

Vitória de Carvalho Brito  
Orientador: Antonio Oseas de Carvalho Filho

**Uma Abordagem Baseada em *Capsule Network*  
para a Identificação de Glaucoma em Imagens  
de Retina**

Picos - PI  
12 de janeiro de 2021

Vitória de Carvalho Brito  
Orientador: Antonio Oseas de Carvalho Filho

## **Uma Abordagem Baseada em *Capsule Network* para a Identificação de Glaucoma em Imagens de Retina**

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção de grau de Bacharel em Sistemas de Informação. Orientador: Prof. Dr. Antonio Oseas de Carvalho Filho.

Universidade Federal do Piauí  
Campus Senador Helvídio Nunes de Barros  
Bacharelado em Sistemas de Informação

Picos - PI  
12 de janeiro de 2021

**FICHA CATALOGRÁFICA**  
**Universidade Federal do Piauí**  
**Campus Senador Helvídio Nunes de Barros**  
**Biblioteca Setorial José Albano de Macêdo**  
**Serviço de Processamento Técnico**

**B862a** Brito, Vitória Carvalho  
Uma abordagem baseada em *Capsule Network* para a identificação de glaucoma em imagens de retina / Vitória Carvalho Brito – 2021.  
  
43 f.; CD-ROM 4 ¾ pol.  
  
Monografia (Graduação em Sistemas de Informação) – Universidade Federal do Piauí, Picos-PI, 2021.  
  
“Orientador: Prof. Antônio Oseas de Carvalho Filho”  
  
1. *Capsule Network*. 2. Glaucoma. 3. Imagens de Retina. I. Título.

**CDD 006.6**

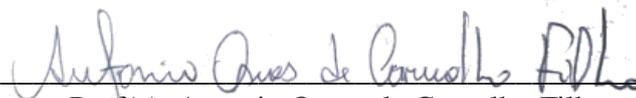
UMA ABORDAGEM BASEADA EM CAPSULE NETWORK PARA A IDENTIFICAÇÃO  
DE GLAUCOMA EM IMAGENS DE RETINA

VITÓRIA DE CARVALHO BRITO

Monografia APROVADA como exigência parcial para obtenção do grau de Bacharel em  
Sistemas de Informação.

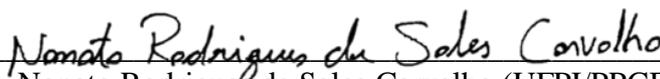
Data de Aprovação

Picos – PI, 20 de janeiro de 2021



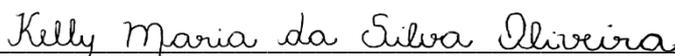
---

Prof(a). Antonio Oseas de Carvalho Filho



---

Nonato Rodrigues de Sales Carvalho (UFPI/PPGEE)



---

Kelly Maria da Silva Oliveira (UFPI/PPGEE)

# Agradecimentos

Ao meu bondoso e amado Deus, por nunca ter desistido de mim, por ter sido o meu auxílio e fortaleza em todos os momentos de alegria e tristeza, por ter me dado forças para enfrentar todos os obstáculos durante a graduação e por ter colocado pessoas incríveis na minha vida durante esse tempo. Glória a Ti, Senhor!

À minha mãe, Rita, e minhas irmãs, Ruthe e Emanuelle, por serem tão especiais na minha vida, por se alegrarem com minhas conquistas, por me aconselharem quando é preciso e por orarem comigo nos momentos difíceis. Agradeço também ao meu pai, Carlos, por todo apoio e cuidado. Deus abençoe por tudo! Amo vocês!

Às minhas tias, Eva e Ivonete, por todo o amor e cuidado de sempre. É meu agradecimento especial por me abrigarem com tanto amor em sua casa durante o tempo de graduação presencial. Amo vocês!

Aos meu amado grupo de amigos do curso, carinhosamente apelidado de "panelinha", vocês são um presente do céu na minha vida. Com vocês eu aprendi não somente coisas do curso, mas aprendi sobre paciência, humildade e compreensão. Obrigada por tudo, eu amo vocês!

Ao meu amigo, Patrick, por seu companheirismo durante todos esses anos, por batalhar comigo no curso, na monitoria, na Iniciação Científica e em todos os outros projetos que estive junto comigo, inclusive esta monografia. Obrigada pela paciência comigo, mesmo quando você tinha certeza que eu estava errada!

Ao meu amigo, Jederson, pela paciência e humildade de me ensinar tantas coisas no curso durante todos esses anos. Obrigada por sempre estar disposto a compartilhar seu conhecimento para ajudar os demais. Deus te abençoe por tudo!

Ao meu orientador, Antonio Oseas, por seu apoio e dedicação comigo e com o Patrick. Só tenho a agradecer pelos puxões de orelha e pela pressão nos prazos dos trabalhos, isso só contribuiu para que aprendêssemos mais. Obrigada também pelo esforço de sempre buscar o melhor para nós. Vou levar o 'Cuida!' no meu vocabulário.

Aos meus queridos professores do curso, em especial Alcilene, Deborah, Patrícia Medyna, Flávio, Romuere e Oseas. Como eu admiro vocês! Obrigada pelo aprendizado, pelas brincadeiras, pelo incentivo e pelo companheirismo!

Aos meus colegas da UFPI, minha família, amigos e todos os que direta ou indiretamente contribuíram para essa conquista na minha vida. Agradeço de coração!

*Mas em todas estas coisas somos mais do que vencedores, por aquele que nos amou.*  
*Romanos 8:37*

# Resumo

O glaucoma é uma doença ocular responsável pela segunda causa mais comum de cegueira no mundo. A necessidade de detectar a doença em seus estágios iniciais é imprescindível, visto que o tratamento tardio pode acarretar na perda de visão. Nesse contexto, métodos computacionais estão sendo desenvolvidos para auxiliar os especialistas na detecção do glaucoma através da análise de imagens oculares, a fim de conferir maior precisão ao diagnóstico e aumentar as chances do paciente receber tratamento em tempo hábil, evitando a perda da visão. Neste trabalho, apresenta-se uma metodologia para a classificação automática do glaucoma em imagens de retina utilizando a *Capsule Network (CapsNet)*, um modelo recente de *Deep Learning (DL)* que analisa as relações espaciais hierárquicas entre as características para representar as imagens, de modo a requerer menos amostras de treinamento do que as *Convolutional Neural Networks (CNNs)* tradicionais para obter um desempenho eficiente na classificação. Antes da execução da *CapsNet*, aplicou-se uma etapa de pré-processamento nas imagens, com o intuito de destacar as características da retina. Os resultados mostraram-se promissores, atingindo *accuracy* de 91,89%, *recall* de 88,52%, *precision* de 96,42%, *f1-score* de 92,30%, *AUC* de 0,92, e *kappa* de 0,83. A principal contribuição desta abordagem está na obtenção de resultados promissores sem a necessidade da aplicação de técnicas para aumentar os dados e segmentar a região do disco óptico. Assim, este estudo comprovou o potencial das cápsulas em identificar as relações entre as características, mesmo diante de um conjunto de dados com apenas 556 imagens de retina.

**Palavras-chaves:** *Capsule Network*, Glaucoma, Imagens de Retina, *Convolutional Neural Networks*.

# Abstract

Glaucoma is an eye disease responsible for the second most common cause of blindness in the world. The need to detect the disease in its early stages is essential since late treatment leads to loss of vision. In this context, computational methods are being developed to assist specialists in the detection of glaucoma through the analysis of ocular images in order to provide greater precision to the diagnosis and increase the chances of the patient receiving treatment in a timely manner, avoiding vision loss. In this work, a methodology for the automatic classification of glaucoma in retinal images using the Capsule Network (CapsNet) is presented, a recent model of Deep Learning (DL) that analyzes the hierarchical spatial relationships between the characteristics to represent the images, in order to require fewer training samples than traditional Convolutional Neural Networks (CNNs) for efficient classification performance. Before the execution of CapsNet, a pre-processing step was applied to the images to highlight the retina's characteristics. The results were promising, reaching an accuracy of 91.89%, recall of 88.52%, precision of 96.42%, f1-score of 92.30%, AUC of 0.92, and kappa of 0.83. This approach's main contribution is to obtain promising results without the need to apply techniques to increase data and segment the region of the optical disc. Thus, this study proved the capsules' potential in identifying the relationships between the characteristics, even in the face of a data set with only 556 images of the retina.

# Lista de ilustrações

Figura 1 – Exemplos de (a) retina saudável e (b) glaucomatosa. Fonte: Ghanem (2020). . . . .	16
Figura 2 – Exemplos de uma imagem de retina (a) antes e (b) depois da equalização de histograma. Imagens disponíveis na base RIM-ONE v2 (FUMERO et al., 2011). . . . .	17
Figura 3 – Ilustração da arquitetura <i>LeNet</i> para classificação de imagens de retina, com suas três principais camadas: convolucionais, de <i>pooling</i> e totalmente conectadas. Adaptado de Araújo et al. (2017). . . . .	18
Figura 4 – Comparação representativa entre a <i>CapsNet</i> e a <i>CNN</i> na classificação de uma imagem de panda antes e depois de modificações. Adaptado de Kumar (2019). . . . .	19
Figura 5 – Arquitetura da <i>CapsNet</i> . Fonte: Sabour, Frosst e Hinton (2017). . . . .	20
Figura 6 – Fluxo da metodologia proposta. Autoria própria. . . . .	25
Figura 7 – Arquitetura da <i>CapsNet</i> aplicada neste trabalho. Autoria própria. . . . .	27
Figura 8 – Gráfico de (a) <i>accuracy</i> e (b) <i>loss</i> de treino e validação do modelo que obteve o melhor resultado no experimento 3. . . . .	32
Figura 9 – Matriz de confusão gerada no teste do melhor modelo do experimento 3. . . . .	32
Figura 10 – Representação dos dois componentes principais gerados pelo <i>PCA</i> para o conjunto de características extraídas pela <i>CapsNet</i> no conjunto de teste do experimento 3. . . . .	33

# Lista de tabelas

Tabela 1 – Categorização do desempenho da classificação, de acordo com o índice $\kappa$ . . . . .	22
Tabela 2 – Resumo dos trabalhos relacionados ao método proposto. . . . .	24
Tabela 3 – Principais parâmetros usados no modelo. . . . .	28
Tabela 4 – Experimento 1: avaliação do modelo em decorrência da variação da dimensão das imagens de entrada. . . . .	30
Tabela 5 – Experimento 2: avaliação do modelo em virtude da variação dos canais da imagem executados com e sem a equalização de histograma na etapa de pré-processamento. . . . .	30
Tabela 6 – Experimento 3: avaliação do modelo em decorrência da variação do número de iterações do roteamento dinâmico. . . . .	31
Tabela 7 – Resultados obtidos com a execução de outras arquiteturas de <i>CNNs</i> . . . . .	35
Tabela 8 – Comparação do método proposto com os trabalhos relacionados. . . . .	35

# Lista de abreviaturas e siglas

<i>Acc</i>	<i>Accuracy</i>
<i>AUC</i>	<i>Area Under ROC Curve</i>
<i>CapsNet</i>	<i>Capsule Network</i>
<i>CNN</i>	<i>Convolutional Neural Network</i>
<i>DCNN</i>	<i>Deep Convolutional Neural Network</i>
<i>DL</i>	<i>Deep Learning</i>
<i>F1</i>	<i>F1-score</i>
FN	Falso Negativo
FP	Falso Positivo
<i>GAN</i>	<i>Generative Adversarial Network</i>
<i>HRT</i>	Tomografia Retiniana de Heidelberg
<i>LBP</i>	<i>Local Binary Pattern</i>
<i>ML</i>	<i>Machine Learning</i>
<i>OCT</i>	Tomografia de Coerência Óptica
<i>PCA</i>	<i>Principal Component Analysis</i>
PIO	Pressão Intraocular
<i>Prec</i>	<i>Precision</i>
<i>RCNN</i>	<i>Regions with Convolutional Neural Network</i>
<i>Rec</i>	<i>Recall</i>
<i>ReLU</i>	<i>Rectified Linear Unit</i>
<i>ROC</i>	<i>Receiver Operating Characteristic</i>
TFP	Taxa de Falso Positivo
TVP	Taxa de Verdadeiro Positivo
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

# Sumário

<b>1</b>	<b>Introdução</b>	<b>12</b>
1.1	Objetivos	13
1.2	Organização do Trabalho	14
<b>2</b>	<b>Referencial Teórico</b>	<b>15</b>
2.1	Glaucoma	15
2.2	Pré-processamento de imagens	16
2.3	<i>Deep Learning</i>	17
2.3.1	<i>Capsule Network</i>	18
2.4	Métricas de Validação	20
<b>3</b>	<b>Trabalhos Relacionados</b>	<b>23</b>
<b>4</b>	<b>Metodologia</b>	<b>25</b>
4.1	Aquisição de Imagens	26
4.2	Pré-processamento	26
4.3	Classificação	26
4.4	Validação	28
<b>5</b>	<b>Resultados e Discussões</b>	<b>29</b>
5.1	Resultados Obtidos pelo Método Proposto	29
5.1.1	Experimento 1	29
5.1.2	Experimento 2	30
5.1.3	Experimento 3	31
5.2	Comparação com Outras Abordagens	33
5.2.1	Testes com Outras Arquiteturas	34
5.2.2	Comparação com os Trabalhos Relacionados	35
5.3	Discussões	36
<b>6</b>	<b>Conclusão</b>	<b>38</b>
<b>7</b>	<b>Publicações</b>	<b>39</b>
	<b>Referências</b>	<b>40</b>

# 1 Introdução

O glaucoma é uma doença ocular caracterizada pela danificação gradual do nervo óptico, podendo causar a perda da visão. Depois da catarata, o glaucoma é a causa mais comum de cegueira no mundo (KINGMAN, 2004). Estima-se que, até 2040, 111,8 milhões de pessoas serão portadoras desta doença (THAM et al., 2014). No caso da catarata, é possível reverter a perda da visão através de cirurgia, por outro lado, o glaucoma provoca cegueira permanente se tratado tardiamente.

A elevação da Pressão Intraocular (PIO) é um dos principais fatores que acarretam no desenvolvimento do glaucoma, pois, quando essa pressão aumenta, pode causar a morte das células ganglionares da retina (JUNIOR et al., 2018). Quando isso acontece, ocorre um aumento na escavação do nervo óptico, que é um espaço central do nervo que não contém fibras nervosas, sendo assim, o aumento desta escavação se torna um indício da presença do glaucoma.

Por ser uma doença assintomática em seus estágios iniciais, o prognóstico favorável do glaucoma está relacionado à detecção precoce, bem como ao tratamento em tempo hábil (JUNIOR et al., 2018). Uma forma comum de diagnosticar o glaucoma é verificando a PIO. No entanto, podem existir casos de glaucoma onde a PIO não é alta. Alguns exames permitem detectar o glaucoma, como a Tomografia de Coerência Óptica (*OCT*, do inglês *Optical Coherence Tomography*) e a Tomografia Retiniana de Heidelberg (*HRT*, do inglês *Heidelberg Retinal Tomography*), que fornecem imagens de alta resolução das estruturas da retina, e a Retinografia, que inclui imagens coloridas do fundo do olho.

As imagens geradas por meio desses exames são analisadas por especialistas que farão o diagnóstico da doença. Contudo, essa análise demanda tempo e é um processo cansativo, tendo em vista a necessidade de analisar com cautela as alterações morfológicas no nervo óptico. Dessa forma, métodos computacionais estão sendo desenvolvidos para auxiliar os especialistas na tomada de decisão durante o diagnóstico do glaucoma, a fim de conferir uma maior precisão aos resultados e aumentar as chances do paciente receber tratamento em tempo hábil, evitando a perda da visão.

Em comparação às abordagens tradicionais, a utilização de métodos computacionais combinada à análise dos especialistas possibilita uma maior precisão na sensibilidade e especificidade dos resultados, mesmo no diagnóstico de doenças assintomáticas, como o glaucoma (ABDEL-ZAHER; ELDEIB, 2016; TSOCHATZIDIS et al., 2017). Dentre as técnicas computacionais utilizadas no auxílio ao diagnóstico do glaucoma, está o *DL*, uma sub-área do *ML*, onde os modelos computacionais identificam padrões nas imagens, para que, no processo de classificação, possam discriminar corretamente as classes em questão (LECUN; BENGIO; HINTON, 2015). As *Convolutional Neural Networks (CNNs)* são exploradas em vários domínios de imagem, inclusive, é o estado da arte em reconhecimento

de imagens (LECUN; BENGIO; HINTON, 2015).

Um estudo recente, proposto por Sabour, Frosst e Hinton (2017), apresenta um novo modelo de *DL*, a *Capsule Network* (*CapsNet*), visando solucionar algumas desvantagens trazidas pelas *CNNs* tradicionais, como a grande quantidade de imagens necessárias para um bom treinamento e as operações de *pooling*, que, durante a redução da dimensionalidade das características, podem acabar perdendo informações importantes. Na *CapsNet*, os principais blocos são as cápsulas, essas estruturas analisam a relação espacial hierárquica das características para tentar representar melhor os dados. Dessa forma, ao contrário das *CNNs* que apenas identificam a presença das características em um objeto, as cápsulas identificam relacionamentos espaciais entre elas.

Estudos recentes que aplicam a *CapsNet* apontam que é possível treinar esse modelo com menos amostras do que as *CNNs* tradicionais e ainda assim obter um bom desempenho (JIMÉNEZ-SÁNCHEZ; ALBARQOUNI; MATEUS, 2018; AFSHAR; MOHAMMADI; PLATANIOTIS, 2018). Além disso, Gaddipati et al. (2019) sugere que a *CapsNet* tem um comportamento mais robusto em relação ao desbalanceamento das classes, sendo esse um problema recorrente em muitas bases de imagens.

No contexto das imagens glaucomatosas, a análise das relações espaciais entre as características da retina, como, por exemplo, tamanho e orientação do disco óptico e distância da escavação para o disco, que é um dos fatores que indicam a presença do glaucoma, podem ser relevantes para a classificação das imagens em glaucomatosas ou saudáveis. Sendo assim, este trabalho visa explorar tais aspectos para a classificação das imagens de retina avaliando a arquitetura *CapsNet* diante de um conjunto de dados público com apenas 556 imagens de retina.

## 1.1 Objetivos

Diante do contexto apresentado, o propósito deste trabalho é explorar a *CapsNet* para a detecção automática de glaucoma em imagens da retina, de modo a avaliar se as cápsulas são capazes de identificar as relações espaciais entre as características da retina de forma que um conjunto reduzido de imagens de treinamento seja suficiente para que o modelo alcance uma classificação satisfatória dos dados.

De maneira específica, pretende-se:

1. Avaliar a viabilidade do uso da arquitetura *CapsNet* para a identificação da doença, utilizando bases públicas de imagens de retina, sem a necessidade de uma segmentação prévia do disco óptico;
2. Aplicar técnicas de pré-processamento nas imagens que possam realçar as características da doença; e

3. Construir uma metodologia para classificação automática do glaucoma, que possa ser aplicada no auxílio ao diagnóstico da doença.

## 1.2 Organização do Trabalho

O trabalho está organizado da seguinte maneira: o Capítulo 2 aborda os conceitos fundamentais para o entendimento da metodologia proposta; o Capítulo 3 discute os trabalhos relacionados ao tema abordado neste estudo; o Capítulo 4 apresenta a metodologia aplicada para alcançar os objetivos apontados neste trabalho; o Capítulo 5 apresenta e discute os resultados obtidos com a execução da metodologia; e, finalmente, o Capítulo 6 conclui este trabalho.

## 2 Referencial Teórico

Este capítulo apresenta os conceitos fundamentais para a compreensão da proposta deste trabalho. O capítulo está dividido em quatro subseções: a primeira relata sobre o Glaucoma, abordando as formas atuais de diagnóstico da doença; a segunda discorre sobre a etapa de pré-processamento de imagens; a terceira contextualiza o *DL*, adentrando no modelo investigado neste trabalho; finalmente, a quarta subseção explana sobre as métricas utilizadas para avaliar os resultados da classificação das imagens.

### 2.1 Glaucoma

O glaucoma é uma doença ocular crônica que danifica progressivamente o nervo óptico, causando perda irreversível da visão periférica se não for tratada. A detecção precoce do glaucoma é muito importante, pois permite que os pacientes busquem tratamento adequado rapidamente, o que pode minimizar os danos e a perda da visão (ARAÚJO et al., 2019). Atualmente, o glaucoma é a segunda maior causa de cegueira no mundo. Há uma estimativa de que, até 2040, 111,8 milhões de pessoas serão afetadas pelo glaucoma (THAM et al., 2014).

Um exame comum para a detecção do glaucoma é a verificação da PIO. Contudo, o aumento da PIO não necessariamente diz respeito a presença de glaucoma, assim como a PIO normal também não exclui a possibilidade da existência da doença. Loewen e Tanna (2014) afirmam que, do ponto de vista da gestão de pacientes, isso significa que o diagnóstico do glaucoma deve ser baseado principalmente na análise do disco óptico, sendo realizada por meio da análise de imagens médicas do fundo da retina.

Além da medição da PIO, o diagnóstico do glaucoma pode ser feito através de exames periódicos, como o *OCT* e o *HRT*, ou através da avaliação manual do disco óptico por meio de imagens digitais do fundo do olho. A doença causa alterações morfológicas na cabeça do nervo óptico e na camada de fibras nervosas que podem ser examinadas para fornecer um diagnóstico (ARAÚJO et al., 2019). A Figura 1 ilustra exemplos de retina saudável e glaucomatosa, onde é possível observar que a elevação da PIO vai danificando progressivamente as células ganglionares da retina, que formam o nervo óptico. Dessa maneira, como essas células constituem a única saída de informações visuais da retina, essa danificação provoca a perda da visão.

O diagnóstico a partir dos exames de imagens é feito através da análise visual e interpretação dos especialistas. Sendo assim, diferentes especialistas podem fornecer interpretações distintas para um mesmo exame. Além disso, a interpretação é uma tarefa repetitiva, requerendo um grande nível de atenção sobre os detalhes presentes na imagem. Diante disso, a análise desses exames através de sistemas computacionais tem sido de grande va-

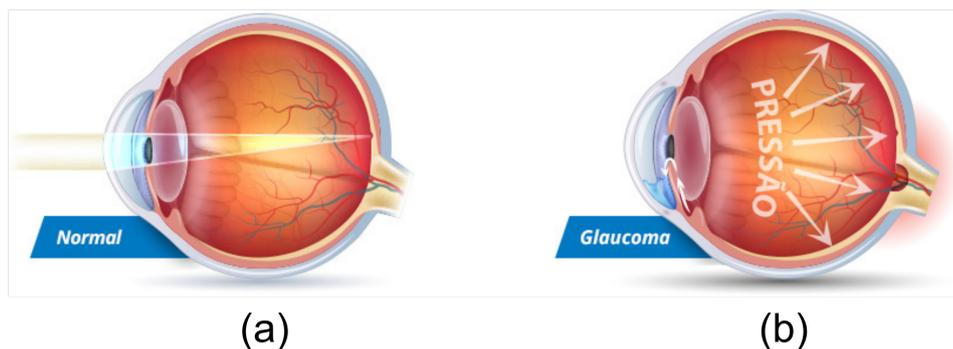


Figura 1 – Exemplos de (a) retina saudável e (b) glaucomatosa. Fonte: [Ghanem \(2020\)](#).

lia para monitorar o desenvolvimento natural da doença e fornecer uma segunda opinião sobre o diagnóstico.

## 2.2 Pré-processamento de imagens

O pré-processamento de imagens é uma etapa presente nos métodos de Visão Computacional, responsável por aplicar procedimentos capazes de proporcionar um melhoramento nos aspectos visuais e estruturais da imagem ([FILHO et al., 2016](#)). Dessa maneira, consegue-se aumentar contraste, eliminar ruídos, etc. Nos problemas de *ML*, é comum a utilização de conjuntos de dados que não estão padronizados, o que muitas vezes pode ser resultado da aquisição dos dados provenientes de fontes distintas. Dessa forma, é crucial que os dados sejam tratados antes de serem aplicados na construção do modelo.

No âmbito das imagens, a etapa de pré-processamento pode ser uma tarefa simples, como o redimensionamento de imagens. Para alimentar uma *CNN* com um conjunto de dados de imagens, por exemplo, é necessário que todas as imagens possuam a mesma dimensão. Outras tarefas de pré-processamento podem incluir transformações geométricas, aplicação de filtros, conversão dos canais da imagem, normalização dos pixels, entre outras.

Neste trabalho, aplicou-se um método de pré-processamento denominado equalização de histograma ([KRUTSCH; TENORIO, 2011](#)), que proporciona uma distribuição uniforme no histograma, de forma a melhorar o contraste da imagem. Isso acontece porque a equalização distribui os valores de intensidade mais frequentes, possibilitando que regiões que antes apresentavam um contraste menor, agora obtenham um maior contraste. Assim, pode-se evitar que a variação acentuada no contraste das imagens influencie na análise das características ([BISNETO; FILHO; MAGALHÃES, 2020](#)). Na Figura 2, tem-se um exemplo de uma imagem de retina antes e depois da aplicação da equalização de histograma, onde é possível observar que o contraste foi realçado e que algumas regiões, principalmente a escavação do disco óptico, ficaram bem destacadas.

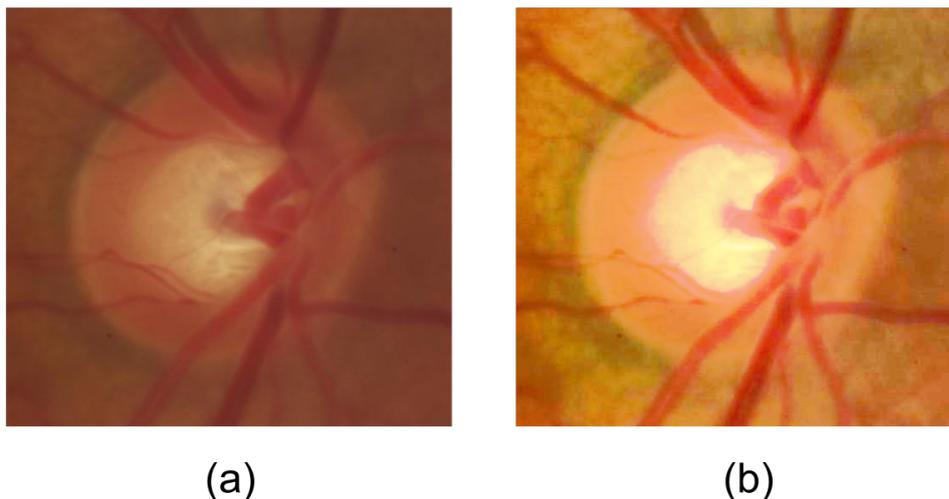


Figura 2 – Exemplos de uma imagem de retina (a) antes e (b) depois da equalização de histograma. Imagens disponíveis na base RIM-ONE v2 (FUMERO et al., 2011).

## 2.3 Deep Learning

O *DL* é sub-área do *ML* na qual modelos computacionais compostos por várias camadas de processamento aprendem representações de dados com vários níveis de abstração. Esses métodos melhoraram o estado da arte em reconhecimento de fala, reconhecimento visual de objetos, detecção de objetos e vários outros domínios, como descoberta de medicamentos e genômica (LECUN; BENGIO; HINTON, 2015).

O *DL* busca identificar uma estrutura complexa em grandes conjuntos de dados usando o algoritmo de *backpropagation* para indicar como uma rede deve alterar seus parâmetros internos para diminuir a taxa de erro durante o aprendizado. As *CNNs* são algoritmos de *DL* que trouxeram avanços no processamento de imagens, vídeo, fala e áudio, enquanto as redes recorrentes se destacaram em dados sequenciais, como texto e fala (LECUN; BENGIO; HINTON, 2015).

Uma *CNN* é uma arquitetura baseada no processamento de dados visuais, capaz de aplicar filtros nesses dados, mantendo a relação de vizinhança entre os pixels da imagem ao longo do processamento da rede (LECUN et al., 1998). Essa operação é conhecida como convolução, onde ocorre a somatória do produto ponto a ponto entre os valores de um filtro e cada posição da vizinhança do pixel de entrada.

Um dos primeiros projetos de *CNNs* foi a *LeNet*, proposta por LeCun et al. (1998), que contribuiu para impulsionar a área de *DL*. Inicialmente, essa arquitetura foi utilizada para reconhecimento de caracteres, tais como código postal e dígitos numéricos. Novas arquiteturas foram propostas nos últimos anos como forma de melhoria da *LeNet*, embora as versões de *CNNs* melhoradas compartilhem os conceitos fundamentais, como observado na Figura 3.

Após uma camada convolucional, geralmente existe uma camada de *pooling*. O obje-

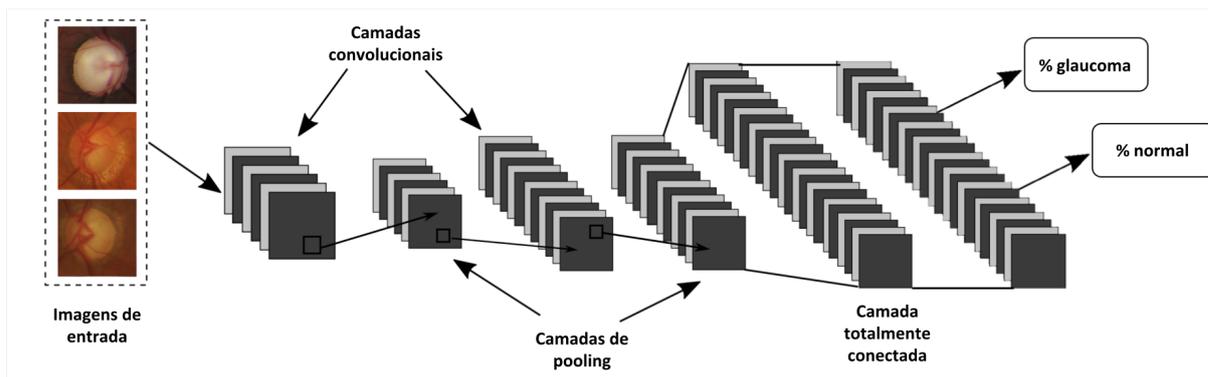


Figura 3 – Ilustração da arquitetura *LeNet* para classificação de imagens de retina, com suas três principais camadas: convolucionais, de *pooling* e totalmente conectadas. Adaptado de Araújo et al. (2017).

tivo dessa camada é reduzir progressivamente a dimensionalidade das imagens de entrada, diminuindo o custo computacional da rede e evitando o *overfitting*. A saída das camadas convolucionais e de *pooling* representam as características extraídas da imagem de entrada. O objetivo das camadas totalmente conectadas é utilizar essas características para classificar a imagem em uma classe previamente especificada (ARAÚJO et al., 2017).

Novas arquiteturas de *CNNs* que surgiram na última década são basicamente formadas pelas mesmas camadas presentes na *LeNet*, mas variam em relação ao propósito da classificação, à quantidade de camadas convolucionais, entre outros fatores. Em meio a diversidade dessas arquiteturas, este projeto investigou a utilização da *CapsNet*, que será detalhada a seguir.

### 2.3.1 Capsule Network

A *CapsNet* foi sugerida por Sabour, Frosst e Hinton (2017) como um novo modelo de rede neural para a tarefa de classificação de imagens. De acordo com Sabour, Frosst e Hinton (2017), este modelo tem como principal vantagem a preservação das relações espaciais hierárquicas dos atributos presentes na imagem, de forma a identificar não somente a presença de uma determinada característica, mas também a sua relação espacial com as demais características do objeto. Isso é feito através de vetores que carregam consigo informações como posição, tamanho e orientação do objeto (SHAHROUDNEJAD et al., 2018). Dessa forma, os autores afirmam que a *CapsNet* pode ser treinada com menos amostras e ainda assim alcançar um desempenho satisfatório.

Como visto em Moraes (2018), a identificação de elementos de interesse nas redes neurais tradicionais é feita através das ativações neuronais. Essas ativações são derivadas das somas ponderadas dos valores de suas arestas incidentes e o resultado desta soma é um único valor escalar, o que permite apenas informar a confiança da ocorrência de um elemento na região coberta pelo neurônio. Em contrapartida, a *CapsNet* utiliza vetores que permitem descrever mais do que somente a confiança de ocorrência dos elementos,

eles carregam consigo informações de instanciação (posição, tamanho, orientação, etc.) das entidades identificadas. Assim, a *CapsNet* propõe eliminar o conceito de invariância com o de equivariância, ou seja, se a entrada sofre transformação espacial, a rede se adapta de acordo com a modificação (SHAHROUDNEJAD et al., 2018).

Na figura 4, tem-se uma representação de comparação entre as predições da *CapsNet* e da *CNN* diante de uma imagem antes e depois da aplicação de modificações espaciais. Como a *CapsNet* é equivariante, se a imagem for transformada de alguma forma, os vetores de ativação também mudam de acordo. Por esse motivo, a imagem deformada não foi identificada como um panda, pois as cápsulas não concordaram que as características estão posicionadas corretamente para que a imagem seja classificada como tal. Já em relação à imagem rotacionada, as características estão posicionadas corretamente na imagem, independente da orientação, por conseguinte, a *CapsNet* identificou a entrada como panda.

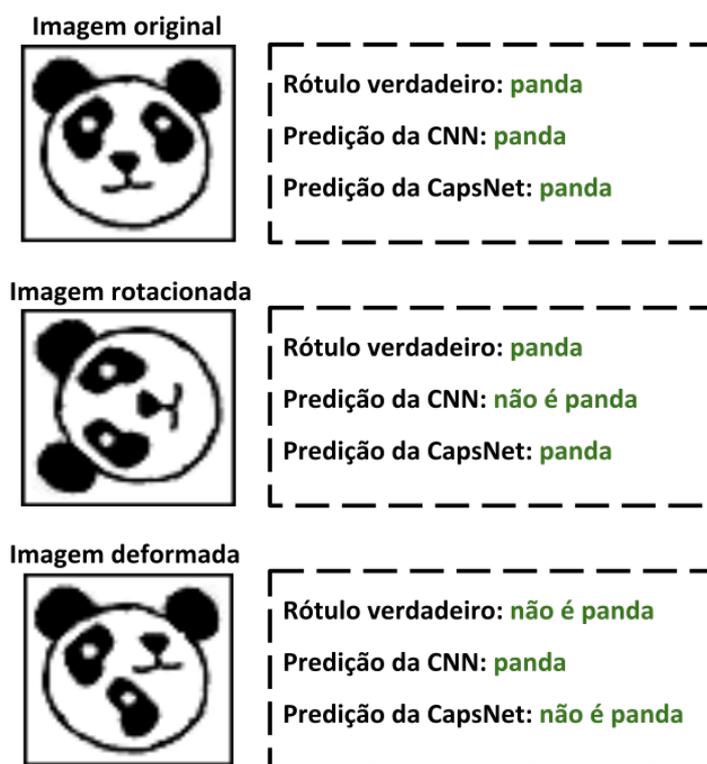


Figura 4 – Comparação representativa entre a *CapsNet* e a *CNN* na classificação de uma imagem de panda antes e depois de modificações. Adaptado de Kumar (2019).

A arquitetura original da *CapsNet* foi utilizada para a classificação de dígitos da base MNIST, que contém imagens de dimensão 28x28. O principal bloco desse modelo é a cápsula, que pode ser definida como um conjunto de neurônios organizados de forma aninhada dentro de uma camada. Esse grupo de neurônios pode ser representado como um vetor que carrega informações de instanciação, como, por exemplo, posição, escala, orientação, textura e cor. Esse vetor de informações passa por uma função *squashing*, onde seu comprimento é ajustado para um valor entre 0 e 1, de modo que esse valor

representará a probabilidade de que a entidade expressa pela cápsula esteja presente na atual entrada da rede.

A arquitetura original, representada pela Figura 5, utiliza uma camada convolucional, uma camada de cápsulas primárias e uma camada totalmente conectada. A primeira camada convolucional é composta por 256 filtros com convoluções de  $9 \times 9$ , passo 1 e função de ativação *Rectified Linear Unit* (*ReLU*). Já a segunda camada convolucional é a *PrimaryCaps*, composta de 32 cápsulas convolucionais com 8 dimensões cada, sendo que cada uma dessas dimensões se trata de uma cápsula convolucional com kernel de  $9 \times 9$  e passo 2. A última camada é a *DigitCaps*, composta por uma cápsula de 16 dimensões por cada classe de dígito, sendo que cada cápsula dessas dimensões é conectada à todas as cápsulas da camada inferior. Por fim, o vetor de probabilidades das classes é calculado através da norma L2, uma regularização onde o somatório de todos os parâmetros ao quadrado é adicionado ao somatório da diferença quadrada entre a saída real e as previsões. O roteamento dinâmico é feito entre as cápsulas da *PrimaryCaps* e *DigitCaps*, é ele que irá possibilitar que uma cápsula que está em um determinado nível receba ativações de algumas cápsulas do nível abaixo e ignore outras.

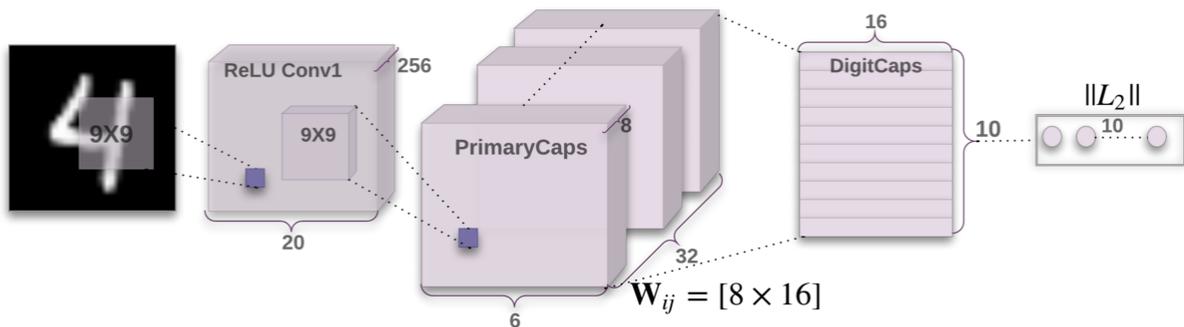


Figura 5 – Arquitetura da *CapsNet*. Fonte: Sabour, Frosst e Hinton (2017).

## 2.4 Métricas de Validação

Para avaliar as previsões fornecidas pelo modelo proposto neste trabalho, foram aplicadas as seguintes métricas de validação: *accuracy* (*Acc*), *recall* (*Rec*), *precision* (*Prec*), *f1-score* (*F1*), *Area Under ROC Curve* (*AUC*) (ERKEL; PATTYNAMA, 1998) e índice *kappa* (COHEN, 1968). O cálculo dessas métricas requer valores fornecidos pela matriz de confusão, que organiza os resultados de predição em quatro classes: Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN). A seguir, está uma breve explicação sobre as fórmulas para calcular as métricas citadas.

A *accuracy*, representada pela Equação 2.1, calcula a capacidade de um teste diagnóstico identificar verdadeiros positivos e verdadeiros negativos em uma amostra, ou seja, a proporção de acertos nos dados classificados.

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

Em um teste diagnóstico, o *recall* (também conhecido como sensibilidade) mede a capacidade de identificar verdadeiros positivos, ou seja, indivíduos que têm a doença. É uma métrica fundamental na análise, pois indica a eficácia do método. A equação 2.2 apresenta a fórmula para o cálculo do *recall*.

$$Rec = \frac{VP}{VP + FN}. \quad (2.2)$$

A *precision*, expressa pela Equação 2.3, mede a proporção de amostras positivas corretamente classificadas entre todas as classificadas como positivas.

$$Prec = \frac{VP}{VP + FP}. \quad (2.3)$$

É possível combinar *precision* e *recall* para medir o desempenho da classificação, e é isso que a *f1-score* faz. Como mostra a Equação 2.4, esta métrica calcula uma média harmônica entre os valores de *precision* e *recall*, sendo uma medida comumente utilizada em problemas de classificação onde as classes são desbalanceadas.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}. \quad (2.4)$$

A métrica *AUC* mede a capacidade do modelo de distinguir as classes usando o gráfico de desempenho fornecido pela curva *ROC* (*Receiver Operating Characteristic*). A curva *ROC* é calculada a partir de diferentes limiares, analisando duas métricas calculadas usando a matriz de confusão: a Taxa de Verdadeiro Positivo (TVP), equivalente ao *recall*, e a Taxa de Falso Positivo (TFP), equivalente a  $1 - especificidade$ , sendo que a especificidade é a métrica que calcula a proporção de verdadeiros negativos corretamente classificados. Quanto mais próximo de 1 é o valor da *AUC*, melhor é o desempenho do método na distinção das classes.

Por fim, também foi utilizado o coeficiente *kappa* para avaliar o modelo de classificação. O *kappa* mede o nível de concordância do experimento, sendo que, quanto mais próximo de 1, melhor é o resultado. A fórmula para calcular o *kappa* está representada pela Equação 2.6, onde *RandomAccuracy* é definida como a soma dos produtos de probabilidade de referência e de resultado para cada classe dividida pelo quadrado da quantidade total de amostras, como observado na Equação 2.5.

$$RandomAccuracy = \frac{(VN + FP) * (VN + FN) + (FN + VP) + (FP + VP)}{Total * Total} \quad (2.5)$$

$$Kappa = \frac{Accuracy - RandomAccuracy}{1 - RandomAccuracy} \quad (2.6)$$

De acordo com o Landis e Koch (1977), há uma categorização dos níveis de desempenho de classificação pelo índice *kappa*, que pode ser encontrada na Tabela 1.

Tabela 1 – Categorização do desempenho da classificação, de acordo com o índice *kappa*.

<b><i>Kappa</i> (K)</b>	<b>Qualidade</b>
$K < 0.2$	Ruim
$0.2 \leq K < 0.4$	Razoável
$0.4 \leq K < 0.6$	Bom
$0.6 \leq K < 0.8$	Muito bom
$K \geq 0.8$	Excelente

### 3 Trabalhos Relacionados

A *CapsNet* vem sendo aplicada em muitos problemas de classificação de imagens, desde que foi introduzida como um novo modelo de aprendizado profundo, em 2017 (PATRICK et al., 2019). Neste sentido, a literatura dispõe de estudos recentes sobre a aplicação da *CapsNet* no contexto da classificação de imagens médicas. No caso do glaucoma, o trabalho de Gaddipati et al. (2019) propõe o uso da *CapsNet* para a classificação de imagens 3D de *OCT* em glaucomatosas e normais, avaliando o método em uma base privada composta por 233 exames. O próprio autor afirma que a análise desses exames não é uma tarefa simples, uma vez que o modelo se torna computacionalmente complexo por lidar com as imagens tridimensionais, além do fato de existir uma baixa disponibilidade de conjuntos de dados de *OCT*.

Por outro lado, arquiteturas tradicionais de *DL* continuam sendo aplicadas no auxílio ao diagnóstico do glaucoma. O trabalho de Phan et al. (2019) comparou 3 modelos de *CNNs* na tarefa de classificação de imagens de retina: a *VGG19*, a *ResNet152* e a *DenseNet201*. Neste experimento, Phan et al. (2019) utilizou um conjunto de 3312 imagens, adquiridas de duas bases distintas, e, durante os testes, os autores perceberam que a qualidade das imagens afetou negativamente o desempenho dos métodos na distinção das classes. Já o estudo de Bajwa et al. (2019) teve os experimentos divididos em dois estágios: um estágio de segmentação do disco óptico, usando uma *Regions with Convolutional Neural Network (RCNN)*; e um estágio de classificação das regiões em glaucomatosas e saudáveis, usando uma *Deep Convolutional Neural Network (DCNN)*. Para tanto, Bajwa et al. (2019) utilizou 780 imagens provindas de três bases distintas.

Pelo fato de ter usado um conjunto de dados com apenas 455 imagens, o trabalho de Maheshwari, Kanhangad e Pachori (2020) aplicou uma técnica de aumento de dados baseada no *Local Binary Pattern (LBP)*, a fim de evitar o problema de *overfitting*. Em seguida, os autores realizaram a classificação das imagens de retina usando um modelo pré-treinado de *CNN*, a *AlexNet*. Outro estudo recente na detecção automática do glaucoma foi proposto por Bisneto, Filho e Magalhães (2020), onde o mesmo aplicou uma *Generative Adversarial Network (GAN)* para segmentar o disco óptico nas imagens e, em seguida, utilizou descritores de textura associados à classificadores tradicionais na literatura para classificar as regiões em glaucomatosas e normais. Os métodos propostos por Bisneto, Filho e Magalhães (2020) foram avaliados em um conjunto de 556 imagens, derivadas de duas bases diferentes que, inclusive, são as mesmas aplicadas neste trabalho.

A Tabela 2 aponta um resumo dos estudos relacionados, onde é possível observar que o problema de classificação do glaucoma foi explorado pela *CapsNet* em apenas um dos trabalhos. Todavia, o modelo proposto por Gaddipati et al. (2019) é uma arquitetura 3D que utiliza volumes de *OCT*, onde o mesmo aponta que há uma baixa disponibilidade

de amostras desse tipo de exame, fator que muitas vezes dificulta a utilização dessas imagens em abordagens computacionais. Em comparação à abordagem descrita, o método proposto neste trabalho utiliza imagens de fundo do olho, através de bases publicamente disponíveis.

Em relação aos outros estudos levantados, percebe-se que as *CNNs* são amplamente aplicadas na detecção do glaucoma, com uma extensiva variedade de abordagens. Contudo, é possível observar que dos métodos que não aplicaram segmentação, um deles precisou utilizar aumento de dados no conjunto de imagens, enquanto o outro não aplicou, mas utilizou o maior conjunto de imagens entre todos os trabalhos apresentados. Isso evidencia o fato que as *CNNs* precisam de muitas amostras de treinamento para alcançar uma capacidade discriminativa satisfatória. Em comparação à esses métodos, a abordagem proposta neste trabalho utilizou um conjunto de apenas 556 imagens, todavia, sem a necessidade de uma segmentação prévia do disco óptico ou técnicas de aumento de dados.

Tabela 2 – Resumo dos trabalhos relacionados ao método proposto.

Trabalho	Método	Objetivo	Bases	Tipo de Imagem	Amostra	Métricas de Validação
Gaddipati et al. (2019)	<i>CapsNet</i>	Classificação	Privada	<i>OCT</i>	233	<i>Accuracy</i> = 94% <i>Sensitivity</i> = 89% <i>Specificity</i> = 96% <i>AUC</i> = 0,973
Phan et al. (2019)	<i>VGG19</i> , <i>ResNet152</i> e <i>DenseNet201</i>	Classificação	YMU e KOSEI	Imagens de fundo do olho	3312	<i>AUC</i> = 0,9
Bajwa et al. (2019)	<i>RCNN</i> e <i>DCNN</i>	Segmentação e Classificação	ORIGA, HRF e OCT & CFI	Imagens de fundo do olho	780	<i>Sensitivity</i> = 71,17% <i>AUC</i> = 0,874
Maheshwari, Kanhangad e Pachori (2020)	<i>LBP</i> e <i>AlexNet</i> pré-treinada	Classificação	RIM-ONE	Imagens de fundo do olho	455	<i>Accuracy</i> = 98,9% <i>Sensitivity</i> = 100% <i>Specificity</i> = 97,5%
Bisneto, Filho e Magalhães (2020)	<i>GAN</i> e características de textura	Segmentação e Classificação	RIM-ONE e Drishti-GS	Imagens de fundo do olho	556	<i>Accuracy</i> = 100% <i>Sensitivity</i> = 100% <i>Specificity</i> = 100% <i>AUC</i> = 1
Método Proposto	<i>CapsNet</i>	Classificação	RIM-ONE e Drishti-GS	Imagens de fundo do olho	556	<i>Accuracy</i> = 91,89% <i>Recall</i> = 88,52% <i>Precision</i> = 96,42% <i>F1-score</i> = 92,30% <i>AUC</i> = 0,92 <i>Kappa</i> = 0,83

## 4 Metodologia

Este capítulo descreve as etapas envolvidas neste trabalho para a classificação automática de glaucoma em imagens de retina usando a *CapsNet*, como levantado na Seção 1.1. Os algoritmos implementados neste trabalho para a execução dessas etapas estão publicamente disponíveis no [GitHub](#). A Figura 6 ilustra o fluxo da metodologia proposta, que terá cada etapa detalhada no decorrer deste capítulo.

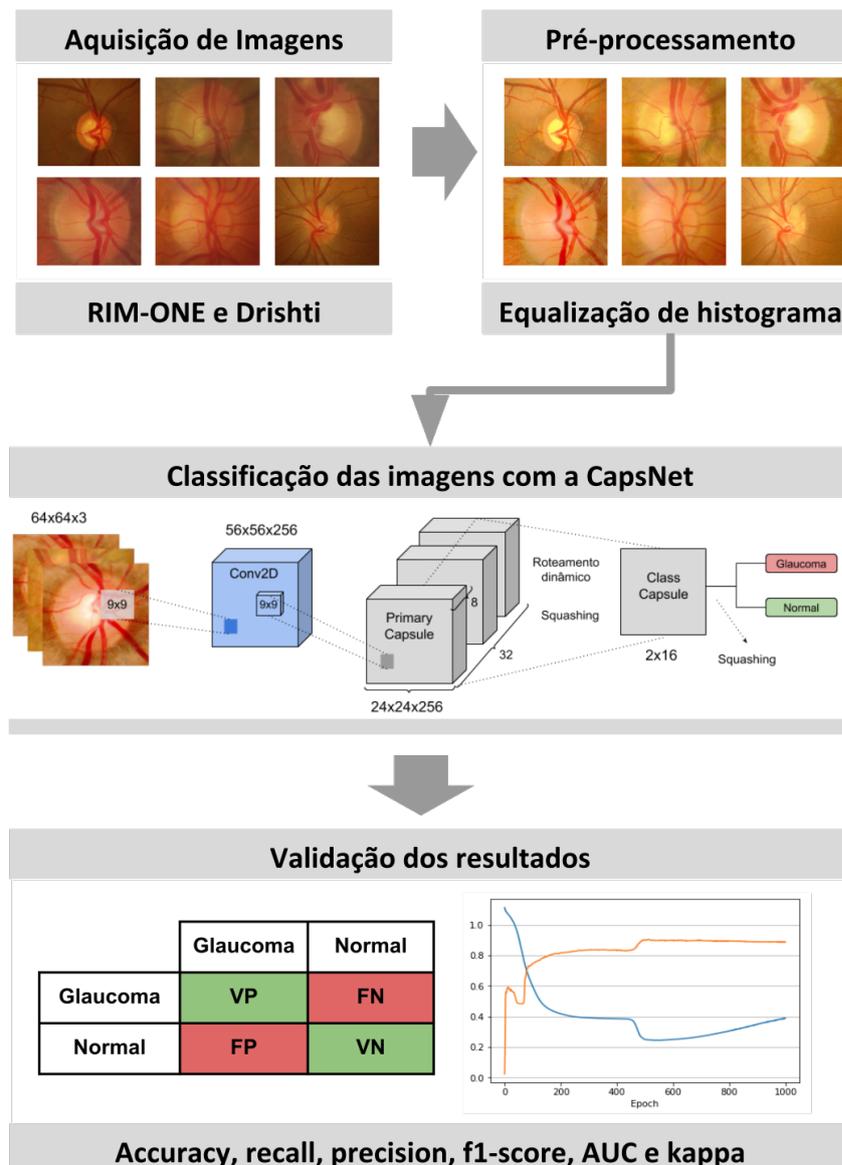


Figura 6 – Fluxo da metodologia proposta. Autoria própria.

## 4.1 Aquisição de Imagens

Para avaliar o modelo de classificação, foram utilizadas as bases de imagens de retina RIM-ONE v2 (FUMERO et al., 2011) e Drishti-GS (SIVASWAMY et al., 2014), onde ambos os conjuntos contêm imagens de disco óptico glaucomatosas e saudáveis. De acordo com Bisneto, Filho e Magalhães (2020), essas bases são comumente utilizadas na comunidade científica, sendo aplicadas em diversos métodos de classificação do glaucoma.

A RIM-ONE v2 consiste em imagens fornecidas por três hospitais: Hospital Universitário de Canarias, Hospital Clínico San Carlos e Hospital Miguel Servet. Ela possui 455 imagens oculares, sendo 200 imagens de retina com glaucoma em diferentes estágios e 255 imagens de retinas saudáveis.

A Drishti-GS é uma base destinada à validação de métodos computacionais com o mesmo propósito deste trabalho e é composta de apenas 101 imagens, onde 70 delas apresentam retinas glaucomatosas e 31 apresentam retinas saudáveis. As imagens da base Drishti-GS foram coletadas e anotadas pelo Aravind Eye Hospital, Madurai, Índia e são de uma única população (SIVASWAMY et al., 2014).

Ambos os conjuntos possuem imagens de diferentes resoluções e dimensões. Todas as imagens são de 8 *bits* e estão no modelo de cores *RGB* (do inglês *red, green, blue*).

## 4.2 Pré-processamento

Como visto na Seção 2.2, antes do treinamento com a *CapsNet*, foi aplicado o método de equalização de histograma (KRUTSCH; TENORIO, 2011) para realçar o contraste das imagens, aprimorando aspectos estruturais e visuais das imagens de retina. De acordo com Bisneto, Filho e Magalhães (2020), como a equalização distribui os valores de intensidade mais frequentes, pode-se evitar que a variação acentuada no contraste das imagens influencie na análise das características. Para a implementação deste método, foi utilizado o módulo *exposure* da biblioteca *skimage* (WALT et al., 2014).

## 4.3 Classificação

Como indicado na Seção 2.3.1, este trabalho explorou um modelo de *DL* denominado *CapsNet* para a classificação das imagens de retina em glaucomatosas e saudáveis. O modelo implementado neste trabalho seguiu as configurações originais da arquitetura, logo, foram utilizadas uma camada convolucional, uma camada de cápsulas primárias e uma camada totalmente conectada, que possui 16 dimensões por classe, isto é, glaucoma e normal. Na Figura 7, tem-se uma representação ilustrativa da arquitetura utilizada, onde é possível notar que há uma camada convolucional (*Conv2D*), uma camada de cápsulas primárias (*Primary Capsule*) e uma camada totalmente conectada (*Class Capsule*).

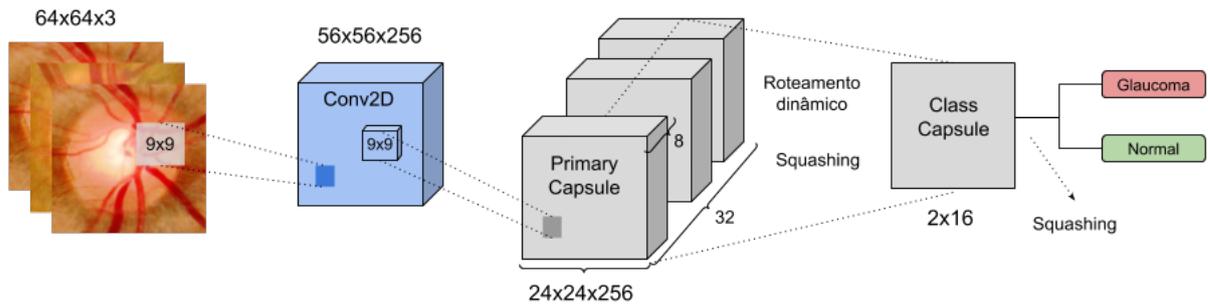


Figura 7 – Arquitetura da *CapsNet* aplicada neste trabalho. Autoria própria.

As duas bases de imagens mencionadas na Seção 4.1 foram utilizadas na execução da metodologia, resultando em um conjunto de 556 imagens (270 glaucomatosas e 286 normais). Esse conjunto foi dividido aleatoriamente em treino, validação e teste, com uma proporção de 70%, 10% e 20%, respectivamente. Sendo assim, utilizou-se 390 imagens para o treinamento, 55 para a validação e 111 para o teste. Uma vez dividido, o mesmo conjunto foi utilizado em todos os testes realizados.

A entrada da rede foi adequada para receber as imagens das bases mencionadas na Seção 4.1, mas antes todas as imagens foram padronizadas para uma mesma dimensão através da função de redimensionamento disponível na biblioteca *pillow* (CLARK, 2015). O próximo capítulo deste trabalho apresenta os resultados obtidos com a execução do modelo em dimensões variadas das imagens, mas o melhor resultado foi alcançado quando o formato das imagens foi definido como 64x64. Em relação ao modelo de cores das imagens, além do *RGB*, também foram testados os canais R, G e B, separadamente, e o canal *gray*, sendo esta conversão implementada através do método *rgb2gray* da biblioteca *skimage* (WALT et al., 2014). Após a aplicação do método de pré-processamento, uma normalização das imagens foi realizada, deixando a distribuição dos valores de pixel entre 0 e 1.

Nos testes iniciais da *CapsNet*, o número de iterações do roteamento dinâmico foi definido como 3, mas o capítulo de resultados mostra a execução do modelo com a atribuição de outros valores para este parâmetro. O valor 3 foi definido como base pelo fato de ter sido o melhor valor encontrado no artigo original da *CapsNet* (SABOUR; FROSST; HINTON, 2017). Contudo, é válido variar este parâmetro, tendo em vista que o problema é diferente do que foi abordado no artigo original. Os outros parâmetros definidos nos testes estão detalhados na Tabela 3. Em relação ao número de épocas, o valor 200 foi selecionado como um ponto de partida por ser um número maior de épocas do que o que normalmente tem sido aplicado nos estudos com a *CapsNet* (GADDIPATI et al., 2019; BUTUN et al., 2020; TSUJI et al., 2020; AFSHAR et al., 2020; JIANG et al., 2020; AFSHAR; PLATANOTIS; MOHAMMADI, 2019), no entanto, convém analisá-lo nos resultados, através dos gráficos de desempenho, para comprovar se o valor é suficiente para o aprendizado. Os outros parâmetros também foram baseados na literatura.

Tabela 3 – Principais parâmetros usados no modelo.

<b>Hiperparâmetros</b>	<b>Valores aplicados</b>
Número de épocas	200
Otimizador	<i>Adam optimizer</i>
Taxa de aprendizado	0,0001
Tamanho do lote	64
Número do roteamento dinâmico	3

## 4.4 Validação

O processo de avaliação da classificação é essencial para a análise do desempenho do modelo e possíveis melhorias. Para avaliar as predições da *CapsNet* nas imagens de retina, foram aplicadas as métricas *accuracy*, *recall*, *precision*, *f1-score*, *AUC* e *kappa*, calculadas a partir da matriz de confusão gerada, como discutido na Seção 2.4. Na implementação dessas métricas, foram utilizados os métodos do módulo *metrics* da biblioteca *sklearn* (PEDREGOSA et al., 2011). Essas métricas são amplamente adotadas na literatura para avaliar modelos de *ML*.

## 5 Resultados e Discussões

Este capítulo apresenta e discute os resultados alcançados através dos experimentos realizados para a classificação das imagens do fundo da retina em glaucomatosas e saudáveis utilizando a arquitetura *CapsNet*, de acordo com a metodologia apontada no Capítulo 4. O capítulo está dividido em duas partes: na primeira, são apresentados os resultados dos experimentos realizados com a *CapsNet*; na segunda, o desempenho da metodologia proposta foi comparado com outras abordagens da literatura.

### 5.1 Resultados Obtidos pelo Método Proposto

Esta Seção apresenta os resultados obtidos com a execução da *CapsNet* em vários casos de teste, divididos em três experimentos. Os experimentos têm como objetivo encontrar a dimensão e a representação de cores ideais para as imagens, analisar o impacto do realce de contraste e encontrar o melhor valor para o roteamento dinâmico entre as cápsulas.

#### 5.1.1 Experimento 1

Como exposto no Capítulo 4, as imagens de retina passaram por um redimensionamento antes de serem impostas como entrada da rede, isso porque a arquitetura requer uma padronização na dimensão das imagens. Para encontrar a melhor dimensão das imagens para a rede, o primeiro experimento foi a execução do modelo com três dimensões diferentes: 64x64, 128x128 e 256x256. Com o aumento da altura e largura das imagens, o modelo vai ficando mais custoso computacionalmente, sendo assim, o cenário ideal seria aquele onde a dimensão da entrada não seja tão extensa, mas ainda assim consiga representar bem as características do problema.

A Tabela 4 aponta os resultados da execução do modelo com as três dimensões de entrada mencionadas. Para esse primeiro teste, optou-se por utilizar o canal *gray* das imagens, que consegue carregar a importância dos três canais em apenas uma matriz. Em relação ao número de iterações do roteamento dinâmico, para esse teste foi definido o valor 3, pelo fato de ter sido o melhor valor encontrado no artigo original da *CapsNet* (SABOUR; FROSST; HINTON, 2017). Quanto ao número de épocas, utilizou-se 200. Por fim, nessa execução, o método de pré-processamento ainda não foi aplicado, visto que o propósito desse primeiro teste é apenas encontrar a melhor dimensão para que, posteriormente, outros parâmetros possam ser variados.

Analisando a Tabela 4, é possível observar que o modelo alcançou melhores resultados quando executado com as imagens na dimensão 64x64. Além de conseguir representar melhor as características da doença dentre todas as dimensões testadas, o valor 64x64

Tabela 4 – Experimento 1: avaliação do modelo em decorrência da variação da dimensão das imagens de entrada.

Dimensão	Acc (%)	Rec (%)	Prec (%)	F1 (%)	AUC	Kappa
<b>64x64</b>	<b>86,48</b>	<b>81,96</b>	<b>92,59</b>	<b>86,95</b>	<b>0,86</b>	<b>0,73</b>
128x128	76,57	72,13	83,01	77,19	0,77	0,53
256x256	76,57	73,77	81,81	77,58	0,76	0,53

proporcionou ao modelo um treinamento mais rápido, uma vez que o custo computacional é intensificado conforme o aumento na dimensionalidade das imagens.

### 5.1.2 Experimento 2

Em posse da dimensão ideal para a entrada do modelo, o segundo experimento realizado foi um conjunto de testes para encontrar o melhor canal para a representação das imagens. Ademais, nesse experimento também foi avaliada a aplicação do método de pré-processamento nas imagens, a fim de comprovar se o realce das características impacta positivamente os resultados. Assim como no experimento anterior, a quantidade de épocas foi definida como 200 e o roteamento dinâmico como 3. Os resultados alcançados neste experimento podem ser visualizados através da Tabela 5.

Tabela 5 – Experimento 2: avaliação do modelo em virtude da variação dos canais da imagem executados com e sem a equalização de histograma na etapa de pré-processamento.

	Canal	Acc (%)	Rec (%)	Prec (%)	F1 (%)	AUC	Kappa
Sem pré-processamento	RGB	86,48	88,52	87,09	87,80	0,86	0,72
	R	81,98	80,32	85,96	83,05	0,82	0,63
	G	81,08	67,21	97,61	79,61	0,82	0,63
	B	82,88	77,04	90,38	83,18	0,83	0,65
	Gray	86,48	81,96	92,59	86,95	0,86	0,73
Com pré-processamento	<b>RGB</b>	<b>88,28</b>	<b>83,60</b>	<b>94,44</b>	<b>88,69</b>	<b>0,88</b>	<b>0,76</b>
	R	74,77	73,77	78,94	76,27	0,74	0,49
	G	87,38	83,60	92,72	87,93	0,87	0,74
	B	82,88	80,32	87,5	83,76	0,83	0,65
	Gray	84,68	86,88	85,48	86,17	0,84	0,69

Visualizando os resultados expostos na Tabela 5, observa-se que o *RGB* mostrou-se a melhor representação de cores para o reconhecimento das imagens de retina pela *CapsNet*. Além disso, os melhores resultados dessa representação foram alcançados, na maioria das métricas, com a aplicação do realce de contraste, obtendo *accuracy* de 88,28%, *recall* de 83,60%, *precision* de 94,44%, *f1-score* de 88,69%, *AUC* de 0,88 e *kappa* de 0,76. Nesse teste, o modelo apresentou resultados superiores aos demais na maioria das métricas de validação, inclusive, obteve um *kappa* muito bom, de acordo com a categorização de Landis e Koch (1977). Diante do exposto, a aplicação do pré-processamento no canal *RGB*, utilizando a dimensão 64x64 nas imagens, mostrou-se a melhor configuração de entrada para a *CapsNet*.

### 5.1.3 Experimento 3

Até o momento, os experimentos realizados permitiram a descoberta da dimensão e canal ideais para as imagens de entrada, além da aplicação do pré-processamento, que mostrou-se relevante para o modelo. Outro aspecto importante para ser avaliado no modelo é o impacto do roteamento dinâmico. Como mencionado anteriormente, os testes realizados utilizaram 3 iterações no roteamento dinâmico, por ter sido o melhor valor encontrado nos experimentos do artigo original da *CapsNet*. No entanto, este trabalho trata-se da exploração de um problema totalmente diferente do contexto do artigo original, tornando-se necessária uma nova avaliação desse parâmetro para a obtenção de melhores resultados.

A Tabela 6 apresenta os resultados da execução do modelo diante da variação do número de iterações do roteamento dinâmico. Os resultados presentes no roteamento 3 (quarta linha da Tabela 6) são os melhores resultados obtidos no experimento anterior, ou seja, com o modelo aplicado ao canal RGB das imagens 64x64 após a equalização de histograma. Os valores do roteamento foram variados para menos e para mais de 3.

Tabela 6 – Experimento 3: avaliação do modelo em decorrência da variação do número de iterações do roteamento dinâmico.

Iterações do Roteamento Dinâmico	Acc (%)	Rec (%)	Prec (%)	F1 (%)	AUC	Kappa
1	87,38	81,96	94,33	87,71	0,87	0,74
2	90,09	86,88	94,64	90,59	0,90	0,80
3	88,28	83,60	94,44	88,69	0,88	0,76
4	87,38	80,32	96,07	87,50	0,88	0,74
5	85,58	85,24	88,13	86,66	0,85	0,70
6	88,28	85,24	92,85	88,88	0,88	0,76
7	88,28	81,96	96,15	88,49	0,88	0,76
<b>8</b>	<b>91,89</b>	<b>88,52</b>	<b>96,42</b>	<b>92,30</b>	<b>0,92</b>	<b>0,83</b>
9	87,38	88,52	88,52	88,52	0,87	0,74
10	88,28	81,96	96,15	88,49	0,88	0,76

Analisando os resultados apontados na Tabela 6, é possível observar que, de fato, modificar o parâmetro do roteamento dinâmico influencia os resultados de predição do modelo. Contudo, os resultados foram bastante variados, não houve um desempenho crescente ou decrescente em função do incremento do número de iterações do roteamento. Quando o roteamento foi definido como 8, o modelo alcançou uma melhor capacidade discriminatória, atingindo *accuracy* de 91,89%, *recall* de 88,52%, *precision* de 96,42%, *f1-score* de 92,30%, *AUC* de 0,92 e *kappa* de 0,83, considerado excelente, consoante a categorização de Landis e Koch (1977).

Em face do exposto, observa-se que o melhor cenário de teste encontrado foi na execução do modelo usando 8 iterações no roteamento dinâmico e recebendo como entrada as imagens *RGB*, redimensionadas para 64x64 e realçadas através da equalização de histograma. Com o intuito de facilitar a compreensão dos resultados, as Figuras 8(a) e 8(b)

apresentam os gráficos de *accuracy* e *loss* do treinamento e validação do modelo que alcançou os melhores resultados neste experimento, enquanto a Figura 9 representa a matriz de confusão gerada no teste.

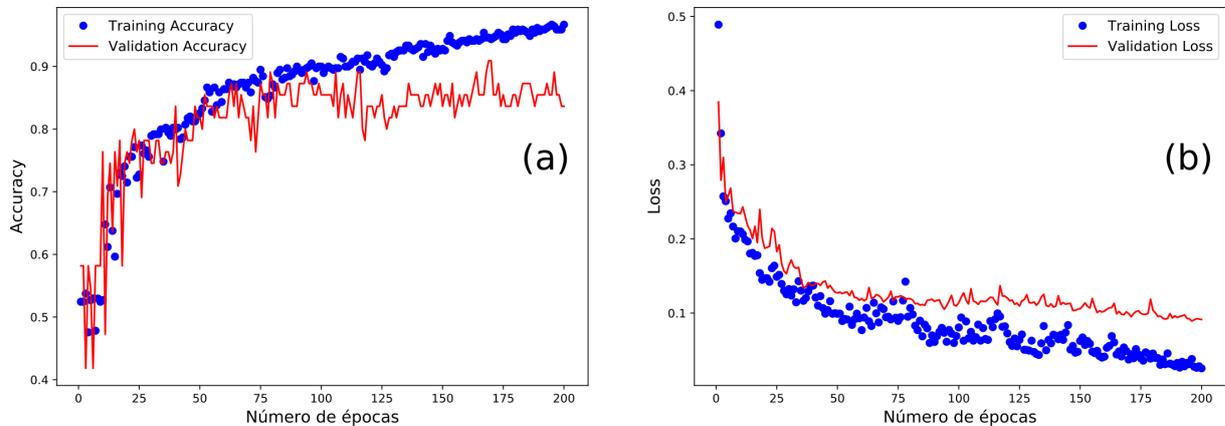


Figura 8 – Gráfico de (a) *accuracy* e (b) *loss* de treino e validação do modelo que obteve o melhor resultado no experimento 3.

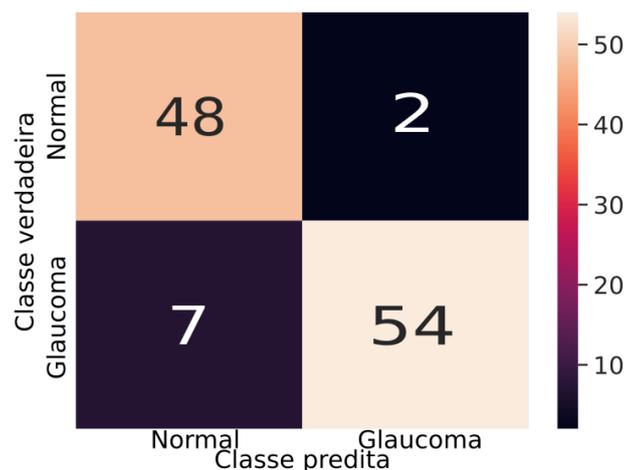


Figura 9 – Matriz de confusão gerada no teste do melhor modelo do experimento 3.

Nas Figuras 8(a) e 8(b) é possível observar que, embora as curvas de aprendizado no treino e validação tenham se distanciado um pouco por volta da época 100, a rede conseguiu aprender, já que errou apenas 8,1% dos dados de teste, com alguns falsos negativos a mais que falsos positivos, como mostra a Figura 9. Para comprovar esta afirmação de que a rede, de fato, está aprendendo, a Figura 10 apresenta as características extraídas da camada *Class Capsule* na predição do conjunto de teste. Com o propósito de facilitar a visualização das características resultantes das cápsulas, foi utilizado o algoritmo *Principal Component Analysis (PCA)* para reduzir a dimensionalidade dos dados sem perder informações importantes. O redimensionamento recebeu como entrada a saída das cápsulas, que possuem formato 2x16, resultando posteriormente ao *PCA* um vetor com dois componentes principais que carregam 99,92% das informações presentes nas cápsulas.

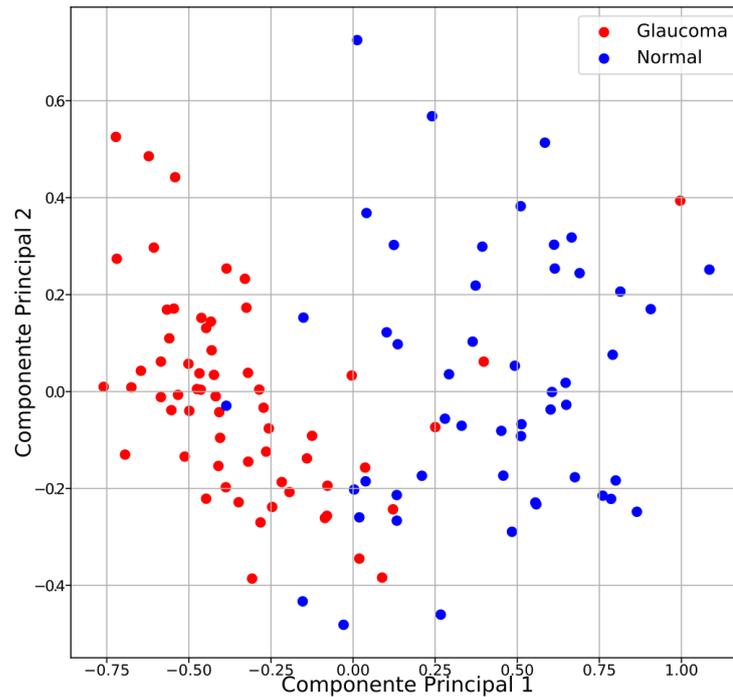


Figura 10 – Representação dos dois componentes principais gerados pelo *PCA* para o conjunto de características extraídas pela *CapsNet* no conjunto de teste do experimento 3.

Analisando a Figura 10, com o auxílio do algoritmo *PCA* para o redimensionamento das características, é possível notar a capacidade de distinção das classes pela *CapsNet*. Os resultados mostram que, apesar de alguns exemplos mais difíceis (principalmente no caso do glaucoma), o treinamento foi capaz de separar as classes de forma que seja possível visualizar uma classificação binária nos dados. Dessa maneira, observa-se que a *CapsNet* mostrou-se promissora para o problema abordado, tendo em vista que, diante das imagens realçadas e redimensionadas para  $64 \times 64$  e com um roteamento dinâmico de 8, o modelo alcançou resultados consistentes e satisfatórios em todas as métricas de validação.

O número de épocas não foi variado neste trabalho pelo fato de que, como observado nas Figuras 8(a) e 8(b), por volta da época 100, a curva de validação passou a se distanciar da curva de treino. Dessa maneira, caso o treinamento continuasse com mais épocas, é provável que esse distanciamento aumentasse, pois o modelo passaria a se ajustar demais aos dados de treinamento, causando o *overfitting*. No cenário atual, percebe-se, através dos resultados apresentados, que a rede obteve um aprendizado satisfatório, uma vez que alcançou uma capacidade discriminatória das imagens de teste.

## 5.2 Comparação com Outras Abordagens

Esta seção compara o desempenho do método proposto neste trabalho com outras abordagens aplicadas ao problema de classificação do glaucoma. Na primeira comparação, foi avaliado o desempenho de três arquiteturas de *CNN* comumente aplicadas na

literatura. Em seguida, os resultados obtidos neste trabalho foram postos em comparação com os resultados dos estudos apresentados no Capítulo 3.

### 5.2.1 Testes com Outras Arquiteturas

Os estudos apresentados no Capítulo 3 usam diferentes modelos de *CNNs* para a categorização de imagens de retina em saudáveis e glaucomatosas. Todavia, a intenção de realizar testes com outras arquiteturas é comprovar a eficácia desses modelos diante do mesmo conjunto de imagens utilizado neste trabalho e sem o uso de técnicas de segmentação e/ou aumento de dados.

Nesse contexto, foram realizados testes com três arquiteturas de *CNN* comumente aplicadas na literatura: a *ResNet-50* (HE et al., 2015), a *DenseNet-121* (HUANG et al., 2017) e a *VGG16* (SIMONYAN; ZISSERMAN, 2014). A *ResNet-50* é estruturada em 50 camadas de profundidade e aplica blocos residuais para otimizar o treinamento. A *DenseNet-121* é composta por camadas estreitas, geralmente 12 filtros por camada, e o classificador final toma a decisão com base em todos os mapas de características que a rede produziu. Por fim, a *VGG16* é um modelo de 16 camadas apresentado como uma evolução da arquitetura *AlexNet*, utilizando filtros de tamanho 3x3.

Os modelos foram executados em dois cenários de teste: usando transferência de aprendizado e ajuste fino. A transferência de aprendizado foi aplicada pelo fato de que as arquiteturas exploradas nesta seção são amplamente aplicadas com a inicialização de pesos treinados em outro conjunto de dados. Nesse teste, foram utilizados os modelos pré-treinados com a ImageNet (DENG et al., 2009), um conjunto de dados com mais de 1 milhão de imagens divididas em 1.000 classes de objetos. A predição dos modelos foi feita nas bases de imagens utilizadas neste trabalho.

Utilizar apenas os modelos pré-treinados pode não proporcionar um bom desempenho na classificação das imagens nesse problema. Em razão disso, optou-se por testar também o ajuste fino, que consiste em continuar o treinamento das imagens a partir dos pesos ajustados para outro problema, como é o caso da ImageNet. Nesse segundo teste, alguns parâmetros aplicados foram os mesmos utilizados na metodologia deste trabalho, como: número de épocas, taxa de aprendizado e otimizador.

É importante ressaltar que os mesmos conjuntos de treinamento, validação e teste foram utilizados tanto na metodologia proposta quanto nos algoritmos de comparação, com a diferença apenas na dimensão das imagens, já que as arquiteturas dessa comparação possuem um formato de entrada diferente da *CapsNet*. Ademais, as imagens foram utilizadas no canal *RGB* e realçadas pela etapa de pré-processamento, uma vez que esse mostrou-se o melhor cenário neste trabalho. Os resultados podem ser conferidos na Tabela 7.

Na Tabela 7, é possível perceber que as *CNNs* tradicionais não conseguiram distinguir com êxito as classes do problema em questão. Nenhum dos resultados mostrou-se eficiente

Tabela 7 – Resultados obtidos com a execução de outras arquiteturas de *CNNs*.

	Arquitetura	Acc (%)	Rec (%)	Prec (%)	F1 (%)	AUC	Kappa
<b>Transferência de aprendizado</b>	ResNet-50	48,64	50,00	24,32	32,00	0,50	0,00
	DenseNet-121	54,95	55,04	55,00	54,86	0,55	0,09
	VGG16	47,74	47,95	47,96	47,00	0,50	0,00
<b>Ajuste fino</b>	ResNet-50	54,95	50,00	27,47	35,00	0,50	0,00
	DenseNet-121	50,45	50,40	50,40	50,00	0,50	0,00
	VGG16	54,95	50,00	27,47	35,00	0,50	0,00

na categorização das imagens, inclusive, todos os modelos apresentaram um *kappa* ruim, de acordo com a categorização de Landis e Koch (1977). Em relação à transferência de aprendizado, os resultados mostraram que apenas os pesos da *ImageNet* não foram suficientes para que a rede identificasse corretamente as classes do problema.

Quanto ao ajuste fino, acredita-se que os resultados ruins são motivados pelo fato de que esses modelos requerem uma quantidade maior de amostras de treinamento para classificar as imagens corretamente. A *DenseNet-121*, por exemplo, possui muitos pesos, sendo assim, seu aprendizado acaba se tornando difícil em decorrência dos poucos exemplos de treinamento. Por esse motivo, as técnicas de aumento de dados são amplamente aplicadas quando o conjunto de treinamento não é vasto. Em contrapartida, diante de um mesmo cenário com poucas imagens de treinamento, o método proposto neste trabalho mostrou-se promissor na tarefa de classificação de imagens, enfatizando a capacidade das cápsulas de representar melhor as entidades das imagens, sem a necessidade de um amplo conjunto de treinamento.

## 5.2.2 Comparação com os Trabalhos Relacionados

A Tabela 8 resume a comparação dos resultados obtidos neste trabalho com os estudos descritos no Capítulo 3. Como *sensitivity* e *recall* apresentam a mesma fórmula, os resultados de *sensitivity* dos trabalhos foram considerados como *recall* na tabela, para padronizar a comparação.

Tabela 8 – Comparação do método proposto com os trabalhos relacionados.

Trabalho	Método	Bases	Amostra	Acc (%)	Rec (%)	Prec (%)	F1 (%)	AUC	Kappa
Gaddipati et al. (2019)	<i>CapsNet</i>	Privada	233	94,00	89,00	-	-	0,97	-
Phan et al. (2019)	<i>VGG19</i> , <i>ResNet152</i> e <i>DenseNet201</i>	YMU e KOSEI	3312	-	-	-	-	0,90	-
	<i>RCNN</i> e <i>DCNN</i>	ORIGA, HRF e OCT & CFI	780	-	71,17	-	-	0,87	-
Maheshwari, Kanhangad e Pachori (2020)	<i>LBP</i> e <i>Alexnet</i> pré-treinada	RIM-ONE	455	98,90	100	-	-	-	-
Bisneto, Filho e Magalhães (2020)	<i>GAN</i> e características de textura	RIM-ONE e Drishti-GS	556	100	100	-	-	1	-
Método Proposto	<i>CapsNet</i>	RIM-ONE e Drishti-GS	556	91,89	88,52	96,42	92,30	0,92	0,83

Analisando os resultados da Tabela 8, nota-se que o método proposto superou os

resultados apenas dos estudos de [Phan et al. \(2019\)](#) e [Bajwa et al. \(2019\)](#). Entretanto, o método proposto obteve um resultado satisfatório sem a necessidade de uma etapa de segmentação, ao contrário das abordagens de [Bajwa et al. \(2019\)](#) e [Bisneto, Filho e Magalhães \(2020\)](#), e sem a necessidade de uma técnica de aumento de dados, ao contrário da metodologia proposta por [Maheshwari, Kanhangad e Pachori \(2020\)](#).

Outro ponto importante é que apenas o trabalho de [Gaddipati et al. \(2019\)](#) aplica a *CapsNet* para a classificação de glaucoma em um conjunto menor de imagens, se comparado ao utilizado neste trabalho. Contudo, esse estudo avalia imagens 3D de *OCT* em sua metodologia, sendo que, conforme mencionado anteriormente, a disponibilidade desse tipo de imagem não é tão ampla quanto no caso das imagens de fundo da retina, além do fato de que o método torna-se computacionalmente complexo por lidar com muitas camadas de imagens.

### 5.3 Discussões

Como mencionado anteriormente, o objetivo deste trabalho foi avaliar se a *CapsNet* seria capaz de identificar as relações entre as características da retina de forma que um conjunto reduzido de imagens de treinamento fosse suficiente para que o modelo atingisse uma classificação satisfatória dos dados. Conforme observado na Seção 5.1, a *CapsNet*, em união ao método de pré-processamento aplicado às imagens, apresentou resultados promissores, mas com espaço para ajustes adicionais a fim de melhorar a separabilidade entre as características. A seguir, destacam-se alguns pontos essenciais sobre este trabalho:

- A abordagem proposta neste trabalho faz uso de um modelo de *DL* proposto com o intuito de solucionar as desvantagens que as *CNNs* tradicionais carregam, mas ainda é um modelo pouco explorado no contexto das imagens oculares, de acordo com os estudos levantados no Capítulo 3;
- Durante os testes, observou-se que a *CapsNet*, no problema das imagens oculares, é sensível ao canal de cor utilizado;
- Uma etapa de pré-processamento foi aplicada nas imagens de retina. Os resultados mostraram que o realce de contraste afetou positivamente a capacidade discriminatória das cápsulas;
- Com base nos resultados obtidos, nota-se que as relações entre as características das retinas são relevantes para a ativação das cápsulas;
- Os testes comparativos apresentados na Seção 5.2.2 mostraram que a *CapsNet* obteve resultados superiores às demais arquiteturas aplicadas ao mesmo problema;

- Em relação à dimensão das imagens, acredita-se que a dimensão 64x64 proporcionou à rede melhores resultados em decorrência do fato de que, por ser uma arquitetura composta por poucas convoluções e nenhuma camada de *pooling*, imagens com dimensões maiores podem desfavorecer o aprendizado da *CapsNet*.

## 6 Conclusão

Este trabalho propôs a aplicação da arquitetura *CapsNet* para a classificação de imagens de retina em glaucomatosas e saudáveis. Na abordagem proposta, foram utilizadas duas bases de imagens da imagem do fundo da retina, que passaram por uma etapa de pré-processamento, onde a equalização de histograma foi aplicada para melhorar o contraste das imagens, destacando as estruturas da retina. Em conjunto com a etapa de pré-processamento aplicada ao canal *RGB* das imagens, a *CapsNet* alcançou resultados robustos e equilibrados em todas as métricas de validação, inclusive, obteve um *kappa* de 0,83, considerado excelente, consoante a categorização de [Landis e Koch \(1977\)](#).

Os experimentos realizados neste trabalho provaram a relevância da análise das relações espaciais hierárquicas entre as características da retina para a separabilidade das classes, utilizando menos amostras de treinamento do que outras abordagens, e sem a necessidade de segmentar a região do disco óptico. Ademais, este estudo apresenta contribuições tanto no campo computacional quanto na área da saúde, em virtude da exploração de técnicas de *DL* e processamento de imagens no contexto de imagens médicas. Os resultados alcançados neste trabalho demonstram o potencial da metodologia para auxiliar o especialista na tomada de decisão sobre o diagnóstico dos exames, contribuindo para um diagnóstico precoce do paciente, conseqüentemente, evitando a perda da visão.

Os resultados mostraram-se robustos, não obstante, ainda há espaço para ajustes no método, de modo a aprimorar a classificação. Como trabalhos futuros, pretende-se: i) adicionar mais convoluções à arquitetura original para destacar regiões importantes das imagens antes de entrar nas cápsulas; e, ii) modificar o número de cápsulas no modelo para avaliar o impacto das mesmas na classificação dos dados.

## 7 Publicações

O método apresentado neste trabalho alcançou uma publicação a nível internacional, a saber:

- P. R. Sales dos Santos, V. de Carvalho Brito, A. O. de Carvalho Filho, F. Henrique Duarte de Araújo, R. de erade Lira Rabêlo e M. Joseph Mathew, "A Capsule Network-based for identification of Glaucoma in retinal images", **2020 IEEE Symposium on Computers e Communications (ISCC)**, Rennes, France, 2020, pp. 1-6, doi: 10.1109/ISCC50000.2020.9219708.

Além desta, a proponente do presente trabalho possui mais dois artigos submetidos nos seguintes jornais com os respectivos títulos:

- Título: COVID-index: a texture-based approach to classification of lung lesions. Submetido em: **Pattern Recognition**.
- Título: Texture descriptors based on species relationships for classification of skin lesions. Submetido em: **Journal of Digital Imaging**.

## Referências

- ABDEL-ZAHER, A. M.; ELDEIB, A. M. Breast cancer classification using deep belief networks. *Expert Systems with Applications*, Elsevier, v. 46, p. 139–144, 2016. Citado na página 12.
- AFSHAR, P. et al. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *arXiv preprint arXiv:2004.02696*, 2020. Citado na página 27.
- AFSHAR, P.; MOHAMMADI, A.; PLATANIOTIS, K. N. Brain tumor type classification via capsule networks. In: IEEE. *2018 25th IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2018. p. 3129–3133. Citado na página 13.
- AFSHAR, P.; PLATANIOTIS, K. N.; MOHAMMADI, A. Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries. In: IEEE. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2019. p. 1368–1372. Citado na página 27.
- ARAÚJO, F. H. et al. Redes neurais convolucionais com tensorflow: Teoria e prática. *SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. III Escola Regional de Informática do Piauí. Livro Anais-Artigos e Minicursos*, Sociedade Brasileira de Computação, v. 1, p. 382–406, 2017. Citado 2 vezes nas páginas 8 e 18.
- ARAÚJO, J. D. L. et al. Glaucoma diagnosis in fundus eye images using diversity indexes. *Multimedia Tools and Applications*, v. 78, n. 10, p. 12987–13004, May 2019. ISSN 1573-7721. Disponível em: <<https://doi.org/10.1007/s11042-018-6429-z>>. Citado na página 15.
- BAJWA, M. N. et al. Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. *BMC medical informatics and decision making*, BioMed Central, v. 19, n. 1, p. 136, 2019. Citado 4 vezes nas páginas 23, 24, 35 e 36.
- BISNETO, T. R. V.; FILHO, A. O. de C.; MAGALHÃES, D. M. V. Generative adversarial network and texture features applied to automatic glaucoma detection. *Applied Soft Computing*, Elsevier, v. 90, p. 106165, 2020. Citado 6 vezes nas páginas 16, 23, 24, 26, 35 e 36.
- BUTUN, E. et al. 1d-cadcapsnet: One dimensional deep capsule networks for coronary artery disease detection using ecg signals. *Physica Medica*, Elsevier, v. 70, p. 39–48, 2020. Citado na página 27.
- CLARK, A. *Pillow (PIL Fork) Documentation*. readthedocs, 2015. Disponível em: <<https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf>>. Citado na página 27.
- COHEN, J. Weighted kappa - nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological bulletin*, v. 70, p. 213–20, 11 1968. Citado na página 20.

- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. *2009 IEEE conference on computer vision and pattern recognition*. [S.l.], 2009. p. 248–255. Citado na página 34.
- ERKEL, A. R. van; PATTYNAMA, P. M. Receiver operating characteristic (roc) analysis: Basic principles and applications in radiology. *European Journal of Radiology*, v. 27, n. 2, p. 88 – 94, 1998. ISSN 0720-048X. Citado na página 20.
- FILHO, A. O. d. C. et al. Métodos para sistemas cad e cadx de nódulo pulmonar baseada em tomografia computadorizada usando análise de forma e textura. Universidade Federal do Maranhão, 2016. Citado na página 16.
- FUMERO, F. et al. Rim-one: An open retinal image database for optic nerve evaluation. In: . [S.l.: s.n.], 2011. p. 1 – 6. Citado 3 vezes nas páginas 8, 17 e 26.
- GADDIPATI, D. J. et al. Glaucoma assessment from oct images using capsule network. In: IEEE. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.], 2019. p. 5581–5584. Citado 6 vezes nas páginas 13, 23, 24, 27, 35 e 36.
- GHANEM, H. de O. S. A. *Glaucoma*. 2020. <https://sadalla.com.br/especialidades/glaucoma/>. Acesso em: 26 de dez. de 2020. Citado 2 vezes nas páginas 8 e 16.
- HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Citado na página 34.
- HUANG, G. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4700–4708. Citado na página 34.
- JIANG, X. et al. Hyperspectral image classification with capsnet and markov random fields. *IEEE Access*, IEEE, v. 8, p. 191956–191968, 2020. Citado na página 27.
- JIMÉNEZ-SÁNCHEZ, A.; ALBARQOUNI, S.; MATEUS, D. Capsule networks against medical imaging data challenges. In: *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. [S.l.]: Springer, 2018. p. 150–160. Citado na página 13.
- JUNIOR, A. S. V. de C. et al. Automatic methods for diagnosis of glaucoma using texture descriptors based on phylogenetic diversity. *Computers & Electrical Engineering*, Elsevier, v. 71, p. 102–114, 2018. Citado na página 12.
- KINGMAN, S. Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization*, SciELO Public Health, v. 82, p. 887–888, 2004. Citado na página 12.
- KRUTSCH, R.; TENORIO, D. Histogram equalization: Technical report. *Microcontroller Solutions Group, Guadalajara*, 2011. Citado 2 vezes nas páginas 16 e 26.
- KUMAR, S. *CNN vs CAPSULE NETWORKS*. 2019. <https://www.linkedin.com/pulse/cnn-vs-capsule-network-sonu-kumar/>. Acessado em 11-01-2021. Citado 2 vezes nas páginas 8 e 19.

- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, [Wiley, International Biometric Society], v. 33, n. 1, p. 159–174, 1977. ISSN 0006341X, 15410420. Citado 5 vezes nas páginas 22, 30, 31, 35 e 38.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015. Citado 3 vezes nas páginas 12, 13 e 17.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Taipei, Taiwan, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 17.
- LOEWEN, N. A.; TANNA, A. P. Glaucoma risk factors: Intraocular pressure. In: \_\_\_\_\_. *Clinical Glaucoma Care: The Essentials*. New York, NY: Springer New York, 2014. p. 1–22. ISBN 978-1-4614-4172-4. Disponível em: <[https://doi.org/10.1007/978-1-4614-4172-4\\_1](https://doi.org/10.1007/978-1-4614-4172-4_1)>. Citado na página 15.
- MAHESHWARI, S.; KANHANGAD, V.; PACHORI, R. B. Cnn-based approach for glaucoma diagnosis using transfer learning and lbp-based data augmentation. *arXiv preprint arXiv:2002.08013*, 2020. Citado 4 vezes nas páginas 23, 24, 35 e 36.
- MORAES, L. Â. G. d. Utilização de redes em cápsulas para classificação de regiões promotoras em sequências de dna. 2018. Citado na página 18.
- PATRICK, M. K. et al. Capsule networks—a survey. *Journal of King Saud University-Computer and Information Sciences*, Elsevier, 2019. Citado na página 23.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 28.
- PHAN, S. et al. Evaluation of deep convolutional neural networks for glaucoma detection. *Japanese journal of ophthalmology*, Springer, v. 63, n. 3, p. 276–283, 2019. Citado 4 vezes nas páginas 23, 24, 35 e 36.
- SABOUR, S.; FROSST, N.; HINTON, G. E. Dynamic routing between capsules. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2017. p. 3856–3866. Citado 6 vezes nas páginas 8, 13, 18, 20, 27 e 29.
- SHAHROUDNEJAD, A. et al. Improved explainability of capsule networks: Relevance path by agreement. In: IEEE. *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. [S.l.], 2018. p. 549–553. Citado 2 vezes nas páginas 18 e 19.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014. Citado na página 34.
- SIVASWAMY, J. et al. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In: IEEE. *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*. [S.l.], 2014. p. 53–56. Citado na página 26.
- 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. Citado 2 vezes nas páginas 12 e 15.
- TSOCHATZIDIS, L. et al. Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach. *Pattern Recognition*, Elsevier, v. 71, p. 106–117, 2017. Citado na página 12.

---

TSUJI, T. et al. Classification of optical coherence tomography images using a capsule network. *BMC ophthalmology*, BioMed Central, v. 20, n. 1, p. 1–9, 2020. Citado na página 27.

WALT, S. van der et al. scikit-image: image processing in Python. *PeerJ*, v. 2, p. e453, 6 2014. ISSN 2167-8359. Disponível em: <<https://doi.org/10.7717/peerj.453>>. Citado 2 vezes nas páginas 26 e 27.



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA  
“JOSÉ ALBANO DE MACEDO”**

**Identificação do Tipo de Documento**

- ( ) Tese
- ( ) Dissertação
- (X) Monografia
- ( ) Artigo

Eu, **Vitória de Carvalho Brito**, autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de 02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação **Uma Abordagem Baseada em Capsule Network para a Identificação de Glaucoma em Imagens de Retina**, de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título de divulgação da produção científica gerada pela Universidade.

Picos-PI, 20 de abril de 2021.

*Vitória de Carvalho Brito*  
Assinatura