

Segmentação de exsudatos em imagens de retina utilizando redes neurais convolucionais

Alexandre M. Bernardo; Evandro L. L. Rodrigues.

*Escola de Engenharia Elétrica, Universidade de São Paulo - EESC
USP, (e-mail: alexandre.bernardo@usp.br)
Escola de Engenharia Elétrica, Universidade de São Paulo - EESC
USP, (e-mail: evandro@sc.usp.br)*

Abstract: Diabetic retinopathy is the main cause of preventable blindness in people of working age. About one in three people who have diabetes mellitus presents some degree of Diabetic Retinopathy and it is estimated that by the year 2045 this disease will have a 48% increase. Expediting, facilitating, and standardizing diagnosis becomes the target of studies to reduce this growth. One of the indicators of diabetic retinopathy is the presence of exudates in retinal images. This work proposes the application of retinal image processing for the automatic segmentation of exudates using a convolutional neural networks. The networks were trained using the IDRiD (*Indian Diabetic Retinopathy Image Dataset*), with retinal images evaluated by specialists who ensured adequate characteristics for the use of the proposed techniques. Using this database, two deep learning architectures were proposed, trained and evaluated. To enable the use of the technique, a subdivision of the original images into smaller was proposed and implemented. In this configuration, the MobileNet network presented a sensitivity of 96.2%, specificity of 72.4%, accuracy of 71.7%. The Unet network presented 44.5% sensitivity, 91.5% specificity, 90.9% accuracy. In this work, descriptions of functioning, network topologies, learning techniques used and comparative results between the architectures are presented.

Resumo: A retinopatia diabética é a principal causa de cegueira evitável na população em idade ativa no mundo. Aproximadamente uma em cada três pessoas portadoras do diabetes mellitus apresenta algum grau de retinopatia diabética e estima-se que até 2045 haja um crescimento de 48% nos casos dessa doença. Agilizar, facilitar e padronizar o diagnóstico do exame de retina torna-se alvo de estudos para reduzir este crescimento. Um dos indicadores da retinopatia diabética é a presença de exsudatos em imagens de retina, desta forma, este trabalho propõe a aplicação de processamento de imagens de retina para a segmentação automática de exsudatos utilizando redes neurais convolucionais. Para treinamento das redes, foi utilizada a base de dados IDRiD (*Indian Diabetic Retinopathy Image Dataset*), com imagens de retinas avaliadas por especialistas que garantiram características adequadas ao uso das técnicas propostas. Neste trabalho, foram propostas, treinadas e avaliadas duas arquiteturas de redes neurais profundas com as referidas imagens. Para viabilizar o uso da técnica, foi proposto e implementado a subdivisão das imagens originais em subimagens menores. Neste configuração, a rede MobileNet apresentou sensibilidade de 96,2%, especificidade de 72,4%, acurácia de 71,7%. A rede Unet apresentou sensibilidade 44,5%, especificidade 91,5%, acurácia de 90,9%. Nesse trabalho, são apresentadas as descrições de funcionamento, topologias das redes, técnicas de aprendizado utilizadas e resultados comparativos entre as arquiteturas.

Keywords: Diabetic Retinopathy; Image Processing; Convolutional Neural Network; Exudates; Segmentation;

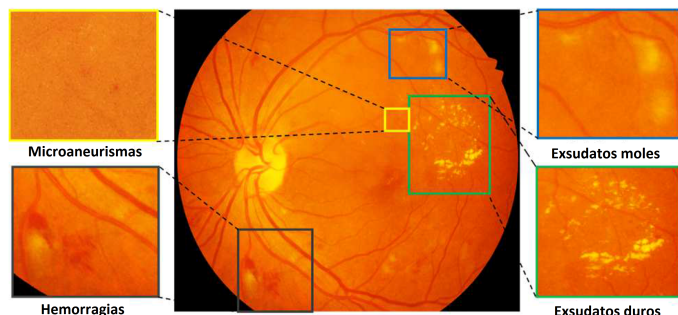
Palavras-chaves: Retinopatia Diabética; Processamento de imagens; Rede Neural Convolutacional; Exsudatos; Segmentação;

1. INTRODUÇÃO

A retinopatia diabética (RD) é a principal causa da perda de visão em adultos em idade ativa (20 a 74 anos) (Cheung et al. (2010)). Aproximadamente uma em cada três pessoas portadoras do diabetes mellitus (DM) tem algum grau de retinopatia diabética e uma em cada 10 perderá a visão (Yau et al. (2012)).

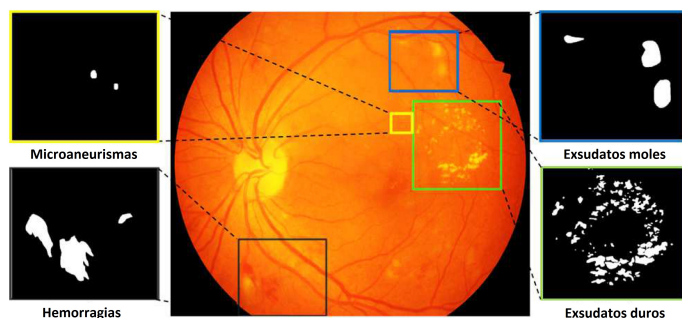
Para detecção da retinopatia diabética, deve se realizar o exame de fundo de olho (retina). Este exame é não invasivo e requer a dilatação da pupila (midríase) para permitir a visualização de toda a retina e inferir a presença e o grau da retinopatia diabética (Patton et al. (2006)). As alterações típicas observadas na retinopatia diabética precoce (não proliferativa) são microaneurismas, hemorragias e exsudatos (Figura 1 e Figura 2).

Figura 1. Imagem de retina apresentando os pontos de interesse (microaneurismas, hemorragias, exsudatos duros e exsudatos moles).



Fonte: Imagem adaptada da base de dados IDRiD Porwal et al. (2018).

Figura 2. Imagem de retina apresentando o *Ground Truth* das regiões de interesse (microaneurismas, hemorragias, exsudatos duros e exsudatos moles).



Fonte: Imagem adaptada da base de dados IDRiD Porwal et al. (2018).

Os recentes avanços tecnológicos em sistemas de comunicação e poder de processamento computacional viabilizam novas oportunidades aos engenheiros biomédicos e cientistas da computação para auxiliar o diagnóstico clínico por meio de processamento de imagens médicas (Shortliffe and Blois (2006); Patton et al. (2006)). As bases de dados com imagens reais e respectivas avaliações facilitam o desenvolvimento, validação, comparação e melhoria dos algoritmos de detecção de lesões da retinopatia diabética utilizados na aplicação clínica (Trucco et al. (2013)).

Como técnica de processamento de imagens para auxílio ao diagnóstico médico, destaca-se a segmentação de imagens, sendo esta uma abordagem de grande relevância na análise de imagens médicas. A fim de fornecer uma base confiável para diagnóstico, pesquisa de patologia e auxiliar em resultados mais precisos, é necessário segmentar as partes das imagens médicas que apresentam informações relevantes (Zeng et al. (2019)). Anotações de nível de pixel precisas de anormalidades associadas a retinopatia diabética permitem avaliação de desempenho de técnicas de segmentação de lesões individuais. Mesmo com o banco de dados de imagens adequado ao processo, identificar exsudatos em imagens de retina é um desafio, pois não apresentam bordas bem definidas, apresentam variações no gradiente da textura, textura visual semelhante ao disco óptico e ruído de aquisição.

Este cenário propicia o uso de redes neurais artificiais (RNA's) para segmentação. RNAs implementam um processo chamado aprendizado de representação, onde diversas camadas de processamento são interligadas e alimentadas com dados para gerar uma saída. Métodos de *deep learning* utilizam redes neurais aperfeiçoadas com maior número de camadas que permitem aprendizado de representação em níveis mais baixos e abstratos (Ravi et al. (2016)).

Litjens et al. (2017) apresenta uma revisão bibliográfica com mais de 300 artigos contemplando ampla variedade de aplicações em *deep learning* para análise de imagens médicas. Os autores analisaram os principais conceitos e resultados de *deep learning* aplicados a classificação de imagens, detecção de estruturas, segmentação e outras tarefas envolvendo diversas aplicações da área de imagens médicas.

Motivado pelos desafios apresentados nas características de imagens de retina e pelos recentes avanços e bons resultados no uso de *deep learning* em imagens médicas, este trabalho tem por objetivo apresentar técnicas de segmentação de exsudatos para auxílio ao diagnóstico médico utilizando redes neurais convolucionais. Deseja-se que esse sistema seja treinado e encontre a melhor relação de pesos sinápticos para obter a melhor predição. São, portanto, propostas duas arquiteturas de redes neurais convolucionais, sendo uma rede com menos camadas (*MobileNet* - Howard et al. (2017)) e uma *Unet* (Falk et al. (2019)).

A escolha das arquiteturas de rede decorrem da busca por redes neurais com reduzido custo computacional de aplicação, pois deseja-se contribuir para que esta tecnologia seja produtificada. A rede *MobileNet* destaca-se pela sua simplicidade e baixo custo computacional. Já a rede *Unet* apresenta maior número de camadas e bons resultados na literatura em segmentações em outras áreas. Essas redes foram alteradas para que sua entrada utilize componentes RGB e possa avaliar através dos pesos sinápticos e treinamento as melhores informações a se extrair de cada canal de cor ou combinação destes. O uso da técnica de subdivisão das imagens no banco de dados IDRiD juntamente a comparação das topologias propostas, permitem a obtenção de resultados não encontrados atualmente na literatura.

2. TRABALHOS RELACIONADOS

Vários trabalhos tem apresentado métodos para uma classificação binária de RD com resultados encorajadores. No trabalho Gardner et al. (1996) foram usados redes neurais e valores de intensidade de pixel para alcançar resultados de sensibilidade e especificidade de 88,4% e 83,5%, respectivamente. Eles utilizaram uma pequena base dados de cerca de 200 imagens e dividiram cada imagem em *patches* e, em seguida, solicitaram que um clínico as classificasse.

As redes neurais também foram usadas na classificação de três classes de RD. No trabalho Nayak et al. (2008) foram utilizadas características como a área dos exsudatos e a área dos vasos sanguíneos, juntamente com parâmetros de textura como entradas em uma rede neural para classificar as imagens em retinopatia normal não proliferativa. Os resultados da detecção foram validados por comparação

com a classificação de oftalmologistas experientes. Eles demonstraram uma acurácia de 93%, sensibilidade de 90% e especificidade de 100%. Os autores utilizaram esse sistema em um conjunto de dados de 140 imagens e a extração de características foi necessária em todas as imagens, tanto no treinamento quanto no teste. Desta forma, exige maior custo computacional e consequentemente maior tempo de resolução.

A maior parte das pesquisas sobre a classificação de cinco classes da retinopatia diabética que foi realizada utilizou máquinas de vetores de suporte (SVMs). No trabalho Acharya et al. (2008) propuseram um método automatizado para identificar as cinco classes. As características, que são extraídas dos dados brutos usando um método de espectros de ordem superior, são inseridas no classificador SVM e capturam a variação nas formas e contornos das imagens. Este método SVM relatou uma precisão média de 82%, sensibilidade de 82% e especificidade de 88%. No trabalho *Computer-based detection of diabetes retinopathy stages using digital fundus images* Acharya et al. (2009) também propuseram um método de classificação de classe especial, calculando as áreas de várias características, como hemorragias, microaneurismas, exsudato e vasos sanguíneos. As características determinadas como as mais relevantes foram então aplicadas ao SVM para classificação. Uma sensibilidade de 82%, especificidade de 86% e precisão de 85,9% foi alcançada usando este sistema. Estes métodos foram treinados em conjuntos de dados relativamente pequenos e a queda na sensibilidade e especificidade foi provavelmente devido à natureza complexa do problema das cinco classes da retinopatia diabética (Pratt et al. (2016)). No trabalho Adarsh and Jeyakumari (2013) utilizaram as técnicas de processamento de imagens para produzir um diagnóstico automatizado para RD através da detecção de vasos sanguíneos da retina, exsudatos, micro-aneurismas e características de textura. A área de lesões e textura da região foram usadas para construir o vetor de característica para o SVM multiclasse. Isso alcançou precisões de 96% e 94,6% nas bases de dados DIARETDB0 e DIARETDB1, respectivamente. Cada um dos cinco métodos de classe anteriores exigia a extração de características das imagens antes de serem inseridas em um classificador SVM que foram validados apenas em pequenos conjuntos de teste de aproximadamente 100 imagens. Outra classe de métodos de segmentação que obteve destaque, explora as diferenças na intensidade da cor entre as várias regiões das imagens. Zubair et al. (2016) usou um processo de limiar dinâmico selecionando o valor do limiar para cada imagem após determinar sua faixa de iluminação. Essa faixa foi calculada considerando os valores mínimo, máximo, média e desvio padrão dos pixels de cada imagem. Com esta técnica foi possível atingir resultados com acurácia de 97,62%, sensibilidade 98,73% e especificidade 98,25% para o conjunto de imagens (MESSIDOR) capturado em diferentes condições de iluminação. No entanto, o cálculo do valor limite para cada imagem ainda permanece uma tarefa de alto custo computacional. O sucesso de classificadores de aprendizado profundo como AlexNet, VGGNet, GoogleNet e ResNet é muito evidente no desafio de classificação do ImageNet. Essas redes forneceram bons resultados de classificação, no entanto, quando estendidos para a segmentação em pixels, eles falharam em fornecer resultados igualmente bons (Sambyal et al.

(2020)), desta forma, esse trabalho se atém a apresentação das duas arquiteturas propostas comparativamente com o fronte de pesquisa. O trabalho Sambyal et al. (2020) apresentam uma arquitetura U-Net modificada com base na rede residual e emprega embaralhamento periódico com convolução de sub-pixel para redimensionar o vizinho mais próximo de convolução. A arquitetura proposta foi treinada e validada para microaneurisma e segmentação de exsudato duro em dois conjuntos de dados disponíveis ao público, IDRiD e e-optha. Para o conjunto de dados IDRiD, a rede obtém 99,88% de acurácia, 99,85% de sensibilidade, 99,95% de especificidade, sendo este o fronte de pesquisa até o presente momento.

3. MÉTODO

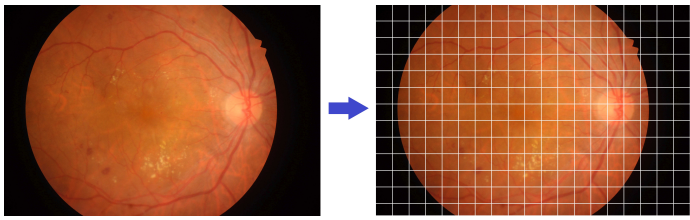
3.1 Base de dados

A base de dados IDRiD consiste de 81 imagens de retina com sinais de retinopatia diabética. As imagens apresentam anotações precisas de nível de pixel indicando anormalidades associadas a microaneurismas, exsudatos moles, exsudatos duros e hemorragias. Também são apresentadas máscaras binárias para avaliação de desempenho de técnicas de segmentação de lesões individuais. As imagens de retina são coloridas e máscaras binárias separadas para cada tipo de lesão. As imagens foram classificadas e avaliadas por especialistas médicos que garantiram a detecção das regiões de interesse, a qualidade adequada, relevância, unicidade e razoável mistura de estratificação da doença representativa da retinopatia diabética (Porwal et al. (2018)). Este conjunto de dados foi disponibilizado como parte do “Desafio da Retinopatia Diabética: Segmentação e Classificação (<http://biomedicalimaging.org/2018/challenges/>)” organizado em conjunto com o *IEEE International Symposium on Biomedical Imaging (ISBI 2018)*, Washington DC. O desafio de dados foi hospedado no *Grand Challenges in Biomedical Imaging Platform* Patton et al. (2006). Para esta aplicação, este trabalho propõe a subdivisão da imagem original em imagens menores o que amplia substancialmente o número de imagens de treinamento e avaliação. Dividir as imagens em subimagens menores também amplia as possibilidades de treinamento evitando erros de generalização, já que nem toda subimagem apresenta *ground truth* com exsudatos.

3.2 Pré-processamento

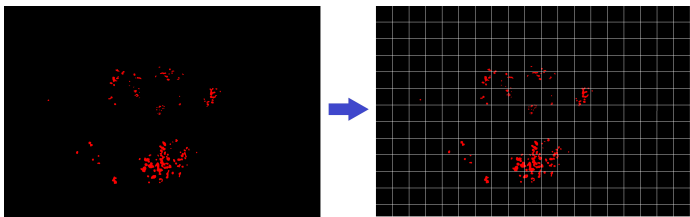
A etapa inicial de desenvolvimento do método foi dedicada à preparação do *dataset* para o treinamento da rede. O pré-processamento consistiu em subdividir as imagens e seus respectivos *ground truth*. Isso foi necessário porque as imagens originais apresentam resolução de 4288 pixels x 2848 pixels x 3 canais (RGB). Treinar a rede convolucional com a imagem completa, sem subdividir, torna-se um processo computacionalmente oneroso por não apresenta memória suficiente. Como solução as imagens foram subdivididas em 224 pixels x 224 pixels por 3 canais (RGB) para a rede *MobileNet* e 512 pixels x 512 pixels por 3 canais (RGB) para a rede *Unet*. A Figura 3 apresenta de forma ilustrativa a subdivisão da imagem original da retina em subimagens. A Figura 4 ilustra a aplicação do processo de subdivisão do *Ground truth* garantindo a correspondência durante o processo de treinamento.

Figura 3. Ilustração da subdivisão da imagem original da retina em subimagens.



Fonte: Elaborada pelo autor.

Figura 4. Ilustração da subdivisão da imagem original do *Ground truth* em subimagens.



Fonte: Elaborada pelo autor.

Como uma única imagem foi subdividida em imagens menores, algumas subimagens passaram a não conter informações relevantes, tanto na imagem original, quanto no *ground truth*, apresentando matrizes nulas (conforme pode ser observado nas Figuras 3 e 4). Neste caso, tanto a subimagem quanto seu respectivo *ground truth* foram descartados com o intuito de acelerar o processo de treinamento, isto é, manter apenas subimagens com informações/relevância. Dessa forma, a partir do conjunto inicial de treinamento de 54 imagens da base de dados IDRiD, tornou-se 9.267 subimagens e o conjunto de testes com 27 imagens, tornou-se 4.784 subimagens. De forma análoga, o conjunto de imagens foi dividido levando em consideração a entrada da rede *Unet* de 512 pixels x 512 pixels, excluindo subimagens totalmente nula. Para a subdivisão da *Unet* foram geradas 1.990 subimagens para treinamento e 1.060 imagens para testes. Estas subdivisões e exclusões das subimagens nulas foram aplicadas no subconjunto de treinamento e de testes, desta forma, seus resultados continuam válidos e aceleram o processo de treinamento e testes.

3.3 Processamento

Foram implementadas duas arquiteturas de redes neurais convolucionais, uma rede com menor número de parâmetros treináveis (*MobileNet*) e outra rede baseada na arquitetura *Unet*. Cada uma possui tamanho específico de entrada, desta forma, a subdivisão das imagens geraram números distintos de subimagens. Este diferença não invalida o resultado final, pois em ambos os casos a correlação entre imagem e respectivo *ground truth* é mantida. Deve-se ressaltar que as subimagens pré-processadas para a rede *MobileNet* possuem tamanho reduzido o que gera maior número de subimagens descartadas, dessa forma, o número total de pixels avaliado é distinto da rede *Unet*. A saída de ambas as redes neurais apresentam gradiente de probabilidade da segmentação correta, desta forma,

antes da análise, todas as imagens foram binarizadas de forma a maximizar os parâmetros de avaliação na etapa de treinamento.

3.4 Configurações dos hiperparâmetros

Os hiperparâmetros são um conjunto de variáveis cujos valores são normalmente selecionados experimentalmente antes do início do processo de treinamento. Isso geralmente inclui taxa de aprendizado, *batch size* e número de épocas. O processo de obtenção de valores otimizados de hiperparâmetros é conhecido como ajuste de hiperparâmetros (Sambyal et al. (2020)). Os valores dos hiperparâmetros usados no trabalho proposto foram determinados experimentalmente utilizando método empírico, através de busca exaustiva destes parâmetros com a finalidade de otimizar as métricas de avaliação (acurácia, especificidade e sensibilidade). Em ambas as redes foram testados ao menos 20 possibilidades com incremento automático do decaimento e do *batch size*. Por fim, foram selecionados os valores dos parâmetros que forneceram os melhores resultados para a segmentação semântica.

3.5 Métricas de avaliação dos resultados

Ambos os conjuntos de imagens resultantes foram avaliados em nível de pixel. As métricas utilizadas foram acurácia, especificidade e sensibilidade. Para as equações são utilizados os parâmetros de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

$$Especificidade = \frac{VN}{(FP + VN)}$$

$$Sensibilidade = \frac{VP}{(VP + FN)}$$

4. EXPERIMENTOS E RESULTADOS

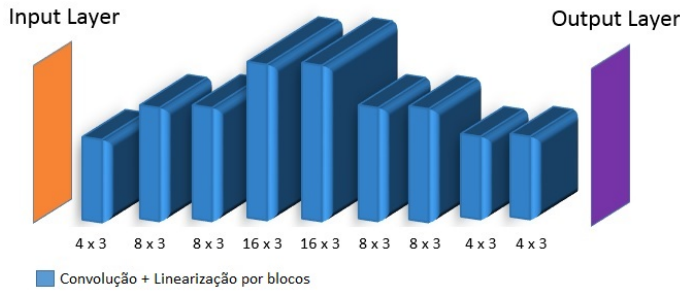
4.1 Rede Neural *MobileNet*

A Tabela 1 apresenta a topologia de rede implementada, com seus parâmetros. A Figura 5 apresenta o diagrama de blocos da rede neural implementada. A rede neural escolhida - *MobileNet* - apresenta grande simplicidade e eficiência (Howard et al. (2017)). Para a rede implementada, foi inserida não linearidade do tipo ReLu com o objetivo de tornar a rede mais esparsa e eficiente (Glorot et al. (2011)). A função de ativação na camada de saída foi a sigmóide com o intuito de gerar um gradiente de probabilidade e melhor avaliar a assertividade da rede, uma vez que sua saída não será binária. A construção dos *padding* das convoluções se deram pela técnica da réplica da borda (*same*) (Dumoulin and Visin (2016)).

Parâmetros de treinamento:

Utilizando toda base de dados de imagens de testes, e realizando o treinamento em 500 épocas, foi possível obter resultados menos precisos, porém com maior generalização.

Figura 5. Diagrama de blocos da rede simplificada implementada - *MobileNet*.



Fonte: Elaborada pelo autor.

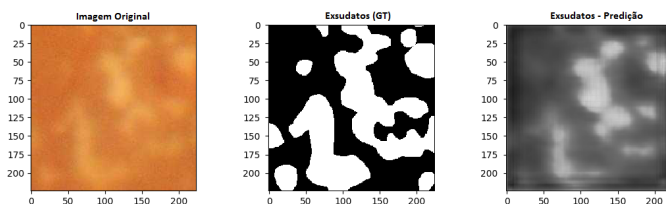
Decaimento	1 e-9
Batch size	18
Total de parâmetros	6.989
Parâmetros treináveis	6.845
Parâmetros fixos	144
Épocas	500

Tabela 1. Arquitetura da rede utilizada - Rede simplificada.

Layer (tipo)	Formato de saída	Parâmetros
input_1 (InputLayer)	(None, 224, 224, 3)	0
conv2d_1 (Conv2D)	(None, 224, 224, 4)	112
batch_normalization_1	(Batch (None, 224, 224, 4)	16
conv2d_2 (Conv2D)	(None, 224, 224, 8)	296
batch_normalization_2	(Batch (None, 224, 224, 8)	32
conv2d_3 (Conv2D)	(None, 224, 224, 8)	584
batch_normalization_3	(Batch (None, 224, 224, 8)	32
conv2d_4 (Conv2D)	(None, 224, 224, 16)	1168
batch_normalization_4	(Batch (None, 224, 224, 16)	64
conv2d_5 (Conv2D)	(None, 224, 224, 16)	2320
batch_normalization_5	(Batch (None, 224, 224, 16)	64
conv2d_6 (Conv2D)	(None, 224, 224, 8)	1160
batch_normalization_6	(Batch (None, 224, 224, 8)	32
conv2d_7 (Conv2D)	(None, 224, 224, 8)	584
batch_normalization_7	(Batch (None, 224, 224, 8)	32
conv2d_8 (Conv2D)	(None, 224, 224, 4)	292
batch_normalization_8	(Batch (None, 224, 224, 4)	16
conv2d_9 (Conv2D)	(None, 224, 224, 4)	148
conv2d_10 (Conv2D)	(None, 224, 224, 1)	37

A Figura 6 ilustra uma das imagens resultantes do teste. Observa-se qualitativamente, que a região de predição não apresenta borda bem definida como o *ground truth*, mas as maiores regiões de exsudatos apresentam ao menos uma marcação em sua predição.

Figura 6. Resultado qualitativo - Predição de exsudatos *MobileNet*.



Fonte: Imagem obtida no processo de treinamento da rede neural aplicada a base de dados IDRiD.

Tabela 2. Matriz de confusão - Rede Neural *MobileNet*

		Predição	
		Com exsudatos	Sem exsudatos
Real	Com exsudatos	929.690	2.582.185
	Sem exsudatos	65.260.904	171.269.205

Tabela 3. Resultados - Rede Neural *MobileNet*

Métrica	Resultado
Sensibilidade	96,2%
Especificidade	72,4%
Acurácia	71,7%

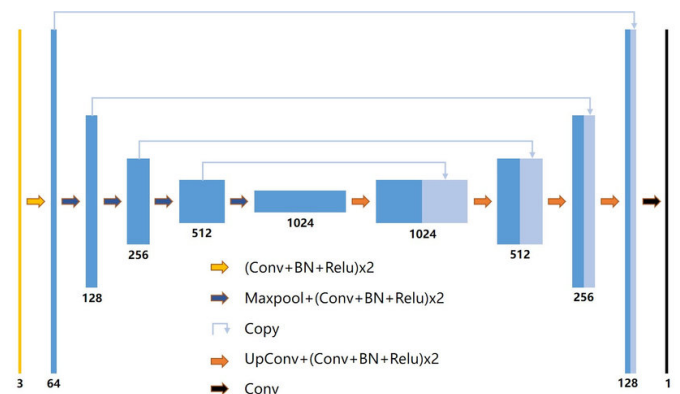
Ao finalizar a etapa de treinamento da rede, esta foi aplicada ao conjunto de imagens de testes para a predição. Avaliando um total de 4.784 subimagens desse subconjunto, foram obtidos os resultados apresentados nas Tabelas 2 e 3.

A Figura 6 ilustra um dos resultados obtidos com a *MobileNet*. Deve-se considerar que essa subimagem representa 0,4% do número total de pixels da imagem. Essa diminuta proporção chama a atenção (qualitativamente) pois o sistema foi capaz de reproduzir consideravelmente os *ground truth*.

4.2 Rede neural Unet

A Tabela 4 apresenta a topologia de rede implementada, com seus parâmetros. A Figura 7 apresenta o diagrama de blocos da rede neural implementada. A função de ativação na camada de saída foi a sigmóide com o intuito de gerar um gradiente de probabilidade e melhor avaliar a assertividade da rede, uma vez que sua saída não será binária. Para esta rede foi utilizado o otimizador Adam e função de perda *BinaryCrossentropy*.

Figura 7. Diagrama de blocos da rede simplificada implementada.



Fonte: Imagem de Zhao et al. (2019).

Parâmetros de treinamento:

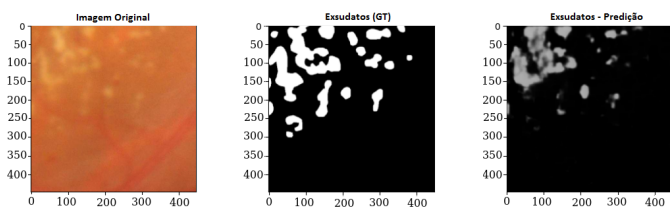
Decaimento	1 e-4
Batch size	32
Função de custo	<i>BinaryCrossentropy</i>
Total de parâmetros	31.032.837
Parâmetros treináveis	31.032.837
Parâmetros fixos	0
Épocas	500

Tabela 4. Arquitetura da rede utilizada - *Unet*

Layer (tipo)	Formato de saída	Parâmetros
input_1 (InputLayer)	(None, 256, 256, 3)	0
conv2d (Conv2D)	(None, 256, 256, 64)	1792
conv2d_1 (Conv2D)	(None, 256, 256, 64)	36928
max_p2d (MaxPooling2D)	(None, 128, 128, 64)	0
conv2d_2 (Conv2D)	(None, 128, 128, 128)	73856
conv2d_3 (Conv2D)	(None, 128, 128, 128)	147584
max_p2d_1 (MaxPooling2D)	(None, 64, 64, 128)	0
conv2d_4 (Conv2D)	(None, 64, 64, 256)	295168
conv2d_5 (Conv2D)	(None, 64, 64, 256)	590080
max_p2d_2 (MaxPooling2D)	(None, 32, 32, 256)	0
conv2d_6 (Conv2D)	(None, 32, 32, 512)	1180160
conv2d_7 (Conv2D)	(None, 32, 32, 512)	2359808
dropout (Dropout)	(None, 32, 32, 512)	0
max_p2d_3 (MaxPooling2D)	(None, 16, 16, 512)	0
conv2d_8 (Conv2D)	(None, 16, 16, 1024)	4719616
conv2d_9 (Conv2D)	(None, 16, 16, 1024)	9438208
dropout_1 (Dropout)	(None, 16, 16, 1024)	0
up_s2d (UpSampling2D)	(None, 32, 32, 1024)	0
conv2d_10 (Conv2D)	(None, 32, 32, 512)	2097664
concatenate (Concatenate)	(None, 32, 32, 1024)	0
conv2d_11 (Conv2D)	(None, 32, 32, 512)	4719104
conv2d_12 (Conv2D)	(None, 32, 32, 512)	2359808
up_s2d_1 (UpSampling2D)	(None, 64, 64, 512)	0
conv2d_13 (Conv2D)	(None, 64, 64, 256)	524544
conc_1 (Concatenate)	(None, 64, 64, 512)	0
conv2d_14 (Conv2D)	(None, 64, 64, 256)	1179904
conv2d_15 (Conv2D)	(None, 64, 64, 256)	590080
up_s2d_2 (UpSampling2D)	(None, 128, 128, 256)	0
conv2d_16 (Conv2D)	(None, 128, 128, 128)	131200
conc_2 (Concatenate)	(None, 128, 128, 256)	0
conv2d_17 (Conv2D)	(None, 128, 128, 128)	295040
conv2d_18 (Conv2D)	(None, 128, 128, 128)	147584
up_s2d_3 (UpSampling2D)	(None, 256, 256, 128)	0
conv2d_19 (Conv2D)	(None, 256, 256, 64)	32832
conc_3 (Concatenate)	(None, 256, 256, 128)	0
conv2d_20 (Conv2D)	(None, 256, 256, 64)	73792
conv2d_21 (Conv2D)	(None, 256, 256, 64)	36928
conv2d_22 (Conv2D)	(None, 256, 256, 2)	1154
conv2d_23 (Conv2D)	(None, 256, 256, 1)	3

Utilizando toda base de dados de imagens de testes, e realizando o treinamento em 500 épocas, foi possível obter resultados menos precisos, porém com maior generalização. A Figura 6 ilustra uma das imagens resultantes do teste. Observa-se qualitativamente, que a região de predição não apresenta borda bem definida como o *ground truth*, mas as maiores regiões de exsudatos apresentam ao menos uma marcação em sua predição e algumas pequenas regiões, diferentemente dos resultados da *MobileNet*, foram segmentadas.

Figura 8. Resultado qualitativo - Predição de exsudatos *Unet*.



Fonte: Imagem obtida no processo de treinamento da rede neural aplicada a base de dados IDRiD.

Tabela 5. Matriz de confusão - Rede Neural *Unet*.

		Predição	
		Com exsudatos	Sem exsudatos
Real	Com exsudatos	1.564.429	1.947.446
	Sem exsudatos	23.442.562	250.918.203

Tabela 6. Resultados - Rede Neural *Unet*

Métrica	Resultado
Sensibilidade	44,5%
Especificidade	91,5%
Acurácia	90,9%

Ao finalizar a etapa de treinamento da rede, esta foi aplicada ao conjunto de imagens de testes para a predição. Avaliando um total de 1.060 subimagens desse subconjunto, foram obtidos os resultados apresentados nas Tabelas 5 e 6. Os resultados obtidos são apresentados abaixo:

A Figura 8 ilustra um dos resultados obtidos com a *Unet*. Deve-se considerar que essa subimagem representa 2% do número total de pixels da imagem (Original). Essa diminuta proporção chama a atenção pois o sistema foi capaz de reproduzir consideravelmente o *ground truth*.

4.3 Resultados comparativos *Unet*, *MobileNet* e frente de pesquisa

Os resultados obtidos nas duas redes propostas nesse trabalho (*Unet* e *MobileNet*) e no frente de pesquisa (Sambyal et al. (2020)) podem ser observados na Tabela 7, onde pode-se observar as métricas de interesse de forma comparativa. Destaca-se a grande diferença de tempo entre as arquiteturas, mesmo com a diferença de *hardware*. A arquitetura *MobileNet* apresenta 19 *layers* compostas de blocos de convolução e *Batch Normalizations*. A arquitetura *Unet* proposta possui 39 *layers* compostas de bloco de convolução, *max pooling*, *dropout*, concatenação e *upsampling*. A rede neural de melhor resultado proposta em Sambyal et al. (2020) apresenta 71 *layers* composta por blocos de *ResNet*, *Unet*, *Middle Convolution* e *layers merge*.

Tabela 7. Resultados comparativos

	MobileNet	Unet	Frente de pesquisa
Sensibilidade	96,2%	44,5%	99,85%
Especificidade	72,4%	91,5%	99,95%
Acurácia	71,7%	90,9%	99,88%
Épocas	500	500	16
Tempo de treinamento por época (s)	0,074	0,151	26,76
Tempo de treinamento total (s)	36,911	75,766	428,16
Hardware (GPU Nvidia)	RTX2060 6 GB	RTX2060 6 GB	1080Ti 11 GB

A diferença de tempo de processamento por época evidencia o baixo custo computacional em detrimento da rede proposta em Sambyal et al. (2020), mesmo sendo necessário substancialmente maior número de épocas para ajuste dos pesos sinápticos, as redes propostas nesse trabalho terminaram o processo de treinamento em menor tempo. Para os resultados obtidos nas redes propostas nesse trabalho foi utilizado um computador convencional

Tabela 8. Computador e sistema utilizado

Processador	Intel Core i5-10400 Hexa-Core 2.9Ghz (4GHz Turbo) 12MB Cache
Memória	16 GB (2x8GB DDR4 3000MHz)
Armazenamento	SSD Sata 6GB/s
Sistema Operacional	Linux - Ubuntu 20.04 LTS
Programação	Python 3.8.10
Biblioteca de deep learning	Keras 2.4.3

descrito na Tabela 8 com sistema operacional e processos convencionais em *background*.

A segmentação de exudatos tem por objetivo auxiliar o diagnóstico médico, uma vez que os sistemas existentes são incapazes de atingir 100% em todas as métricas. Nesta perspectiva, admite-se a existência de maior número de falsos positivos em detrimento de falsos negativos. Desta forma, a rede *Unet* proposta se apresenta como potencial solução para auxílio médico na identificação/segmentação de exudatos, com baixo custo computacional.

5. CONCLUSÃO

A retinopatia diabética é a maior causa da cegueira evitável da população em idade ativa. Como agravante, esta doença apresenta grande perspectiva de crescimento, o que motiva o desenvolvimento de pesquisas que auxiliem a detecção e consequentemente a redução dessa perspectiva de crescimento. Buscando contribuir com técnicas de processamento de imagens que auxiliem o diagnóstico, esse trabalho aborda a identificação de pontos de interesse em imagens de retina para auxílio ao diagnóstico da retinopatia diabética. Historicamente, diversas técnicas e ferramentas foram empregadas para classificar a retina como portadora de retinopatia diabética ou saudável, entretanto, há um número reduzido de trabalhos que busquem a segmentação de pontos. Como os sistemas propostos não apresentam acurácia acima de 99%, não há garantia de um correto diagnóstico, desta forma, há a necessidade de uma avaliação e diagnóstico final realizado por um médico especialista da área. Nesse sentido, o trabalho proposto, alia a identificação e segmentação de pontos de risco buscando auxiliar a avaliação e diagnóstico médico, não sendo utilizado para uma resposta conclusiva sobre a existência de retinopatia diabética.

A maior parte dos trabalhos é voltada ao desenvolvimento da técnica com grande número de operações de pré-processamento. Neste trabalho, foram implementadas duas redes neurais com arquiteturas distintas, onde o pré-processamento se restringiu a subdivisões das imagens originais. Estas subdivisões viabilizaram o uso de redes neurais por adequar o tamanho de entrada das redes e por aumentar substancialmente o número de imagens, como consequência, torna a rede neural com maior número de amostras, contribuindo para a generalização de predições.

Este trabalho apresentou duas topologias de redes neurais, a primeira dela, *MobileNet*, mais simples com resultados de sensibilidade, especificidade e acurácia de 96,2%, 72,4% e 71,7% respectivamente.

A segunda rede neural utilizada para segmentação de exudatos foi a *Unet*, com maior número de camadas e

épocas de treinamento. Os resultados de sensibilidade, especificidade e acurácia obtidos neste topologia foram de 44,5%, 91,5% e 90,9%

O fronte de pesquisa desta doença, que utiliza segmentação para detecção de retinopatia diabética, apresentou resultados de sensibilidade, especificidade e acurácia de 99,85%, 99,95% e 99,88% respectivamente.

Apesar do resultado abaixo do fronte de pesquisa, vale ressaltar, que o sistema proposto requer baixo custo computacional nas etapas de pré-processamento e aplicação ou testes. Este trabalho também apresentou o uso comparativo das redes *MobileNet* e *Unet* com entradas RGB em subimagens do banco de dados IDRiD o que permitiu a obtenção de resultados não encontrados atualmente na literatura. Estes resultados motivam a utilização de redes mais complexas e profundas como *Fully Convolutional Residual Network* (FCRN) proposta por Laina et al. (2016), arquitetura esta que foi inspirada na rede ResNet-50 He et al. (2016), e espera-se prosseguir nas pesquisas para alcançar resultados cada vez mais próximos aos do fronte, tendo como meta sempre o baixo custo computacional.

O objetivo atual desse trabalho foi atingido por permitir ampliar substancialmente o número de imagens através da subdivisão de imagens permitindo maior número de treinamento e teste dentre os trabalhos apresentados. Espera-se que esse sistema apresente maior robustez devido ao maior volume de treinamento e testes. Sua simplicidade, por não requerer pré-processamento sofisticado e baixo custo computacional para aplicação, viabiliza a utilização desse sistema mesmo em locais com poucos recursos, o que pode auxiliar ao médico no diagnóstico da retinopatia diabética.

REFERÊNCIAS

- Acharya, R., Chua, C.K., Ng, E., Yu, W., and Chee, C. (2008). Application of higher order spectra for the identification of diabetes retinopathy stages. *Journal of Medical Systems*, 32(6), 481–488.
- Acharya, U.R., Lim, C.M., Ng, E.Y.K., Chee, C., and Tamura, T. (2009). Computer-based detection of diabetes retinopathy stages using digital fundus images. *Proceedings of the institution of mechanical engineers, part H: journal of engineering in medicine*, 223(5), 545–553.
- Adarsh, P. and Jeyakumari, D. (2013). Multiclass svm-based automated diagnosis of diabetic retinopathy. In *2013 International Conference on Communication and Signal Processing*, 206–210. doi:10.1109/iccsp.2013.6577044.
- Cheung, N., Mitchell, P., and Wong, T.Y. (2010). Diabetic retinopathy. *Lancet*, 376(9735), 124–136.
- Dumoulin, V. and Visin, F. (2016). A guide to convolution arithmetic for deep learning. 1–31. doi:10.1051/0004-6361/201527329. URL <http://arxiv.org/abs/1603.07285>.
- Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäkel, Z., Seiwald, K., et al. (2019). U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1), 67–70.
- Gardner, G., Keating, D., Williamson, T.H., and Elliott, A.T. (1996). Automatic detection of diabetic retino-

- pathy using an artificial neural network: a screening tool. *British journal of Ophthalmology*, 80(11), 940–944.
- Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315–323.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. doi:10.1109/CVPR.2016.90. URL <http://ieeexplore.ieee.org/document/7780459/>.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., and Navab, N. (2016). Deeper depth prediction with fully convolutional residual networks. In *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, 239–248. doi:10.1109/3DV.2016.32.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., and Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60–88.
- Nayak, J., Bhat, P.S., Acharya, R., Lim, C.M., and Kagathi, M. (2008). Automated identification of diabetic retinopathy stages using digital fundus images. *Journal of medical systems*, 32(2), 107–115.
- Patton, N., Aslam, T.M., MacGillivray, T., Deary, I.J., Dhillon, B., Eikelboom, R.H., Yogesan, K., and Constable, I.J. (2006). Retinal image analysis: concepts, applications and potential. *Progress in retinal and eye research*, 25(1), 99–127.
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudhe, V., and Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. *Data*, 3(3), 25.
- Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., and Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90, 200–205.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G.Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1), 4–21.
- Sambyal, N., Saini, P., Syal, R., and Gupta, V. (2020). Modified u-net architecture for semantic segmentation of diabetic retinopathy images. *Biocybernetics and Biomedical Engineering*, 40(3), 1094–1109. doi:<https://doi.org/10.1016/j.bbe.2020.05.006>. URL <https://www.sciencedirect.com/science/article/pii/S0208521620300747>.
- Shortliffe, E.H. and Blois, M.S. (2006). The computer meets medicine and biology: emergence of a discipline. In *Biomedical Informatics*, 3–45. Springer.
- Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J.P., Al-Diri, B., Cheung, C.Y., Wong, D., Abramoff, M., et al. (2013). Validating retinal fundus image analysis algorithms: issues and a proposal. *Investigative ophthalmology & visual science*, 54(5), 3546–3559.
- Yau, J.W., Rogers, S.L., Kawasaki, R., Lamoureux, E.L., Kowalski, J.W., Bek, T., Chen, S.J., Dekker, J.M., Fletcher, A., Grauslund, J., Haffner, S., Hamman, R.F., Ikram, M.K., Kayama, T., Klein, B.E., Klein, R., Krishnaiah, S., Mayurasakorn, K., O’Hare, J.P., Orchard, T.J., Porta, M., Rema, M., Roy, M.S., Sharma, T., Shaw, J., Taylor, H., Tielsch, J.M., Varma, R., Wang, J.J., Wang, N., West, S., Xu, L., Yasuda, M., Zhang, X., Mitchell, P., Wong, T.Y., Yau, J.W., Rogers, S.L., Kawasaki, R., Lamoureux, E.L., Wong, T.Y., Kowalski, J.W., Mahabhashyam, S., Yeh, W.S., Aung, T., Saw, S.M., Tay, W., Wong, W., Panero, F., Porta, M., Bruno, G., Caengow, S., Somthip, N., Chuikarat, N., Wani-chsuwan, M., Mayurasakorn, K., Chen, S.J., Cheng, C.Y., Chou, P., Hsu, W.M., Liu, J.H., Chakravarthy, U., Cotch, M.F., Vingerling, J., De Jong, P., Ikram, M., Zavrelava, H., Nijpels, G., Dekker, J.M., Fletcher, A., Grauslund, J., Sjølie, A.K., Bek, T., Stern, M., Haffner, S., Hamman, R.F., Yasuda, M., Ishibashi, T., Kiyohara, Y., Jensen, R.A., Jonas, J.B., Kato, T., Yamashita, H., Kayama, T., Munoz, B., Katz, J., West, S., Tielsch, J.M., Friedman, D., Klein, R., Klein, B.E., Krishnaiah, S., Lehman, D.M., McCarty, C., Miller, R.G., Orchard, T., Pradeepa, R., Mohan, R., Mohan, V., O’Hare, J.P., Raymond, N.T., Polak, B.C., Wang, J.J., Mitchell, P., Roychchina, E., Roy, M.S., Raman, R., Sharma, T., Torres, M., Varma, R., Seland, J., Vioque, J., Wang, F.H., Wang, N.L., Liang, Y.B., You, Q.S., Xu, L., Zhang, X.Y., Wang, Y.X., Young, I., Zhang, X., Taylor, H.R., Siscovick, D.S., Stehouwer, C.D., Rahu, M., Soubrane, G., Tomazzoli, L., Topouzis, F., Shaw, J., and Zimmet, P. (2012). Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 35(3), 556–564.
- Zeng, Z., Xie, W., Zhang, Y., and Lu, Y. (2019). Ricunet: An improved neural network based on unet for nuclei segmentation in histology images. *Ieee Access*, 7, 21420–21428.
- Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., and Zhang, D. (2019). Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. *Sensors*, 19(18), 3859.
- Zubair, M., Ali, H., and Javed, M.Y. (2016). Automated segmentation of hard exudates using dynamic thresholding to detect diabetic retinopathy in retinal photographs. *J. Multim. Process. Technol.*, 7(3), 109–116.