do *ImageNet*, foi necessário realizar um ajuste na última camada das arquiteturas citadas acima. O *ImageNet* possui 1.000 categorias diferentes, por isso a última camada possui 1.000 saídas, por estarem no formato de representação *hot-encoding*. Então para o caso binário do diagnóstico de glaucoma, esta última camada foi removida e foi adicionada uma camada com duas saídas, uma para glaucoma e outra para normal.

O algoritmo de otimização *SGD* foi utilizado no treinamento de todas as 8 redes, sendo que os parâmetros utilizados com o *SDG* foram: taxa-aprendizado=0.0001, momentum=0.9 e decaimento=0.0005.

Para a validação durante o treinamento, 20% do conjunto de dados de treinamento foi separado. O treinamento foi configurado para ocorrer no máximo em 500 épocas, mas caso o classificador não aumentasse o desempenho com relação a acurácia entre 30 iterações (maior que 0.001), o treinamento era interrompido.

No capítulo seguinte são discutidos os resultados obtidos adicionando os mapas de espessura de RNFL sintéticos ao conjunto de dados reais.

6 AVALIAÇÃO DOS RESULTADOS

A avaliação do desempenho das redes neurais profundas para o diagnóstico de glaucoma foi realizada medindo os valores de *AROC* para os conjuntos de dados formados por mapas de espessura de RNFL reais e sintéticos em diferentes proporções.

Para gerar os mapas de espessura de RNFL sintéticos utilizados nesses conjuntos de dados, a *GAN* foi treinada utilizando o conjunto de dados de treinamento. O gráfico das funções de erro durante o treinamento da *GAN* é apresentado na figura 22 e indica os erros obtidos em cada uma das 5 mil épocas de treinamento. Neste gráfico são exibidos tanto as funções de erro da rede discriminadora quanto da rede geradora para o conjunto de dados de treinamento e para o conjunto de validação formado por 20% do conjunto de dados de treinamento. O gráfico apresenta a instabilidade das redes durante as primeiras épocas, onde os erros variam muito de época para época. A partir da época 800, já é possível observar uma estabilidade entre a rede geradora e a rede discriminadora. Mesmo após a estabilidade, as redes continuaram convergindo, evidenciado pelas imagens da figura 23, onde ocorre a evolução da rede geradora, produzindo mapas de espessura de RNFL com as características estruturais presentes no conjunto de dados de treinamento. Mesmo depois da época 1000, ainda houve melhora na qualidade visual das imagens produzidas.

Para avaliar os mapas de espessura de RNFL produzidos pela rede geradora, foram extraídos os valores de espessura de RNFL de uma região circular de 3,4 mm de diâmetro com centro posicionado ao centro da imagem. Para montar o gráfico normativo por faixa etária foi extraído de um exame do OCT o gráfico relativo à faixa etária entre 60 e 69 anos presente na figura 24.

Figura 22: Funções de erro das redes geradora e discriminadora depois de 5 mil épocas.

Figura 23: Espessuras de RNFL produzidas pela rede geradora e convertidas em imagens

Figura 24: Gráfico do banco de dados normativo para faixa de 60 à 69 anos

Figura 25: Extração dos valores das espessuras de RNFL peripapilares das espessura sintéticas geradas

Foram gerados mapas de espessura de RNFL sintéticos utilizando a *GAN*. Dois destes mapas são apresentados na figura 25. A primeira imagem de mapa de espessura de RNFL foi gerada condicionando a rede geradora para produzir um mapa de espessura de RNFL de um paciente com glaucoma. A segunda imagem de mapa de espessura de RNFL foi produzida condicionando a rede geradora para produzir um mapa de espessura de RNFL de um paciente normal. Ao lado direito das duas imagens estão os gráficos normativos para a faixa de 60 a 69 anos onde a linha preta indica os valores de espessura extraídos das imagens à esquerda.

A avaliação quantitativa dos mapas de espessuras produzidos pela *GAN* se deu através da comparação do desempenho dos classificadores treinados com estes dados. O desempenho foi medido através da comparação do valor de AROC para a predição dos mapas de espessura de RNFL do conjunto de dados de validação. O treinamento dos classificadores foi realizado a partir de conjuntos de dados compostos por mapas de espessura de RNFL reais e sintéticos em diferentes proporções.

O experimento *E1* consistiu em treinar todas as 8 redes somente com mapas de es-

Tabela 2: Mapas de espessuras do conjunto de dados de treinamento + dados sintéticos para os experimentos E1, E2, E3 e E4. Conjunto de dados de treinamento é composto por 275 mapas de espessura de RNFL.

Sintéticos	E1: (0%)	E2: 10% (27 mapas)	E3: 50% (137 mapas)	E4: 100% (275 mapas)
Reais + sintéticos	275 mapas	302 mapas	412 mapas	550 mapas

Tabela 3: Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 10%, 50% e 100%.

Arquiteturas	E1	E2 (+10%)	E3 (+50%)	E4 (+100%)
DenseNet	0.797	0.726	0.793	0.812
Inception ResNet V2	0.788	0.732	0.802	0.774
Inception V3	0.767	0.778	0.717	0.774
ResNet 152 V2	0.779	0.812	0.773	0.785
ResNet 50	0.798	0.818	0.778	0.796
VGG 16	0.842	0.856	0.851	0.833
VGG 19	0.839	0.816	0.811	0.820
Xception	0.747	0.683	0.735	0.730

peçura de RNFL reais do conjunto de dados de treinamento. Logo em seguida foram adicionados mapas de espessura de RNFL sintéticos gerados pela *GAN* ao conjunto de dados de treinamento durante o treinamento das redes. Para o treinamento de cada classificador foram adicionados dados sintéticos ao conjunto de dados nas proporções de 10% (experimento E2), 50% (experimento E3) e depois 100% (experimento E4) da quantidade original de mapas de espessura de RNFL do conjunto de dados de treinamento, formando 3 novos conjuntos de dados. As quantidades de mapas de espessura de RNFL em cada um dos novos conjuntos de dados são apresentadas na tabela 2.

Os valores de AROC para os classificadores treinados nos experimentos E1, E2, E3 e E4 são apresentados na tabela 3 e podem ser comparados também na figura 26. Em relação ao experimento E1, que foi realizado integralmente com dados reais, alguns classificadores tiveram desempenho melhor isoladamente em algum experimento, como a *DenseNet* em E4 (+1,8%), mas não houve melhora significativa consistente entre os experimentos E2, E3 e E4.

Em seguida, analisou-se o desempenho dos classificadores quando o número de mapas de espessura de RNFL sintéticos é maior que a quantidade de mapas de espessura de

Figura 26: Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 10%, 50% e 100%.

Tabela 4: Mapas de espessuras do conjunto de dados de treinamento + dados sintéticos para os experimentos E5, E6, E7 e E8. Conjunto de dados de treinamento é composto por 275 mapas de espessura de RNFL.

Sintéticos	E5: 2x (550 mapas)	E6: 10x (2.750 mapas)	E7: 20x (5.500 mapas)	E8: 50x (13.750 mapas)
Reais + sintéticos	825 mapas	3.025 mapas	5.775 mapas	14.025 mapas

RNFL reais no conjunto de dados. Foram formados novos conjunto de dados adicionando mapas de espessura de RNFL sintéticos ao conjunto de dados de treinamento nas seguintes proporções: 2 vezes (experimento E5), 10 vezes (experimento E6), 20 vezes (experimento E7) e por último 50 vezes (experimento E8) a quantidade total de mapas de espessura de RNFL do conjunto de dados de treinamento. A tabela 4 apresenta a quantidade de mapas de espessura de RNFL dos novos conjuntos de dados.

Os valores de AROC obtidos nos experimentos E5, E6 e E7, juntamente com o valor de E1 são apresentados na tabela 5 e podem ser comparados na figura 27. Conforme o número de mapas de espessura de RNFL sintéticos aumentou, não houve melhora significativa

Figura 27: Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 2x, 10x, 20x e 50x.

nos valores de AROC. *Inception ResNet V2* (-21,60%) e *Xception* (-24,09%) tiveram piora significativa quando a quantidade dos dados sintéticos foram 50x maiores do que a quantidade de dados reais.

Por último foi investigado se mantendo a quantidade de mapas de espessura de RNFL do conjunto de dados de treinamento igual a quantidade de mapas sintéticos, resultaria em um desempenho melhor dos classificadores. Então os mapas de espessura de RNFL do conjunto de dados de treinamento foram repetidos até a quantidade de mapas reais ser

Tabela 5: Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados sintéticos: 2x, 10x, 20x e 50x.

Arquiteturas	E1	E5 (+2x)	E6 (+10x)	E7 (+20x)	E8 (+50x)
DenseNet	0.797	0.814	0.753	0.788	0.780
Inception ResNet V2	0.788	0.728	0.766	0.755	0.572
Inception V3	0.767	0.764	0.806	0.800	0.667
ResNet 152 V2	0.779	0.803	0.804	0.820	0.801
ResNet 50	0.798	0.795	0.829	0.763	0.814
VGG 16	0.842	0.842	0.835	0.835	0.854
VGG 19	0.839	0.840	0.807	0.851	0.778
Xception	0.747	0.755	0.789	0.722	0.567

Tabela 6: Mapas de espessuras do conjunto de dados de treinamento aumentado + dados sintéticos para os experimentos E9, E10 e E11. Conjunto de dados de treinamento é composto por 275 mapas de espessura de RNFL.

Sintéticos	E9: 10x (2.750 mapas)	E10: 20x (5.500 mapas)	E11: 50x (13.750 mapas)
Reais aumentados + sintéticos	5.500 mapas	11.000 mapas	27.500 mapas

igual a quantidade de mapas sintéticos dos experimentos E6, E7 e E8, formando novos conjunto de dados com 50% de dados reais e 50% de dados sintéticos. Foram escolhidos os experimentos E6, E7, e E8 por conterem as maiores quantidades de mapas entre os novos conjuntos de dados criados até agora. As quantidades de mapas de espessura de RNFL para estes 3 novos conjuntos de dados são apresentadas na tabela 6.

Os valores de AROC dos experimentos E9, E10 e E11 juntamente com o experimento E1 são apresentados na tabela 7 e também podem ser comparados na figura 28. Para a rede *DenseNet* quanto maior a quantidade de dados sintéticos e dados reais, maior foi o valor de AROC. Porém nem todos os classificadores se comportaram da mesma forma. Utilizando a rede *VGG16* houve piora nos experimentos E9 (-0,47%) e E10 (-0,95%) para os valores de AROC. Os melhores desempenhos foram notados utilizando os conjunto de dados com a maior quantidade de dados sintéticos e reais. O experimento E11 obteve os 3 maiores valores de AROC entre todos os experimentos, onde o maior valor de AROC foi a do classificador utilizando a rede *VGG16*, com aumento de 2,65% entre o experimento E11 e o E1.

Como o experimento com o maior número de mapas de espessura de RNFL, E11, foi o que apresentou a maior quantidade de melhores AROCs entre os experimentos, foram gerados, para cada um destes classificadores, os valores da matriz de confusão apresentados na tabela 8. Esses valores são correspondentes a predição do conjunto de dados de validação quando o rótulo positivo é igual 'glaucoma'. A rede *VGG16* foi a rede que mais identificou mapas de espessura de RNFL com glaucoma corretamente (48), com a menor quantidade de falsos positivos (8), tendo uma acurácia de 79,4%. Enquanto

Figura 28: Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados reais aumentados + dados sintéticos: 2x, 10x, 20x e 50x.

Tabela 7: Valores de AROC dos classificadores para conjunto de dados de treinamento (E1) e para diferentes quantidade de dados reais aumentados + dados sintéticos: 2x, 10x, 20x e 50x.

Arquiteturas	E1	E9 (+10x)	E10 (+20x)	E11 (+50x)
DenseNet	0.797	0.807	0.808	0.815
Inception ResNet V2	0.788	0.769	0.780	0.829
Inception V3	0.767	0.835	0.806	0.818
ResNet 152 V2	0.779	0.805	0.786	0.793
ResNet 50	0.798	0.828	0.781	0.792
VGG 16	0.842	0.838	0.834	0.865
VGG 19	0.839	0.797	0.829	0.824
Xception	0.747	0.729	0.730	0.770

isso a rede *Inception ResNet V2* foi a que identificou corretamente a maior quantidade de mapas de espessura de RNFL de pacientes normais.

Por fim, apesar da qualidade visual das imagens dos mapas de espessura de RNFL sintéticos ainda não respeitar todas as características do conjunto de dados real, elas indicam que a *GAN* aprendeu algumas das características estruturais dos mapas de espessura de RNFL reais. O uso dos mapas produzidos pela rede geradora nos conjuntos de dados, trouxe melhora no desempenho medido através do valor de *AROC* para os conjuntos que apresentaram grande quantidade de mapas reais e sintéticos. Somente a adição de uma grande quantidade de mapas sintéticos ao conjunto de dados original não tornou o desempenho dos classificadores melhor nos experimentos. Foi necessário aumentar a quantidade de mapas reais para acompanhar a proporção de mapas sintéticos adicionados.

Tabela 8: Desempenhos dos classificadores no experimento E11. VN=Verdadeiros Negativos. FP=Falsos Positivos. FN=Falsos Negativos. VP=Verdadeiros Positivos. Acc=Acurácia. P=*Precision*. R=*Recall*.

Classificadores	VN	FP	FN	VP	Ac	P	R
DenseNet	32	12	14	44	0.745	0.758	0.785
Inception ResNet V2	34	12	12	44	0.764	0.785	0.785
Inception V3	29	11	17	45	0.725	0.725	0.803
ResNet 152 V2	30	11	16	45	0.735	0.737	0.803
ResNet 50	28	11	18	45	0.715	0.714	0.803
VGG16	33	8	13	48	0.794	0.786	0.857
VGG19	32	17	14	39	0.696	0.735	0.696
Xception	29	11	17	45	0.725	0.725	0.803

7 CONSIDERAÇÕES FINAIS

Neste trabalho de pesquisa foi analisado o uso de *GANs* para a geração de mapas de espessura de RNFL sintéticos para o treinamento de classificadores para o diagnóstico de glaucoma. Com a utilização de conjuntos de dados compostos em parte por mapas de espessura de RNFL reais e em parte por mapas sintéticos, esperou-se melhorar o desempenho de classificadores de redes neurais profundas treinados com conjunto de dados composto por poucas amostras.

Nos experimentos realizados foi utilizada uma GAN com características baseadas na *AC-GAN*, permitindo o uso de informações condicionais para a geração de mapas de espessura de RNFL de pacientes com glaucoma ou de pacientes normais.

A avaliação da *GAN* foi realizada através do desempenho dos classificadores para o diagnóstico de glaucoma treinados, medido através dos valores de AROC.

Com o intuito de não contaminar o treinamento dos classificadores com imagens de espessura de RNFL produzidas por um gerador treinado pelo mesmo conjunto de dados, foram utilizados dois conjuntos de dados distintos para treinamento e validação. O conjunto de dados de treinamento foi utilizado no treinamento da GAN e no treinamento dos classificadores, enquanto o conjunto de dados de validação foi utilizado somente na validação dos classificadores para o diagnóstico de glaucoma.

Foram escolhidas oito tipos de arquitetura de redes neurais profundas para serem treinadas com o conjunto de dados de treinamento. Posteriormente utilizando a *GAN* treinada com o mesmo conjunto de dados, foram gerados vários novos conjuntos de dados

compostos por mapas de espessura de RNFL reais e o mapas sintéticos em diferentes proporções.

Primeiro os classificadores foram treinados somente com o conjunto de dados de treinamento. Posteriormente foram criados 3 novos conjuntos de dados, adicionando as quantidades de 10%, 50% e 100% de dados sintéticos ao conjunto de dados de treinamento para treinamento. Também foi investigado o comportamento ao aumentar significativamente o número de dados sintéticos em relação ao tamanho do conjunto de dados real. Foram criados mais 4 conjuntos de dados com 2 vezes, 10 vezes, 20 vezes e 50 vezes o tamanho do conjunto de dados de treinamento. Investigamos também o balanceamento entre os mapas de espessura de RNFL reais e sintéticos. Os mapas reais do conjunto de dados de treinamento foram copiados para ficarem com a mesma quantidade de mapas sintéticos. Então foram gerados mais 3 conjuntos de dados adicionando 10 vezes, 20 vezes e 50 vezes mais imagens reais ao conjuntos de dados que já continham os mapas sintéticos.

Os melhores valores para AROC encontrados foi quando o tamanho do conjunto de dados sintético foi 50 vezes o tamanho do conjunto de dados de treinamento original, mas que respeitasse uma proporção de 50% entre dados reais e sintéticos. Isso aconteceu para as redes para as redes *DenseNet*, *VGG16*, *Xception* e *Inception ResNet V2*.

Diante dos resultados obtidos, pode-se concluir que o uso de dados sintéticos produzidos por *GANs* contribuiu para a melhora do desempenho de classificadores para diagnóstico de glaucoma construídos a partir de redes neurais profundas e treinados com conjunto de dados com poucas amostras. Mas ainda foi necessário manter o equilíbrio da quantidade de mapas de espessura de RNFL reais e sintéticos dentro de um mesmo conjunto de dados. Apesar disso, a qualidade visual das espessuras de RNFL produzidas pela *GAN* ainda pode melhorar a ponto de representar com mais fidelidade as características estruturais dos mapas no conjunto de dados real.

7.1 Contribuições do Trabalho

Este trabalho teve como contribuição investigar o uso de *GANs* para a geração de mapas de espessura de RNFL sintéticos. Estes mapas podem ser utilizados para aumentar o conjunto de dados existente no treinamento de classificadores para o diagnóstico de glaucoma. Classificadores baseados em redes neurais profundas exigem grande quantidade de dados para não sofrerem de problemas de *overfitting*. Diminuir os custos e principalmente o tempo associados na captura e formação de conjunto de dados no treinamento de classificadores para diagnóstico de várias patologias é de interesse da área médica.

7.2 Trabalhos Futuros

Como trabalho futuro propõe-se:

- Analisar o uso de outros tipos de *GANs* como *Wasserstein GAN* e *WGAN-GP* na geração de mapas de espessura de RNFL sintéticos. É necessário investigar outras arquiteturas de *GANs* para gerar mapas de espessura de RNFL com mais características estruturais do que os apresentados neste trabalho.
- Empregar métodos de avaliação quantitativos como *Inception Score* (IS) e *Frechet Inception Distance* (FID) no comparativo entre *GANs* na geração de espessuras de RNFL. Na análise de várias *GANs* é necessário uma forma de medir a qualidade dos mapas de espessura de RNFL produzidos pela rede geradora. Estas duas métricas são as mais comuns para realizar uma análise quantitativa entre *GANs*.
- Avaliar o desempenho de outros métodos generativos como Variational Autoencoders na geração de mapas de espessuras de RNFL sintéticos. Existem outros modelos generativos, como o *VAE*, que podem ser comparados com a *GAN* na geração de mapas de espessura de RNFL. O objetivo é melhorar a qualidade dos mapas produzidos com relação às características estruturais dos mapas reais.

- Incorporar aos parâmetros condicionais da *GAN*, mais informações como dados clínicos ou mais características estruturais oculares. É de interesse das pesquisa poder controlar mais aspectos dos mapas produzidos pela *GAN*. Adicionando informações como a idade, por exemplo, pode-se estudar a interpolação entre seus valores e a relação com a progressão da doença.

REFERÊNCIAS

- ALI, M.; RAMOS-ESTEBAN, J. C.; KATZ, L. J. Clinical examination of glaucoma. In: MD, M. Y.; MD, J. S. D. (Ed.). *OPHTHALMOLOGY*. 4. ed. [S.l.]: Elsevier, 2014. p. 1019:1028.
- AMATO, F. et al. Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, v. 11, n. 2, p. 47 – 58, 2013. ISSN 1214-021X.
- ASAOKA, R. et al. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*, v. 123, n. 9, p. 1974 – 1980, 2016. ISSN 0161-6420.
- BARELLA, K. A. et al. Glaucoma diagnostic accuracy of machine learning classifiers using retinal nerve fiber layer and optic nerve data from sd-oct. *Journal of ophthalmology*, Hindawi Publishing Corporation, v. 2013, 2013.
- BENGTSSON, B.; BIZIOS, D.; HEIJL, A. Effects of input data on the performance of a neural network in distinguishing normal and glaucomatous visual fields. *Investigative Ophthalmology & Visual Science*, v. 46, n. 10, p. 3730, 2005.
- BISNETO, T. R. V.; de Carvalho Filho, A. O.; MAGALHÃES, D. M. V. Generative adversarial network and texture features applied to automatic glaucoma detection. *Applied Soft Computing*, v. 90, p. 106165, 2020. ISSN 1568-4946. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1568494620301058>>.
- BIZIOS, D.; HEIJL, A.; BENGTSSON, B. Trained artificial neural network for glaucoma diagnosis using visual field data: A comparison with conventional algorithms. *Journal of Glaucoma*, v. 16, n. 1, p. 20–28, jan. 2007.
- BOURNE, R. R. The optic nerve head in glaucoma. *Community Eye Health Journal*, v. 19, n. 59, p. 44–45, jul. 2006.
- BOURNE, R. R. A. et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, Elsevier, v. 5, n. 9, p. e888–e897, 2021/01/18 2017. Disponível em: <[https://doi.org/10.1016/S2214-109X\(17\)30293-0](https://doi.org/10.1016/S2214-109X(17)30293-0)>.
- BOWD, C. et al. The retinal nerve fiber layer thickness in ocular hypertensive, normal, and glaucomatous eyes with optical coherence tomography. *Archives of Ophthalmology*, American Medical Association, v. 118, n. 1, p. 22–26, 2000.
- BRADLEY, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, Elsevier, v. 30, n. 7, p. 1145–1159, 1997.
- BURGANSKY-ELIASH, Z. et al. Optical coherence tomography machine learning classifiers for glaucoma detection: A preliminary study. *Investigative Ophthalmology & Visual Science*, v. 46, n. 11, p. 4147–4152, 2005.

- CHEN, X. et al. Glaucoma detection based on deep convolutional neural network. In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. [S.l.: s.n.], 2015. p. 715–718. ISSN 1557-170X.
- CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017.
- DIAS, M. *Geração de Populações Artificiais para Treinamento de Classificadores de Apoio ao Diagnóstico de Glaucoma*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2009.
- DUMOULIN, V.; VISIN, F. A guide to convolution arithmetic for deep learning. *ArXiv*, 2016.
- FAWCETT, T. An introduction to {ROC} analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006. ISSN 0167-8655. {ROC} Analysis in Pattern Recognition.
- FOSTER, P. J. et al. The definition and classification of glaucoma in prevalence surveys. *British Journal of Ophthalmology*, v. 86, n. 2, p. 238–242, 2002.
- FUJIMOTO, J. G. et al. Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. *Neoplasia*, Elsevier, v. 2, n. 1, p. 9–25, 2000.
- GOEL RENATA G PICCIANI, R. K. L. M.; BHATTACHARYA, S. K. Aqueous humor dynamics: A review. *Open Ophthalmol Journal*, v. 4, p. 52–59, 2010.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: The MIT Press, 2016. ISBN 0262035618.
- GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680.
- GOODFELLOW, I. J. NIPS 2016 tutorial: Generative adversarial networks. *CoRR*, abs/1701.00160, 2017.
- GOOGLE (Ed.). *All Our N-gram are Belong to You*. 2006. Disponível em: <<https://research.googleblog.com/2006/08/all-our-n-gram-are-belong-to-you.html>>.
- GU, J. et al. Recent advances in convolutional neural networks. *Pattern Recognition*, v. 77, p. 354 – 377, 2018. ISSN 0031-3203.
- HARTLINE, H. K. Visual receptors and retinal interaction. *Science*, v. 164, n. 3877, p. 270–278, 1969.
- He, K. et al. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. ISSN 1063-6919.
- HINTON, G. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, v. 29, n. 6, p. 82–97, Nov 2012. ISSN 1053-5888.

- HUANG, D. et al. Optical coherence tomography. *Science*, American Association for the Advancement of Science, v. 254, n. 5035, p. 1178–1181, 1991.
- Huang, G. et al. Densely connected convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 2261–2269. ISSN 1063-6919.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, Wiley Online Library, v. 195, n. 1, p. 215–243, 1968.
- HUMAN Eye Anatomy. 2012. [Http://www.eyesandeyesight.com/2009/02/anatomy-of-the-eye](http://www.eyesandeyesight.com/2009/02/anatomy-of-the-eye).
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. [S.l.]: JMLR.org, 2015. (ICML'15), p. 448–456.
- Isola, P. et al. Image-to-image translation with conditional adversarial networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 5967–5976.
- JACOBS, D. *Correlation and convolution. Class Notes for CMSC 426*. 2005. University of Maryland.
- Karras, T. et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *ArXiv e-prints*, out. 2017.
- KHALIL, T.; KHALID, S.; SYED, A. Review of machine learning techniques for glaucoma detection and prediction. In: *Science and Information Conference (SAI), 2014*. [S.l.: s.n.], 2014. p. 438–442.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: BENGIO, Y.; LECUN, Y. (Ed.). *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. [S.l.: s.n.], 2015.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.
- Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 105–114. ISSN 1063-6919.
- LIU, M. M. et al. Comparison of time-and spectral-domain optical coherence tomography in management of diabetic macular edemaa comparison of td-and sd-oct for dme. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 55, n. 3, p. 1370–1377, 2014.

- MAREK, B. M.; HARRIS, A.; SIESKY, B. Optic nerve blood flow measurement. In: MD, M. Y.; MD, J. S. D. (Ed.). *OPHTHALMOLOGY*. 4. ed. [S.l.]: Elsevier, 2014. p. 1047:1049.
- MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.
- MITCHELL, T. M. *Machine Learning*. 1. ed. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997.
- MURPHY, K. P. *Machine learning: a probabilistic perspective*. [S.l.]: MIT press, 2012.
- MURTHI, A.; MADHESWARAN, M. Enhancement of optic cup to disc ratio detection in glaucoma diagnosis. In: *Computer Communication and Informatics (ICCCI), 2012 International Conference on*. [S.l.: s.n.], 2012. p. 1–5.
- NG, A. et al. *Unsupervised Feature Learning and Deep Learning Tutorial*. 2013. Disponível em: <http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial>.
- ODENA, A.; OLAH, C.; SHLENS, J. Conditional image synthesis with auxiliary classifier gans. In: JMLR. ORG. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. [S.l.], 2017. p. 2642–2651.
- OPHIR, A. First-visit diagnosis of preperimetric glaucoma. *The open ophthalmology journal*, Bentham Science Publishers, v. 4, p. 22, 2010.
- POWERS, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. Bioinfo Publications, 2011.
- RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.
- ROGERS, D. Fifty years of monte carlo simulations for medical physics. *Physics in medicine and biology*, IOP Publishing, v. 51, n. 13, p. R287, 2006.
- ROJAS, R. *Neural networks: a systematic introduction*. [S.l.]: Springer Science & Business Media, 2013.
- ROSENBLATT, F. *The Perceptron: A Perceiving and Recognizing Automaton*. [S.l.], 1957.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Cognitive modeling*, v. 5, n. 3, p. 1, 1988.
- SAEEDI, O. J.; RAMULU, P.; FRIEDMAN, D. S. Epidemiology of glaucoma. In: *OPHTHALMOLOGY*. 4. ed. [S.l.]: Elsevier, 2014. p. 1001:1006.
- SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural Networks*, v. 61, p. 85 – 117, 2015. ISSN 0893-6080.
- SCHUMAN, J. S. Spectral domain optical coherence tomography for glaucoma (an aos thesis). *Transactions of the American Ophthalmological Society*, American Ophthalmological Society, v. 106, p. 426, 2008.

- SCHUMAN, J. S. et al. *Optical coherence tomography of ocular diseases*. [S.l.]: Slack Thorofare, NJ, 2004.
- SHIGUEOKA, L. et al. Sensitivity and specificity of machine learning classifiers using sd-oct and standard automated perimetry compared to glaucoma specialists, general ophthalmologists, the oct database and anderson's visual field (vf) criteria. *Investigative Ophthalmology & Visual Science*, The Association for Research in Vision and Ophthalmology, v. 57, n. 12, p. 384–384, 2016.
- SILVA, F. R. et al. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using spectral domain oct and standard automated perimetry. *Arquivos brasileiros de oftalmologia*, SciELO Brasil, v. 76, n. 3, p. 170–174, 2013.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. 2014.
- SOMMER, A. et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Archives of ophthalmology (Chicago, Ill. : 1960)*, v. 109 1, p. 77–83, 1991.
- SRIVASTAVA, N. et al. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, v. 15, p. 1929–1958, 2014.
- SZEGEDY, C. et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI conference on artificial intelligence*. [S.l.: s.n.], 2017.
- SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.]: IEEE, 2016.
- Szegedy, C. et al. Going deeper with convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1–9. ISSN 1063-6919.
- TAN, J. C.; KAUFMAN, P. L. Primary open-angle glaucoma. In: MD, M. Y.; MD, J. S. D. (Ed.). *OPHTHALMOLOGY*. 4. ed. [S.l.]: Elsevier, 2014. cap. 1052:1056.
- THAM, Y.-C. et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, Elsevier, v. 121, n. 11, p. 2081–2090, 2014.
- VIDOTTI, V. G. et al. Sensitivity and specificity of machine learning classifiers and spectral domain oct for the diagnosis of glaucoma. *European journal of ophthalmology*, p. 0–0, 2012.
- WEINREB, R. N.; KHAW, P. T. Primary open-angle glaucoma. *The Lancet*, v. 363, n. 9422, p. 1711–1720, 2004. ISSN 0140-6736.
- WOLLSTEIN, G.; SCHUMAN, J. S. Optic nerve analysis. In: *OPHTHALMOLOGY*. 4. ed. [S.l.]: Elsevier, 2014. p. 1040:1047.
- YAQOOB, Z.; WU, J.; YANG, C. Spectral domain optical coherence tomography: a better oct imaging strategy. *Biotechniques*, v. 39, 2005.
- ZOU, K. H. *Receiver Operating Characteristic (ROC) Literature Research*. 2007. Disponível em: <<https://www.spl.harvard.edu/archive/spl-pre2007/pages/ppl/zou/roc.html>>.

APÊNDICE A

A.1 Preprocessamento das imagens dos conjuntos de dados

O conjunto de dados de treinamento consiste em 137 imagens de espessuras de RNFL com glaucoma e 138 imagens de espessuras de RNFL de olhos normais. O conjunto de dados de validação consiste em 56 imagens de espessuras de RNFL de pacientes com glaucoma e 46 imagens de espessuras de RNFL de pacientes normais. O conjunto de dados de treinamento possui imagens de espessuras de RNFL dos dois olhos do paciente enquanto o conjunto de dados de validação possui somente um por paciente.

As imagens de espessuras de RNFL foram extraídas dos exames no formato JPEG. Pela falta de qualidade, talvez decorrente do formato da imagem, elas possuíam discrepâncias entre as cores apresentadas na imagem e as cores presentes na escala de cores que as acompanham. Algumas cores presentes em uma imagem não existiam na imagem de escala de cores.

Para se ter uniformidade de cores nas imagens dos conjuntos de dados, foi montado uma nova escala de cores permitindo converter uma cor em um determinado valor de espessura para cor e também ser possível o processo inverso.

Para se obter uma nova escala de cores que representassem as espessuras de RNFL adequadamente, foi necessário extrair da escala original somente uma coluna de pixels. Foi escolhida a coluna 38 da imagem da escala original. As imagens originais de espessura de RNFL ainda possuem mais dois tons de cinza para representar o centro do nervo óptico, mas que não se encontram na escala original que acompanha a imagem. Sendo assim a

nova escala criada possui mais as duas cores RGB(169,169,169) e RGB(128,129,129) em sua composição.

Com essa nova escala criada, foi possível obter novas imagens de espessura de RNFL aproximando os pixels das imagens originais para a nova escala. Cada pixel da imagem original foi aproximado de uma cor da nova escala criada utilizando distância euclidiana, e através dessa cor foi possível estimar o valor da espessura de RNFL para aquele pixel. Dessa maneira é possível navegar entre imagens de espessura de RNFL e valores de espessura com exatidão e vice-versa como na figura 29 onde a imagem original 29A é convertido para a nova escala, resultando na figura 29B.

Figura 29: Conversão do imagem de espessura de RNFL para uma escala construída

Outra alteração necessária é o redimensionamento dessas novas imagens de espessura de RNFL para terem tamanho de 128x128 pixels a fim de se enquadrarem nas características das arquiteturas das redes apresentadas no trabalho, assim como foi utilizado em Odena, Olah e Shlens (2017).

Além disso, algumas imagens apresentavam áreas em preto proveniente do equipamento utilizado nos exames, como na figura 30A. Para remover estas áreas em preto, foi criada uma imagem média baseada nas imagens das espessuras de pacientes normais do conjunto de dados de treinamento, conforme a figura 30B. As áreas em preto então foram substituídas por pixels da imagem média. O resultado da substituição pode ser observado na figura 30C.

Por último as imagens das espessuras de RNFL para o olho direito foram espelhadas horizontalmente para seguir o mesmo padrão de espessura do olho esquerdo.

Figura 30: Imagem a partir da média das imagens de pacientes normal do conjunto de dados de treinamento