UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

GILBERTO LUIS DE CONTO JUNIOR

AVALIAÇÃO DE REDES NEURAIS CONVOLUCIONAIS COM TRANSFERÊNCIA DE APRENDIZADO PARA CLASSIFICAÇÃO DE IMAGENS DE EDEMA MACULAR DIABÉTICO

PONTA GROSSA

2024

GILBERTO LUIS DE CONTO JUNIOR

AVALIAÇÃO DE REDES NEURAIS CONVOLUCIONAIS COM TRANSFERÊNCIA DE APRENDIZADO PARA CLASSIFICAÇÃO DE IMAGENS DE EDEMA MACULAR DIABÉTICO

Evaluation of convolutional neural networks with transfer learning for images classification of diabetic macular edema

Dissertação apresentada como requisito para obtenção do título de Mestre em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Ionildo José Sanches.

PONTA GROSSA

2024

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Ponta Grossa

GILBERTO LUIS DE CONTO JUNIOR

AVALIAÇÃO DE REDES NEURAIS CONVOLUCIONAIS COM TRANSFERÊNCIA DE APRENDIZADO PARA CLASSIFICAÇÃO DE IMAGENS DE EDEMA MACULAR DIABÉTICO

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Ciência Da Computação da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Sistemas E Métodos De Computação.

Data de aprovação: 08 de Março de 2024

Ionildo Jose Sanches, - Universidade Tecnológica Federal do Paraná

Dr. Erikson Freitas De Morais, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Mauren Abreu De Souza, Doutorado - Pontifícia Universidade Católica do Paraná (Pucpr)

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 08/03/2024.

RESUMO

O Diabetes Mellitus é uma doença que acomete aproximadamente 422 milhões de pessoas ao redor do mundo. A Retinopatia Diabética é uma doença que afeta os pequenos vasos da retina, sendo uma das mais frequentes complicações do diabetes, podendo progredir para a cegueira. A doença pode atingir a área central da retina, a mácula, causando o edema macular diabético. Este edema pode ser diagnosticado através de um exame de Tomografia de Coerência Óptica (OCT) para observar as camadas da retina e identificar os sinais da doença como vazamento de fluidos na região. Este trabalho tem por objetivo a utilização de modelos pré-treinados de redes neurais convolucionais, com a aplicação de transferência de aprendizado, para classificação de imagens de OCT e realizar uma análise comparativa dessas redes. Foram utilizadas as arquiteturas de rede ResNet-50, *Inception* V3, AlexNet, GoogLeNet e VGG-19. Para o treinamento das redes foram utilizadas 22.696 imagens de OCT, igualmente divididas em exames normais e com edema macular diabético. Os modelos foram testados com 250 imagens de cada classe conforme já definidas previamente no *dataset*. Após cada modelo ser treinado e aplicado na base de teste, percebeu-se que a rede GoogLeNet foi a que apresentou melhor desempenho na classificação. Os resultados obtidos com a GoogLeNet foram uma acurácia de 99,20%, uma precisão de 98,43%, um *recall* de 100%, um *F-measure* de 99,21% e uma especificidade de 98,40%.

Palavras-chave: edema macular diabético; tomografia de coerência óptica; processamento digital de imagens; rede neural convolucional; transferência de aprendizado.

ABSTRACT

Diabetes Mellitus is a disease that affects approximately 422 million people around the world. Diabetic Retinopathy is a condition that impacts the small vessels of the retina, being one of the most common complications of diabetes, which can progress to blindness. The disease can affect the central area of the retina, the macula, causing diabetic macular edema. This edema can be diagnosed through an Optical Coherence Tomography (OCT) exam to observe the layers of the retina and identify signs of the disease, such as fluid leakage in the area. This work aims to use pre-trained models of convolutional neural networks, applying transfer learning for the classification of OCT images, and to perform a comparative analysis of these networks. The network architectures used were ResNet-50, Inception V3, AlexNet, GoogLeNet, and VGG-19. For the training of the networks, 22,696 OCT images were used, equally divided between normal exams and those with diabetic macular edema. The models were tested with 250 images of each class as previously defined in the dataset. After each model was trained and applied to the test base, it was observed that the GoogLeNet network presented the best performance in classification. The results obtained with GoogLeNet were an accuracy of 99.20%, a precision of 98.43%, a recall of 100%, an F-measure of 99.21%, and a specificity of 98.40%.

Keywords: diabetic macular edema; optical coherence tomography; digital image processing; convolutional neural network; transfer learning.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ADAM	*Adaptive Moment Estimation* (Estimativa de Momento Adaptativo)
CNN	*Convolutional Neural Network* (Rede Neural Convolucional)
DM	*Diabetes Mellitus*
DME	*Diabetic Macular Edema* (Edema Macular Diabético)
IA	Inteligência Artificial
MLP	*Multi-Layer Perceptrons* (*Perceptron* Multicamadas)
OCT	*Optical Coherence Tomography* (Tomografia de Coerência Óptica)
OpenCV	*Open Source Computer Vision Library* (Biblioteca de Visão Computacional de Código Aberto)
ReLU	*Rectified Linear Unit* (Unidade Linear Retificada)
RMSProp	*Root Mean Squared Propagation* (Propagação da Raiz Quadrada Média)
VEGF	*Vascular Endothelial Growth Factor* (Fator de Crescimento Endotelial Vascular)

SUMÁRIO

1 INTRODUÇÃO

O Diabetes Mellitus (DM) é uma doença caracterizada pela elevação da glicose no sangue (hiperglicemia). Esta doença se manifesta em dois tipos, o tipo 1 no qual o pâncreas não produz insulina ou produz muito pouco e o tipo 2 ocorre quando o organismo se torna resistente à insulina. Estima-se que uma em cada dez pessoas com idade entre 20 e 79 anos têm diabetes. Isso equivale a aproximadamente 537 milhões de pessoas no mundo segundo o Atlas do Diabetes da Federação Internacional de Diabetes (*International Diabetes Federation* - IDF) (IDF, 2021).

Esta doença traz complicações para o portador, incluindo a Retinopatia Diabética que pode causar cegueira ao afetar os vasos sanguíneos da retina, resultando em vazamentos e pontos de visão turva, bem como manchas escuras semelhantes a teias de aranha, de acordo com o Instituto Nacional do Olho (*National Eye Institute* - NEI) (NEI, 2021). A Retinopatia Diabética é uma doença que por si só pode causar severos danos à saúde ocular do portador e em seus estágios mais avançados pode levar à outras complicações como descolamento de retina, glaucoma neovascular e edema macular diabético (*Diabetic Macular Edema* - DME). Acomete uma em cada quinze pessoas diabéticas e é a principal causa de perda de visão em pessoas com Retinopatia Diabética. O Edema Macular Diabético ocorre quando vazamentos de fluidos da retina atingem a região da mácula causando uma visão borrada no paciente (NEI, 2019).

Para o diagnóstico do edema macular diabético são comumente usados exames como a Angiofluoresceinografia (exame que avalia o fluxo dos vasos da retina através de fotografias do fundo do olho), e principalmente a Tomografia de Coerência Óptica ou OCT (*Optical Coherence Tomography*). Na OCT um aparelho provê imagens detalhadas das camadas da retina para que o especialista avalie a incidência da doença verificando a presença ou não de fluidos na região macular (AAO, 2020).

O processamento digital de imagens envolve técnicas de transformação de imagens, como transformações geométricas, quantização, suavização e realce e operações morfológicas, visando melhorar suas características visuais para a interpretação humana e o processamento de cenas para percepção de máquinas (visão computacional) (GONZALEZ; WOODS, 2008).

As redes neurais convolucionais (CNNs) que são uma das mais importantes estruturas no campo do *deep learning*, tem atraído muita atenção da indústria e da academia nos últimos anos (LI *et al.,* 2022). Esta estrutura é um tipo de rede neural útil especialmente em tarefas relacionadas a imagens como classificação e segmentação semântica (WU, 2017).

Uma forma de aprimorar os resultados de uma rede neural convolucional é utilizar o método de transferência de aprendizado. Este procedimento consiste na transferência dos recursos aprendidos por uma rede neural aplicada a outra base de dados em uma nova, resultando na melhoria da capacidade de generalização da nova rede neural (FAWAZ *et al.,* 2018).

Para o auxílio na classificação das imagens se propõe neste trabalho a aplicação de redes neurais convolucionais em uma base de exames de OCT contendo imagens normais da retina e imagens com edema macular diabético. Inicialmente aplica-se técnicas de pré-processamento visando eliminar ruídos e evidenciar a região de interesse para em seguida ser treinada utilizando modelos de redes neurais pré-treinados. As redes neurais pré-treinadas propostas neste trabalho são a ResNet-50, *Inception* V3, AlexNet, GoogLeNet e VGG-19. Após os modelos treinados e realizada as classificações na base de teste são mensuradas métricas de desempenho como precisão, acurácia, especificidade, *recall* e *F-measure* utilizando as imagens da base de teste para avaliar como cada rede se comportou na classificação. Para a validação dos resultados obtidos são feitas comparações gráficas com outros resultados disponíveis na literatura.

1.1 Justificativa

A aplicação de Inteligência Artificial na área de saúde é vista com um bom prognóstico partindo da premissa que esta não substituiria o trabalho do médico especialista, mas aprimoraria seu trabalho (BOHR; MEMARZADEH, 2020). Devido ao grande número de dados gerados pelos exames modernos a Inteligência Artificial (IA), bem como seus ramos, podem auxiliar os diagnósticos por um especialista. Desta forma, permite que através de uma imagem ou de um conjunto de dados haja uma diminuição do tempo necessário para se obter um diagnóstico mais preciso (KRITTANAWONG *et al.,* 2017).

O fato de a utilização de inteligência artificial estar em constante expansão, juntamente com os grandes conjuntos de dados disponíveis na era digital, trouxe avanços para as tarefas que precisam analisar dados de forma precisa, apoiando o diagnóstico de doenças como cânceres e complicações advindas do diabetes (ALUGUBELLI, 2016).

Alugubelli (2016) é incisivo ao dizer que com o advento do aprendizado de máquina/Inteligência Artificial no sistema de saúde, espera-se uma expansão da automação na tomada de decisões clínicas. Estima-se que a utilização de Inteligência Artificial para análise de imagem e auxílio de diagnósticos médicos chegue a um investimento de aproximadamente 2.5 bilhões de dólares até 2025 (*Health IT Analytics*, 2017).

Como o edema macular diabético é diagnosticado através do exame de Tomografia de Coerência Óptica, no qual o médico avalia as camadas internas da retina, o uso de uma inteligência artificial para auxílio do processo de diagnóstico se torna viável com o objetivo de reduzir o erro humano e diminuir o custo e o tempo do diagnóstico além de abrir um precedente para, no futuro, levar o diagnóstico da doença para áreas remotas através da telemedicina.

Este diagnóstico pode ser melhorado ainda mais com a manipulação das imagens de entrada realçando suas regiões de interesse e permitindo que o sistema consiga abstrair mais dados relevantes para realizar o treinamento e a classificação das imagens.

Um estudo comparativo permitiu conhecer quais são as redes que melhor performam para o problema de classificação de imagens com edema macular diabético, de forma a haver um ponto de partida nesta proposta, aproveitando o conhecimento já exposto na literatura.

1.2 Objetivos

Nesta seção serão expostos o objetivo geral e os objetivos específicos deste trabalho dividido em subseções.

1.2.1 Objetivo geral

O objetivo deste trabalho é aplicar e avaliar o desempenho de redes neurais convolucionais com transferência de aprendizado para classificação de edema macular diabético em imagens de tomografia de coerência óptica.

1.2.2 Objetivos específicos

Os objetivos específicos deste trabalho, são apresentados a seguir:

- Identificar bases de dados de imagens de OCT de retinas normais e com edema macular diabético;
- Utilizar *data augmentation* para aumentar a quantidade de imagens dos conjuntos do *dataset*;
- Implementar e aplicar técnicas de processamento digital de imagens que permitam melhorar o treinamento e a classificação das imagens;
- Implementar um sistema que permita o treinamento e a classificação de imagens de OCT utilizando modelos pré-treinados de redes neurais convolucionais com transferência de aprendizado;
- Validar e testar o sistema implementado;
- Analisar quais arquiteturas CNNs apresentam melhores resultados para a classificação de edema macular diabético utilizando imagens de OCT.

1.3 Organização do trabalho

Este documento está organizado em seis capítulos. No Capítulo 2 são apresentados os fundamentos teóricos para entendimento do trabalho, tais como o sistema visual humano, o diabetes mellitus, a retinopatia diabética, o edema macular diabético, técnicas de processamento digital de imagens, aprendizado de máquina, transferência de aprendizado, modelos de arquiteturas de redes pré-treinadas, métricas de desempenho e o estado da arte. No Capítulo 3 é apresentada a metodologia utilizada para o desenvolvimento do trabalho. No Capítulo 4 são expostos os resultados obtidos. No Capítulo 5 é apresentada uma análise dos resultados e no Capítulo 6 a conclusão e propostas para trabalhos futuros.

2 REFERENCIAL TEÓRICO

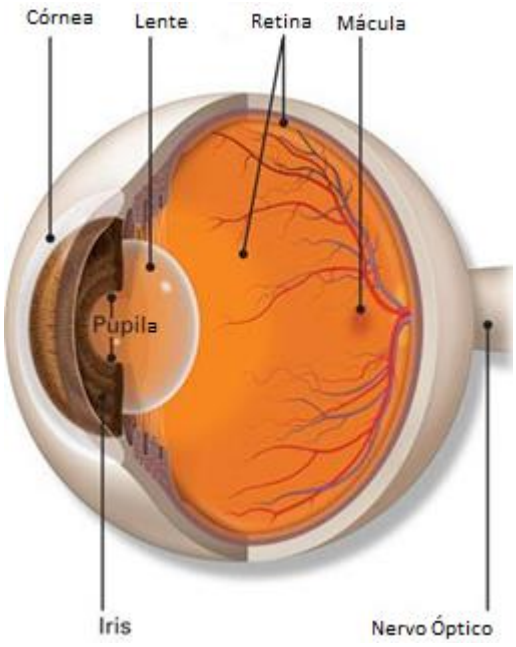
Nesta seção serão abordados os temas pertinentes ao trabalho de acordo com a visão teórica de alguns autores.

2.1 Sistema visual humano

O ser humano, assim como 95% das espécies animais, possui em seu sistema visual um conjunto de órgãos sensíveis à luz e especializados em obter suas informações como detalha a Figura 1 (HELENE; HELENE, 2011).

As partes gerais do sistema visual humano incluem o olho, o cérebro e os nervos ópticos. O olho é o ponto de entrada da luz que aciona todo o sistema visual sendo assim o sensor utilizado na definição de todo o processo. A percepção da imagem é um processo abrangente que envolve capturá-la com o olho, reconhecê-la e interpretar o conteúdo no cérebro (LYAPUNOV, 2017). O nervo óptico transmite o impulso de luz percebido pelo olho ao cérebro para interpretação. A coordenação dos vários níveis de imagens é possibilitada pela transmissão da luz no olho e a possível interpretação no cérebro.

Figura 1 - Órgãos oculares



Fonte: Adaptada de AAO - American Academy of Ophthalmology (2020).

De acordo com a ilustração apresentada na Figura 1, é possível observar que o olho humano é composto por sete elementos fundamentais, tais como a córnea, a íris, a pupila, a lente, a retina, a mácula e o nervo óptico. A retina é a estrutura responsável por formar as imagens que o olho é capaz de perceber, as quais são transmitidas instantaneamente para o cérebro através do nervo óptico. Já a mácula é uma parte da retina, responsável pela visão central. Uma vez que a região da mácula é obstruída, seja por decorrência do Diabetes ou pela própria idade do paciente, pode causar borramento ou perda de visão.

2.1.1 Retina

A retina é uma estrutura formada por células sensíveis à luz, cones e bastonetes, que absorvem as imagens projetadas, bem como neurônios, que mandam o que foi absorvido ao cérebro por meio do nervo óptico (BOSCO *et al.,* 2005).

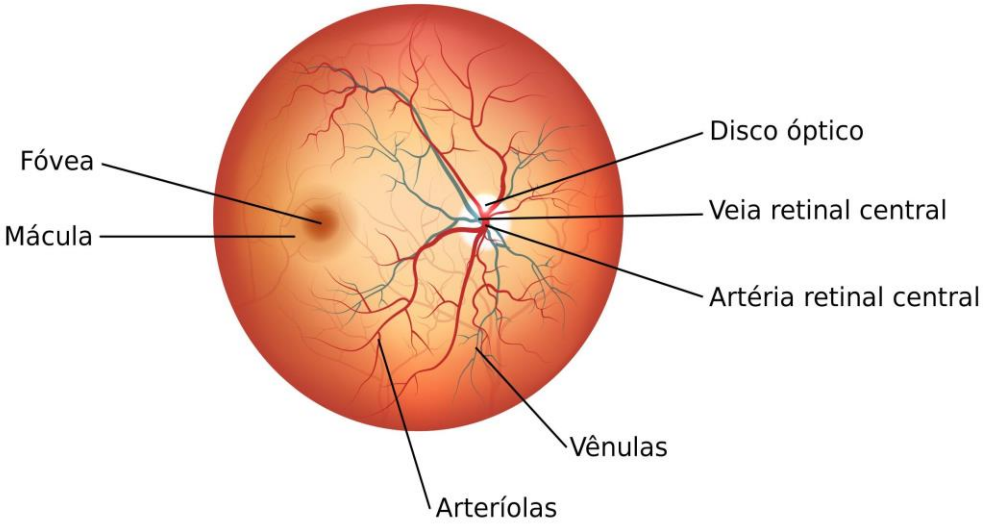
A superfície da retina humana é revestida por numerosos fotorreceptores. Os fotorreceptores são de grande ajuda no estabelecimento da conversão da luz em reações eletroquímicas. Estas reações eletroquímicas que ocorrem na retina são transmitidas ao cérebro para o processamento da imagem (MASHTAKOV, A., 2019). Percebe-se, portanto, que a retina desempenha um papel chave no sistema visual humano. Sua ineficiência seja por qualquer motivo, como obstrução, significa que a qualidade dos impulsos eletroquímicos para o cérebro para interpretação será pior, levando a dificuldades de visão.

De acordo com Silva (2018), a mácula é uma região muito importante da retina, por conter maior concentração de cones, fazendo com que as imagens projetadas sejam nítidas. Os cones são especializados na visão diurna e no reconhecimento das cores. A fóvea é uma pequena depressão presente na mácula.

O disco óptico é a zona de entrada do nervo óptico, ou seja, o ponto que o nervo óptico se expande na retina tendo o aspecto de um disco esbranquiçado, em que se desenham os vasos sanguíneos que suprem o olho (Figura 2).

Os bastonetes estão ausentes na mácula e se concentram em maior parte na periferia da retina. Em ambientes pouco iluminados, conseguem captar imagens, diferenciar tonalidades de claro e escuro, além de diferenciar movimentos e formas das imagens (SILVA, 2018).

Figura 2 - Estrutura da retina



Fonte: Silva (2018).

Percebe-se através da Figura 2 que a região do nervo óptico é largamente irrigada por veias e artérias. Por isso, há muita probabilidade de vazamento de fluidos nesta região podendo obstruir a visão do paciente haja vista que as obstruções e morte das células da retina causam cegueira.

2.2 Diabetes mellitus

Diabetes Mellitus (DM) é um grupo de doenças metabólicas caracterizadas por hiperglicemia, ou seja, excesso de glicose no sangue, consequência de defeitos na secreção ou ação da insulina. A hiperglicemia crônica da diabetes está relacionada com danos duradouros, disfunções e falha em vários órgãos e sistemas, especialmente os olhos, rins, coração e vasos sanguíneos (TARR *et al.,* 2013).

Dados da Organização Mundial de Saúde (OMS) apontaram, em 2014, que 16 milhões de brasileiros têm DM. A taxa de incidência da doença cresceu 61,8% nos anos entre 2004 e 2014, em que o Brasil ocupa o 4º lugar no ranking dos países com o maior número de casos (BEAGLEY *et al.,* 2014).

São muitos os processos patogênicos envolvidos no desenvolvimento do DM. Podem ser desde a destruição autoimune das células β pancreáticas, com consequente deficiência na produção de insulina, até anormalidades no metabolismo de carboidratos, gorduras e proteínas, resultando em resistência à ação da insulina.

Essa ação deficiente da insulina é resultado da secreção inadequada e/ou diminuição da resposta dos tecidos à insulina em um ou mais pontos da ação hormonal (TARR *et al.,* 2013).

> Complicações no longo prazo de diabetes incluem retinopatia com grandes chances de perda de visão; nefropatia levando a insuficiência renal; neuropatia periférica com risco de úlceras nos pés, amputações e articulações de Charcot; e neuropatia autonômica causando sintomas gastrointestinais, geniturinários e cardiovasculares. Pacientes com DM apresentam uma maior incidência de doença cardiovascular, arterial periférica e cerebrovascular, bem como hipertensão e anormalidades do metabolismo das lipoproteínas (TARR *et al.,* 2013, p.11).

A má gestão da condição de diabetes leva a complicações de saúde importantes, como a retinopatia diabética que na maioria dos casos é uma condição que leva a problemas de visão. A complicação pode levar à cegueira se os cuidados não forem tomados corretamente (SCHEEN, 2021).

A maioria dos casos de DM pertence a duas categorias: No tipo 1, a causa é a destruição das células β e uma total deficiência na secreção de insulina. No tipo 2, que é mais recorrente, tem como causa uma combinação de resistência à ação da insulina e uma inadequada resposta secretora de insulina (GUARIGUATA *et al.,* 2014).

2.3 Retinopatia diabética

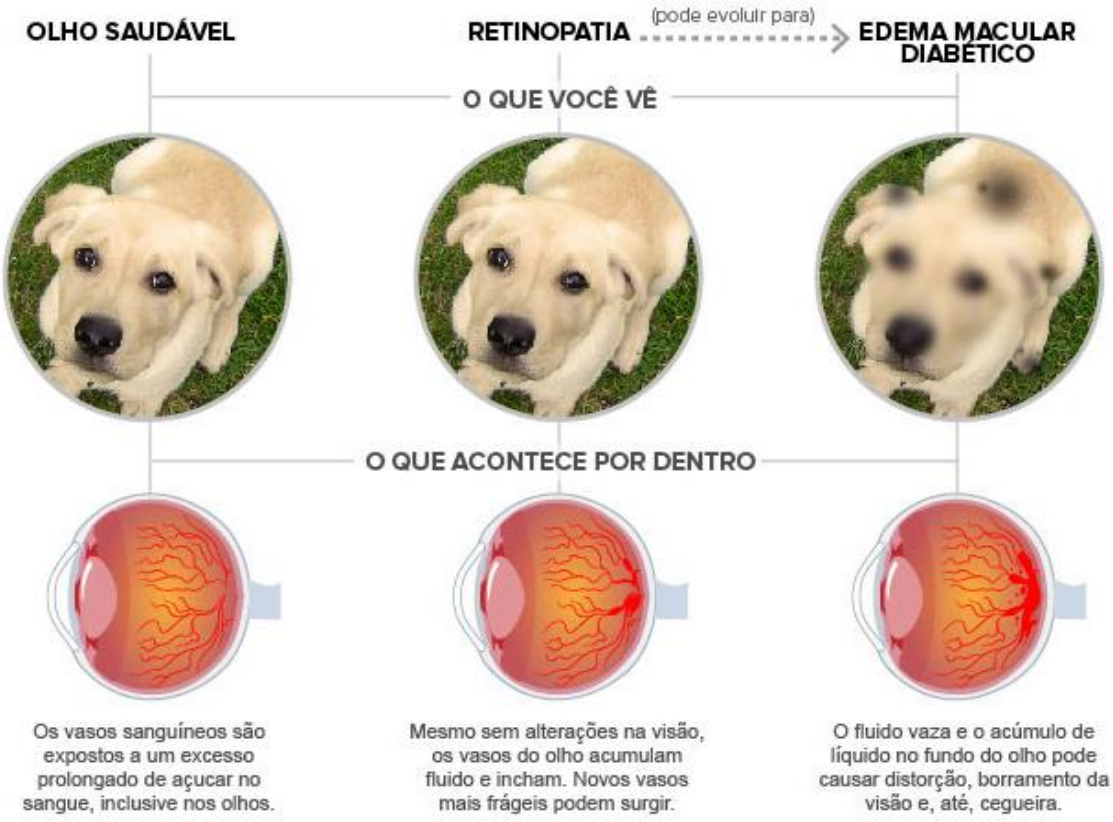
A retinopatia diabética é uma doença que atinge os pequenos vasos sanguíneos existentes na retina. Tarr *et al.,* (2013) destacam que é a causa mais frequente de casos de cegueira em adultos entre 20 e 74 anos. Durante as primeiras duas décadas de doença, quase todos os pacientes com diabetes tipo 1 e mais de 60% dos pacientes com diabetes tipo 2 têm retinopatia diabética (BOELTER *et al.,* 2003).

> A retinopatia diabética é a manifestação retiniana de uma microangiopatia sistêmica generalizada que pode ser observada na forma de edema de retina, exsudatos e hemorragias. Hemorragias intra-retinianas apresentam-se em formatos variados e são características desta fase (DORCHY, 1993, p.1848).

A retinopatia diabética traz complicações crônicas para os portadores de diabetes apesar do vasto conhecimento acumulado na literatura sobre os fatores de risco e estratégias de redução dos efeitos. Evidências apontam para a participação de fatores genéticos na surgência e evolução da retinopatia diabética, tanto como fator

de risco quanto como proteção contra a doença (ESTEVES, 2009). A Figura 3 mostra detalhes da doença.

Figura 3 - Exemplificação da progressão de retinopatia diabética



Fonte: Motta; Coblentz; Melo (2008).

Percebe-se na Figura 3 como a Retinopatia afeta a retina, podendo avançar para o edema macular diabético. O avanço da condição pode levar anos até que possa causar um grande problema de visão, porém é possível atingir o nível de cegueira se o diagnóstico adequado e tratamento precoce não for definido (MARASHI, 2016).

2.3.1 Estágios da retinopatia diabética

Os estágios progressivos da retinopatia diabética podem ser identificados clinicamente. No estágio inicial a doença é caracterizada por apresentar micro aneurismas capilares, hemorragias e exsudatos. A fase seguinte é a pré-proliferativa, caracterizada por exsudatos algodonosos ou áreas de infarto retiniano com isquemia progressiva. A fase proliferativa, que é a mais severa e pode levar ao Edema Macular

Diabético, é caracterizada pela neovascularização da retina, disco óptico e íris. Essa neovascularização desencadeia complicações como hemorragia vítrea e descolamento tracional da retina que levam à cegueira (SCHEFFEL *et al.,* 2004).

Nos estágios menos avançados as características da doença são a aparição de micro aneurismas que são pequenas dilatações vasculares, hemorragias e vasos sanguíneos obstruídos, fazendo com que várias regiões da retina não recebam oxigênio e nutrientes levados pelo sangue. No estágio mais avançado, ocorre um crescimento excessivo de novos vasos sanguíneos, os neovasos, na superfície da retina. Isso ocorre devido à obstrução ou ruptura de vasos sanguíneos antigos. (CIULLA; AMADOR; ZINMAN, 2003).

Os neovasos não causam nenhum sintoma ou perda de visão, mas por serem muito finos e frágeis, podem romper e causar hemorragias e, como consequência, levar à cegueira. (CIULLA; AMADOR; ZINMAN, 2003).

2.3.2 Tratamento da retinopatia diabética

Existem modalidades de tratamento que podem prevenir ou retardar o início da retinopatia diabética, bem como prevenir a perda da visão em grande parte dos pacientes com diabetes. O controle glicêmico e da pressão arterial pode prevenir e retardar a progressão da RD em pacientes com DM. A terapia oportuna de fotocoagulação a laser também pode prevenir a perda de visão em uma grande proporção de pacientes (FONG *et al.,* 2004).

Segundo Fong *et al.,* a foto coagulação a laser de argônio é o primeiro tratamento e deve ser instituído precocemente, antes que a doença se torne sintomática. A fotocoagulação focal ou a fotocoagulação pan-retiniana podem reduzir o risco de perda da visão em pacientes com retinopatia diabética.

Pacientes que apresentam edema macular, retinopatia não proliferativa moderada ou grave e qualquer retinopatia proliferativa devem ser encaminhados a um especialista experiente na área, pois além da fotocoagulação a laser, são necessários métodos terapêuticos adicionais, como agentes anti-inflamatórios, antiproliferativos, por exemplo, infusão paralímbica trans escleral de triamcinolona intra-hialoidea, e nos casos mais avançados, a cirurgia vitreorretiniana retinopexia/vitrectomia para

recuperação da perda visual iminente ou já instalada, como na hemorragia vítrea ou descolamento de retina (FONG *et al.,* 2004).

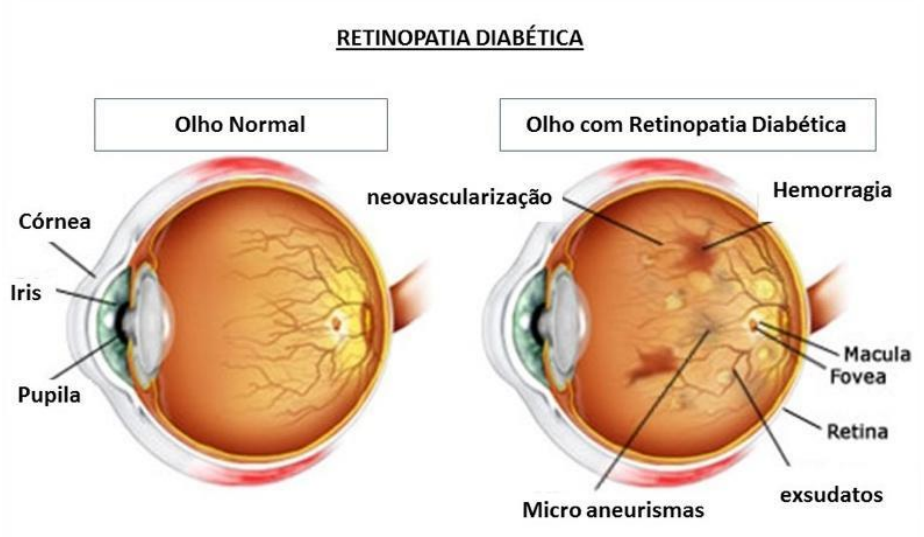
2.3.3 Detecção da retinopatia diabética

O método de documentação validado para a triagem da retinopatia diabética é a documentação fotográfica do fundo de olho, que apresenta muitas vantagens logísticas. Duas a quatro fotos de vários ângulos do fundo da retina de ambos os olhos são examinadas (ALDINGTON *et al.,* 1995). A avaliação inclui:

- A determinação do tipo morfológico do elemento presente (micro aneurismas; hemorragias; exsudatos duros e/ou algodonosos; anormalidades vasculares; edema macular exudativo ou isquêmico; rosário venoso; proliferação vascular; tecido fibroso; e outros);
- A localização desses elementos;
- O número aproximado desses elementos;
- Outros procedimentos como biomicroscopia da retina com lâmpada de fenda e/ou angiografia com fluoresceína devem ser julgados pelo oftalmologista (ALDINGTON *et al.,* 1995).

A Figura 4 apresenta a diferença entre uma retina saudável e uma retina acometida pela retinopatia diabética.

Figura 4 - Diferença entre retina saudável e retina com retinopatia diabética



Fonte: Agrizzi (2020).

Como é possível observar na Figura 4, a retina acometida pela doença apresenta vários pontos hemorrágicos, exsudatos e micro aneurismas que são os principais fatores levados em consideração na hora do diagnóstico pois prejudicam ou impedem a visão saudável do paciente.

2.4 Edema macular diabético

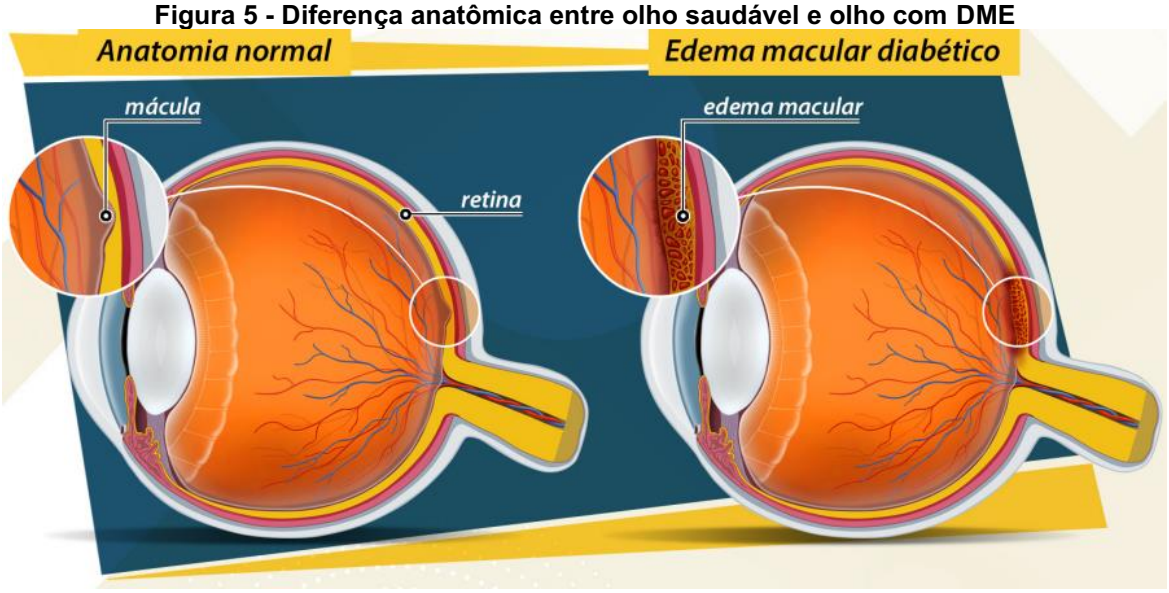
Edema macular é uma doença causada pelo depósito de fluidos e proteínas, exsudatos duros, sobre ou sob a mácula do olho, tornando-a mais espessa e inchada. É a principal causa de perda de visão de pacientes com diabetes mellitus (COSCAS; CUNHA-VAZ; SOUBRANE, 2010).

Isto é consequência do excesso prolongado de glicose no sangue pelo diabetes. Fazendo com que os vasos sanguíneos dos olhos absorvam mais líquido, inchando a retina e afrouxando as junções entre as células das paredes dos vasos sanguíneos. Ocorrendo um extravasamento de sangue e/ou gordura para os tecidos em volta da retina, quando o vazamento é próximo a mácula há maior risco de edema macular (VIEIRA, 2013).

A hiperglicemia é responsável pelo aumento da permeabilidade dos vasos sanguíneos, acumulando liquido e depósitos de proteínas na retina ou mácula causando edema macular. O fator de crescimento vascular é uma proteína que causa permeabilidade vascular enfraquecendo as junções ente as células das paredes dos vasos sanguíneos. À medida que as junções enfraquecem, os líquidos dentro dos vasos sanguíneos vazam para o tecido retiniano incluindo a mácula, formando o edema macular diabético (COSTA, 2017).

No início da doença o paciente não sente dificuldade para enxergar, mas com o tempo pode ocorrer borrões ou distorção da visão. Outros sintomas são a sensibilidade alterada do contraste, mudança na visualização de cores, alterações no campo de visão, bem como fotofobia, ou seja, sensibilidade a luz (VIEIRA, 2013). A Figura 5 mostra a diferença anatômica entre um olho saudável e um olho com edema macular diabético.

Figura 5 - Diferença anatômica entre olho saudável e olho com DME



Fonte: Momento Diabetes (2020).

Por ser uma complicação da diabete mellitus, o edema macular diabético, muitas vezes, não está atrelado à retinopatia diabética, mas que devem ser tratadas, pois o edema macular diabético se não for diagnosticado e tratado adequadamente pode causar cegueira irreversível. Considerando-se que o fator de risco no desenvolvimento do edema macular diabético é a duração do diabetes (MOHAMED; GILLIES; WONG, 2007).

De acordo com um estudo de Suñer *et al*., (2017), o edema macular é um fluido que se acumula na mácula, uma área no centro da retina. O estudo aponta que os sintomas da doença incluem visão embaçada ao olhar para objetos próximos. O edema macular também pode causar desbotamento das cores e alguns pacientes podem sofrer perda de visão. De acordo com pesquisa realizada por de Mendes *et al*. (2020), a tomografia de coerência óptica é uma técnica desenvolvida para realizar imagens transversais não invasivas em sistemas biológicos. Mendes *et al*. (2020) também argumentam que "A técnica de OCT usa interferometria de baixa coerência para produzir uma imagem bidimensional de espalhamento óptico de microestruturas de tecido interno de uma forma que é análoga à imagem de pulso-eco ultrassônica." As técnicas de imagens tomográficas incluem imagens de ultrassom, imagens de ressonância magnética e imagens de ultrassom.

Quando diagnosticado precocemente, o edema macular diabético pode ser revertido e a visão perdida pode ser recuperada. São necessários exames frequentes

dos olhos para indivíduos com diabetes mellitus, pelo risco de desenvolverem retinopatia diabética e edema macular diabético pois se detectados no início o tratamento será mais eficaz (MOHAMED; GILLIES; WONG, 2007).

Há que se considerar que o tratamento disponível, há pouco tempo atrás, era o laser, em que se queimam os vasos para estancar o vazamento do líquido. Atualmente há terapias modernas para o tratamento do edema macular diabético, que incluem os inibidores de anti-VEGF (do inglês: *Vascular Endothelial Growth Factor*) e os polímeros de liberação lenta de esteroides. Ambos administrados através de aplicações dentro do olho, chamadas de intravítreas (MATHEW; YUNIRAKASIWI; SANJAY, 2015).

2.4.1 Tratamento do edema macular diabético

Um tratamento de DME visa dificultar ou reduzir a celeridade do avanço da doença e, se viável, acurar os efeitos morfológicos visuais, visando reverter o máximo possível da perda de visão.

O tratamento do DME e da retinopatia diabética em geral, se constitui inicialmente no monitoramento sistêmico, especificamente da glicemia, hemoglobina glicada (HbA1c), níveis pressóricos, lípidos séricos, a função renal e índice de massa corporal, associado a exercício físico e alimentação adequada (BANDELLO *et al.,* 2014).

De acordo com Iverson e Clark (2017) as opções terapêuticas para o tratamento de DME são:

- Terapias baseadas em laser: são a fotocoagulação e pan-coagulação, em que é usado calor de um laser para selar os vasos sanguíneos na retina;

- Injeção intravítrea de substância anti-VEGF, é um procedimento indolor, gotas anestésicas são aplicadas no olho, e uma agulha fina e curta é usada para injetar medicação no gel vítreo, o fluido no centro do olho. Esse tratamento interrompe a atividade do VEGF postergando o aumento do EM;

- Vitrectomia, procedimento cirúrgico efetuado quando o EM é gerado por vítreo, ou seja, o gel que preenche a área entre a lente e a retina, movendo a mácula. Este procedimento remove o gel vítreo, amenizando a tração da mácula, bem como

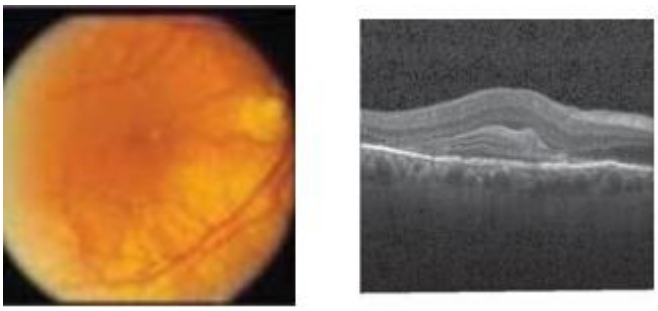
para retirar o sangue coletado no vítreo ou para corrigir a visão quando outros tratamentos são malogrados;

- Corticosteróides (esteróides), diminuem a inflamação, principal tratamento para o EM suscitado por doenças oculares inflamatórias. São medicamentos anti-inflamatórios, administrados por meio de colírios, pílulas, implantes ou injeções de corticosteróides de liberação sustentada dentro ou ao redor do olho.

2.4.2 Detecção do edema macular diabético

A avaliação ocorre por meio de um exame oftalmológico como exame de fundo de olho (ou retinografia), angiografia fluoresceínica (AF) e exame de Tomografia de Coerência Óptica (OCT). A Figura 6 apresenta o exame de fundo de olho (Figura 6 (a)) e o exame de OCT (Figura 6 (b)).

Figura 6 - Exames para diagnóstico de DME. (a) Exame de fundo de olho. (b) Exame de OCT.



(a) (b)

Fonte: Adaptado de Kaymak; Serener, (2018).

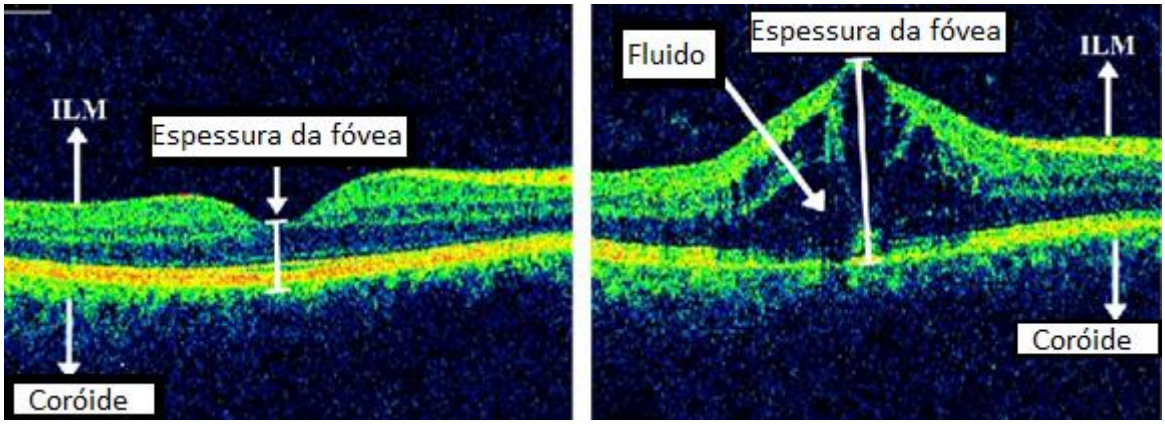
Um dos principais exames é o de OCT que é um modo de imagem da retina não-invasivo, com alta resolução, que usa diferentes comprimentos de onda de luz infravermelha refletida pelas estruturas retinianas para construir bidimensionais, tridimensionais ou seccionais que possibilitam analisar detalhadamente as diferentes camadas retinianas e inferir dados sobre volume e espessura (SANTOS, 2019).

Trata-se de uma tecnologia de imageamento proposta no início dos anos 1990 que foi desenvolvida para detectar falhas em fibras ópticas, mas a capacidade de diferenciar as estruturas do olho e outros tecidos biológicos, pouco tempo depois foi reconhecida (SCHMITT; KNÜTTEL; BONNER, 1993).

A técnica OCT tem por base a interferometria de baixa coerência, uma técnica que usa o interferômetro de *Michelson*, em sua construção (MICHELSON, 1995). A interferometria é um método de medição que usa o fenômeno de interferência das ondas, que representam a superposição de duas ou mais ondas em um mesmo ponto (MARQUES, 2012).

A Figura 7 mostra a diferença entre uma retina saudável e uma retina acometida pelo edema macular diabético através do exame de OCT.

Figura 7 - Diferença entre retinas. (a) Retina saudável e (b) Retina com edema macular diabético



(a) **(b)**

Fonte: Adaptado de Hassan; Raja (2016).

Como é possível se observar na Figura 7(b), a retina acometida apresenta grande espessura na região da fóvea indicando um massivo vazamento de fluidos, ao contrário da retina normal (Figura 7(a)) cujo exame não mostra nenhum tipo de aumento de espessura entre camadas.

Observa-se também na Figura 7 que as imagens de OCT geralmente apresentam ruído e imperfeições. Portanto, a aplicação de técnicas de pré-processamento de imagens pode ser útil para projetos que utilizam e/ou manipulam estas imagens, visto que pode melhorar o aspecto da mesma e consequentemente melhorar os resultados obtidos em uma eventual classificação. Na próxima seção serão apresentadas técnicas de processamento de imagens.

2.5 Processamento de imagens

Uma imagem pode ser definida matematicamente como uma função bidimensional $f(x, y)$, onde x e y são coordenadas espaciais e a amplitude de f em qualquer par de coordenadas (x, y) é chamado intensidade ou nível de cinza da imagem naquele ponto (GONZALEZ; WOODS, 2008). Uma imagem digital é composta de um número finito de elementos denominados *pixels* (*picture elements* ou elementos de imagem) representados na forma de uma matriz bidimensional M x N, onde M representa o número de linhas e N o número de colunas.

A área de processamento de imagens tem como finalidade principal analisar, modificar e melhorar as imagens digitais. Isso é alcançado através de algoritmos e técnicas matemáticas que visam extrair informações importantes das imagens, tornando-as mais nítidas, precisas e úteis para várias aplicações, incluindo a medicina e a segurança (PRATT, 1991).

Uma técnica crucial no processo de processamento de imagens é a segmentação, que tem como objetivo dividir uma imagem em diferentes partes ou objetos para simplificar a análise e modificação futura. A segmentação é executada através de métodos, como o *thresholding*, que consiste em definir um limiar para separar as diferentes áreas da imagem, e algoritmos baseados em agrupamento, que classificam pixels similares em regiões (GONZALEZ; WOODS, 2008). A seleção do método de segmentação a ser utilizado depende da natureza da imagem e dos objetivos específicos do processamento.

A eliminação ou redução do ruído é outro aspecto importante no processamento de imagens. Esse ruído pode ser introduzido durante a aquisição de imagem ou por causa de condições adversas de captura. A redução de ruído é realizada através de filtros espaciais ou temporais que suavizam a imagem, mantendo ao mesmo tempo os detalhes importantes (GONZALEZ; WOODS, 2008). Além disso, técnicas de restauração de imagem também podem ser utilizadas para reduzir o ruído, como por exemplo, a técnica de suavização por filtro mediano.

2.5.1 Processamento morfológico

O processamento morfológico é uma área que apresenta muitas técnicas de processamento de imagem que operam com as características de forma em uma imagem específica.

Segundo Gonzalez e Woods (2008) a essência básica da morfologia matemática é obter informações sobre a geometria e topologia de uma imagem por meio de transformações com outro conjunto definido, citado como elemento estruturante. Entende-se, portanto, que a teoria de conjuntos é a base da morfologia matemática. As operações morfológicas são aplicadas na remoção de imperfeições introduzidas durante o processo de segmentação da imagem.

Muitas vezes, o resultado da segmentação de imagens não é adequado. Para corrigir os defeitos residuais, na etapa denominada de pós-processamento, utiliza-se as técnicas da morfologia matemática. As operações morfológicas básicas mais utilizadas são dilatação e erosão. Essas operações podem ser aplicadas em imagens binárias e imagens em tons de cinza. A operação de erosão, representada pelo símbolo $\ominus$, de uma imagem A pelo elemento estruturante B é representada matematicamente pela Equação 1 onde z representa cada ponto no espaço Euclidiano E em que a erosão está sendo aplicada (GONZALEZ; WOODS, 2008).

$$A \ominus B = \{z \in E | B_z \subseteq A\} \tag{1}$$

Já a dilatação, representada pelo símbolo $\oplus$, de uma imagem A pelo elemento estruturante B é representada matematicamente pela Equação 2. Neste processo, aplicado em cada ponto z do espaço Euclidiano E, a imagem é expandida ao adicionar pontos z ao resultado sempre que a interseção entre A e o elemento estruturante transladado e refletido não é vazia $(\emptyset)$.

$$A \oplus B = \{z \in E | (\hat{B}_z \cap A) \neq \emptyset\} \tag{2}$$

Nas imagens binárias, uma das aplicações de morfologia matemática é a extração de componentes da imagem que sejam úteis na representação e na descrição de formas, como mostra a Figura 8.

Os contornos de uma imagem podem ser obtidos através da aplicação da operação de erosão com um elemento estruturante adequado (determina a espessura do contorno), em seguida subtraindo o resultado da imagem original.

Figura 8 - Exemplo de extração de borda

(a) (b)

Fonte: Silva e Netto (2017).

De acordo com Gonzalez e Woods (2008), para realizar a extração de fronteiras (Figura 8), é usada a Equação 3 na qual é executado a erosão de *A* (conjunto da imagem) por *B* (elemento estruturante) e calculado a diferença entre a imagem A e sua erosão.

$$\beta(A) = A - (A\theta B) \tag{3}$$

2.5.2 Filtro gaussiano

O filtro gaussiano, também é conhecido como suavização gaussiana, é o resultado do desfoque de uma imagem por uma função gaussiana, em homenagem ao matemático e cientista Carl Friedrich Gauss.

O filtro gaussiano é um efeito muito usado em *softwares* gráficos, para reduzir o ruído da imagem e os detalhes. O efeito visual desta técnica é um desfoque suave semelhante ao de ver a imagem através de uma tela translúcida (JESUS; COSTA JUNIOR, 2015). Segundo Deng e Cahill (1993), o filtro gaussiano em imagens bidimensionais é baseado na Equação 4.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{(2\sigma^2)}} \tag{4}$$

Pode se perceber que na Equação 4 $G(x, y)$ representa o valor do filtro na posição (x, y), o símbolo σ é o desvio padrão do filtro gaussiano que controla a extensão do efeito de desfoque do filtro e o símbolo e representa a base do logaritmo natural (DENG; CAHILL, 1993).

Segundo Jesus e Costa Junior (2015) a aplicação do filtro gaussiano é feita pela mesma forma que a imagem é convoluta com uma função gaussiana, ou seja, servindo como um filtro passa-baixa. A Figura 9 apresenta as máscaras do filtro gaussiano com σ = 1 e *kernel* de 3x3 (Figura 9(a)) e 5x5 (Figura 9(b)).

Figura 9 - Máscaras do filtro gaussiano. (a) Máscara 3x3 e (b) Máscara 5x5

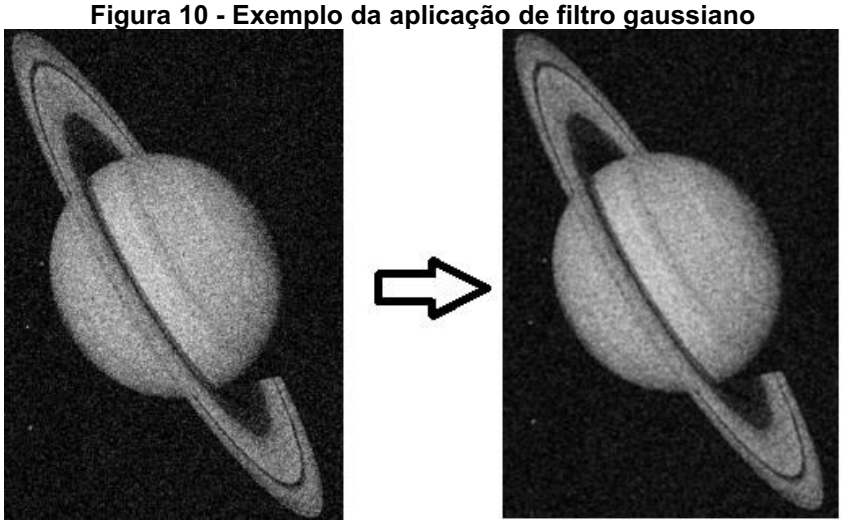
$$\frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix} \qquad \frac{1}{273}\begin{bmatrix}1 & 4 & 7 & 4 & 1\\ 4 & 16 & 26 & 16 & 4\\ 7 & 26 & 41 & 26 & 7\\ 4 & 16 & 26 & 16 & 4\\ 1 & 4 & 7 & 4 & 1\end{bmatrix}$$

(a) (b)

Fonte: Fisher *et al.* (2003).

A Figura 9 apresenta um exemplo da aplicação do filtro gaussiano em uma imagem.

Figura 10 - Exemplo da aplicação de filtro gaussiano



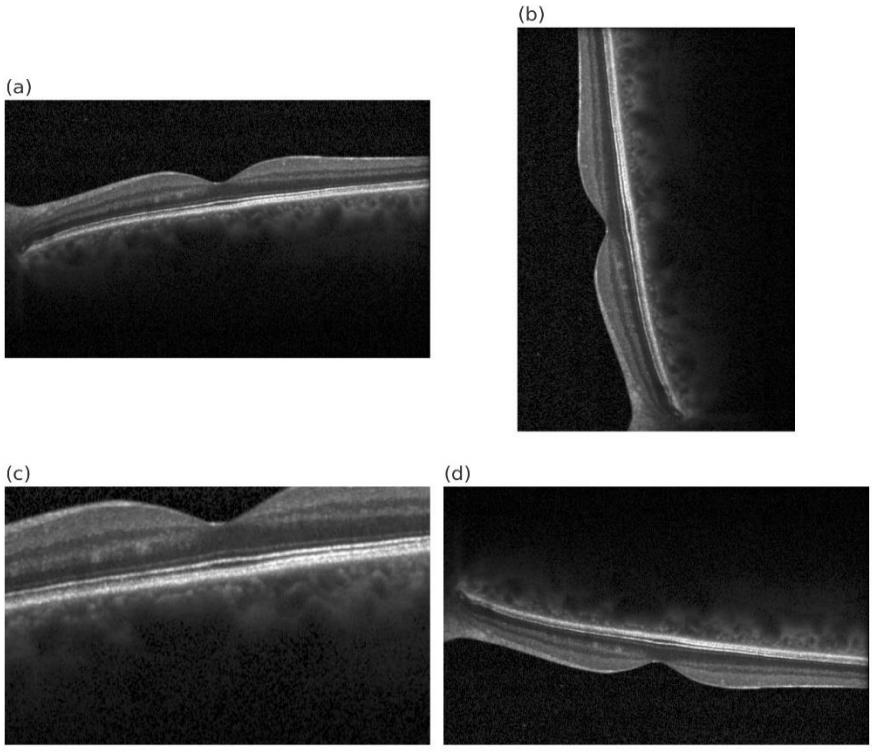
Fonte: Adaptado de Jesus e Costa Junior (2015).

Pode-se observar na Figura 10 o resultado da aplicação de um filtro gaussiano com σ = 1 e máscara 3x3. A imagem resultante apresenta uma suavização do ruído com relação à imagem original. A suavização com filtro passa-baixa tende a minimizar ruídos, mas faz a imagem perder nitidez (*blur*).

2.5.3 Data augmentation

Aumento de dados, ou *data augmentation*, é uma técnica da área de aprendizado de máquina onde a exigência por grandes quantidades de dados de treinamento frequentemente é um fator limitante. Esta técnica refere-se ao processo de gerar novas amostras de treinamento a partir das existentes, aplicando uma série de transformações que preservam as etiquetas de classe (SHORTEN; KHOSHGOFTAAR, 2019). Essas transformações podem incluir modificações variadas, tais como rotações, escalonamentos, recortes, espelhamentos ou a adição de ruído. A Figura 11 apresenta um exemplo destas transformações.

Figura 11 - Exemplos de transformações para aumento de dados. (a) Imagem normal. (b) Rotação. (c) Recorte. (d) Espelhamento.



Fonte: Autoria própria (2024).

Observa-se na Figura 11(a) a figura original, tendo as transformações realizadas na Figura 11(b) com a rotação de 90º no sentido anti-horário; na Figura 11(c) com o recorte e na Figura 11(d) com o espelhamento vertical. Estes métodos contribuem para o objetivo do aumento de dados que é aumentar a diversidade de dados disponíveis para o treinamento dos modelos, mesmo não contando com tantos dados. Além de ampliar o tamanho do conjunto de dados, também introduz um nível de variação que pode melhorar a robustez do modelo e sua capacidade de generalização. Ao simular diversas transformações que podem ocorrer em cenários do mundo real, o aumento de dados auxilia na redução do *overfitting*, um problema comum onde um modelo tem bom desempenho com os dados de treinamento, mas não com dados novos e não vistos (MUMUNI; MUMUNI, 2022).

Na prática, a implementação do aumento de dados é realizada por meio de várias técnicas. Uma abordagem comum são as transformações geométricas, que incluem girar a imagem em diferentes ângulos, ampliar ou reduzir o tamanho, transladar a imagem em diferentes direções e espelhar a imagem na horizontal ou vertical (como representado na Figura 11).

O aumento de dados é amplamente utilizado em tarefas que requerem reconhecimento visual, como classificação de imagens, detecção de objetos e segmentação. Por exemplo, em tarefas de classificação de imagens, é típico aumentar o conjunto de treinamento com imagens que foram aleatoriamente recortadas, redimensionadas ou invertidas horizontalmente. Isso garante que o modelo não esteja aprendendo a reconhecer características específicas da imagem que não são relevantes para a tarefa de classificação, como a posição do objeto na imagem (SHORTEN; KHOSHGOFTAAR, 2019).

2.6 Aprendizagem de máquina

O Aprendizado de Máquina – ou *machine learning* - faz parte de uma área de Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas que adquiram conhecimento automaticamente. É um subconjunto da inteligência artificial, em que os computadores são ensinados para aprender com os dados e ampliar suas

experiências e não apenas serem programados para realizar trabalhos previamente determinados (DUNJKO; BRIEGEL, 2018).

A inteligência artificial é a tecnologia implícita que ampara os subconjuntos de aprendizado de máquina sob seu termo abrangente, assim sendo, um subconjunto de aprendizado de máquina está incluso no primeiro nível. Um subconjunto de aprendizado profundo (do inglês *deep learning)* no segundo nível e um subconjunto de redes neurais inclui-se no terceiro nível.

Um sistema de aprendizado é um programa de computador que assume decisões por meio de experiências reunidas através da solução bem sucedida de problemas anteriores. Os sistemas de aprendizado de máquina são diversos e dispõem de características particulares e comuns que proporcionam a classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado. Os autores citam que o aprendizado de máquina tem por objetivo treinar algoritmos para encontrar padrões e relacionamentos em grandes conjuntos de dados e, posteriormente, fazer os melhores julgamentos e previsões possíveis (DUNJKO; BRIEGEL, 2018).

O aprendizado de máquina está relacionado à inteligência artificial, como também suas subdivisões, como aprendizado profundo e redes neurais se encaixam como subconjuntos concêntricos de inteligência artificial. Utilizando métodos de aprendizado de máquina a inteligência artificial avalia dados usando-os para aprender e se tornar mais inteligente (KHAN, 2022).

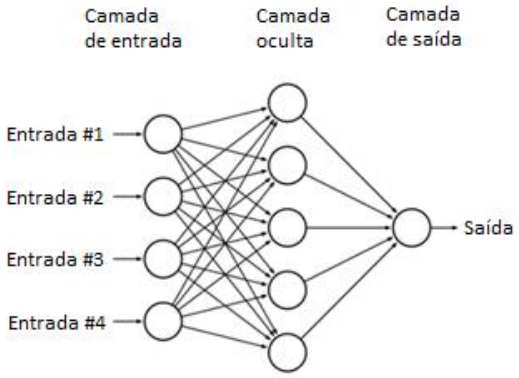
O aprendizado de máquina é composto por uma diversidade de tipos diferentes de modelos e, cada um deles formado por meio de um método algorítmico distinto dos outros tipos. Pode-se utilizar um dos quatro modelos de aprendizagem diferentes, de acordo com a natureza dos dados e do resultado almejado. Os modelos são aprendizado supervisionado, não supervisionado, aprendizado semissupervisionado e aprendizado por reforço. (BURRELL, 2016). Pode ser que uma ou mais técnicas algorítmicas sejam empregadas em cada um desses modelos, consoante aos conjuntos de dados que são utilizados e dos resultados almejados. Os algoritmos de aprendizado de máquina são aptos em advir uma extensa sucessão de tarefas, abarcando catalogar objetos, detectar arquétipos, agourar episódios e realizar pareceres de especialistas, entre outros. Ao tratar com dados difíceis e imprevisíveis, o uso de algoritmos, sozinhos ou em conformidade, podem ser utilizados para conseguir maior eficácia na classificação (BURRELL, 2016).

2.6.1 Redes neurais

A rede neural artificial é o vocábulo utilizado quando um *software* de computador se remete a replicar os neurônios em um cérebro biológico. Os neurônios biológicos são chamados de nós e são dispostos em classes que atuam paralelamente umas às outras (WANG, 2018).

O modelo mais comum de aprendizado com Redes Neurais, o *Perceptron*, foi definido em 1957 por Frank Rosenblatt e soluciona problemas simples que são claramente separáveis. (ROSENBLATT, 1957). É composto por uma estrutura com uma única camada, tendo como unidades básicas neurônios MLP (*Multi-Layer Perceptrons)* e uma regra de aprendizado. O algoritmo de aprendizagem do *Perceptron* usa a correção de erros (diferença entre a resposta desejada e a resposta da rede) como base. Para fazer o aprendizado da Rede Neural há um período de treinamento e um de teste do algoritmo. Os parâmetros da rede (pesos) são mudados a cada apresentação de um novo exemplo à rede. Depois do acerto dos parâmetros, na fase de teste, o sistema é avaliado. A Figura 12 esquematiza seu funcionamento.

Figura 12 - Um exemplo hipotético de rede m*ultilayer perceptron*



Fonte: Adaptado de Hassan *et al.* (2015).

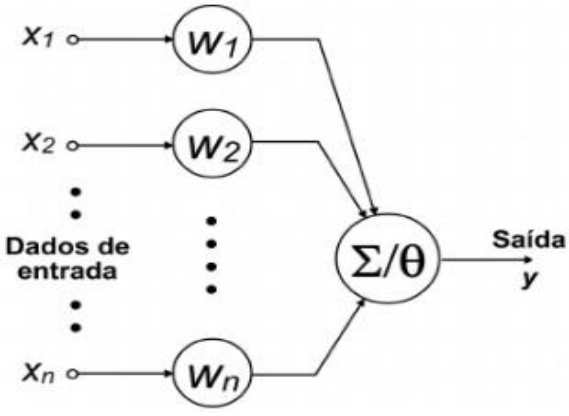
Para resolver problemas mais complexos, são necessárias redes de *Perceptrons* organizadas em múltiplas camadas, conhecidas como MLP (*Multi-Layer Perceptrons*). O algoritmo mais utilizado para treinar as MLPs é o *Backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986). Essa abordagem resolveu vários problemas, mas sem a garantia de alcançar sempre o mínimo global da função de custo, devido à possibilidade de encontrar mínimos locais. Enquanto isso, outras técnicas, como as SVM (*Support Vector Machines*), demonstraram melhores

resultados que as Redes Neurais em diversos problemas. Um exemplo notável é o reconhecimento de imagens, onde as Redes Neurais Profundas (*Deep Neural Networks*) — com várias camadas ocultas —, têm sido especialmente bem-sucedidas (LECUN; BENGIO; HINTON, 2015).

As redes neurais conseguem expandir e se capacitar com o tempo por utilizar dados de treinamento em seus cálculos que representam os valores pelos quais se espera uma classificação. Algoritmos de aprendizado que foram adaptados para precisão podem ter um efeito relevante na ciência da computação e na inteligência artificial (FAWAZ *et al*., 2019). Ao equiparar com a identificação manual de especialistas, as funções de reconhecimento de voz ou imagem podem ser efetuadas em minutos em vez de horas.

As redes neurais artificiais são compostas, essencialmente, por três tipos de camadas: uma camada de entrada, uma de saída, e pelo menos uma camada oculta entre elas. Estas camadas são conectadas por nós, formando uma rede interligada. Inspirados na estrutura dos neurônios humanos, os nós (ou neurônios artificiais) ativam-se em resposta a estímulos ou dados de entrada. Essa ativação se propaga pela rede, imitando o processo de transmissão de sinais entre neurônios no cérebro através de conexões sinápticas. À medida que o sinal avança pelas camadas da rede, ele é submetido a várias operações de processamento, que podem modificar sua intensidade ou forma. Esse processo permite que a rede neural artificial processe informações e responda de maneira adaptativa aos dados, assemelhando-se ao modo como o cérebro processa estímulos (OLIVEIRA; BARBAR; SOARES, 2019).

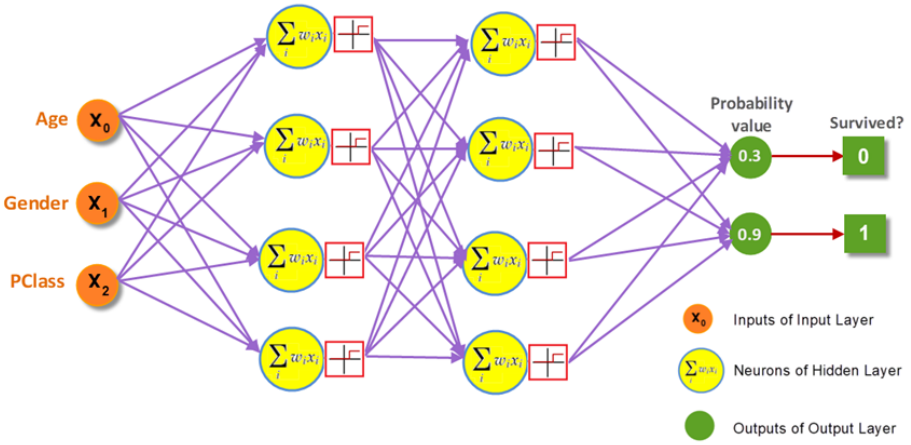
Figura 13 - Representação de um neurônio artificial



Fonte: Fiorin *et al.,* (2011).

A Figura 13 mostra um neurônio artificial, o qual executa uma soma ponderada de suas entradas e aplica uma função de ativação produzindo uma saída. Assim que os neurônios são associados com um problema a ser concluído ou um pedido a ser resolvido, eles iniciam os cálculos matemáticos através de funções de ativação para verificar se a informação provida é cabível ou não para seguir para a próxima célula da cadeia de comando. Para melhor ou pior, uma rede neural examina todos os fatos e indica onde em cada nó individual há uma conexão melhor. Ele é ligado quando a quantidade total de dados obtidos ultrapassa um determinado valor limite e os neurônios que estão acionados a ele também se tornam ativos (OLIVEIRA; BARBAR; SOARES, 2019). A Figura 14 ilustra este funcionamento.

Figura 14 - Exemplo de estrutura básica de uma rede neural



Fonte: Ponraj (2020).

Como verifica-se na Figura 14, neste exemplo de rede há 3 campos de entrada (*age, gender* e *PClass*) e posteriormente estes dados passam pelos nós – ou neurônios – da primeira camada oculta que são ativados ou não dependendo do cálculo da função de ativação pré-estabelecida. Após percorrido os neurônios das camadas ocultas a rede gera, neste caso, uma probabilidade que varia entre 0 e 1, sendo até 0,5 pertencente a classe 0 (não sobreviveu) e de 0,5 a 1 pertencente a classe(sobreviveu). Desta forma a saída da rede será a classificação se a pessoa correspondente aos dados de entrada sobreviveu ou não.

Uma das funções de ativação é a função *sigmoid* que foi introduzida por Verhulst (1845) como representado na Equação 5.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (5)$$

A função *sigmoid,* representada pela função logística de Verhulst (1845) (Equação 5) é caracterizada por uma curva em forma de 'S', na qual define sua saída como um valor entre 0 e 1. No contexto de aprendizado de máquina, este conceito é compreendido como a distinção de predição entre duas classes, a classe 0 – sendo assim classificada com uma saída entre [0; 0,5] - e a classe 1 – sendo assim classificada com uma saída entre a classe (0,5; 1].

Dentro do processo, destaca-se também os otimizadores, que são algoritmos que tem por objetivo minimizar a função de perda e ajustar os parâmetros do modelo. Alguns exemplos de otimizadores são o *Adam,* e o *RMSProp*. O *RMSProp* (*Root Mean Squared Propagation*) foi introduzido por Hinton, Srivastava e Swersky (2012) que tem como características a eficiência em problemas não convexos e a taxa de aprendizado adaptativa, no qual utiliza o quadrado médio dos gradientes recentes para ajustar a taxa de aprendizado. Já o *Adam (Adaptive Moment Estimation),* é um dos otimizadores mais populares e foi introduzido por Kingma e Ba (2017) e combina os benefícios do RMSProp com uma correção de viés que garante ajustes precisos logo nas primeiras iterações. A Equação 6, segundo o trabalho de Kingma e Ba (2017), descreve a atualização da média móvel dos gradientes do Adam.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (6)$$

Como mostra a Equação 6, m_t representa o primeiro momento estimado no tempo 't'. β_1 é o fator de decaimento que controla a ponderação dos gradientes antigos em relação aos mais recentes. Já m_{t-1} é o primeiro momento do passo anterior enquanto g_t é o gradiente no tempo 't'. A Equação 7 descreve a média móvel não centrada dos quadrados dos gradientes no otimizador Adam.

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (7)$$

De acordo com a Equação 7, v_t é o segundo momento estimado no tempo 't', β_2 representa o fator de decaimento, v_{t-1} é o valor do segundo momento no tempo

anterior e g_t^2 é o quadrado do gradiente no tempo 't'. Kingma e Ba (2017) também apresentam a equação de atualização dos pesos do otimizador Adam, representado pela Equação 8.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}} + \epsilon} \cdot \hat{m}_t \quad (8)$$

Observa-se na Equação 8 que θ_{t+1} e θ_t representam os parâmetros do modelo nos tempos $t+1$ e t, respectivamente. η é a taxa de aprendizagem, $\hat{v}$ é a estimativa corrigida do segundo momento dos gradientes, ϵ é um número adicionado para evitar divisão por zero e $\hat{m}_t$ é a estimativa corrigida do primeiro momento.

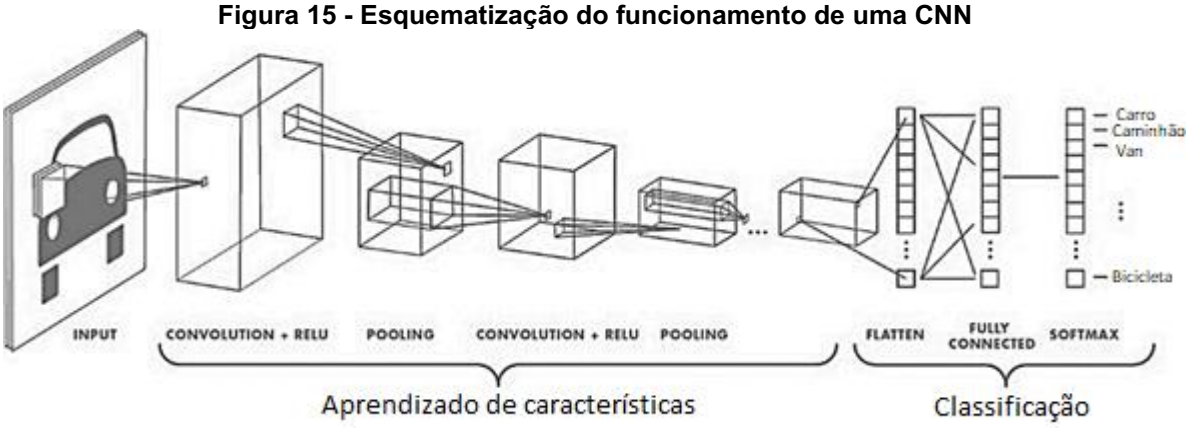
Outra técnica importante a ser ressaltada para melhorar o desempenho da rede é o *EarlyStopping*. Este método monitora a métrica de desempenho do modelo após cada época; uma vez que esta não melhora a partir de um número parametrizável de épocas, o treinamento para, evitando assim o *overfitting* e diminuindo o tempo necessário para treinamento ideal da rede (PRECHELT, 1998).

2.6.2 Redes neurais convolucionais

As redes neurais convolucionais (CNN – *Convolutional Neural Networks)* executam a entrada em um ajuste correlato a uma grade, o que as torna essenciais para verificar imagens visuais quando comparadas a outros tipos de redes neurais (ZHANG *et al.,* 2014). A análise de imagens visuais pode se favorecer da utilização de um tipo de rede neural *feed-forward* nominado como redes neurais convolucionais. Utilizam-se redes neurais convolucionais para identificar e classificar objetos visuais em imagens.

A rede neural convolucional é construída para aprender hierarquias espaciais de recursos automaticamente e de forma adaptável com retropropagação, usando blocos de construção como camadas de convolução, *pooling* e totalmente conectadas (YAMASHITA *et al.*, 2018). A Figura 15 esquematiza o funcionamento de uma rede neural convolucional.

Figura 15 - Esquematização do funcionamento de uma CNN



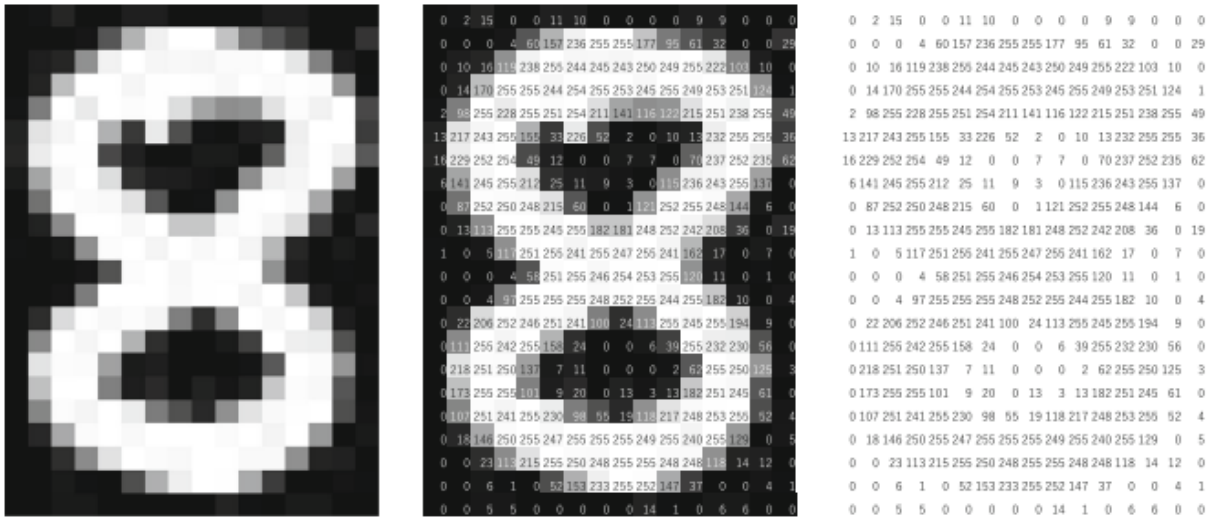
Fonte: Adaptado de Alamsyah, Saputra e Masrury (2019).

Como se pode observar no exemplo mostrado pela Figura 15, a rede neural convolucional tem a mesma estrutura vista no tópico anterior, sendo dividida entre a entrada, as camadas ocultas e a saída ou classificação final. A entrada se dá por meio da imagem do veículo que passa pelas camadas ocultas convolucionais e pela função de ativação ReLU (unidade linear retificada, do inglês *Rectified Linear Unit*). Após passar por cada camada convolucional, é apresentada a camada de *pooling* que reduz a saída da camada convolucional através da aplicação de um filtro.

Em seguida o modelo apresenta uma camada *flatten* que transforma o conteúdo que antes era em formato matricial para um formato linear que em seguida será aplicada à uma camada Densa. Por fim a função de ativação *softmax* é aplicada predizendo à qual classe pertence a imagem de entrada. É importante notar que diferentemente de uma rede neural básica, a convolucional requer que suas camadas ocultas sejam compostas inicialmente por camadas convolucionais e de *pooling*, isto é, camadas capazes de realizar a leitura e a manipulação das imagens.

A Figura 16 explicita que os valores de pixel da imagem a ser lida pelo sistema são mantidos em um *array* bidimensional sendo que cada valor corresponde ao brilho do pixel, variando de 0 (preto) a 255 (branco). Portanto, os valores a serem trabalhados pelas próximas camadas convolucionais de uma hipotética rede que receba esta imagem (Figura 16(a)), são os valores compostos na Figura 16(c).

Figura 16 - Sequencia entre o que um ser humano enxerga e o que um computador enxerga numa imagem



(a) **(b)** **(c)**

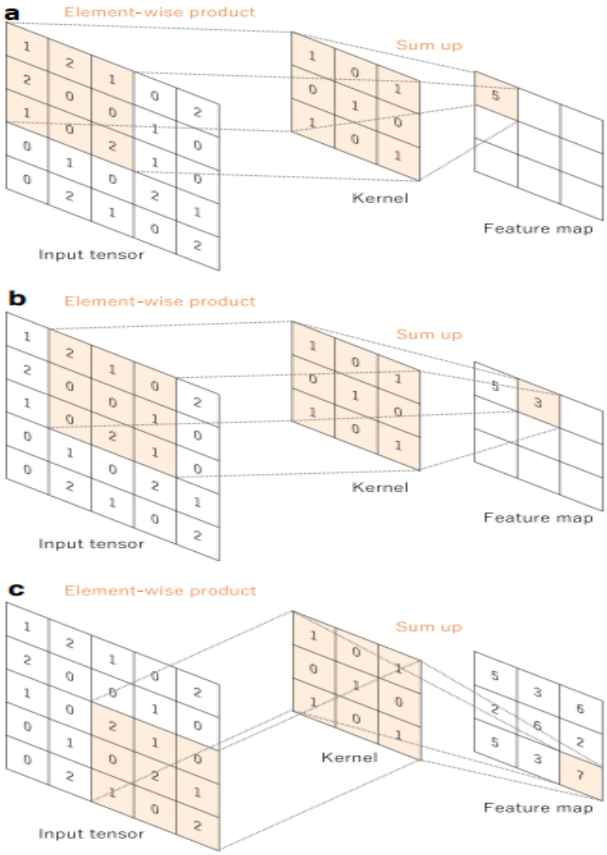
Fonte: Yamashita *et al*., (2018).

A camada convolucional é a primeira camada oculta da rede convolucional e sua função é aprender os parâmetros da rede. A cada nova camada, a complexidade da CNN, propaga, possibilitando que ela identifique mais informações na imagem sempre que aprende mais sobre a mesma. Por exemplo, a coloração e bordas são tratadas na fase inicial do processo de design, enquanto a entrada da imagem segue por meio das camadas CNN, o objeto alvo é relativamente distinguido como a entrada da imagem (HUANG *et al*., 2019).

> A convolução é um tipo especializado de operação linear usada para extração de recursos, onde uma pequena matriz de números, chamada de *kernel*, é aplicada na entrada, que é uma matriz de números, chamada de tensor. O produto entre cada elemento do *kernel* e o tensor de entrada é calculado em cada local do tensor e somado para obter o valor de saída na posição correspondente do tensor de saída, chamado de *feature map*. YAMASHITA *et al*., 2018, p. 613).

A Figura 17 mostra um exemplo de convolução com um *kernel* de 3x3.

Figura 17 - Exemplo de convolução com kernel de 3x3



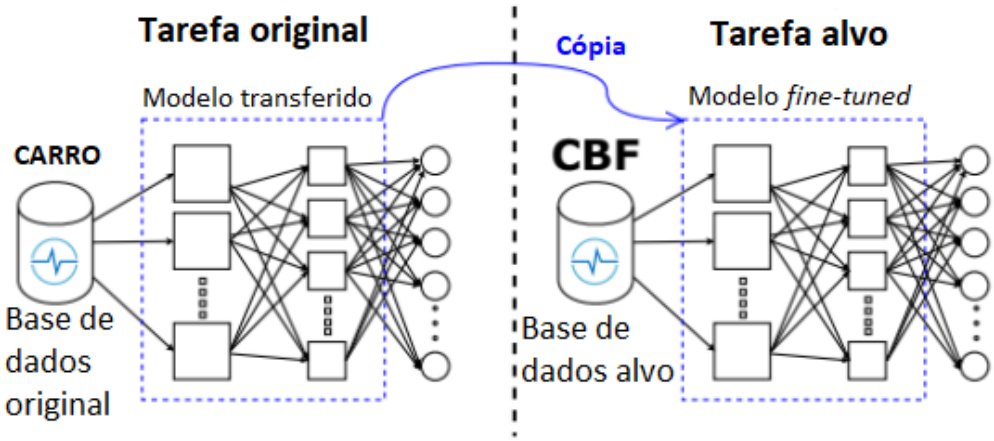
Fonte: Yamashita *et al*. (2018).

Pode-se observar na Figura 17 que as matrizes da esquerda (*element-wise product*) têm um exemplo de matriz de entrada, ou seja, o *array* de valores de uma imagem, no centro é aplicado um *kernel* de 3x3 que percorre por toda a matriz de entrada (*sum up*) e gera o *feature map* de saída.

2.6.3 Transferência de aprendizado

Com base em Fawaz *et al*., (2018) o método de transferência de aprendizado é uma forma de melhorar os resultados de uma rede transferindo os pesos e *features* que foram gerados por uma rede primária para uma rede secundária que não necessariamente tem o mesmo *dataset*. A Figura 18 esquematiza este funcionamento.

Figura 18 - Esquematização do funcionamento da transferência de aprendizado



Fonte: Adaptado de Fawaz *et al*. (2018).

Como pode-se verificar a rede da Figura 18(a) foi treinada com o *dataset* “Carro” e seus pesos das camadas gerais foram transferidos para a rede da Figura 18(b) que usa outro *dataset*, o “CBF”. Este processo é conhecido como transferência de aprendizado. Ele permite que uma rede neural melhore sua capacidade de generalização ao reaproveitar conhecimentos adquiridos anteriormente por outra rede. Isso é feito mantendo-se as camadas iniciais, que aprendem características gerais das imagens, e ajustando as últimas camadas, que são específicas para cada problema. Essa abordagem evita a necessidade de treinar a rede do zero.

Esta é uma forma também de evitar um problema comum em redes neurais chamado *overfitting* que ocorre quando o modelo perde a capacidade de generalização tendo uma excepcional performance durante o treino e uma péssima performance em dados externos (COGSWELL; AHMED; GIRSHICK, 2016).

Portanto, redes treinadas com um *dataset* pequeno, que inicialmente poderia haver incidência de *overfitting* podem utilizar-se da transferência de aprendizado para controlar esta situação.

2.7 Arquiteturas de redes pré-treinadas

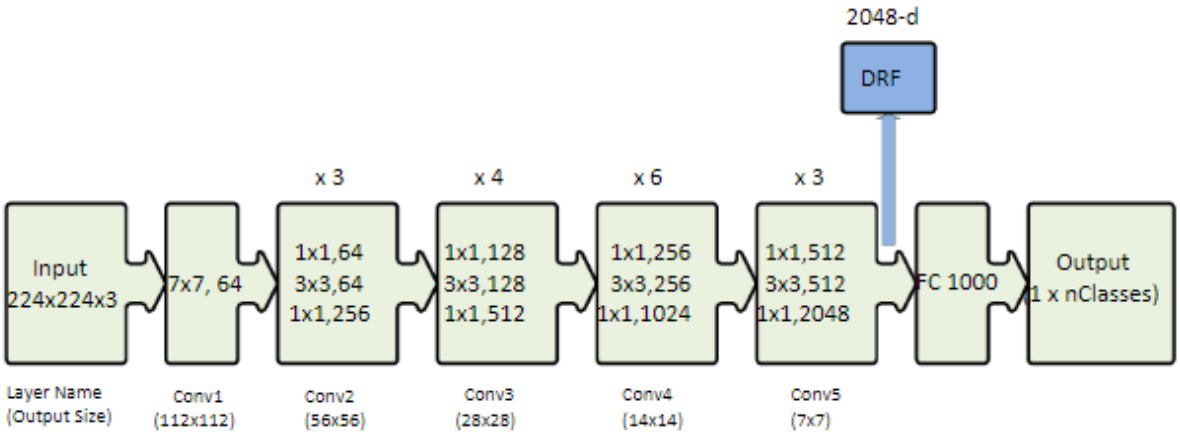
Nesta seção serão expostas algumas redes neurais a serem utilizadas neste projeto.

2.7.1 ResNet-50

Um exemplo de modelo pré-treinado muito utilizado é a rede ResNet (*Residual Network*) que é habitualmente usada em tarefas de visão computacional. Possibilita instruir redes neurais profundas com um amplo número de camadas evitando o problema de gradientes de desaparecimento. Este problema ocorria depois do treinamento de certa quantidade de camadas o incremento de mais camadas não aumentava a qualidade dos modelos (HE *et al.*, 2016).

Esta rede adota o conceito de resíduos, ou seja, a adição de saltos de conexão diretos entre camadas, permitindo que a informação flua diretamente ao longo da rede, preservando assim o gradiente da saída para a entrada. Esta técnica permite que a ResNet treine redes com muito mais facilidade e eficiência do que as arquiteturas convencionais. A Figura 19 representa a arquitetura do modelo ResNet-50. Nela, observa-se que a ResNet-50 é uma versão específica da ResNet com 50 camadas consecutivas. Ela é composta por vários blocos de camadas de diferentes tamanhos e *kernels*, cada um com uma ou mais camadas de convolução e normalização, seguidas de uma camada residual.

Figura 19 - Arquitetura da rede ResNet-50

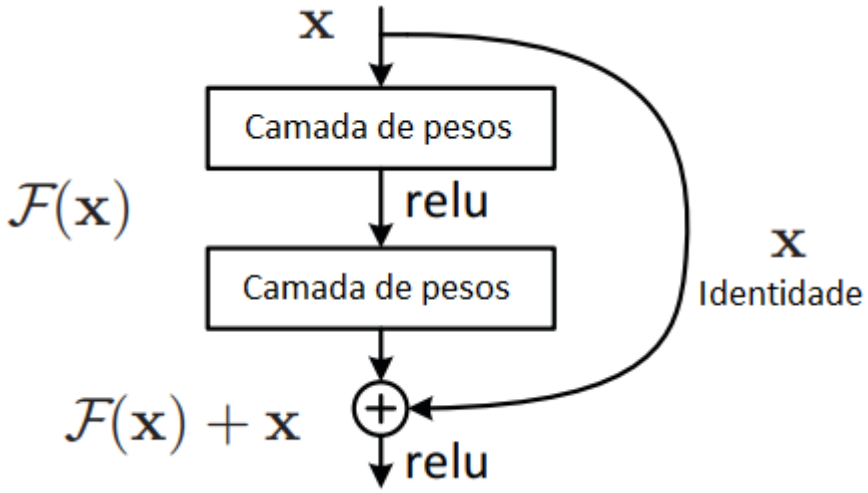


Fonte: He *et al.*, (2016).

Esses blocos de camadas expostos pela Figura 19 são repetidos várias vezes, permitindo que a ResNet-50 capture complexos padrões entre os dados de entrada.

Em azul (DRF) percebe-se onde entram as camadas densa (com ativação ReLU) e *flatten.* Na Figura 20 é mostrado a representação do bloco residual da ResNet.

Figura 20 - Representação do bloco residual



Fonte: Adaptado de He *et al.,* (2016).

A ResNet-50 é amplamente utilizada em aplicações de visão computacional, como reconhecimento de objetos, segmentação de imagens e extração de características. Além disso, esta arquitetura é responsável por uma das redes neurais mais eficientes e de alto desempenho disponíveis, tornando-a uma escolha popular para aplicações em tempo real (He *et al.*, 2016).

Esta arquitetura foi submetida a diversas tarefas de relacionadas à visão computacional e aprendizado profundo, e sua eficiência foi confirmada. Como exemplo disso, o trabalho de Xie et al. (2017) explorou variações na formação de blocos residuais e demonstrou melhorias significativas nos resultados em tarefas de classificação de imagens.

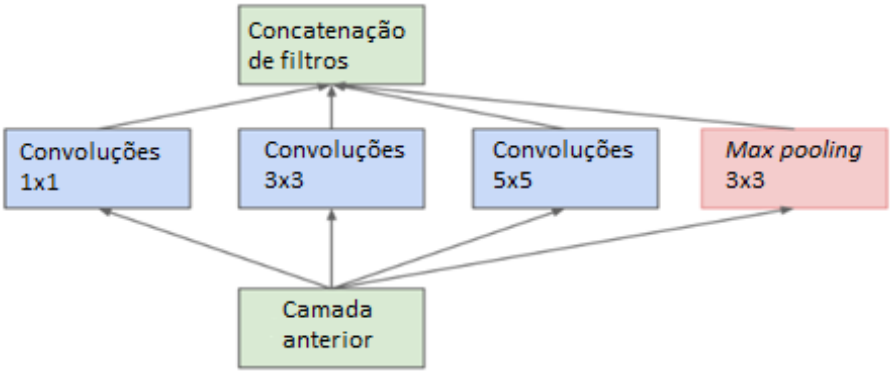
Além da ResNet-50 que é uma rede com 50 camadas, há também as ResNets 34 (de 34 camadas) e 101 (com 101 camadas).

2.7.2 *Inception* V3

A *Inception* V3 é outro exemplo de modelo pré-treinado. É uma arquitetura de rede neural profunda projetada para a tarefa de classificação de imagens (SZEGEDY *et al*., 2015). A rede *Inception* V3 adota o conceito de módulos *Inception*, que permitem

a extração de diversos níveis de características de uma imagem de entrada, ao mesmo tempo em que mantém uma boa eficiência computacional. Os módulos *Inception* consistem em várias camadas paralelas de convolução e *pooling*, cada uma responsável por capturar diferentes padrões. Os resultados dessas camadas paralelas são concatenados e passados para a próxima camada. A *Inception* V3 também inclui camadas completamente conectadas no final da rede para a classificação final das imagens. A Figura 21 apresenta o módulo *Inception* que constitui parte chave da arquitetura da rede.

Figura 21 - Diagrama do módulo Inception



Fonte: Adaptado de Szegedy *et al.*, (2015).

Percebe-se na Figura 21 que o módulo *Inception* apresenta convoluções de tamanhos 1x1, 3x3 e 5x5 que são aplicadas na mesma camada de entrada para extrair diferentes níveis de características, enquanto reduz-se a dimensionalidade com a operação de *max pooling* 3x3.

A arquitetura da *Inception* V3 se destaca por sua eficiência computacional e precisão. Ela introduziu o conceito de módulos de fatorização, que ajudam a reduzir a quantidade de cálculos necessários sem comprometer a eficiência da rede. Esses módulos permitem que a rede mantenha uma abordagem computacionalmente acessível enquanto aumenta a sua profundidade e largura. Além disso, a *Inception* V3 implementou a normalização em lote, uma técnica que acelera o treinamento da rede e melhora sua eficiência (SZEGEDY *et al*., 2016).

Canziani, Paszke e Culurciello (2017) avaliaram a precisão e a eficiência de arquiteturas, incluindo a *Inception* V3, destacando sua superioridade em comparação com outras arquiteturas. Além disso, a *Inception* V3 foi utilizada em uma variedade de

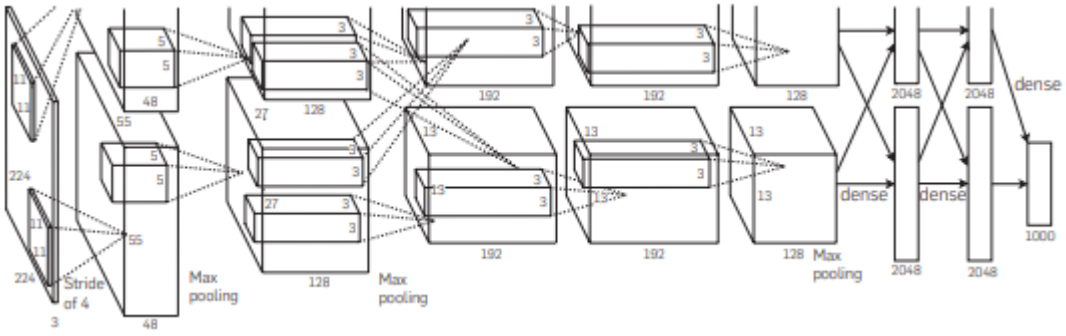
domínios não apenas da visão computacional como também em análises de dados médicos e biológicos. Nestes estudos ressaltou-se a capacidade da rede de capturar padrões em várias escalas, demonstrando sua eficácia nestas tarefas (LITJENS *et al.,* 2017).

2.7.3 AlexNet

Outra rede a ser considerada é a AlexNet que possui 8 camadas, sendo 5 camadas convolucionais e 3 densas. Esta rede utiliza a função de ativação ReLU e faz uso também de camadas *Dropout* para redução do *overfitting* (LI *et al.,* 2021).

A AlexNet, desenvolvida por Alex Krizhevsky, Ilya Sutskever e Geoffrey Hinton (2017), representou um marco na área de redes neurais convolucionais. Apresentada no *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) de 2012, a AlexNet demonstrou um desempenho significativamente superior aos modelos anteriores, reduzindo a taxa de erro top-5 em quase 50% comparada ao segundo colocado. Este avanço desencadeou um renovado interesse em CNNs e aprendizado profundo na comunidade de pesquisa em visão computacional e inteligência artificial (KRIZHEVSKY; SUSTKEVER; HINTON, 2017; RUSSAKOVSKY *et al.,* 2015).

Figura 22 - Arquitetura da rede AlexNet



Fonte: Krizhevsky, Sustkever e Hinton. (2017).

Observa-se na Figura 22 que a arquitetura da AlexNet é composta por oito camadas aprendíveis, incluindo cinco camadas convolucionais e três camadas totalmente conectadas. Um aspecto notável da AlexNet é o uso da função de ativação ReLU, que foi fundamental para acelerar o treinamento da rede. A rede também implementou o conceito de *dropout* para reduzir o *overfitting* nas camadas totalmente conectadas. Além disso, a AlexNet utilizou a técnica de *data augmentation* para

aumentar a diversidade do conjunto de dados de treinamento, melhorando assim a generalização do modelo. Essas inovações técnicas foram fundamentais para o sucesso da AlexNet no ILSVRC-2012 (KRIZHEVSKY; SUSTKEVER; HINTON, 2017; DENG *et al*., 2009).

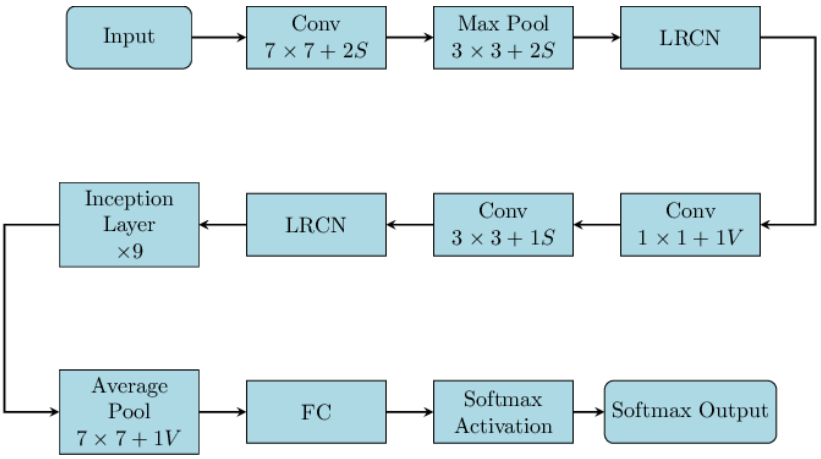
A AlexNet é considerada um divisor de águas na história do aprendizado profundo e visão computacional. Sua vitória no ILSVRC-2012 demonstrou o potencial das CNNs em tarefas de classificação de imagens em larga escala, incentivando o desenvolvimento de redes mais profundas e eficientes. A influência da AlexNet estende-se até hoje, inspirando inúmeras arquiteturas e pesquisas subsequentes, e estabelecendo CNNs como uma abordagem padrão em diversas aplicações práticas de inteligência artificial (RUSSAKOVSKY *et al*., 2015; LECUN; BENGIO; HILTON, 2015).

2.7.4 GoogLeNet

A GoogLeNet, também conhecida como *Inception V1,* é uma arquitetura de rede neural profunda projetada especificamente para a tarefa de classificação de imagens (SZEGEDY *et al*., 2015). A rede é composta por vários módulos *Inception*, que são combinados de forma a criar uma rede profunda com muitos níveis de abstração. A inovação central da GoogLeNet é o módulo *Inception*, que permite à rede escolher entre múltiplos tamanhos de filtros convolucionais em cada bloco. Essa estrutura composta por múltiplos módulos *Inception* empilhados, intercalados com camadas de *max-pooling*, permitiu à GoogLeNet alcançar uma eficiência computacional notável, reduzindo a quantidade de parâmetros sem sacrificar a profundidade ou a precisão da rede e mantendo o custo computacional (SZEGEDY *et al*., 2014). Além disso, a GoogLeNet adota uma estratégia de ensino supervisionado distribuído, que permite que a rede aprenda características importantes a partir de muitos níveis diferentes de abstração.

Observa-se na Figura 23 a arquitetura do GoogLeNet, que por ser proveniente do *Inception*, também utiliza o módulo *Inception* presente no *Inception V3*.

Figura 23 - Arquitetura da rede GoogLeNet



Fonte: Singh e Kishore (2015).

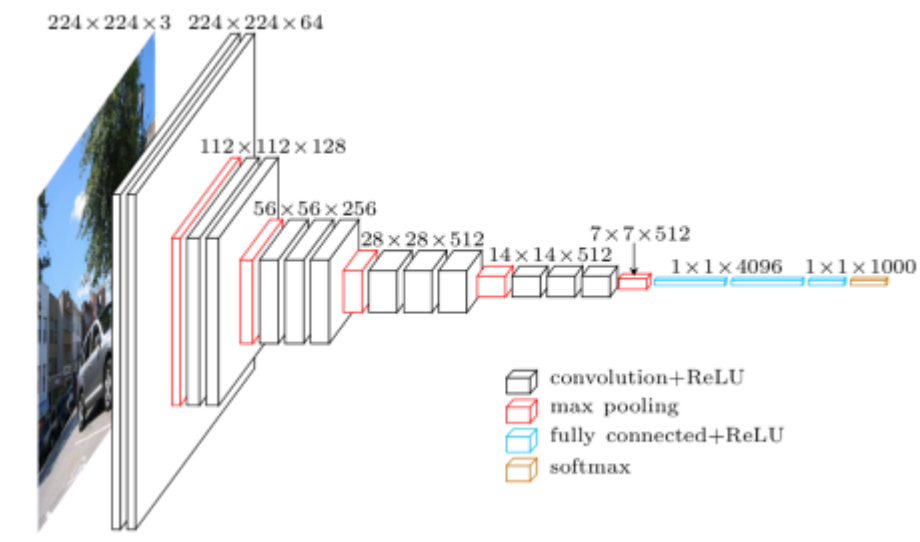
A GoogLeNet é considerada uma rede versátil, tendo sido aplicada em uma ampla variedade de tarefas como classificação de imagens, detecção de objetos e análise de vídeo (RUSSAKOVSKY *et al.,* 2015).

2.7.5 VGG-19

A rede VGG-19 é outra opção de modelo pré-treinado apresentada por Simonyan e Zisserman (2015) que se destaca pelo seu desempenho na classificação de imagens. Ela é composta de 19 camadas de processamento e utiliza filtros de 3x3 para extrair características locais e regionais da imagem, aumentando a sua profundidade através do aumento do número de camadas.

A Figura 24 mostra a arquitetura base do VGG-16 que conta com 16 camadas, a diferença para o VGG-19 é que este segundo possui mais 3 camadas convolucionais em sua arquitetura (SIMONYAN; ZISSERMAN, 2015).

Figura 24 - Arquitetura da rede VGG-16



Fonte: Simonyan e Zisserman (2015).

A VGG-19 foi aplicada em alguns contextos diversos como transferência de estilo artístico e detecção de objetos. Johnson, Alahi e Li (2016) realizaram testes com a VGG-19 na transferência de estilo de uma imagem para outra, mas mantendo o conteúdo da segunda, numa espécie de mesclagem.

2.8 Métricas de desempenho

Para realizar a avaliação, algumas métricas utilizadas em problemas de classificação são utilizadas.

2.8.1 Matriz de confusão

A matriz de confusão é uma ferramenta importante, com a qual se pode analisar com rapidez o desempenho de cada sistema. Os valores que compõem a matriz são obtidos fornecendo os segmentos do conjunto de teste ao método de classificação e comparando sua predição com a classe correta de cada segmento (SOUZA, 2009). O Quadro 1 caracteriza a estrutura de uma matriz de confusão.

Quadro 1 - Estrutura da matriz de confusão

	Real 'Sim'	**Real 'Não'**
Predito 'Sim'	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Predito 'Não'	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Autoria própria (2022).

Os itens que compõe uma matriz de confusão podem ser classificados em quatro opções:

1. Verdadeiro Positivo (VP): segmentos que pertencem a classe positiva e foram classificados como positivos. No caso desse trabalho, são imagens que contém edema macular diabético e assim foram preditas pela rede;

2. Falso Positivo (FP): segmentos que pertencem a classe negativa e foram classificados como positivos. Para este trabalho seriam as imagens que não apresentam a doença porem sua predição foi como se tivesse;

3. Falso Negativo (FN): segmentos que pertencem a classe positiva e foram classificados como negativos. Seriam imagens que contém a doença, mas foram preditas como normais;

4. Verdadeiro Negativo (VN): segmentos que pertencem a classe negativa e foram corretamente classificados como negativos. Neste caso seriam as imagens sem a doença que assim foram preditas;

2.8.2 Métricas de qualidade

As métricas são usadas na comparação de modelos de classificação. Cada uma delas busca avaliar um aspecto diferente do modelo de acordo com os resultados obtidos de verdadeiros e falsos positivos e negativos.

1. Acurácia: É uma das mais básicas métricas que conforme a Equação 9 explicita, avalia a proporção de itens classificados corretamente para o número total de classificações;

$$Acurácia = \frac{VP+VN}{VP+FP+FN+VN} \quad (9)$$

2. Precisão: É a quantidade de predições que a rede previu como sendo positiva e de fato é. Na Equação 10, a precisão é a divisão dos verdadeiros positivos (VP) pela soma dos verdadeiros positivos com os falsos positivos (FP);

$$Precisão = \frac{VP}{VP+FP} \tag{10}$$

3. *Recall*: A Equação 11 mostra que o recall é os verdadeiros positivos (VP) dividido pela soma dos verdadeiros positivos com os falsos negativos (FN), ou seja, dos exemplos que são positivos quantos foram preditos como tal;

$$Recall = \frac{VP}{VP+FN} \tag{11}$$

4. *F-Measure*: é uma métrica que considera tanto a taxa de precisão quanto o *recall* como é possível observar na Equação 12. O seu valor é a média harmônica entre a precisão e o *recall*;

$$F - Measure = \frac{2*Precisão*Recall}{Precisão+Recall} \tag{12}$$

5. Especificidade: Métrica que segundo a Equação 13 avalia a performance da rede na predição correta de casos negativos. É a divisão dos verdadeiros negativos (VN) pela soma dos falsos positivos (FP) com os verdadeiros negativos.;

$$Especificidade = \frac{VN}{FP+VN} \tag{13}$$

Com estas métricas é possível avaliar os resultados da rede em vários aspectos englobando desde a capacidade de gerar predições positivas confiáveis como a capacidade geral da rede considerando todos os acertos comparados com os erros.

2.9 Trabalhos relacionados

Inteligência artificial é a ciência que faz com que os computadores aprendam a realizar atividades da mesma forma que os humanos. É uma inteligência

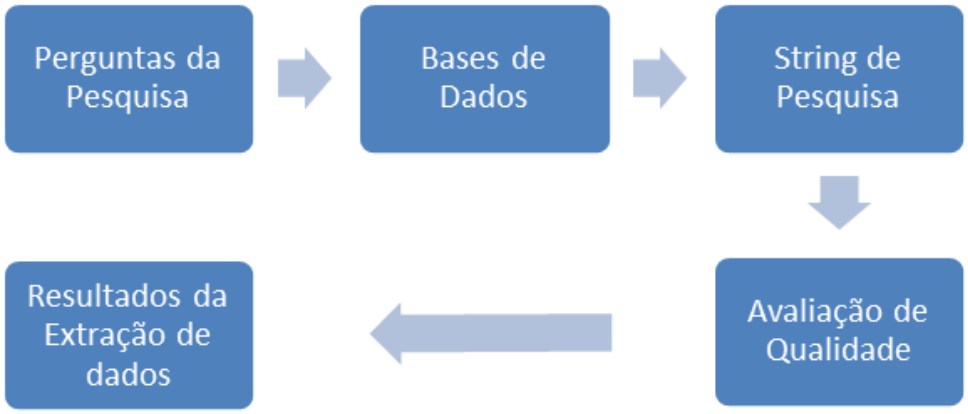
demonstrada por máquinas, em vez da inteligência natural demonstrada por animais. A aplicação clínica da inteligência artificial inclui diagnóstico por imagem e atenção crescente para melhorar a precisão e velocidade no diagnóstico de muitas condições clínicas. De acordo com um estudo de Oren, Gersh e Bhatt (2020), o uso de inteligência artificial, especificamente processamento e classificação de imagens, melhora a precisão, sensibilidade e especificidade para detectar anormalidades radiográficas menores e tem o potencial de melhorar a saúde pública. Por exemplo, o aprendizado profundo (*deep learning*) é em muitos casos utilizado para classificação de edema macular usando uma rede neural artificial.

Rodriguez *et al*., (2017) realizaram estudos sobre a construção de uma ferramenta de telemedicina para auxiliar os médicos na detecção de DME usando imagens do exame de fundo do olho. Nos estudos realizados, observou-se que a rede neural teve uma precisão de 73,5% na detecção de edema macular diabético. Já Zhang *et al*., (2020) atingiram 94% de acurácia e 97,5% de precisão em seu estudo fazendo o uso de imagens de Tomografias de Coerência Óptica, como é a proposta do presente trabalho.

2.9.1 Plano de mapeamento

O plano de mapeamento a seguir mostra a identificação da literatura existente sobre a aplicação da inteligência artificial em exames médicos. O plano de mapeamento categoriza as obras existentes da literatura e analisa-as para determinar sua relevância para a questão de pesquisa. O objetivo principal deste estudo sistemático é obter uma visão abrangente da questão de pesquisa, apresentar uma avaliação imparcial da literatura atual sobre o tópico e identificar lacunas de pesquisa nas obras existentes da literatura para coletar evidências suficientes para auxiliar nesta pesquisa. A Figura 25 apresenta as etapas a serem seguidas.

Figura 25 - Etapas do mapeamento sistemático



Fonte: Autoria própria (2022).

Conforme ilustrado na Figura 25, nas próximas seções serão abordadas as perguntas da pesquisa, as bases de dados que serão consultadas, a *string* de pesquisa, a forma de avaliação de qualidade e por fim os resultados da extração de dados.

2.9.2 Perguntas da pesquisa

As aplicações da IA na medicina incluem dados clínicos, triagem em massa, diagnóstico por imagem e eletrodiagnóstico. Esta pesquisa se especializou no uso de processamento de imagens, um ramo da IA, no diagnóstico por imagem de edema macular diabético. Nesta pesquisa procurou-se obter as respostas para as seguintes perguntas:

1. Qual é a utilidade da inteligência artificial no diagnóstico médico?
2. O uso de redes neurais artificiais melhorou a precisão do diagnóstico de edema macular?
3. O uso de inteligência artificial no diagnóstico melhora a velocidade do diagnóstico e levando ao diagnóstico precoce?

2.9.3 Bases de dados de pesquisa

A pesquisa se baseia em bancos de dados específicos que fornecem informações sobre aprendizado profundo. Entender o aprendizado profundo é fundamental para compreender como as redes neurais artificiais processam imagens, permitindo a execução de tarefas complexas como reconhecimento facial, detecção de objetos e análise de imagens médicas. Os bancos de dados usados para obter informações sobre inteligência artificial incluem *Academic Search Complete*, que contém mais de 5.100 periódicos revisados por pares*, Applied Science & Technology Full Text* que contém 750 dos principais periódicos em ciência e tecnologia. Outros bancos de dados que são fontes de informação para esta pesquisa incluem *BioMed Central External*, que contém informações sobre os testes controlados atuais e tópicos da biologia moderna. *PubMed Central (PMC)* é um arquivo de literatura de periódicos biomédicos e de ciências biológicas do Instituto Nacional de Saúde dos EUA (NIH). O arquivo *PubMed Central (PMC)* é desenvolvido e gerenciado pelo *National Center for Biotechnology Information (NCBI)* do NIH na *National Library of Medicine*.

2.9.4 *String* de busca

As melhores *strings* de pesquisa para o estudo são aquelas que contêm palavras-chave que fornecem os resultados mais relevantes do mesmo pois o objetivo de seu uso é a obtenção de artigos relevantes, portanto, as *strings* de pesquisa devem pesquisar e recuperar o conteúdo mais relevante nas bases de dados de inteligência artificial e diagnósticos médicos. As frases a seguir contêm várias cadeias de caracteres de pesquisa que podem ser usadas juntamente com suas combinações usando operadores “e” booleanos para melhorar os resultados da pesquisa.

1. ‘*deep learning’* e *‘optical coherence tomography’*;
2. *‘transfer learning’* e *‘diabetic macular edema*;

2.9.5 Avaliação de qualidade

De forma a avaliar qualitativamente os artigos lidos, foi elaborado uma lista de relevância dos artigos utilizados neste trabalho na ordem do mais relevante, para o

menos relevante levando em consideração pontos como similaridade do artigo ao trabalho aqui proposto e importância no processo de resposta das perguntas.

2.9.6 Resultados da extração de dados

De forma a avaliar qualitativamente os artigos lidos, foi elaborado uma lista de relevância dos artigos utilizados neste trabalho na ordem do mais relevante, para o menos relevante levando em consideração pontos como similaridade do artigo ao trabalho aqui proposto e importância no processo de resposta das perguntas.

A aplicação do aprendizado de máquina na medicina melhorou o atendimento social à saúde de pessoas em todo o mundo, embora ainda não sejam amplamente utilizados. Solórzano *et al*. (2020) afirmou que o aprendizado de máquina facilitou o processamento de imagens durante o diagnóstico clínico de doenças crônicas. O aprendizado de máquina usa segmentação de imagem para particionar imagens digitais em vários segmentos de pixels que representam estruturas discretas. Os algoritmos de aprendizado de máquina usados para conduzir o processamento de imagens em imagens clínicas incluem *K-Nearest-Neighbour*, *Support Vector Machines*, redes neurais artificiais e redes neurais convolucionais.

De acordo com um estudo de Wang *et al*., (2020), a inteligência artificial despertou um enorme interesse global nas últimas três décadas. A inteligência artificial tem sido amplamente adotada para processamento de imagens e processamento de linguagem natural, o que tem mostrado um impacto positivo significativo nos sistemas de saúde. Os autores citam que a inteligência artificial tem sido aplicada para fotografias de fundo de olho, tomografia de coerência óptica de campos visuais em oftalmologia e tem registrado desempenho de alto nível na detecção de edema macular diabético.

O Quadro 2 representa a classificação dos artigos envolvidos na pesquisa.

Quadro 2 - Relação dos artigos pesquisados e suas especificidades

Autor (es)	Doença/Exame	Abordagem	Nº de Imagens	Resultados
Kermany, Zhang e Goldbaum (2018)	DME / OCT	*Transfer Learning* com *Inception* V3	108.312	Acurácia: 96,6% *Recall*: 97,8% Especificidade: 97,4%

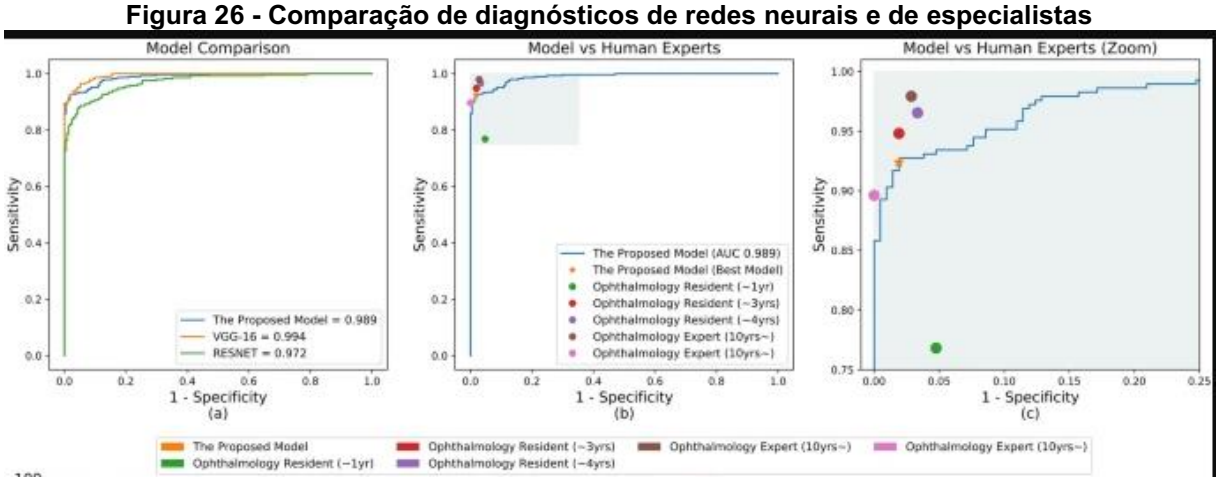
Autor (es)	Doença/Exame	Abordagem	Nº de Imagens	Resultados
Chen *et al.,* (2021)	DME / OCT	*Transfer Learning com VGG-19 / ResNet-50 / ResNet-101*	83.484	Acurácia: 99,4% Precisão: 99,4% *Recall*: 99,4% *F-measure: 99,4%* Especificidade: 99,8%
Kaymak, S. e Serener, A. (2018)	DME / OCT	*Transfer Learning* com AlexNet	83.484	Especificidade: 99,60% Acurácia: 99,80% *Recall*: 100%
Choudary *et al.,* (2023)	DME / OCT	*Transfer Learning* com VGG-19	84.568	Acurácia: 99,17% *Recall:* 99,00% Especificidade: 99,50%
Zhang *et al*., (2020)	DME / OCT	*Transfer Learning* com *Inception* V3	38.057	Acurácia: 94,5% Precisão: 97,2% Especificidade: 97,0% Sensibilidade: 97,7%
Sun, Zhang e Yao (2020)	DME / OCT	*Transfer Learning* com ResNet-50	1.294	Acurácia: 98,00% *Recall:* 95,65% Especificidade: 99,26%
Prahs *et al*., (2017)	DME / OCT	*Transfer Learning* com *Inception* V1	183.402	Acurácia: 95,5% Especificidade: 96,2% Sensibilidade: 90,1%
Alsaih *et al.,* (2017)	DME / OCT	*Support Vector Machine* (SVM)	3.800	Especificidade: 87,5% Sensibilidade: 87,5%
Yoon *et al.,* (2020)	Coriorretinopatia serosa / OCT	Rede neural convolucional (CNN) sem *Transfer Learning*	2.360	Acurácia: 93,8% Especificidade: 99,1% Sensibilidade: 90,0%
Rodriguez *et al*., (2017)	DME / Exame de fundo de olho	AlexNet	1.399	Especificidade: 61,0% Sensibilidade: 70,0%

Fonte: Autoria própria (2023).

O resultado obtido por Chen *et al.,* (2021) foi um dos mais notáveis, se aproximando de 100%. Os autores utilizaram os modelos ResNet-50, ResNet-101 e VGG-19. Em seu estudo, Yoon *et al*. (2020) descobriram que algoritmos de aprendizado profundo de rede neural convolucional podem ser usados para

desenvolver um modelo de aprendizado profundo baseado em tomografia de coerência óptica para detectar coriorretinopatia serosa central. O modelo de rede neural convolucional criado teve 90% de precisão na classificação de olhos normais e 98% de precisão na classificação de um olho afetado. Portanto, o modelo é mais preciso na classificação de olhos infectados e olhos normais do que os especialistas humanos. A Figura 26 mostra comparações entre diferentes modelos e especialistas humanos.

Figura 26 - Comparação de diagnósticos de redes neurais e de especialistas



Fonte: Yoon *et al*. (2020).

Chen *et al*. (2018) desenvolveram uma rede neural artificial para prever automaticamente resultados visuais em pacientes tratados com ranibizumabe intravítreo com edema macular diabético. As redes neurais artificiais tiveram coeficientes de correlação positivos significativos para predizer o prognóstico. O modelo pode ser uma excelente ferramenta clínica para prever o sucesso dos tratamentos e diagnosticar novos pacientes.

O estudo de Zhang *et al.,* (2020) foi conduzido por uma equipe de pesquisa da Universidade de Nankai com o objetivo de identificar DME e outros distúrbios oculares através de imagens de OCT. Foram utilizadas 38.057 imagens, incluindo amostras de treinamento e validação. O estudo implementa um algoritmo de aprendizado por transferência que inclui um módulo de detecção e realce de bordas, seguido pela classificação das doenças. Esse processo, segundo os autores, permite uma classificação precisa das imagens, mesmo aquelas de baixa qualidade, por meio do realce das bordas e da representação intuitiva dos alvos de detecção. Os

resultados demonstraram uma precisão de 94,5%, com 97,2% de precisão, 97,7% de sensibilidade e 97% de especificidade no conjunto de teste.

A pesquisa de Kaymak e Serener (2018) utilizou o banco de dados de Kermany, Zhang e Goldbaum (2018) para propor uma metodologia baseada no treinamento de um modelo de rede neural profunda com a AlexNet. Os autores também focaram na detecção de DME através de imagens de OCT. Foram utilizadas 83.484 imagens OCT. Na classificação entre imagens saudáveis e com DME, a precisão foi de 99,8%, com sensibilidade de 100% e especificidade de 99,6%.

A pesquisa de Choudary *et al.,* (2023) desenvolveu uma rede neural para diagnóstico de edema macular diabético, utilizando uma rede neural convolucional profunda VGG-19 aplicando transferência de aprendizado. O modelo foi treinado com um conjunto de 84.568 imagens de OCT da base de dados de Kermany, Zhang e Goldbaum (2018). Os resultados demonstraram alta precisão do modelo, com 99,17% de acurácia, 99,50% de especificidade e 99,00% de *recall*.

Kaymak *et al.,* (2018), Choudary *et al.,* (2023) e Sun, Zhang e Yao (2020) obtiveram resultados próximos a 100% utilizando três redes diferentes no *transfer learning.* O primeiro utilizou a AlexNet com o mesmo *dataset* que o segundo, porém este utilizou a rede VGG-19. Por sua vez, Sun, Zhang e Yao (2020) obtiveram os mesmos altos resultados com a ResNet-50 mesmo tendo menos dados de treinamento.

Convém citar que os trabalhos de Chen *et al.,* (2021), Kaymak *et al.,* (2018) e Choudary *et al.,* (2023), utilizaram o mesmo *dataset* de Kermany, Zhang e Goldbaum (2018), além dos próprios autores.

O estudo de Alsaih *et al*., (2017) se aproxima mais da proposta deste projeto. Nele, os autores realizam um experimento de classificação de DME através de exames de OCT no qual utiliza-se de um pré-processamento com combinação de blocos, filtragem 3D, achatamento e corte da região de interesse. Foi utilizado extração da região de interesse por pirâmide de multi-resolução e por fim realizada a classificação através de *Support Vector Machines*. Seus resultados, como podem ser vistos na Quadro 2, foram de 87,5% de sensibilidade e especificidade. Outro estudo semelhante, porém com melhores resultados foi o de Zhang *et al*. (2020) que utilizou a mesma base de imagens que o presente trabalho propõe, porém, mantendo suas proporções originais. O modelo utilizado foi o *Inception* V3 e como pré-processamento, os autores utilizaram detecção de bordas para delimitar a região a

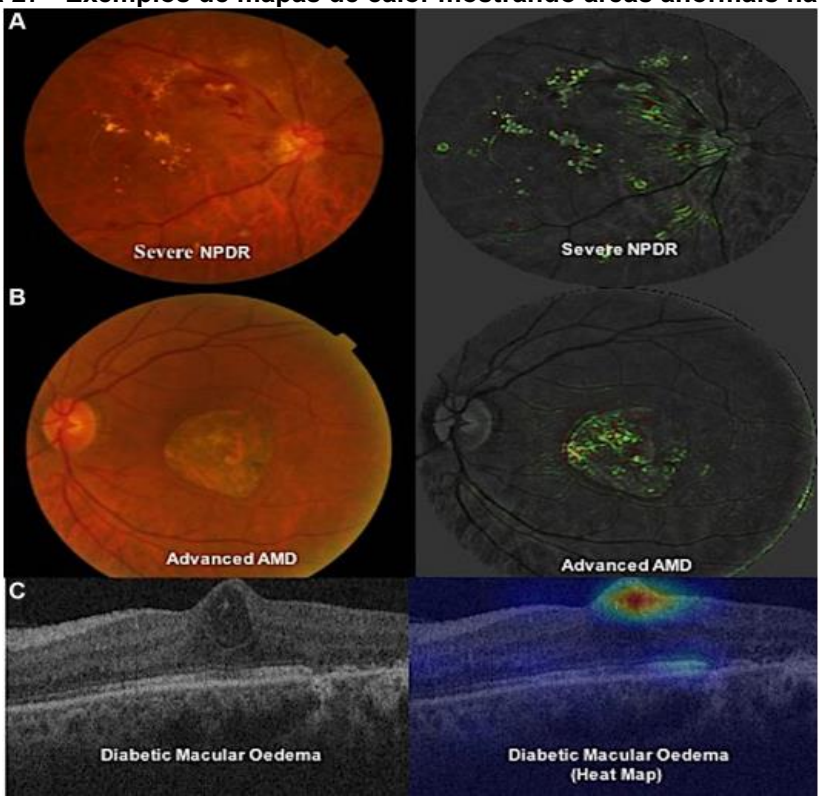
qual os fluidos típicos de retinas acometidas se localizam, conseguindo assim, uma acurácia de 94,5% e 97,5% de precisão.

Ainda dentro do âmbito específico de classificação através do exame de OCT, Prahs *et al.,* (2017) descreve um experimento que não visa a classificação do edema macular diabético propriamente dito, mas sim a indicação ou não de tratamento ocular com medicamentos anti-VEGF (*Vascular Endothelial Growth Factor*). Além disso, os autores fizeram uso de um número maior de imagens em relação aos outros artigos verificados, totalizando 183.402 imagens ao todo. Após sua classificação ser realizada os autores obtiveram uma taxa de 95,5% de acurácia.

Como sugestão de possíveis pesquisas Ting *et al.,* (2017) sugere a construção de mapas de calor que os mesmos supõem ser o primeiro passo para a evidenciação de regiões anormais tanto do exame de fundo de olho, para verificação de Retinopatia Diabética como o OCT para diagnóstico de DME.

A Figura 27 mostra a abordagem de Ting *et al.,* (2017) sobre os mapas de calor. A Figura mostra imagens de retinopatia diabética não proliferativa, degeneração macular e edema macular diabético, bem como o resultado do pré-processamento dos autores nestas imagens.

Figura 27 - Exemplos de mapas de calor mostrando áreas anormais na retina



Fonte: Ting *et al*. (2017).

Na Figura 27 pode-se observar como uma região de interesse no exame de OCT entra em evidência com o uso do mapa de calor.

2.9.7 Considerações finais

A primeira pergunta desta pesquisa que indaga sobre a utilidade da inteligência artificial no diagnóstico médico se correlaciona com a terceira que questiona sobre a melhora da velocidade no diagnóstico e se esta melhora pode levar a um diagnóstico precoce. Como observa-se na Quadro 2 a grande maioria dos trabalhos pesquisados que fizeram algum tipo de classificação sobre alguma doença obtiveram uma percentagem de acerto de mais de 80% com algumas métricas se aproximando de 100%. Prahs *et al.,* (2017) são enfáticos ao dizer que “Redes neurais apresentam desempenho impressionante em classificações com OCT da retina. Após o treinamento em dados clínicos históricos, os métodos de aprendizado de máquina podem oferecer ao clínico suporte no processo de tomada de decisão...”. Os autores também deixam a ressalva para que não se tome o resultado da rede neural como verdade absoluta, ou seja, a predição da rede neural é apenas uma predição e não deve ser confundida com o apropriado diagnóstico médico.

Em todos os artigos que fazem menção a este tópico, encaixam o aprendizado de máquina como muito importante e em crescente expansão, principalmente na área médica. Segundo Ting *et al.,* (2018) práticas com *deep learning* aplicadas a doenças oculares apresentaram um desempenho “aceitável” clinicamente. Ressaltando ainda que nas próximas décadas o *deep learning* terá um enorme impacto na medicina, e que há muito espaço para pesquisas futuras que avaliem a implantação e custo benefício desta tecnologia.

Ainda em relação à pergunta 3, conforme resultados obtidos na pesquisa de Chen (2018) que avaliou a indicação ou não de tratamento intravítreo na retina, detecções precoces de doenças no geral, mas principalmente da retina aliado a um tratamento agressivo. Assim que descoberta a doença pode sim melhorar a perspectiva de pacientes portadores de DME diminuindo a probabilidade de haver prejuízos à visão.

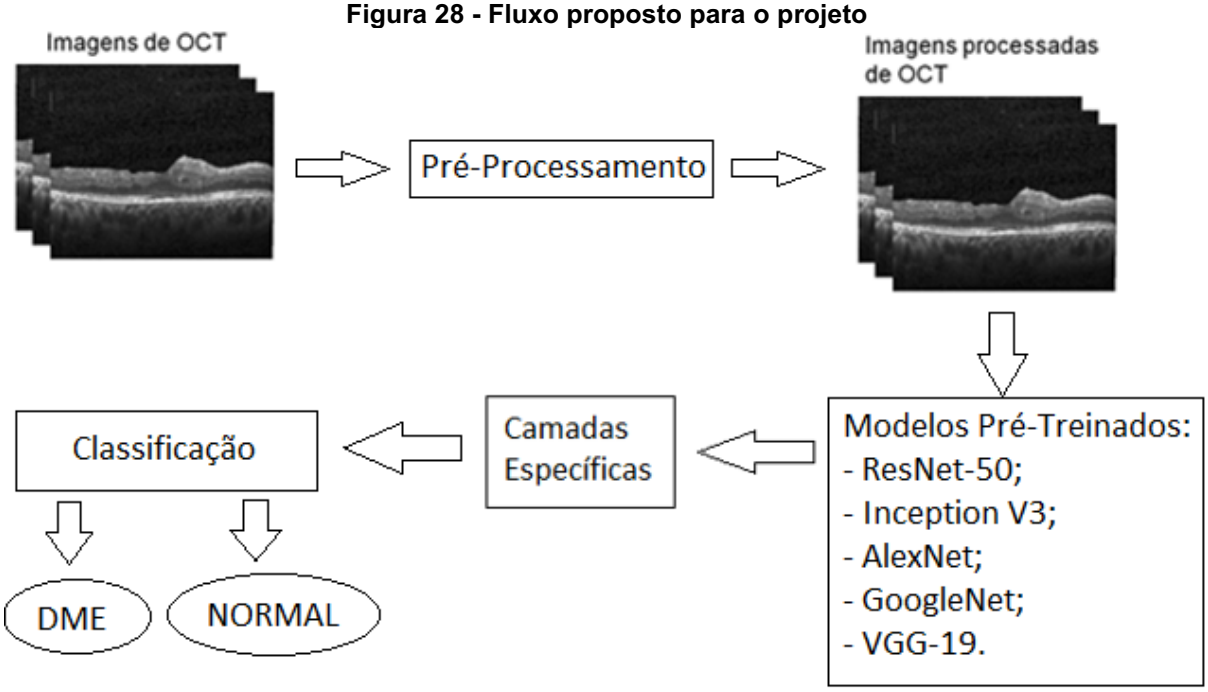
A pergunta 2, por sua vez, visava saber se o diagnóstico de Edema Macular é melhorado com o apoio da inteligência artificial / redes neurais. O modelo proposto

por Yoon *et al.,* (2020) apresentou melhor sensibilidade e acurácia do que os oftalmologistas na tarefa de distinguir entre coriorretinopatia serosa central aguda e crônica. A análise que fica aqui é o quanto o diagnóstico do especialista será melhorado caso tenha um sistema de aprendizado de máquina capaz de auxiliá-lo na tomada de decisão, diminuindo assim consideravelmente o erro humano, tempo de diagnóstico e quem sabe até o mesmo o custo deste diagnóstico. Pois um sistema que prediz corretamente mais de 90% dos casos que lhe é exposto deve ser objeto de grande pesquisa e desenvolvimento para que no futuro esta contribuição seja ainda maior.

3 METODOLOGIA

Nesta seção serão abordadas as etapas para realização do projeto proposto. A metodologia proposta neste trabalho é representada na Figura 28.

Figura 28 - Fluxo proposto para o projeto



Fonte: Autoria própria (2024).

O primeiro passo é a aquisição das imagens de OCT da base de dados que será explicitada na subseção 3.2. Após ser delimitada a base de dados, estas imagens passarão por um pré-processamento (seção 3.3) que visa realçar a imagem de modo que a rede consiga performar melhor.

Com as imagens devidamente processadas, o próximo passo, conforme detalha a seção 3.4, é realizar o *fine tunning* dos modelos neurais com as arquiteturas pré-treinadas *ResNet-50, Inception* V3, GoogLeNet, VGG-19 e AlexNet. Após inserido os pesos da rede pré treinada e as camadas específicas (explicado nas próximas seções rede por rede) é realizado o treinamento da rede e posterior classificação das imagens de teste. Estes resultados estarão sendo exibidos no Capítulo 4 de resultados obtidos com a aplicação das métricas de desempenho abordadas no referencial teórico.

Nas próximas seções será detalhado a construção e aplicação da metodologia proposta que foi previamente esquematizada pela Figura 28.

3.1 Ferramentas de desenvolvimento

Para este trabalho foram selecionadas as ferramentas que melhor se alinham com a natureza técnica demanda. Como linguagem de programação foi utilizado o *Python* que é amplamente reconhecido pela sua eficácia e disponibilidade de recursos para trabalhos relacionados a inteligência artificial. O *Python* também dispõe de suporte a bibliotecas especializadas como a *OpenCV* (*Open Source Computer Vision Library*), que foi escolhido para a manipulação e processamento de imagens visto de dispõe de uma ampla gama de métodos e operações em imagens. O *MatLab* também foi utilizado para as operações morfológicas pois é um ambiente altamente otimizado para operações numéricas e algorítmicas.

Para a construção e treinamento das redes neurais foi utilizado o *Keras* que é um *framework* de alto nível de redes neurais. Sua interface é simplificada e roda em cima de outras bibliotecas como o *Tensorflow*. O projeto também integra cinco arquiteturas de redes neurais pré-treinadas. São elas a ResNet-50, a *Inception* V3, a AlexNet, a GoogLeNet e a VGG-19.

Como Ambiente de Desenvolvimento Integrado (*Integrated Development Environment – IDE)* foi utilizado o *Spyder*, que trabalha com a linguagem *Python* e oferece recursos como depuração integrada e ferramentas úteis para tarefas de *deep learning*.

As especificações do *hardware* no qual foi realizado o experimento são um processador Intel I7 7500U de 2,7 GHZ, 8GB de memória RAM e 4GB de memória de vídeo com a placa dedicada AMD Radeon R7 M445.

3.2 Base de dados

A base de dados para este trabalho foi obtida de Kermany, Zhang e Goldbaum (2018). A base é dividida por imagens para treino e para teste. A Tabela 1 mostra a distribuição das imagens no *dataset*.

Tabela 1 - Distribuição das imagens na base de dados

	Treino	Teste	Treino selecionado	Teste selecionado
Normal	51.140	250	11.348	250
Edema macular diabético	11.348	250	11.348	250
Neovascularização coroidal	37.205	250	0	0
Drusas	8.616	250	0	0
TOTAL	**108.309**	**1000**	**22.696**	**500**

Fonte: Kermany, Zhang e Goldbaum (2018).

De acordo com a Tabela 1, para treinamento, encontra-se 108.309 imagens divididas desproporcionalmente entre as classes que são: “normal”, “edema macular diabético”, “neovascularização coroidal” e “drusas”. Já para teste existem exatamente 250 imagens de cada classe. Como o objetivo deste trabalho é diferenciar os exames que apresentam o edema macular diabético dos normais, foram selecionadas todas as 11.348 imagens do conjunto ‘DME’ mais 11.348 imagens aleatórias do conjunto ‘Normal’ para realizar o treinamento. Para a etapa de teste foram mantidos os dois conjuntos ‘DME’ e ‘Normal’ como aparecem no *dataset* onde cada classe possui um total de 250 imagens.

Entre as imagens de treino foi delimitado uma divisão de 80% para treinamento e 20% para validação do treinamento. Esta divisão foi realizada com o parâmetro ‘*validation_split = 0,2*’ do *ImageDataGenerator.* A Tabela 2 apresenta a quantidade de imagens para treinamento, validação e teste utilizadas no trabalho.

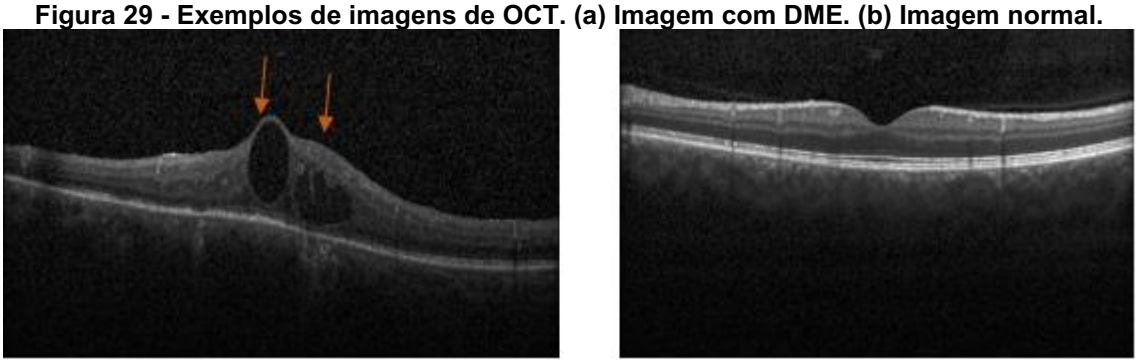
Tabela 2 - Quantidade de imagens da base de dados utilizadas

Procedimento	Número de imagens (total)
Treinamento	18.158
Validação	4.538
Teste	500
Total	**23.196**

Fonte: Autoria própria (2024).

A Figura 29(a) apresenta um exemplo de imagem com DME enquanto a Figura 29(b) explicita um exemplo normal.

Figura 29 - Exemplos de imagens de OCT. (a) Imagem com DME. (b) Imagem normal.



(a) **(b)**

Fonte: Kermany, Zhang e Goldbaum (2018).

Pode-se perceber nas duas figuras a diferença entre uma tomografia de retina acometida pela doença e outra saudável. Há vários pontos com vazamentos de fluidos no entorno do disco óptico indicando a presença de DME (setas laranjas na Figura 29(a)).

3.3 Pré-processamento de imagens

O *framework Keras* possui uma função chamada *flow_from_directory* que percorre os arquivos do diretório, passada por parâmetro, identificando as classes que serão tratadas na rede neural. Neste projeto há o diretório ‘dme’ e o diretório ‘normal’, que serão as classes a serem identificadas. A partir disso o algoritmo percorre estes diretórios armazenando todas suas respectivas imagens em variáveis que posteriormente serão introduzidas à rede neural para treinamento.

Após obtidas as imagens do diretório, nos testes iniciais foram realizados alguns métodos preliminares de pré-processamento como processamento morfológico, filtro gaussiano e recorte da região de interesse (ROI). Foi realizado também o recorte da região de interesse indicando os pixels limiares do centro da imagem através do *imcrop.* A binarização das imagens com o auxílio do método *im2bw,* que converte a imagem para *grayscale* (escala de cinza) e, em seguida realiza a limiarização. Em seguida, através do método *bwmorph*, realiza-se a abertura morfológica (operação morfológica de erosão seguida de dilatação). Estes métodos foram executados através da ferramenta *MatLab* sendo mostrado nas Figuras 30 e 31. Porém estes métodos de pré-processamento não performaram bem após os testes serem realizados.

Portanto, foi realizada e mantida na versão final do projeto a aplicação do desfoque gaussiano com o método *'cv2.GaussianBlur'* com um *kernel* de tamanho 5x5 e desvio padrão de 0, permitindo que o *OpenCV* calcule automaticamente o valor ideal baseado no *kernel*. Outra técnica utilizada foi o *data augmentation* através do método *ImageDataGenerator* do *Keras,* com as seguintes transformações:

- *width_shift_range* e *height_shift_range*: As imagens são transladadas horizontal e verticalmente em até 20% do total da largura e altura, respectivamente;
- *horizontal_flip* e *vertical_flip:* As imagens são espelhadas horizontal e verticalmente;
- Aplicação do filtro gaussiano: As imagens geradas pelo *data augmentation* assim como as fixas da base de dados passam pela aplicação do filtro gaussiano;
- Normalização: Todas as imagens (geradas e originais) passam por uma normalização que se dá pelo parâmetro *rescale = 1,0 / 255,0.*

O método de aumento de dados do *Keras* aplica transformações aleatórias seguindo os parâmetros citados acima em cada época de treinamento. Desta forma a aplicação gera mais exemplos automaticamente dentro do conjunto de dados, o que auxilia a generalização da rede.

3.4 Redes neurais

Para confecção das redes neurais foi utilizada a técnica de transferência de aprendizado que funciona transferindo conhecimento aprendido de uma rede para outra. O objetivo desta técnica é melhorar o desempenho da rede a qual se está aplicando através do reaproveitamento de *features* aprendidas pela 'rede-mãe'. Isso ocorre obtendo os pesos das camadas mais gerais de aprendizado da 'rede-mãe' e manipulando apenas as últimas camadas da nova rede de modo a abordar suas especificidades nela como o formato de saída.

Em todas as redes foi utilizado a função do *framework Keras* chamada *EarlyStopping* que interrompe o treinamento quando a acurácia não estiver melhorando por um número definido de épocas (neste trabalho foi definido para 10 épocas). Com o auxílio desta técnica foi delimitado que todos os treinamentos terão

até 150 épocas com um *batch size* de 32, visto que a partir do momento que o modelo não conseguir mais generalizar e aumentar a acurácia, o treinamento é parado.

Outra técnica aplicada para todos os modelos foi o *ReduceLROnPlateau* que reduz a taxa de aprendizado (*learning_rate)* quando a acurácia não melhora após 6 épocas ajustando-a por um fator de 0,2 e definindo como limite mínimo para o *learning rate* o valor de 0,0001.

Depois das especificidades de cada modelo, que serão discutidas nos próximos subcapítulos, o próximo passo foi ajustar algumas configurações para a compilação como o otimizador e o *learning rate*. O otimizador utilizado em todas as redes foi o Adam que é um método de gradiente estocástico descendente baseado na estimativa adaptativa de momentos de primeira e segunda ordem (KINGMA; BA, 2017). É um dos otimizadores mais utilizados do *Keras*. Já o *learning rate*, ou taxa de aprendizado, foi mantido em 0,001 em todas as redes pois a diminuição da taxa pode melhorar significativamente a precisão da generalização enquanto ajuda a evitar o *overfitting*.

Tendo inserido todas as camadas, otimizadores e taxa de *learning rate* foi utilizado o método do *Keras* chamado *fit_generator* que é responsável por obter as configurações previamente definidas e realizar o treinamento da rede.

Em todos os modelos foi inserido por fim uma camada densa com ativação sigmoide que é responsável pela predição final dos modelos e se utilizando da função sigmoidal, retornará um valor entre 0 e 1. Como a classificação é binária (ou DME ou normal) se o resultado for entre [0; 0,5], pertence a classe 1, que neste caso é ‘DME’, e se for entre (0,5; 1] pertence a classe 2 que é ‘normal’, ou seja, quanto mais perto dos extremos, maior a certeza da rede nesta predição.

3.4.1 Arquitetura final do ResNet-50

O modelo pré-treinado *ResNet-50*, que é uma arquitetura com 50 camadas em seu estado original ao contrário das *ResNets* 101, de 101 camadas e a *ResNet-34*, de 34 camadas. Para utilização do modelo foi selecionado os pesos da base de dados *ImageNet* que conta com aproximadamente 14 milhões de imagens.

O primeiro passo na personalização do modelo envolve a aplicação de uma *GlobalAveragePooling2D* na saída do modelo base. Essa técnica reduz a

dimensionalidade dos mapas de características, convertendo-os em um vetor de características único por mapa, o que ajuda a minimizar o risco de *overfitting* e reduz a complexidade computacional.

Após o *pooling* global, são adicionadas sequencialmente três camadas densas com 256, 128 e 32 neurônios, respectivamente, cada uma seguida por uma função de ativação ReLU. Essas camadas densas são responsáveis por interpretar as características globais extraídas pela ResNet50, permitindo que o modelo faça distinções mais finas necessárias para a classificação.

Para combater ainda mais o *overfitting* uma camada de *dropout* com taxa de 0,1 é introduzida após as camadas densas. Isso ajuda a garantir que o modelo não dependa excessivamente de qualquer característica isolada, aumentando assim sua capacidade de generalização para dados não vistos.

A construção do modelo é finalizada com a adição de uma camada densa de saída com uma única unidade e uma função de ativação sigmoide. Esta camada converte as interpretações de alto nível das camadas anteriores em uma probabilidade, indicando a confiança do modelo na classificação das imagens em uma das duas categorias possíveis.

3.4.2 Arquitetura final do *Inception* V3

Para o treinamento da rede baseada no modelo *Inception* V3 são introduzidas uma série de camadas personalizadas para adaptar ainda mais o modelo à tarefa em questão. O processo se inicia com a etapa de *Flatten*, onde a saída do modelo base é convertida de mapas de características multidimensionais em um vetor unidimensional. Isso é fundamental para a transição suave das camadas convolucionais para as camadas *fully connected* que se seguirão.

Em seguida, três camadas são introduzidas, de 256, 128 e 32 neurônios respectivamente, todas elas com a função de ativação ReLU. Essas camadas são responsáveis por analisar e interpretar as características extraídas pelo modelo base, gradualmente focando nas nuances e padrões que são relevantes para a tarefa de classificação em questão.

Para mitigar o risco de *overfitting*, é incorporada uma camada de *dropout* com taxa de 0,1, que funciona excluindo aleatoriamente um subconjunto de características durante o treinamento. Isso promove a robustez do modelo, incentivando-o a aprender

características mais generalizadas que não dependem excessivamente de um pequeno conjunto de neurônios.

Por fim, a camada de saída do modelo é definida, consistindo em uma única unidade com uma função de ativação sigmoide.

3.4.3 Arquitetura final do AlexNet

A construção da AlexNet adaptada começa com uma camada convolucional inicial dotada de 96 filtros e um tamanho de kernel de 11×11, com um passo de 4 e sem preenchimento ("*padding*"). Esta configuração é específica para capturar características de alto nível em uma escala maior, adequada para a entrada de imagem especificada. Segue-se uma camada de *MaxPooling* para reduzir a dimensionalidade e concentrar-se nas características mais importantes.

A rede então se aprofunda com camadas convolucionais adicionais, aumentando o número de filtros para 256 e posteriormente para 384, mantendo os kernels menores e um *padding* que permite a preservação das dimensões espaciais das características. Estas camadas são projetadas para extrair uma rica hierarquia de características visuais, das mais simples às mais complexas.

Após as camadas convolucionais, a rede emprega uma camada de *Flatten* para transformar os mapas de características multidimensionais em um vetor único, facilitando a transição para o processamento por meio das camadas densas. Essas camadas densas, com 4.096 unidades cada, são cruciais para a interpretação das características visuais em um nível de abstração mais elevado, permitindo que a rede faça previsões sobre a classificação das imagens.

Para mitigar o risco de *overfitting*, são incorporadas camadas de *Dropout* com uma taxa de 0,5 após cada camada densa. A rede culmina com uma camada densa de saída com uma função de ativação sigmoide.

3.4.4 Arquitetura final do GoogLeNet

A arquitetura GoogLeNet é notável pela sua abordagem inovadora na construção de redes neurais profundas, introduzindo os "Módulos *Inception*" que permitem que a rede se adapte às escalas dos recursos visuais de maneira eficiente.

A personalização inicia com a aplicação de uma camada de *GlobalAveragePooling2D*, seguida por uma camada *Flatten*. Este processo reduz a dimensionalidade dos dados, transformando os mapas de características complexos em um vetor simples, facilitando a transmissão dessas informações para as camadas subsequentes. Este passo é essencial para condensar as informações extraídas pelas múltiplas camadas e módulos *Inception* de uma forma que possa ser efetivamente utilizada para a classificação.

Para fortalecer a capacidade de interpretação do modelo, uma camada de *Dropout* com uma taxa de 0,1 é introduzida para mitigar o risco de *overfitting*. A personalização termina com a adição de uma camada densa de saída que utiliza uma função de ativação sigmoide, tornando-a adequada para tarefas de classificação binária. Esta camada traduz as representações aprendidas e condensadas pelo modelo em uma probabilidade, refletindo a confiança do modelo na classificação das entradas em uma das duas categorias possíveis.

3.4.5 Arquitetura final do VGG-19

O processo de personalização da rede VGG-19 começa com o "achatamento" da saída do modelo base utilizando uma camada *Flatten*. Esse passo converte os mapas de características 3D em um vetor 1D, facilitando a transição para camadas densas que seguem.

Após a etapa de achatamento, duas camadas densas são adicionadas com 256 e 32 neurônios, respectivamente, ambas ativadas pela função ReLU. Estas camadas atuam como interpretadoras das características achatadas, permitindo uma análise mais profunda e a capacidade de fazer distinções mais finas que são cruciais para a tarefa de classificação em questão.

Para prevenir *overfitting*, especialmente considerando que o VGG19 é uma rede profunda e pode se tornar propensa a memorizar os dados de treinamento em detrimento da generalização, uma camada de dropout com uma taxa de 0,1 é introduzida. Este passo é estratégico para promover a robustez do modelo, forçando-o a aprender representações mais generalizáveis ao desativar aleatoriamente uma fração das características durante o treinamento.

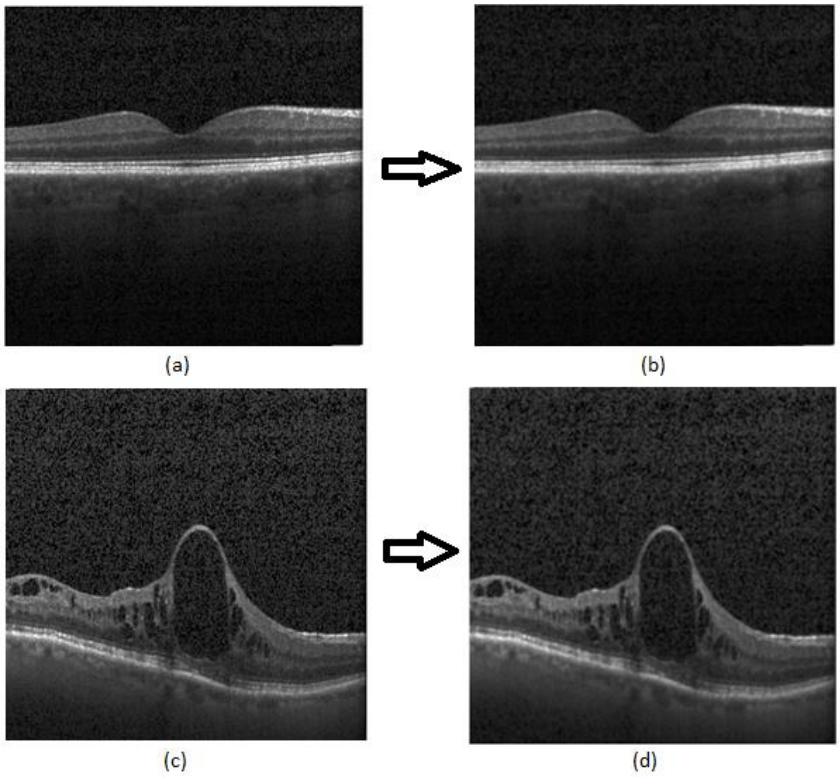
Por fim, a construção do modelo é concluída com uma camada densa de saída que possui uma única unidade com uma função de ativação sigmoide. Esta configuração é ideal para tarefas de classificação binária, pois a função sigmoide mapeia as saídas da rede para um intervalo entre 0 e 1, representando a probabilidade de a entrada pertencer a uma das duas classes.

4 RESULTADOS

O treinamento das redes previamente explicado culminou em um modelo neural de cada arquitetura de rede. Uma vez obtido os modelos treinados, é possível submetê-los a imagens externas (que não fizeram parte do processo de treinamento) para avaliar o desempenho real da rede. Neste caso a base de dados utilizada possui 250 imagens de teste para cada classe, estas portanto, não foram incluídas no treinamento.

Como resultado do pré-processamento a Figura 30 mostra o resultado do filtro gaussiano.

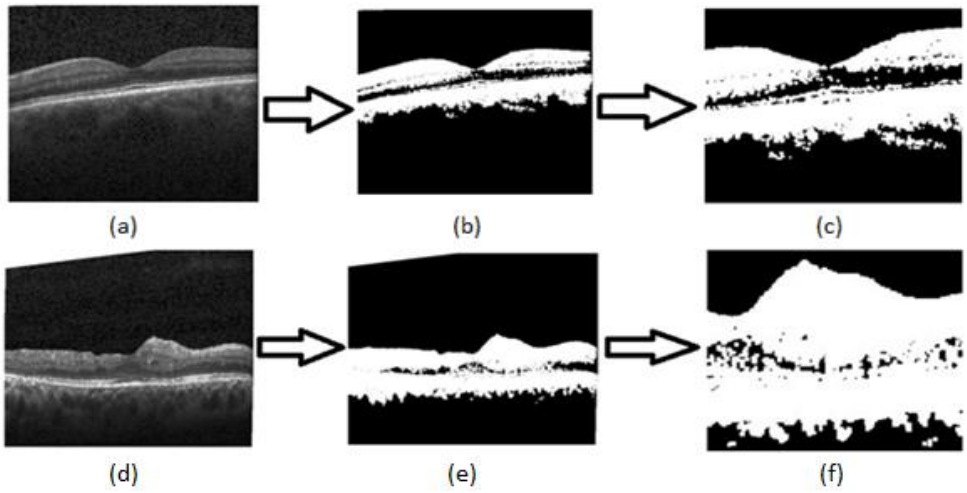
Figura 30 - Resultado do filtro gaussiano



Fonte: Autoria própria (2024).

A Figura 30 (a) mostra a imagem da classe normal sem filtro, e na Figura 30 (b) é apresentado a mesma imagem após a aplicação do filtro gaussiano. Já na Figura 30 (c) é exibido uma imagem da classe DME sem filtro e na Figura 30 (d) é mostrado o resultado do filtro gaussiano aplicado nesta imagem. É possível perceber o efeito do filtro gaussiano suavizando levemente os ruídos das imagens. Na Figura 31 é apresentado o resultado do processamento morfológico.

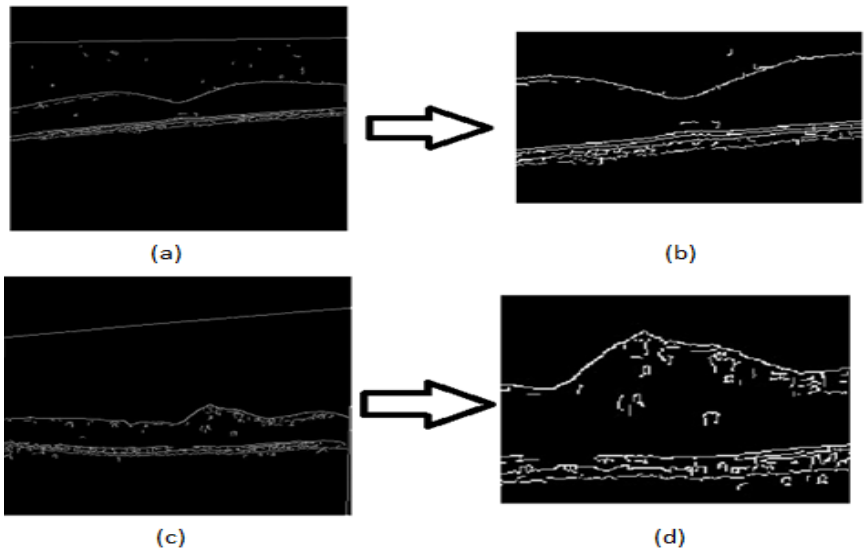
Figura 31 - Resultado do processamento morfológico



Fonte: Autoria própria (2024).

A Figura 31 (a) mostra a imagem da classe normal sem filtro, e na Figura 31 (b) é apresentado a mesma imagem após a aplicação do processamento morfológico enquanto na Figura 31 (c) é mostrado o recorte da região central da imagem após processamento. Já na Figura 31 (d) é exibido uma imagem da classe DME sem filtro e na Figura 31 (e) é mostrado o resultado do processamento morfológico aplicado nesta imagem enquanto na Figura 31 (f) é exibido o recorte da região central da imagem após processamento. Na Figura 32 é apresentado o resultado do filtro de bordas canny.

Figura 32 - Resultado da aplicação do filtro de bordas Canny



Fonte: Autoria própria (2024).

A Figura 32 (a) mostra a imagem da classe normal sem filtro, e na Figura 32 (b) é apresentado a mesma imagem após a aplicação do filtro canny. Já na Figura 32 (c) é exibido uma imagem da classe DME sem filtro e na Figura 32 (d) é mostrado o resultado do filtro canny aplicado nesta imagem.

No entanto, a aplicação do processamento morfológico, bem como filtro canny não contribuiu positivamente para a melhora dos resultados. Desta forma, somente o filtro gaussiano foi considerado na etapa de pré-processamento.

Nas próximas seções serão apresentados os resultados obtidos após a classificação das imagens da base de teste em cada rede neural pré-treinada.

4.1 ResNet-50

A Tabela 3 demonstra a matriz de confusão obtida após a aplicação do modelo ResNet-50 na base de teste.

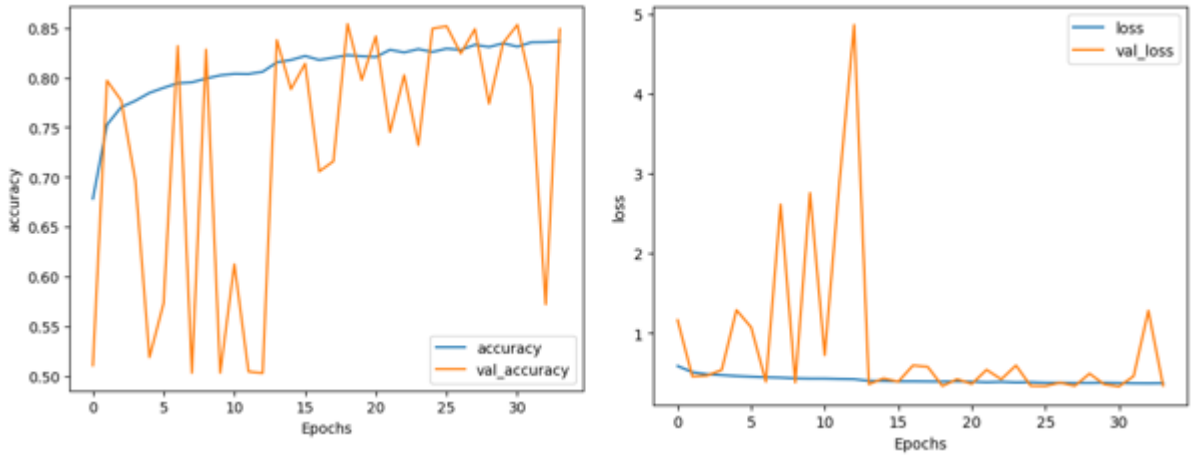
Tabela 3 - Matriz de Confusão obtida com a ResNet-50

	Real "DME"	**Real "Normal"**
Predito "DME"	243 (VP)	10 (FP)
Predito "Normal"	7 (FN)	240 (VN)

Fonte: Autoria própria (2024).

A matriz de confusão da ResNet-50 mostra um alto equilíbrio entre verdadeiros positivos e verdadeiros negativos, com poucos falsos positivos e falsos negativos. Este modelo alcançou uma acurácia de 96,60%, indicando uma alta taxa de classificações corretas. A precisão de 96,05% reflete a capacidade do modelo de prever corretamente a classe DME, enquanto o *recall* de 97,2% destaca sua habilidade em detectar a maioria dos casos reais de DME. O *F-Measure* de 96,62% e a especificidade de 96,00% reforçam a eficiência equilibrada do modelo em classificar ambas as classes corretamente. Para atingir estes resultados o modelo foi treinado e o Gráfico 1 mostra os dados obtidos através do treinamento.

Gráfico 1 - Dados de treinamento da rede ResNet-50



(a) **(b)**

Fonte: Autoria própria (2024).

Ao analisar as imagens com os dados de treinamento da rede ResNet-50, é possível observar a evolução da acurácia do treinamento e validação e também da função de perda (*Loss*) ao longo de 34 épocas (*Epochs*). No Gráfico 1(a), que representa a acurácia, observa-se que a acurácia de treinamento (azul) flutua em torno de valores altos, enquanto a acurácia de validação (laranja) apresenta variações mais significativas, com tendência de aumento até estabilizar-se.

No Gráfico 1(b), que ilustra a função de perda, a perda de treinamento (azul) diminui consistentemente, indicando que o modelo está aprendendo e se ajustando aos dados de treinamento. No entanto, a perda de validação (laranja) apresenta picos mais significativos no início das épocas.

4.2 Inception V3

Na Tabela 4 é apresentada a matriz de confusão obtida após a aplicação do modelo *Inception* V3 na base de teste.

Tabela 4 - Matriz de Confusão obtida com a Inception V3

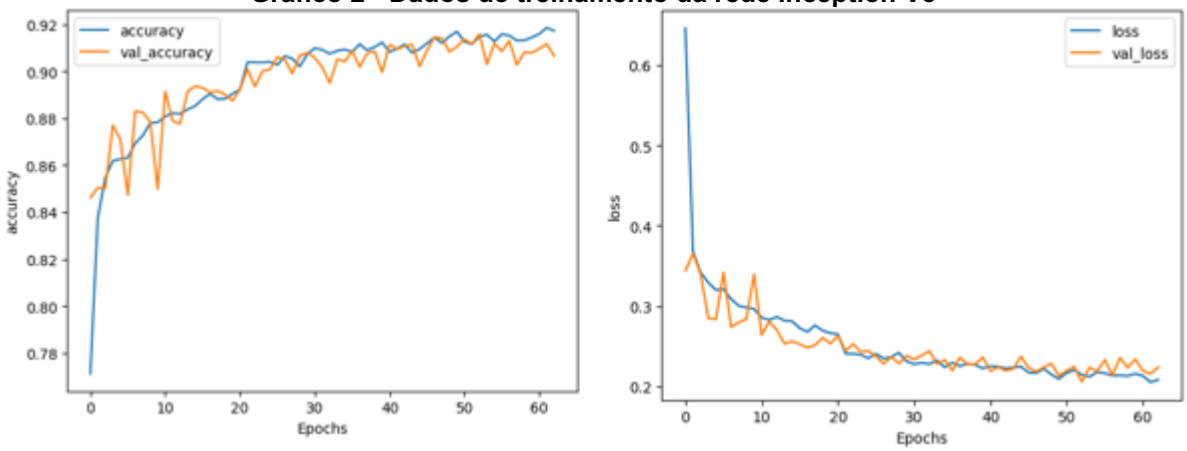
	Real "DME"	**Real "Normal"**
Predito "DME"	247 (VP)	9 (FP)
Predito "Normal"	3 (FN)	241 (VN)

Fonte: Autoria própria (2024).

Observa-se na Tabela 4 a matriz de confusão exibindo uma performance com altos valores de VP e VN e mínimos erros de FP e FN. A Acurácia atingiu 97,60%,

mostrando uma eficácia notável na classificação geral. Uma Precisão de 96,48% significa que quase todas as previsões positivas estavam corretas, e o *Recall* de 98,80% indica uma forte capacidade de identificar positivos reais. Outras métricas como *F-Measure* e especificidade obtiveram 97,63% e 96,40% respectivamente.

Gráfico 2 - Dados de treinamento da rede Inception V3



(a) **(b)**

Fonte: Autoria própria (2024).

O Gráfico 2 apresenta as curvas de aprendizado para o modelo ao longo de 63 épocas. A curva do Gráfico 2(a) mostra a acurácia do modelo (*'accuracy'*) e a acurácia de validação (*'val_accuracy'*). Observa-se que a acurácia de treinamento começa em aproximadamente 78% e aumenta de forma estável, alcançando uma precisão máxima próxima a 92% por volta da época 50. A acurácia de validação segue uma trajetória similar sem oscilações bruscas.

Já o Gráfico 2(b) ilustra a perda de treinamento e a perda de validação. Esta métrica inicia em um valor próximo de 0,6, reduzindo rapidamente e estabilizando em torno de 0,2. A perda da validação diminui de maneira similar.

4.3 AlexNet

É exibido na Tabela 5 a matriz de confusão obtida após a aplicação do modelo AlexNet na base de teste.

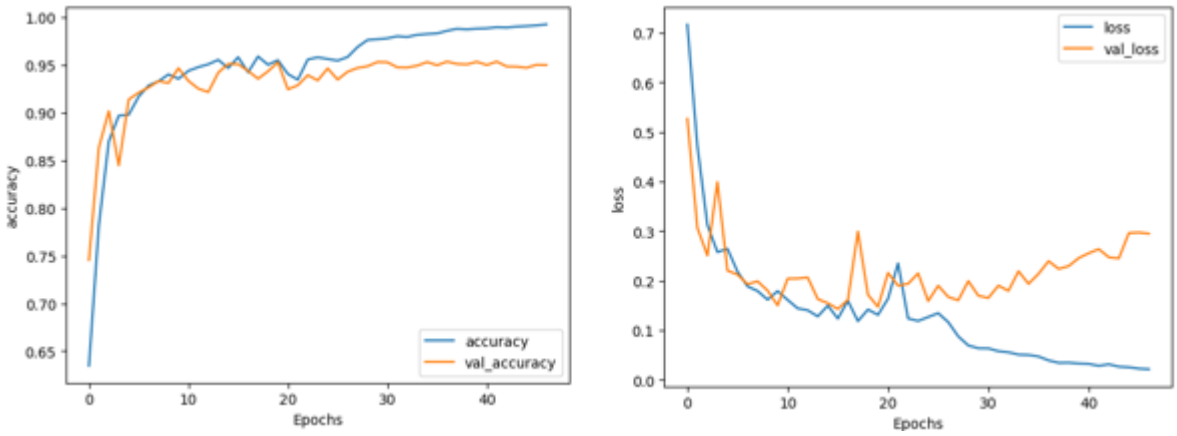
Tabela 5 - Matriz de Confusão obtida com a AlexNet

	Real “DME”	**Real “Normal”**
Predito “DME”	250 (VP)	6 (FP)
Predito “Normal”	0 (FN)	244 (VN)

Fonte: Autoria própria (2024).

A AlexNet, com sua matriz de confusão exposta na Tabela 5, demonstra um desempenho notável, especialmente na identificação de classes negativas, como indicado pela ausência de Falsos Negativos (FN). Isso se reflete no *recall* perfeito de 100%, destacando a eficácia do modelo em classificar corretamente todas as imagens com DME. Embora a Precisão seja de 97,66%, indicando que todas as previsões positivas estavam corretas, a presença de 6 falsos positivos afeta a especificidade, que é de 97,60%. O *F-Measure* de 98,81% ainda assim indica um equilíbrio robusto entre Precisão e *Recall*, refletindo a eficiência global do modelo na classificação. O Gráfico 3 mostra os dados de treinamento do modelo.

Gráfico 3 - Dados de treinamento da rede AlexNet



(a) **(b)**

Fonte: Autoria própria (2024).

Como observa-se no Gráfico 3 a evolução do treinamento do modelo se deu ao longo de 47 épocas. O Gráfico 3(a), se percebe que ambas as métricas começam com valores substancialmente diferentes, mas convergem rapidamente, sugerindo que o modelo está aprendendo de forma efetiva. A acurácia do treinamento inicia em torno de 65% e atinge a estabilidade acima de 95%, enquanto a precisão de validação acompanha de perto após as primeiras épocas

Já no Gráfico 3(b), as curvas mostram a perda durante o treinamento e validação. A perda de treinamento começa em torno de 0,7 e decresce rapidamente, nivelando-se próximo a 0,1. A perda de validação, por outro lado, segue uma tendencia de subida distinta, como pode se notar um pico de aumento na perda de validação por volta da época 25.

4.4 GoogLeNet

Na Tabela 6 será exposta a matriz de confusão obtida com a rede GoogLeNet.

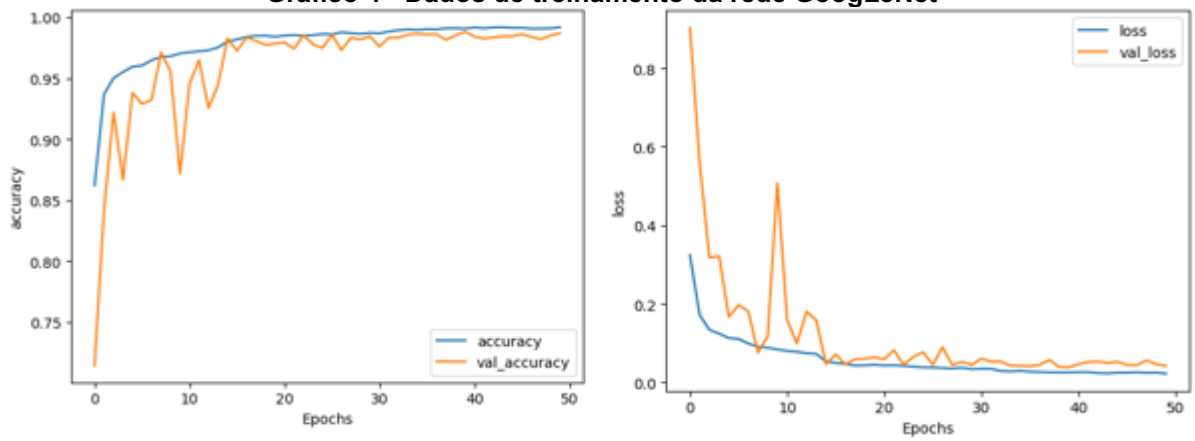
Tabela 6 - Matriz de Confusão obtida com a GoogLeNet

	Real "DME"	**Real "Normal"**
Predito "DME"	250 (VP)	4 (FP)
Predito "Normal"	0 (FN)	246 (VN)

Fonte: Autoria própria (2024).

A GoogLeNet apresenta uma matriz de confusão exemplar com o maior número de verdadeiros positivos e negativos, além da inexistência de falsos negativos. Foi a rede que obteve o menor número de predições incorretas (4 falsos positivos), obtendo 98,43% de precisão. Sua acurácia de 99,20% é a mais alta entre os modelos, refletindo um desempenho superior na classificação correta de ambas as classes. O *Recall* perfeito de 100% demonstra a ausência total de falsos negativos, enquanto a especificidade de 98,40% mostra uma excelente capacidade de detecção de casos positivos. O *F-Measure* de 99,21% confirma a eficácia balanceada e notável do modelo. O Gráfico 4 exibe os dados de treinamento da rede.

Gráfico 4 - Dados de treinamento da rede GoogLeNet



(a) **(b)**

Fonte: Autoria própria (2024).

No Gráfico 4(a), observa-se a trajetória da precisão de treino e de validação ao longo de 50 épocas. A acurácia do treinamento inicia acima de 85% e rapidamente atinge valores acima de 95%, demonstrando uma aprendizagem eficiente e consistente do modelo. A acurácia de validação acompanha esta tendência com uma ligeira queda após a época 10, seguida de uma estabilização.

O Gráfico 4(b) mostra as métricas de perda e perda de validação. A perda no treinamento começa alta e decai acentuadamente, em seguida se estabilizando abaixo dos 20%. A perda de validação segue uma linha similar à de treinamento, com algumas variações por volta da época 10.

4.5 VGG-19

É mostrado na Tabela 7 a matriz de confusão obtida após treinamento e teste com a rede VGG-19.

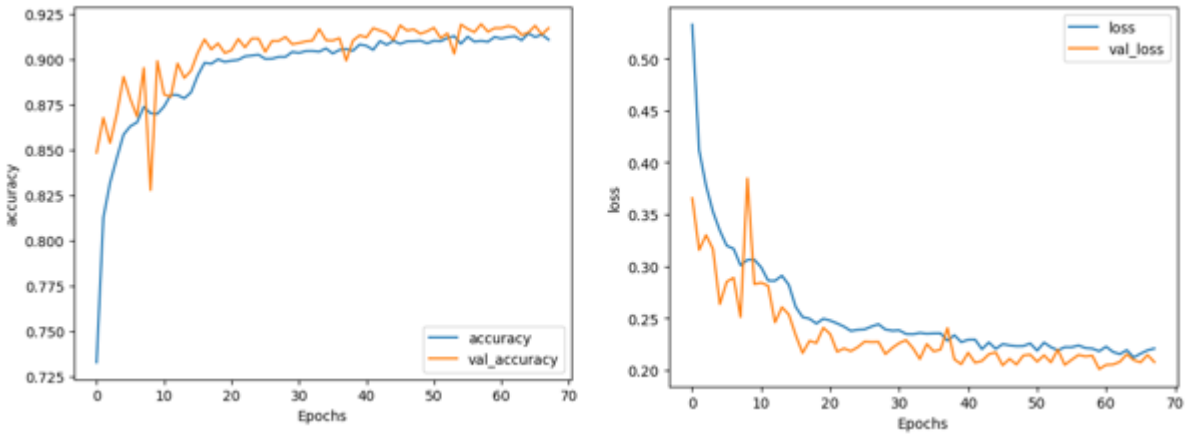
Tabela 7- Matriz de Confusão obtida com a VGG-19

	Real "DME"	**Real "Normal"**
Predito "DME"	244 (VP)	7 (FP)
Predito "Normal"	6 (FN)	243 (VN)

Fonte: Autoria própria (2024).

Conforme a Tabela 7, a rede VGG-19 mostra também um equilíbrio em sua matriz de confusão, com altos valores de VP e VN e baixos FP e FN. Este modelo alcançou uma Acurácia de 97,40%, refletindo uma eficácia geral elevada na classificação. A Precisão de 97,21% sugere que a maioria das previsões positivas são corretas, enquanto o Recall de 97,60% indica uma alta capacidade de detectar positivos reais. Com um *F-Measure* de 97,41% e Especificidade de 97,20%. A VGG-19 destaca-se pelo seu desempenho equilibrado e confiável na classificação de ambas as classes.

Gráfico 5 - Dados de treinamento da rede VGG-19



(a) **(b)**

Fonte: Autoria própria (2024).

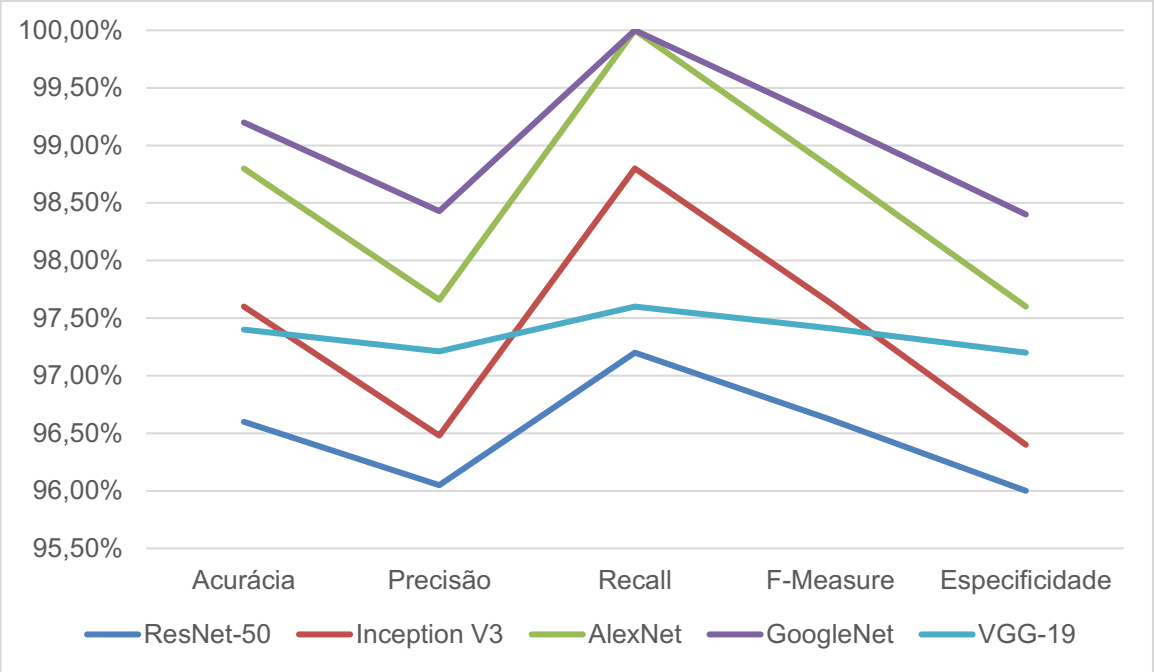
O Gráfico 5(a) destaca o desempenho do modelo em termos de precisão e precisão de validação ao longo de 68 épocas. A acurácia segue uma tendência ascendente inicial, estabilizando-se acima de 87,5%. A acurácia de validação espelha essa progressão, com uma ligeira variação até a 10º época.

No Gráfico 5(b), as curvas de perda de treino e de validação são apresentadas. Ambas iniciam com valores mais altos, com a perda de treinamento caindo rapidamente e a perda de validação apresentando uma queda substancial antes de se estabilizar.

4.6 Resultados gerais

O conjunto de métricas de desempenho aplicadas nos resultados obtidos foram os especificados na seção 2.7, ou seja, Acurácia, Precisão, *Recall*, *F-measure* e especificidade. Como já abordado, todas as métricas são calculadas através da manipulação dos verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN). Levando em consideração estas métricas e as matrizes de confusão, obteve-se o Gráfico 6 com os resultados de desempenho das redes:

Gráfico 6 - Comparação das métricas obtidas por cada modelo



Fonte: Autoria própria (2024).

Observa-se no Gráfico 6 que todos os modelos performaram de forma similar com uma tendência de melhor *recall* e pior precisão. Os dados são apresentados detalhadamente em formato de Tabela na Tabela 8 ordenados pelos modelos com melhores resultados.

Tabela 8 - Resultado das métricas de desempenho em cada modelo

	Acurácia	**Precisão**	***Recall***	***F-Measure***	**Especificidade**
GoogLeNet	99,20%	98,43%	100,00%	99,21%	98,40%
AlexNet	98,80%	97,66%	100,00%	98,81%	97,60%
***Inception* V3**	97,60%	96,48%	98,80%	97,63%	96,40%
VGG-19	97,40%	97,21%	97,60%	97,41%	97,20%
ResNet-50	96,60%	96,05%	97,20%	96,62%	96,00%

Fonte: Autoria própria (2024).

Conforme mostra a Tabela 8, todas as métricas estão compreendidas entre 96 e 100%. Em verde são destacados os melhores valores de cada métrica, sendo todos eles pertencentes a rede GoogLeNet, juntamente com o *Recall* da AlexNet. Na Tabela 9 são apresentados os tempos que cada rede demorou para realizar o treinamento, bem como as épocas.

Tabela 9 - Quantidade de tempo e de épocas para cada treinamento

	Tempo de treinamento (em horas)	Número total de épocas	Média de tempo por época (em segundos)
ResNet-50	01:20:49	34	143
***Inception* V3**	02:36:24	63	149
AlexNet	01:49:13	47	139
GoogLeNet	02:05:17	50	150
VGG-19	02:43:33	68	144

Fonte: Autoria própria (2024).

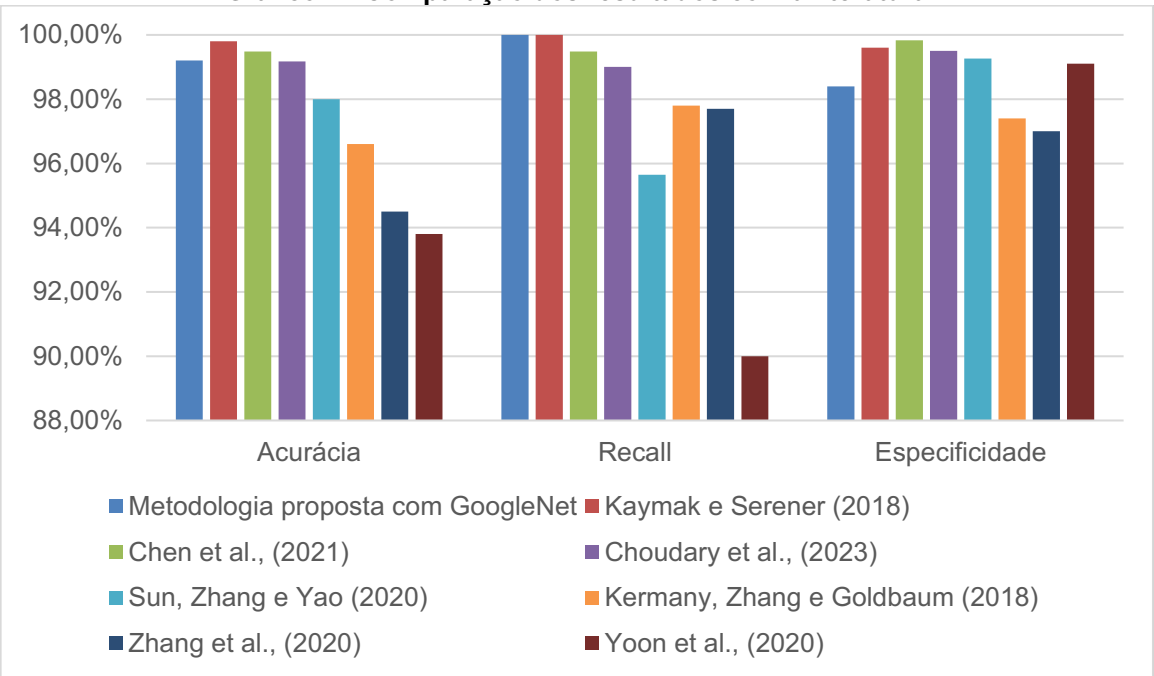
Observa-se na Tabela 9 os tempos de treinamento em horas, o número total de épocas e a média de tempo por época em segundos.

5 ANÁLISE DE RESULTADOS

Inicialmente foi realizado um primeiro teste de treinamento e classificação com a rede ResNet-50, sem divisão de imagens para validação e sem a aplicação de técnicas de processamento. Após realizados os cálculos de métricas de desempenho, o modelo em questão obteve um valor médio de 96,47% nas métricas aplicadas.

Porém com a aplicação do filtro gaussiano, todas as redes se comportaram melhor na classificação, com destaque para a GoogLeNet e a AlexNet que atingiram 100% em algumas métricas. Os resultados obtidos com a GoogLeNet foram muito similares aos encontrados na literatura (Gráfico 7).

Gráfico 7 - Comparação dos resultados com a literatura



Fonte: Autoria própria (2024).

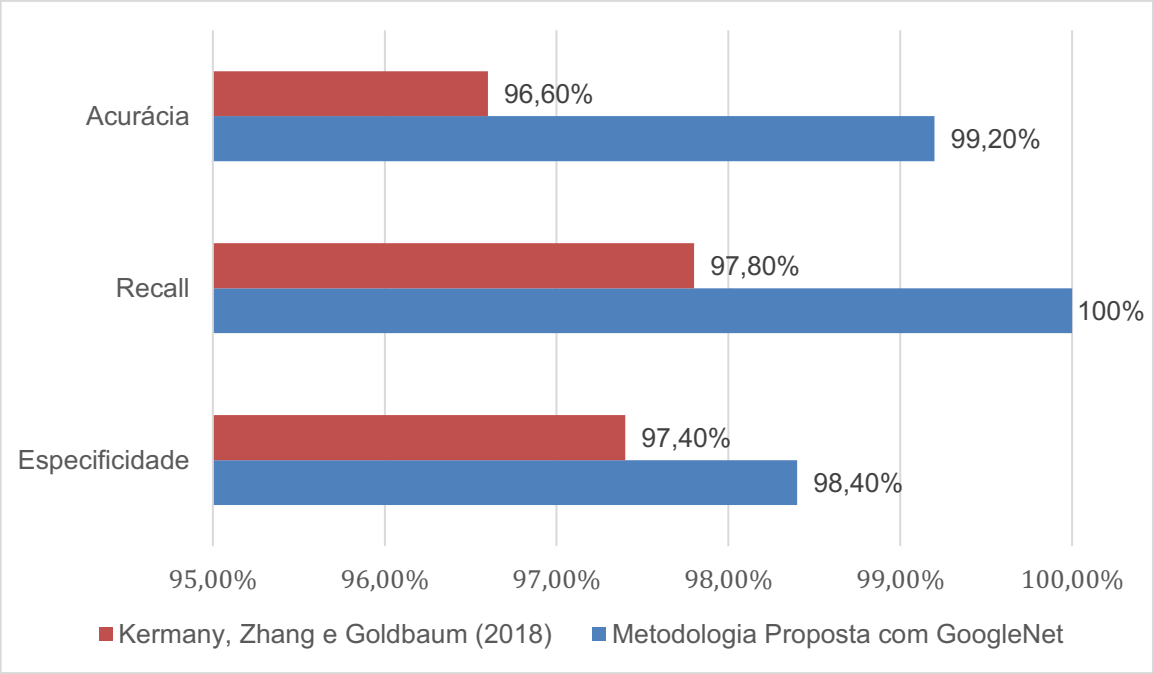
No Gráfico 7 pode-se observar que em alguns trabalhos como os de Chen *et al.,* (2021) e Kaymak e Serener (2018) apresentaram um resultado superior em algumas métricas. É notório o resultado de Kaymak e Serener (2018) que quando comparados com os resultados desta metodologia, se mostra melhor em acurácia e especificidade.

Ainda no Gráfico 7 observa-se que a metodologia proposta com a GoogLeNet se sobressai significativamente. A acurácia desta abordagem chega à marca de 99,20%, indicando uma grande eficiência na classificação correta das imagens. Esta

superioridade é também refletida no *recall* de 100%, uma métrica que mede a capacidade da rede em identificar verdadeiros positivos.

Por outro lado, o estudo de Kermany, Zhang e Goldbaum (2018) exibe uma acurácia de 96,60%. Na métrica de *recall* a abordagem dos autores também obtém resultados inferiores com relação à metodologia deste projeto, com 97,80%. O Gráfico 8 expõe um comparativo entre os resultados da metodologia proposta com o uso da rede GoogLeNet com os resultados do trabalho original de Kermany, Zhang e Goldbaum (2018).

Gráfico 8 - Comparação entre os resultados obtidos com a GoogLeNet e o trabalho original de Kermany, Zhang e Goldbaum (2018)



Fonte: Autoria própria (2024).

Conforme mostrado no Gráfico 8 os resultados obtidos neste projeto são melhores do que os de Kermany, Zhang e Goldbaum (2018) em todas as métricas divulgadas por ele. De qualquer forma, um resultado preliminar acima de 95% é um bom indicador da capacidade que estas técnicas têm de fornecer um resultado muito conciso sobre um paciente ter ou não edema macular diabético. A Tabela 10 apresenta os resultados obtidos pelas redes em formato de porcentagem de acertos.

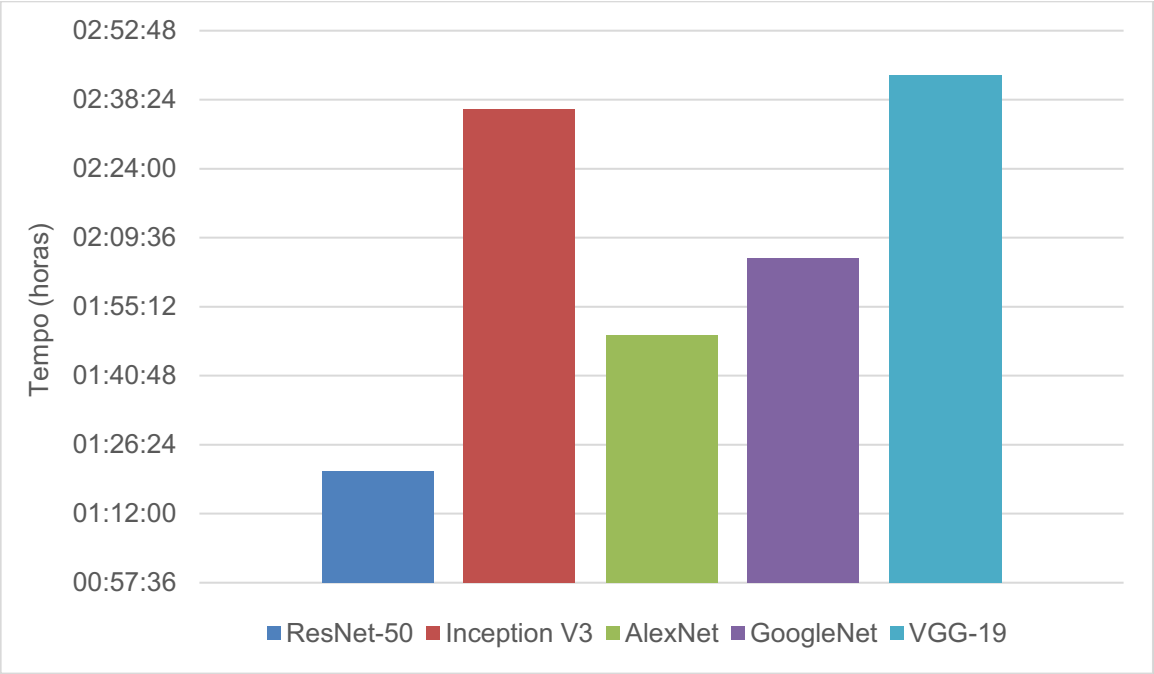
Tabela 10 - Comparação de acertos dos modelos em porcentagem

	ResNet-50	Inception V3	AlexNet	GoogLeNet	VGG-19
Predito Normal corretamente	96,00%	98,77%	100%	98,40%	97,20%
Predito DME corretamente	97,20%	96,48%	97,66%	100%	97,60%

Fonte: Autoria própria (2024).

A Tabela 10 mostra com qual porcentagem cada modelo acertou as predições de cada classe, e se torna mais clara o quanto as predições do sistema foram próximas de 100%, se assemelhando aos melhores trabalhos presentes na literatura. O Gráfico 9 apresenta quanto tempo cada rede levou para realizar o treinamento.

Gráfico 9 - Tempo para treinamento de cada rede



Fonte: Autoria própria (2024).

Conforme o Gráfico 9, observa-se que a rede ResNet-50 foi a que realizou o treinamento em menos tempo, completando a tarefa em 1 hora, 20 minutos e 49 segundos. Já a rede VGG-19 foi a que dispendeu mais tempo, levando 2 horas, 43 minutos e 33 segundos. A rede que obteve os melhores resultados, GoogLeNet, concluiu o treinamento em 2 horas, 5 minutos e 17 segundos.

Vale destacar a rede AlexNet que obteve o menor tempo de treinamento por época (Tabela 9), tendo realizado a tarefa em 139 segundos por época, enquanto a GoogLeNet foi a que mais levou tempo por época, 150 segundos.

6 CONCLUSÃO

A proposta do trabalho era avaliar a utilização de rede neural convolucional, com transferência de aprendizado na classificação de edema macular diabético em imagens de OCT. Os resultados obtidos foram promissores e superaram em várias métricas de desempenho alguns trabalhos presentes na literatura.

Foi selecionado a base de dados de Kermany, Zhang e Goldbaum (2018) que conta com um número significativo de imagens de retinas com DME (11.348). Foi utilizado a técnica de aumento de dados para contribuir com a capacidade de generalização das redes, assim como o filtro gaussiano que contribuiu para uma melhora no aspecto da imagem.

Foram realizados testes em cinco redes neurais pré-treinadas, a ResNet-50, *Inception* V3, AlexNet, GoogLeNet e VGG-19. Todas as redes foram treinadas e testadas nas imagens da base de teste disponível no *dataset* original, gerando dados de resultado e métricas de desempenho que permitiram a comparação entre esta metodologia e os trabalhos similares da literatura.

Observou-se que todas as redes demonstraram um desempenho significativo acima de 96% em todas as métricas. Convém citar em destaque a GoogLeNet, que obteve o menor número de predições falsas, e a AlexNet, que atingiram o valor de 100% na métrica de *recall* indicando que o modelo identificou todos os exames com DME corretamente. Ambas as redes se comportaram muito bem para o problema de identificação de edema macular, tendo acertado grande maioria de suas predições de forma similar ao que é abordado pela literatura.

A rede GoogLeNet foi a que obteve os melhores resultados, registrando uma acurácia de 99,20%, uma precisão de 98,43%, um *recall* de 100%, um *F-measure* de 99,21% e uma especificidade de 98,40%. Vale considerar que a rede GoogLeNet levou cerca de 2 horas e 5 minutos para realizar o treinamento.

É importante citar que os autores do *dataset* utilizado, Kermany, Zhang e Goldbaum (2018), utilizaram a rede *Inception* V3 em seu experimento e obtiveram um resultado de 96,60% em acurácia, 97,80% em *recall* e 97,40% em especificidade. Levando em consideração os resultados da GoogLeNet percebe-se que esta metodologia obteve 2,69% de melhoria em acurácia com relação ao trabalho original de Kermany, Zhang e Goldbaum (2018). Houve melhoria também com relação ao

recall na ordem de 2,25%, enquanto a especificidade foi 1,03% melhor neste projeto em relação ao trabalho de Kermany, Zhang e Goldbaum (2018).

Em conclusão, este trabalho cumpriu todos os seus objetivos atingindo resultados significativos que contribuem para a literatura do tema de classificação de imagens através de redes neurais. Estes resultados servem como ponto de partida para trabalhos futuros visando aprimorar ainda mais os resultados obtidos.

6.1 Trabalhos futuros

Há trabalhos na literatura nos quais os autores obtiveram resultados melhores do que os desta proposta em algumas métricas, o que indica que há margem para melhorias e trabalhos futuros. Como exemplo convém citar o trabalho de Kaymak e Serener (2018) no qual os autores obtiveram uma precisão 0,60% melhor e uma especificidade 1,20% acima da desta proposta.

Para trabalhos futuros pode-se explorar a realização de alguns outros experimentos como:

- A aplicação diferente formas de pré-processamentos afim de conseguir eliminar a pequena discrepância de predições incorretas;
- Utilização de outros tipos de otimizadores aliados à outras formas de pré-processamento tais como: *Adagrad*, *Adamax*, *Adafactor* e *Nadam*;
- Teste com um número maior do que as 250 imagens para cada classe previamente selecionadas no *dataset,* pois neste trabalho foi mantido as mesmas quantidades de imagens para teste do *dataset* original de Kermany, Zhang e Goldbaum (2018) de modo a conseguir comparar os resultados do método proposto com a literatura de forma direta;
- Aplicar a metodologia com outras redes neurais pré-treinadas que não foram consideradas neste trabalho.

REFERÊNCIAS

AAO (American Academy of Ophthalmology). **Macular edema diagnosis.** 2020. Disponível em: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy. Acesso em: 18 jan. 2023.

AGRIZZI, M. D. **Retinopatia diabética**. 2020. Disponível em: https://drmiltonagrizzi.com/retinopatia-diabetica/. Acesso em: 18 jan. 2023.

ALAMSYAH, A.; SAPUTRA, M. A. A.; MASRURY, R. A. J. Object detection using convolutional neural network to identify popular fashion product. **Journal of Physics**. 1192 012040, 2019. DOI: https://doi.org/10.1088/1742-6596/1192/1/012040.

ALDINGTON, S. J. *et al.* Methodology for retinal photography and assessment of diabetic retinopathy: the EURODIAB IDDM complications study. **Diabetologia**, v. 38, p. 437–444, 1995. DOI: https://doi.org/10.1007/BF00410281.

ALSAIH, K. *et al.* Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. **Biomedical Engineering Online,** vol. 16,1 68, 2017. DOI: https://doi.org/10.1186/s12938-017-0352-9.

ALUGUBELLI, R. Exploratory study of artificial intelligence in healthcare**.** **International Journal of Innovations in Engineering Research and Technology**, v. 3, n. 1, p. 1-10, 2016.

BANDELLO, F. *et al.* **Clinical strategies in the management of diabetic retinopathy**. A step-by-step guide for ophthalmologists. 1. ed. Berlin: Springer-Verlag, 2014.

BEAGLEY, J. *et al.* Global estimates of undiagnosed diabetes in adults. **Diabetes Research and Clinical Practice**. v. 103, n. 2, p. 150-160. PMID: 24300018, fev. 2014. DOI: https://doi.org/10.1016/j.diabres.2013.11.001.

BOELTER, M. C. *et al.* Fatores de risco para retinopatia diabética. **Arquivos Brasileiros de Oftalmologia**, v. 66, n. 2, 2003. DOI: https://doi.org/10.1590/S0004-27492003000200024.

BOHR, A.; MEMARZADEH, K. The rise of artificial intelligence in healthcare applications. *In:* **Artificial Intelligence in Healthcare**, p. 25-60, 2020. DOI: https://doi.org/10.1016/B978-0-12-818438-7.00002-2.

BURRELL, J. How the machine ‘thinks’: understanding opacity in machine learning algorithms. **Big Data & Society**, v. 3, n. 1, 2016. DOI: https://doi.org/10.1177/2053951715622512.

BOSCO, A. *et al.* Retinopatia diabética. **Arquivos Brasileiros de Endocrinologia & Metabologia**, v. 49, n. 2, p. 217-227, 2005. DOI: https://doi.org/10.1590/S0004-27302005000200007.

CANZIANI, A.; PASZKE, A.; CULURCIELLO, E. **An analysis of deep neural network models for practical applications**. arXiv:1605.07678 [cs.CV], 2017.

CHEN, S. C. *et al.* A novel machine learning algorithm to automatically predict visual outcomes in intravitreal ranibizumab-treated patients with diabetic macular edema. **Journal of Clinical Medicine**, v. 7, n. 12, p. 475, 2018.

CHEN, Y. M. *et al.* Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning. **BMC Bioinformatics,** v. 22, p. 99, 2021. DOI: https://doi.org/10.1186/s12859-021-04001-1.

CHOUDHARY, A. *et al.* A deep learning-based framework for retinal disease classification. **Healthcare**, v. 11, n. 2, Jan. 2023. DOI: https://doi.org/10.3390/healthcare11020212.

CIULLA, T. A.; AMADOR, A. G.; ZINMAN, B. Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies. **Diabetes Care.** v. 26, n. 9, p. 2653-2664, 2003. DOI: https://doi.org/10.2337/diacare.26.9.2653.

COGSWELL, M.; AHMED F.; GIRSHICK, R. **Reducing overfitting in deep networks by decorrelating representations.** ICLR 2016, arXiv:1511.06068v4 [cs.LG], 10 Jun 2016.

COSCAS, G.; CUNHA-VAZ, J.; SOUBRANE, G. Macular edema: definition and basic concepts. **Developments in Ophthalmology**. v. 47, p. 1-9, 2010. DOI: https://doi.org/10.1159/000320070.

COSTA, D. S. **Detecção automática de edema macular diabético e de degeneração macular relacionada à idade em imagens de tomografia de coerência óptica.** (dissertação). São Luís, MA, 2017.

DENG, G.; CAHILL, L. W. An adaptive gaussian filter for noise reduction and edge detection. *In:* **Nuclear Science Symposium and Medical Imaging Conference**. 1993 IEEE Conference Record, 1993. DOI: https://doi.org/10.1109/NSSMIC.2018.8824723.

DENG, J. *et al.* **Imagenet: a large-scale hierarchical image database**. CVPR09, 2009.

DORCHY, H. Characterization of early stages of diabetic retinopathy. **Diabetes Care**. v. 16, n. 8, p. 1212-1214, 1993. DOI: https://doi.org/10.2337/diacare.16.8.1212.

DUNJKO, V.; BRIEGEL, H. J. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. **Reports on Progress in Physics.** v. 81, n. 7, p. 074001, 2018.

ESTEVES, J. **Complicações oculares nos pacientes com diabetes mellitus tipo 1**. dissertação (mestrado) – Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, 2009.

FAWAZ, H. I. *et al.* **Transfer learning for time series classification**. IRIMAS, Universite de HauteAlsace, Mulhouse, France, arXiv:1811.01533v1 [cs.LG], 2018.

FAWAZ, H. I. *et al.* Deep learning for time series classification: a review. **Data Mining and Knowledge Discovery**. v. 33, p. 917–963, 2019. DOI: https://doi.org/10.1007/s10618-019-00619-1.

FIORIN, D. *et al.* Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física.** v. 33, p. 1309, 2011. DOI: https://doi.org/10.1590/S1806-11172011000100009.

FISHER, R. *et al.* **Gaussian smoothing.** 2003. Disponível em: https://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm. Acesso em: 5 fev. 2023.

FONG, D. S. *et al.* Retinopathy in Diabetes. **Diabetes Care**. American Diabetes Association. v. 27, Suppl. 1, p. S84-7, jan. 2004. DOI: https://doi.org/10.2337/diacare.27.2007.S84.

GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais**. Editora Pearson, ISBN 9788576054016, 3. ed., 2008.

GUARIGUATA, L. *et al.* Global estimates of diabetes prevalence for 2013 and projections for 2035. **Diabetes Research and Clinical Practice**.103(2):137-49, Epub 2013 Dec 1. PMID: 24630390, 2014. DOI: https://doi.org/10.1016/j.diabres.2013.11.002.

HASSAN, B.; RAJA, G. Fully automated assessment of macular edema using optical coherence tomography (oct) images. **International Conference on Intelligent Systems Engineering (ICISE),** p. 5-9, 2016.

HASSAN, H. M. *et al.* Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes: case study el burullus lake. **International Water Technology Journal**, v. 5, 2015.

Health IT Analytics. **Health IT Analytics News**. Xtelligent HealthCare Media. 2017.

HE, K. *et al.* Deep residual learning for image recognition. *In:* **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. p. 770-778, 2016.

HELENE, O; HELENE, A. F. Alguns aspectos da óptica do olho humano. **Revista Brasileira de Ensino de Física**. São Paulo (SP), v. 33, n. 3, p. 3312, 2011.

HINTON, G.; SRIVASTAVA, N.; SWERSKY, K. **Neural networks for machine learning.** lecture 6a overview of mini-batch gradient descent, v. 14, n. 8, p. 2, 2012.

HUANG, G. *et al.* Convolutional networks with dense connectivity. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2019.

IDF Diabetes Atlas. **International Diabetes Federation**, 10th edition committee, 2021. Disponível em: https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf. Acesso em 10 jan. 2023.

ImageNet. Stanford Vision Lab, Stanford University, Princeton University. 2021. Disponível em: https://image-net.org/about.php. Acesso em: 02 fev. 2023.

IVERSON, S. M. CLARK, W. L. Update on the management of diabetic macular edema. **US Ophthalmic Review.** v. 10, n. 01, p. 52, 2017.

JESUS, E. O.; COSTA JUNIOR, R. A utilização de filtros gaussianos na análise de imagens digitais. **Proceeding Series of the Brazilian Society of Applied and Computational Mathematics**. v. 3, n. 1, 2015.

JOHNSON, J.; ALAHI, A.; LI, F. F. **Perceptual losses for real-time style transfer and super-resolution**. arXiv:1603.08155 [cs.CV], 2016. DOI: https://doi.org/10.48550/arXiv.1603.08155.

KAYMAK, S.; SERENER, A. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. *In:* **IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP).** Cluj-Napoca, Romania, p. 265-269, 2018. DOI: https://doi.org/10.1109/ICCP.2018.8516635.

Keras. **API reference.** Disponível em: https://keras.io/api/. Acesso em: 04 abril 2023.

KERMANY, D.; ZHANG, K.; GOLDBAUM, M. **Labeled optical coherence tomography (OCT) and chest x-ray images for classification**. Mendeley data, V3, 2018. DOI: https://doi.org/10.17632/rscbjbr9sj.3.

KHAN, S. **Contour integration in artifical neural networks**. University of Waterloo, Ontario, Canada, 2022.

KINGMA, D. P.; BA, J. **Adam:** a method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2017.

KRITTANAWONG, C. *et al.* Artificial intelligence in precision cardiovascular medicine. **Journal of the American College of Cardiology**. v. 69, n. 21, p. 2657-2664, PMID: 28545640, 2017. DOI: https://doi.org/10.1016/j.jacc.2017.03.571.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. **Communications of the ACM**. New York, NY, USA: association for computing machinery, v. 60, n. 6, p. 84–90, 2017. DOI: https://doi.org/10.1145/3065386.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature.** v. 521, p. 436–444, 2015. DOI: https://doi.org/10.1038/nature14539.

LI, S. *et al.* Image classification algorithm based on improved AlexNet. **Journal of Physics**: Conference Series, v. 1813, p. 012051, 2021. DOI: https://doi.org/10.1088/1742-6596/1813/1/012051.

LI, Z. *et al.* A survey of convolutional neural networks: analysis, applications, and prospects. *In:* **IEEE transactions on neural networks and learning systems**, v. 33, n. 12, p. 6999-7019, 2022.

LITJENS, G. *et al.* A survey on deep learning in medical image analysis. **Medical Image Analysis**. v. 42, p. 60-88, PMID: 28778026, 2017. DOI: https://doi.org/10.1016/j.media.2017.07.005.

LYAPUNOV, S. I. Visual acuity and contrast sensitivity of the human visual system. **Journal of Optical Technology**, v. 84, n. 9, p. 613, 2017. DOI: https://doi.org/10.1364/jot.84.000613.

MARASHI, A. Clinical pathology of diabetic retinopathy and macular edema. **Open Access Journal of Ophthalmology**. v. 1, n. 1, 2016. DOI: https://doi.org/10.23880/oajo-16000105.

MARQUES, C. M. G. **Confiabilidade metrológica da tomografia por coerência óptica em aplicações biomédicas**. tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, 2012.

MASHTAKOV, A. P. Sub-riemannian geometry in image processing and modeling of the human visual system. **Nonlinear Dynamics.** v. 15, n. 4, p. 561-568, 2019. DOI: https://doi.org/10.20537/nd190415.

MATHEW, C.; YUNIRAKASIWI, A., SANJAY, S. Updates in the management of diabetic macular edema. **Journal of Diabetes Research**. PMID: 25984537; PMCID: PMC4423013, 2015. DOI: https://doi.org/10.1155/2015/794036.

MENDES, O. L. C. *et al.* Automatic segmentation of macular holes in optical coherence tomography images: a review. **Journal of artificial intelligence and systems**, v. 1, n. 1, p. 163-185, 2020.

MICHELSON, A. A. **Studies in Optics**. 1 ed. Mineola, NY: Editora Dover, 1995.

MOHAMED, Q.; GILLIES, M. C.; WONG, T. Y. Management of diabetic retinopathy: a systematic review. **JAMA.** v. 298, n. 8, p. 902–916, 2007. DOI: https://dx.doi.org/10.1001/jama.298.8.902.

MOTTA M. M. S.; COBLENTZ J.; MELO L. G. N. Aspectos atuais na fisiopatologia do edema macular diabético. **Revista Brasileira de Oftalmologia**. v. 67, n. 1, p. 45-49, 2008.

Momento Diabetes. **Você sabe o que é edema macular diabético e como prevenir?,** 2020. Disponível em: https://www.momentodiabetes.com.br/voce-sabe-o-que-e-edema-macular-diabetico-e-como-prevenir/. Acesso em: 05 fev. 2023.

MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. **Array**. v. 16, 100258, 2022.

NEI (National Eye Institute). **Diabetic retinopathy**. At a glance: diabetic retinopathy. jul. 2021. Disponível em: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy. Acesso em: 18 jan. 2023.

NEI (National Eye Institute). **Macular edema**. At a glance: macular edema. jul. 2019. Disponível em: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/macular-edema. Acesso em: 18 jan. 2023.

OLIVEIRA, T. P.; BARBAR, J. S.; SOARES, A. S. Previsão de tráfego de rede de computadores: uma comparação entre redes neurais tradicionais e de aprendizado profundo. **International Journal of Big Data Intelligence**. v. 3, n. 1, p. 28-37, 2019.

OREN, O.; GERSH, B. J.; BHATT, D. L. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. **The Lancet Digital Health**. v. 2, n. 9, p. e486-e488, 2020.

PONRAJ, A. **Artificial neural network explained with a regression example.** 22 nov. 2020. Disponível em: https://devskrol.com/2020/11/22/388/. Acesso em: 25 maio 2023.

PRAHS, P. *et al.* OCT-based deep learning algorithm for the evaluation of treatment indication with anti-vascular endothelial growth factor medications. **Graefe's Archive for Clinical and Experimental Ophthalmology**. v. 256, n. 1, p. 91–98, 2017. DOI: https://doi.org/10.1007/s00417-017-3839-y.

PRATT, W. K. **Digital image processing**. John Wiley & Sons, 1991.

PRECHELT, L. Early stopping — but when?. *In:* **Neural networks: tricks of the trade**. 1. ed. Elsevier, Berlin, Heidelberg, 1998.

RODRIGUEZ, F. J. J. *et al.* Neural network for detection of diabetic macular edema in fundus color images. **Investigative Ophthalmology & Visual Science**. v. 58, n. 8, p. 688-688, 2017.

ROSENBLATT, F. **The perceptron - a perceiving and recognizing automaton.** Cornell Aeronautical Laboratory, Inc, 1957.

RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. **Nature.** v. 323, p. 533–536, 1986. DOI: https://doi.org/10.1038/323533a0.

RUSSAKOVSKY, O. *et al.* ImageNet large scale visual recognition challenge. **International Journal of Computer Vision**. v. 115, n. 3, p. 211–252, 2015.

SANTOS, V. S. **Estrutura interna dos olhos**. Rede educação, rede Omnia, 2019.

SCHEEN, A. J. Exciting breakthroughs in the management of Diabetes mellitus. **Diabetes Epidemiology and Management.** v. 1, p. 100005. Elsevier BV, 2021. DOI: https://doi.org/10.1016/j.deman.2021.100005.

SCHEFFEL, R. S. *et al.* Prevalência de complicações micro e macrovasculares e de seus fatores de risco em pacientes com diabetes melito do tipo 2 em atendimento ambulatorial. **Revista da Associação Médica Brasileira.** v. 50, n. 3, p. 263-267, 2004. DOI: https://doi.org/10.1590/S0104-42302004000300031.

SCHMITT, J. M.; KNÜTTEL, A.; BONNER, R. F. Measurement of optical properties of biological tissues by low-coherence reflectometry. **Applied Optics**. v. 32, n. 30, p. 6032-6042, 1993. DOI: https://doi.org/10.1364/AO.32.006032.

SHORTEN, C., KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**. v. 6, p. 60, 2019. DOI: https://doi.org/10.1186/s40537-019-0197-0.

SILVA, E. A. B.; NETTO, S. L. **Processamento digital de imagens** – análise de imagens. Laboratório de sinais, multimídia e telecomunicações UFRJ, 2017. Disponível em: https://docplayer.com.br/52411054-Processamento-digital-de-imagens-analise-de-imagens.html. Acesso em: 01 fev. 2023.

SILVA, M. **Retina.** Infoescola 2018. Disponível em: https://www.infoescola.com/visao/retina/. Acesso em: 01 fev. 2023.

SIMONYAN, K.; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition**. arXiv preprint arXiv:1409.1556, 2015.

SINGH, S.; KISHORE, A. Natural language image descriptor. In: **Proceedings of the RAICS**. 2015. DOI: https://doi.org/10.1109/RAICS.2015.7488398.

SOLÓRZANO, C. A. P. *et al.* Findings from machine learning in clinical medical imaging applications–lessons for translation to the forensic setting. **Forensic Science International**. 2020.

SOUZA, C. **Análise de poder discriminativo através de curvas roc.** 2009. Disponível em: http://crsouza.com/2009/07/13/analise-de-poder-discriminativo-atraves-de-curvas-roc/. Acesso em: 28 fev. 2022.

SUN, Y.; ZHANG, H.; YAO, X. Automatic diagnosis of macular diseases from oct volume based on its two-dimensional feature map and convolutional neural network with attention mechanism. **Journal of Biomedical Optics**. v. 25, n. 9, PMID: 32940026; PMCID: PMC7493033, 2020. DOI: https://doi.org/10.1117/1.JBO.25.9.096004.

SUÑER, I. J. *et al.* Responsiveness of the national eye institute visual function questionnaire-25 to visual acuity gains in patients with diabetic macular edema: evidence from the ride and rise trials. **The Journal of Retinal and Vitreous Diseases**. v. 37,6 p. 1126-1133, 2017. DOI: https://www.doi.org/10.1097/IAE.0000000000001316.

SZEGEDY, C. *et al.* **Going deeper with convolutions**. arXiv:1409.4842 [cs.CV], 2014.

SZEGEDY, C. *et al.* Going deeper with convolutions. *In:* **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. p. 1-9, 2015.

SZEGEDY, C. *et al.* Rethinking the inception architecture for computer vision. *In:* **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. p. 2818-2826, 2016.

TARR, J. M. *et al.* Pathophysiology of diabetic retinopathy. **ISRN Ophthalmol**. PMCID: PMC3914226, 2013. DOI: https://doi.org/10.1155/2013/343560.

TING, D. S. W. *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. **JAMA**. v. 318, n. 22, p. 2211-2223, 2017. DOI: https://doi.org/10.1001/jama.2017.18152.

VERHULST, P. F. **Recherches mathématiques sur la loi d'accroissement de la population**. nouveaux mémoires de l'Académie Royale Des Sciences et Belles-Lettres de Bruxelles, v. 18, p. 1-38, 1845.

VIEIRA, A. **Edema macular diabético: sintomas, tratamentos e causas**. Sociedade Brasileira de Oftalmologia, Sociedade Brasileira de Endocrinologia, diretrizes brasileiras de diabetes (SBD), 2013.

WANG, Y. **Convolutional neural network based malignancy detection of pulmonary nodule on computer tomography.** (doctoral dissertation, University of Saskatchewan), 2018.

WANG, Z. *et al.* Artificial intelligence and deep learning in ophthalmology. **Intelligence in Medicine.** p. 1-34, 2020.

WU, J. **Introduction to convolutional neural networks.** National key lab for novel software technology, Nanjing University, China, 2017.

XIE, S. *et al.* Aggregated residual transformations for deep neural networks. In: **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu, HI, EUA, p. 5987-5995, 2017. DOI: https://doi.org/10.1109/CVPR.2017.634.

YAMASHITA, R. *et al.* Convolutional neural networks: an overview and application in radiology. **Insights Into Imaging**. v. 9, n. 4, p. 611-629, 2018. DOI: https://doi.org/10.1007/s13244-018-0639-9.

YOON, J. *et al.* Optical coherence tomography-based deep-learning model for detecting central serous chorioretinopathy. **Scientific Reports.** v. 10, p. 18852, 2020. DOI: https://doi.org/10.1038/s41598-020-75816-w.

ZHANG, J. *et al.* Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. **Artificial Intelligence Review**. v. 56, n. 2, p. 1013-1070, 2014. DOI: https://doi.org/10.1007/s10462-022-10192-7.

ZHANG, Q. *et al.* Identifying diabetic macular edema and other retinal diseases by optical coherence tomography image and multiscale deep learning. **Diabetes, Metabolic Syndrome and Obesity**. v. 13 p. 4787-4800, 2020. DOI: https://doi.org/10.2147/dmso.s288419.