

Segmentação Semântica do Câncer de Pele Utilizando Aprendizado Profundo

João Victor M. da Silva¹, Kelson R. T. Aires¹, Alan R. F. dos Santos¹,
Rodrigo de M. S. Veras¹, Laurindo de S. B. Neto¹,
¹ Leonardo P. de Sousa¹, Francisco das C. I. Filho¹

¹Núcleo de Computação de Alto Desempenho (NCAD),
Universidade Federal do Piauí (UFPI) - Teresina - Piauí - Brasil

Abstract. *Skin cancer is one of the major public health problems, and the use of Deep Learning can allow the classification of skin lesions in images. In this context, this work aims to develop a skin lesion segmentation method to facilitate the classification of lesions. In this sense, the DeepLab3+ architecture associated with global threshold was used, refined in three specific models: (1) only for malignant lesions, (2) only for benign lesions and (3) for all types of lesions. The experiments used four public databases, HAM10000, ISIC 2016, ISIC 2017 and PH². The best results reached a Dice of 94.42% in the HAM10000 base, 91.68% in the ISIC 2016 base, 87.19% in the ISIC 2017 base and 92.12% in the PH² base. The best results were achieved with the trained model for all types of lesions.*

Resumo. *O câncer de pele é um dos grandes problemas enfrentados pela saúde pública, e a utilização de Aprendizado de Profundo pode permitir a classificação de lesões de pele em imagens. Nesse contexto, este trabalho tem o objetivo de desenvolver um método de segmentação de lesões de pele para facilitar a classificação de lesões. Nesse sentido, foi utilizado a arquitetura DeepLab3+ associada à limiarização global, refinado em três modelos específicos: (1) somente para lesões malignas, (2) somente para lesões benignas e (3) para todos os tipos de lesões. Os experimentos utilizaram quatro bases públicas, HAM10000, ISIC 2016, ISIC 2017 e PH². Os melhores resultados atingiram um Dice de 94,42% na base HAM10000, 91,68% na base ISIC 2016, 87,19% na base ISIC 2017 e 92,12% na base PH². Os melhores resultados foram alcançados com o modelo treinado para todos os tipos de lesões.*

1. Introdução

O câncer de pele é uma doença que está presente em todo mundo. Muitos casos são registrados todos os anos no Brasil [de Oliveira Santos et al. 2023]. O melanoma é um tipo de câncer de pele de menor incidência entre todos. No entanto, a sua ocorrência diminui significativamente as chances de vida do paciente em caso de metástase [Chan et al. 2020]. As pessoas de pele clara possuem uma maior chance de desenvolver a doença, pois ela está relacionada, geralmente, à exposição excessiva aos raios solares [Ferlay et al. 2021].

O diagnóstico inicial do câncer de pele pode ser realizado por imagens digitais. O médico dermatologista analisa as características das lesões e determina o diagnóstico [Dildar et al. 2021]. A análise visual das imagens requer a definição de um padrão de características. O padrão mais conhecido para essa análise é o ABCDE (Assimetria, Bordas,

Cores, Diâmetro e Evolução). As lesões que possuem partes assimétricas, bordas irregulares, aspecto de cor variável, diâmetro acima de 5 mm e evolução rápida são tendenciosas ao câncer maligno [Ferlay et al. 2021].

Apesar da simplicidade do padrão ABCDE, não é tão simples diagnosticar o câncer de pele. Antes de tudo, é necessário delimitar as lesões de pele nas imagens, separando-as da área de pele saudável [Bagheri et al. 2021]. Essa tarefa se torna desafiadora devido a alguns fatores como baixo contraste, diferenças de textura, iluminação, localização, cores e excesso de pelos [Azad et al. 2020]. Esses fatores impulsionam o desenvolvimento de técnicas computacionais que sejam cada vez mais eficazes para essa tarefa.

O aprendizado profundo é uma metodologia computacional que vem se tornando popular para problemas que envolvem imagens médicas [Dildar et al. 2021]. Principalmente, para as tarefas de segmentação e classificação. Essa metodologia implementa operações que conseguem extrair dados dos *pixels* das imagens, aprendidos posteriormente. Já a segmentação semântica é um princípio amplamente aplicado com essa metodologia, sendo ela a atribuição de significado a esses *pixels*. Com isso, é possível delimitar bordas de objetos de forma mais precisa, independente da quantidade de objetos em uma imagem [Azad et al. 2020].

Uma arquitetura de aprendizado profundo que vem ganhando destaque para segmentação semântica é a *DeepLab*. Essa arquitetura e suas variantes têm o objetivo de gerar um mapa de saída com a rotulação dos *pixels* [Azad et al. 2020]. Isso quer dizer que essa arquitetura utiliza operações de convolução e deconvolução para prever as classes de objetos na imagem [Ronneberger et al. 2015]. A primeira parte dessa arquitetura é conhecida como codificadora, sendo responsável por diminuir os mapas de características. Com isso, a vizinhança de *pixels* é reduzida, possibilitando a predição pela invariância de classes. Já a segunda parte é chamada de decodificadora, que expande a predição com o aumento da resolução dos mapas de características [Bagheri et al. 2021].

É perceptível que o médico dermatologista possui dificuldades no diagnóstico preciso do câncer de pele por imagens. Por outro lado, o diagnóstico precoce de lesões malignas é primordial para o melhor tratamento e cura [Codella et al. 2018]. Logo, as metodologias de aprendizado profundo são essenciais no desenvolvimento de soluções que sugiram uma ajuda na segmentação e no diagnóstico. Nesse sentido, o presente trabalho tem o objetivo de avaliar a utilização de uma variante da arquitetura *DeepLab*, conhecida como *DeepLabV3+*, na operação de segmentação automática de lesões de pele. Essa avaliação considera uma análise entre classes de lesões malignas e benignas no processo de separação. Isso tem o intuito de avaliar a qualidade de segmentação individual, com apenas um tipo de lesão, e coletiva, com todos os tipos de lesões, direcionando a produção de modelos robustos que possam atender ambas as classes com melhores resultados.

As arquiteturas de segmentação semântica convencionais utilizam informações contextuais multiescala com filtros, agrupando-os em mapas de características e captando recursos desnecessários na subamostragem de imagens. Essas arquiteturas não conseguem lidar com muitos detalhes de bordas no processo de criação das máscaras de segmentação de lesões de pele [Chen et al. 2018]. Além disso, consomem muitos recursos computacionais em sua execução. Logo, a principal justificativa deste trabalho é

a aplicação da correção de subamostragem com a arquitetura *DeepLabV3+*, que permite apenas a captação de recursos de bordas mais relevantes no processo de criação dessas máscaras.

O diferencial da proposta deste trabalho está na utilização de segmentação semântica com convoluções combinadas à dilatação, para contornar a perda de informações espaciais presentes no processo de subamostragem. Isso tem o intuito de combinar várias características para obter mais detalhes de bordas para segmentação de lesões de pele. Além disso, a proposta também possui uma análise de segmentação individual e coletiva das lesões de pele para verificar a variabilidade de resultados.

2. Trabalhos Relacionados

A segmentação de lesões de pele em imagens digitais é um problema abrangente. Muitas pesquisas já investigaram métodos computacionais que proporcionaram definições próximas da marcação do médico especialista. Essas marcações são de extrema importância para facilitar a delimitação das lesões nas imagens e possibilitar a criação de métodos automatizados de diagnóstico do câncer maligno [Dildar et al. 2021].

Muitos trabalhos na literatura já utilizaram técnicas convencionais para realizar a tarefa de segmentação de lesões de pele, por exemplo, *Threshold* [Otsu 1979], *Otsu* [Fan et al. 2017], *K-means* [MacQueen 1967] e *Fuzzy K-means* [Isa et al. 2009]. Porém, nos últimos anos, o Aprendizado Profundo vem proporcionando novas metodologias mais eficazes [Dildar et al. 2021]. Em [Ronneberger et al. 2015] e [Alom et al. 2018] são apresentados alguns avanços dessa perspectiva com a segmentação baseada no contexto de expansão da lesão nas imagens, mostrando resultados animadores se comparado com as técnicas convencionais.

Em [Goyal et al. 2019] foi desenvolvida uma sistemática mais avançada que [Ronneberger et al. 2015], pois implementa *ensemble* de arquiteturas de Aprendizado Profundo para realizar a segmentação da base de imagens PH² [Mendonça et al. 2013]. Inclusive, os melhores resultados alcançaram um *Dice* de 90,7%. Já em [Bagheri et al. 2021] foi desenvolvido algo semelhante a [Ronneberger et al. 2015]. No entanto, essa técnica desenvolve uma combinação entre as arquiteturas *Retina-Deeplab* e *Mask-RCNN*, que detectam simultaneamente as lesões de pele, combinando os resultados ao final. Os resultados dessa pesquisa mostraram um *Dice* de 94,6% para a base HAM10000 [Tschandl et al. 2018] e um *Dice* de 93,6% para a base MoleMap [Ge et al. 2017].

Uma sistemática de segmentação semântica diferenciada é apresentada em [Hong et al. 2022], que utiliza uma arquitetura convolucional semi-supervisionada para extrair regiões nas imagens de lesões de pele. Em seguida, esses dados são utilizados como entrada na identificação de áreas com *superpixels*. Os resultados indicaram um índice *Jaccard* de 73% e um *Dice* de 90% para a base *International Skin Imaging Collaboration (ISIC) 2017* [Codella et al. 2018]. Além disso, outros resultados demonstraram um *Dice* de 83,11% e um índice *Jaccard* de 74,56% somente para a segmentação de lesões malignas.

Em [Kaur et al. 2022] é implementada uma arquitetura convolucional personalizada para segmentação câncer de pele. Essa nova arquitetura possui ajuste de hiper-

parâmetros e camadas convolucionais com função *LeakyReLU*, que é uma versão melhorada da ReLU. Essa arquitetura foi testada com as bases de imagens ISIC 2016, ISIC 2017 e ISIC 2018, alcançando um índice *Jaccard* de 90,4% na ISIC 2016, 81,8% na ISIC 2017 e 89,1% nos conjuntos de dados ISIC 2018, respectivamente, o que já foi registrado como um dos melhores resultados da competição ISIC.

Por fim, em [Karri et al. 2023], foi desenvolvido uma abordagem de segmentação de lesões de pele em duas fases, semelhante a [Bagheri et al. 2021], denominada de nSknRSUNet. O diferencial dessa arquitetura é seu funcionamento, baseado em segmentação semântica com módulos de atenção. Esses módulos proporcionam ênfase a dados importantes de bordas, combinando-os ao final da arquitetura. Além disso, é utilizado a técnica de transferência de aprendizado entre domínios com os conjuntos ImageNet, que é de imagens naturais, MoleMap e HAM10000, que são imagens de lesões de pele. Os melhores resultados apresentaram um *Dice* de 93,6%, com acurácia de 97%.

É perceptível que a utilização de arquitetura de aprendizado profundo com princípios de segmentação semântica proporcionam resultados animadores. Logo, é extremamente importante aplicá-las ao problema de segmentação do câncer maligno. É o caso dos princípios da arquitetura de [Chen et al. 2017], que já foi utilizada em [Goyal et al. 2019], [Kaur et al. 2022] e [Karri et al. 2023]. Nesse sentido, uma nova sistemática foi sugerida em [Chen et al. 2018], que implementa novas correções e promete melhores resultados de segmentação. Essa arquitetura é a *DeepLabV3+*, sendo um dos objetos de estudo da proposta deste trabalho.

3. Materiais e Métodos

A proposta deste trabalho é a análise da utilização da arquitetura *DeepLabV3+* para a segmentação semântica de lesões de pele. Para isso, foram selecionadas as principais bases de imagens gratuitas, disponíveis em projetos dermatológicos. Os experimentos realizados utilizaram o *backbone ResNet50*, previamente treinado com a base de imagens ImageNet [Deng et al. 2009], que extrai características dos *pixels* das imagens. Por fim, as máscaras obtidas com o modelo sugerido foram comparadas com as máscaras dos especialistas.

3.1. Bases de Imagens

As bases selecionadas para a execução da proposta deste trabalho são amplamente utilizadas no desenvolvimento de métodos de segmentação, são elas HAM10000 [Tschandl et al. 2018], ISIC 2016 [Gutman et al. 2016], ISIC 2017 [Codella et al. 2018], PH² [Mendonça et al. 2013], as quais foram divididas em três subconjuntos, apenas com lesões malignas, apenas com lesões benignas, com lesões benignas e malignas. O principal pré-requisito para a escolha é a disponibilização das máscaras de segmentação, desenvolvidas por médicos dermatologistas. A Tabela 1 mostra a distribuição de lesões de pele presente nas bases.

3.2. *DeeplabV3+*

A *DeepLabV3+* é uma arquitetura de segmentação semântica robusta, utilizada em uma diversidade de problemas complexos. As arquiteturas desse tipo concentram-se na classificação dos *pixels*, possibilitando mapas ótimos de separação. O funcionamento

Tabela 1. Distribuição das Bases de Imagens Sugeridas

Bases de Imagens	Lesões Benignas	Lesões Malignas	Total
HAM10000	8.902	1.113	10.015
ISIC 2016	727	173	900
ISIC 2017	1.626	374	2000
PH ²	160	40	200
	11.415	1.700	13.115

básico dessas arquiteturas está na utilização de um Codificador (*Encoder*) e um Decodificador (*Decoder*) [Chen et al. 2018].

Em outras palavras, a *DeepLabv3+* aplica duas operações conhecidas como *Downsampling* e *Upsampling*. No *Encoder*, são utilizadas camadas de convolução com a função *ReLU*, e sequencialmente a subamostragem com a *Max-Pooling*. Isso proporciona mapas de características (*features maps*) com resolução decrescente (*Downsampling*). Já a decodificação faz o oposto, com operações de convolução que promovem o aumento dos mapas de característica de forma crescente, até atingir a resolução da imagem de entrada (*Upsampling*) [Chen et al. 2018].

A utilização de *Downsampling* e *Upsampling* na arquitetura *DeepLabV3+* faz com que características invariantes sejam captadas na segmentação entre *Encoder* e *Decoder*. Com a diminuição da resolução dos mapas de características, as classes de *pixels* que representam objetos se tornam estáveis, por haver um menor tamanho de vizinhança por *pixel*. Após a definição de regiões, há apenas a extrapolação do tamanho dos mapas de características e a concatenação de áreas relevantes que representam objetos, mantendo a mesma sistemática de previsão de *pixels* por classe [Chen et al. 2018].

Um problema comum em segmentação semântica é o *downsampling* múltiplo. Isso quer dizer que a resolução dos mapas de características tem perda significativa de dados na convolução. O *downsampling* é comum em redes neurais convolucionais para segmentação, por evitar um alto consumo de memória com imagens de alta resolução [Azad et al. 2020]. No entanto, na arquitetura *DeepLabv3+* isso é tratado com uma espécie de convolução combinada com dilatação (*atrous convolution*). Isso faz com que haja um aumento gradativo dos mapas e muitas perdas na classificação dos *pixels* [Chen et al. 2018].

As arquiteturas baseadas na *DeepLab* necessitam de um *backbone* como estrutura de entrada. Esse nada mais é do que uma rede neural convolucional responsável por extrair características de uma lesão. Essa arquitetura proporciona uma possível classificação dos *pixels*, após a identificação das possíveis classes na imagem [Azad et al. 2020]. Geralmente, essas arquiteturas são pré-treinadas para facilitar outros processos. Por exemplo, a transferência de aprendizado e o ajuste fino [Alom et al. 2018].

Após a predição na *DeepLabV3+*, é necessário aplicar uma operação de limiarização, pois a imagem resultante não é binária. Assim, é possível produzir uma máscara destacando apenas as regiões de interesse na imagem. No problema de segmentação de lesões de pele essa etapa é extremamente necessária, pois as métricas de

avaliação do modelo dependem desse resultado [Karri et al. 2023].

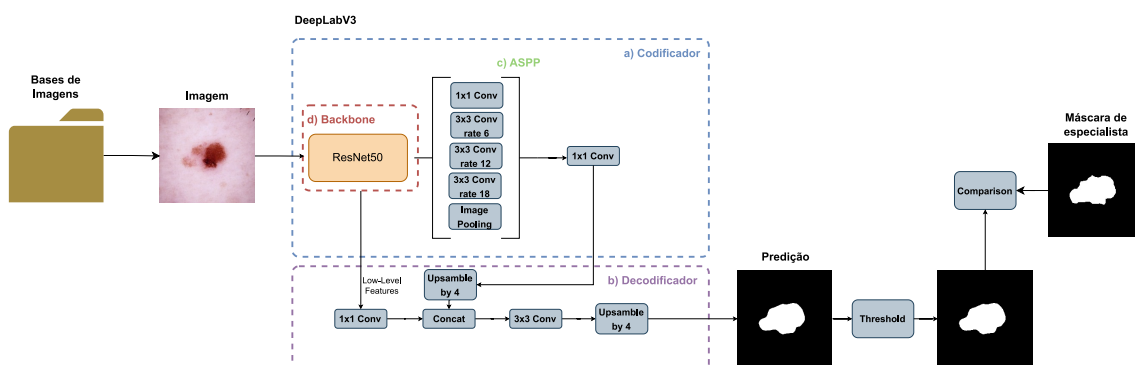


Figura 1. Proposta de Sistemática de Segmentação Semântica (Adaptado de [Chen et al. 2018]).

3.3. Métricas de Avaliação

As métricas sugeridas para avaliar a sistemática proposta é baseada na comparação entre máscaras. Isso quer dizer que as máscaras geradas nos modelos desenvolvidos serão comparadas com as máscaras disponibilizadas nas bases de imagens, e como dito antes, foram marcadas por médicos dermatologistas [Ronneberger et al. 2015]. Essa comparação é realizada *pixel a pixel* entre ambas as máscaras, proporcionando dados matemáticos que podem ser utilizados para avaliar o desempenho da proposta.

O produto da comparação entre as máscaras é a matriz de confusão. Essa matriz é composta pelas métricas de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN). A partir das métricas da matriz de confusão, pode-se obter outras métricas como, índice *Dice*, índice *Jaccard*, precisão, sensibilidade, especificidade e acurácia [Monard and Baranauskas 2003]. Essas métricas são as mais usadas na avaliação dos modelos de segmentação de lesões de pele. As próximas fórmulas matemáticas descrevem essas métricas.

$$Dice = \frac{2VP}{2VP + FP + FN} \quad (1)$$

$$Jaccard = \frac{VP}{VP + FP + FN} \quad (2)$$

$$Precisão = \frac{VP}{VP + FP} \quad (3)$$

$$Sensibilidade = \frac{VP}{VP + FN} \quad (4)$$

$$Especificidade = \frac{VN}{VN + FP} \quad (5)$$

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (6)$$

3.4. Configurações dos Experimentos e Treinamento

A arquitetura *DeepLabV3+* é arquitetura utilizada nos experimentos. O seu treinamento permite a criação de modelos específicos, que servem para segmentar lesões de pele. Dessa forma, o *backbone* sugerido nessa arquitetura é o *ResNet50* pré-treinado com o conjunto ImageNet [Deng et al. 2009]. Esse *backbone* mantém bom desempenho em diversas sistemáticas de segmentação semântica. Portanto, o treinamento realizado é baseado em transferência de aprendizado, a qual é o ajuste da estrutura ao problema específico.

A base utilizada para o treinamento foi a HAM10000, pois dentre todas sugeridas, possui uma maior quantidade de imagens por classes: maligna e benigna. O parâmetro utilizado para a definição dos conjuntos de treino, validação e teste é baseado na técnica de validação cruzada *holdout*. Para esse parâmetro, foram definidas 60% para treino, 20% para validação e 20% para teste. Dessa forma, é possível uma maior confiabilidade nos resultados gerados tanto no teste quanto na validação. As demais bases sugeridas foram utilizadas apenas como conjuntos de testes, com o principal propósito de avaliar os modelos gerados e suas eficácias. O método proposto foi treinado e avaliado em três cenários: (1) somente com lesões malignas, (2) somente com lesões benignas e (3) com lesões malignas e benignas.

Inicialmente, o treinamento foi definido para 100 épocas. Além disso, houve o monitoramento de dados utilizando o parâmetro *Dice loss*. Em outras palavras, esse parâmetro é o mesmo que $1 - Dice$. Já para a otimização dos modelos foi selecionado o *Adam*, com uma taxa inicial de aprendizado de 10^{-4} que pode decair até 10^{-7} . A definição do melhor modelo ocorreu com os melhores resultados da função de perda do conjunto de validação. Por fim, o melhor valor encontrado para *mini-batch* é 8 (*batch size*), pois esse valor maximizou o aprendizado dos modelos.

Por fim, após a execução do treinamento dos modelos, os melhores foram selecionados utilizando a *Dice loss* supracitada para os conjuntos de validação. A Figura 2 apresenta os gráficos obtidos nessa tarefa. Na Seção 4 são apresentados os resultados obtidos com o conjunto de teste da base HAM10000 e os subconjuntos das bases ISIC 2016, ISIC 2017 e PH².

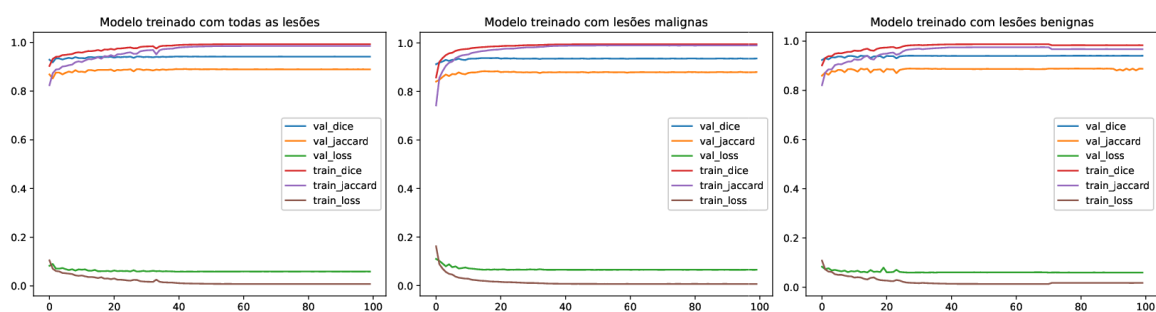


Figura 2. Gráficos de Treinamento dos Modelos

4. Resultados e Discussões

A partir das métricas sugeridas, das configurações de parâmetros e do treino dos modelos, os conjuntos de imagens foram escalonados e testados. Esses modelos foram executados

com lesões malignas (LM), com lesões benignas (LB), e com todas as lesões (TL) (malignas e benignas), pois devido à grande variedade de aparências presente nas lesões malignas, as quais podem apresentar bordas irregulares e difusas, tornando o contraste entre a pele ao redor e a lesão bem baixo e dificultando assim a segmentação [Yuan et al. 2017] como demonstrado na Figura 3.

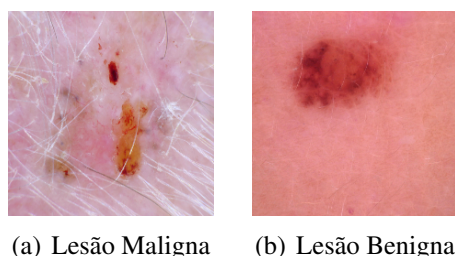


Figura 3. Comparativo entre os tipos de lesões

Na Tabela 2 é possível perceber que o modelo obteve bons resultados. Inclusive, obtivemos a média de *Dice* igual a 95,34% para todas as lesões, com uma média *jaccard* de 91,52% e acurácia média 96,92%. Além desses resultados, também é perceptível que esse modelo obteve melhores resultados sempre utilizando todos os tipos de lesões juntas. Isso é comprovado pela precisão maior que 95% e desvio padrão entre 0,05 e 0,14.

Tabela 2. Resultados de Segmentação no Conjunto de Teste da Base HAM10000.

Dataset	Modelo	<i>Dice</i> (%)	<i>Jaccard</i> (%)	Precisão(%)	Sensitividade(%)	Especificidade(%)	Acurácia(%)
HAM10000 Teste TL	TL	94,42 ± 0,08	90,18 ± 0,11	95,46 ± 0,07	94,38 ± 0,09	98,02 ± 0,05	97,21 ± 0,04
	LM	92,55 ± 0,10	87,29 ± 0,13	93,52 ± 0,09	93,11 ± 0,12	97,76 ± 0,05	96,29 ± 0,06
	LB	94,42 ± 0,08	90,20 ± 0,11	95,44 ± 0,07	94,42 ± 0,09	98,01 ± 0,05	97,19 ± 0,04
HAM10000 Teste LM	TL	95,34 ± 0,05	91,52 ± 0,08	95,87 ± 0,06	95,38 ± 0,07	96,95 ± 0,10	96,92 ± 0,05
	LM	94,34 ± 0,07	89,97 ± 0,10	95,28 ± 0,06	94,38 ± 0,09	96,95 ± 0,09	96,39 ± 0,05
	LB	94,79 ± 0,06	90,66 ± 0,10	95,83 ± 0,06	94,50 ± 0,09	97,01 ± 0,09	96,59 ± 0,05
HAM10000 Teste LB	TL	94,31 ± 0,08	90,01 ± 0,11	95,41 ± 0,07	94,26 ± 0,09	98,16 ± 0,05	97,24 ± 0,04
	LM	92,33 ± 0,10	86,95 ± 0,13	93,30 ± 0,10	92,95 ± 0,12	97,86 ± 0,04	96,27 ± 0,06
	LB	94,38 ± 0,08	90,15 ± 0,11	95,39 ± 0,08	94,40 ± 0,10	98,13 ± 0,05	97,26 ± 0,04

Na Tabela 3 são apresentados os resultados obtidos com o modelo nos subconjuntos de imagens da base ISIC 2016. Para esses experimentos, os resultados mostram um *Dice* médio entre 87% e 88,5% para todas as lesões. Já para os subconjuntos de lesões malignas e benignas, foram obtidos resultados próximos, mesmo com um desbalanceamento entre esses conjuntos. Inclusive, é notável os resultados de acurácias médias maiores que 94%. No entanto, os resultados de precisão são reduzidos, mantendo-se entre 83% e 87%.

Já a Tabela 4 apresenta os resultados alcançados com todos os conjuntos de imagens da base ISIC 2017. No caso de segmentação dessa base, os resultados alcançados foram melhores, se comparado com a ISIC 2016. Na divisão das classes de lesões benignas e malignas, foram obtidos um *Dice* médio de 90% e 92,6%, com desvio padrão entre 0,1 e 0,2. Já para todas as lesões, é notável uma precisão 86,98%. A acurácia média se manteve entre 94% e 95%.

Por fim, a Tabela 5 mostra os resultados obtidos com os modelos na base de imagens PH². O melhor resultado alcançado mostra um *Dice* médio de 93,11%, com desvio

Tabela 3. Resultados de Segmentação dos Modelos com as Imagens da Base ISIC 2016.

Dataset	Modelo	<i>Dice</i> (%)	<i>Jaccard</i> (%)	Precisão(%)	Sensitividade(%)	Especificidade(%)	Acurácia(%)
ISIC 2016 TL	TL	91,68 ± 0,08	85,43 ± 0,11	93,20 ± 0,10	91,75 ± 0,11	97,50 ± 0,05	96,05 ± 0,05
	LM	89,80 ± 0,10	82,72 ± 0,13	91,56 ± 0,11	90,16 ± 0,14	97,80 ± 0,03	95,22 ± 0,07
	LB	89,01 ± 0,16	82,74 ± 0,18	90,28 ± 0,17	89,88 ± 0,17	97,09 ± 0,05	95,72 ± 0,05
ISIC 2016 LM	TL	91,67 ± 0,08	85,42 ± 0,11	93,99 ± 0,08	91,08 ± 0,12	96,15 ± 0,06	94,43 ± 0,06
	LM	90,92 ± 0,08	84,17 ± 0,11	93,99 ± 0,07	89,83 ± 0,13	97,01 ± 0,04	93,67 ± 0,07
	LB	91,62 ± 0,08	85,35 ± 0,11	93,80 ± 0,08	91,22 ± 0,12	96,21 ± 0,05	94,28 ± 0,06
ISIC 2016 LB	TL	91,68 ± 0,08	85,44 ± 0,11	93,01 ± 0,10	91,90 ± 0,10	97,83 ± 0,04	96,44 ± 0,05
	LM	89,53 ± 0,11	82,38 ± 0,14	90,98 ± 0,11	90,23 ± 0,14	97,99 ± 0,03	95,58 ± 0,07
	LB	88,39 ± 0,18	82,12 ± 0,19	89,45 ± 0,18	89,56 ± 0,18	97,30 ± 0,05	96,07 ± 0,05

Tabela 4. Resultados de Segmentação dos Modelos com as Imagens da Base ISIC 2017.

Dataset	Modelo	<i>Dice</i> (%)	<i>Jaccard</i> (%)	Precisão(%)	Sensitividade(%)	Especificidade(%)	Acurácia(%)
ISIC 2017 TL	TL	87,19 ± 0,16	79,89 ± 0,19	85,35 ± 0,19	93,82 ± 0,11	96,29 ± 0,08	95,69 ± 0,07
	LM	84,16 ± 0,20	76,38 ± 0,22	82,88 ± 0,22	90,45 ± 0,18	96,83 ± 0,05	95,56 ± 0,06
	LB	85,44 ± 0,19	78,12 ± 0,21	83,28 ± 0,22	93,10 ± 0,14	95,61 ± 0,09	95,12 ± 0,08
ISIC 2017 LM	TL	88,46 ± 0,12	81,01 ± 0,16	86,98 ± 0,16	93,39 ± 0,11	95,33 ± 0,07	94,67 ± 0,06
	LM	87,87 ± 0,13	80,15 ± 0,16	87,07 ± 0,15	91,93 ± 0,13	96,15 ± 0,05	94,65 ± 0,06
	LB	88,26 ± 0,14	80,96 ± 0,17	86,76 ± 0,16	93,24 ± 0,12	95,18 ± 0,07	94,60 ± 0,07
ISIC 2017 LB	TL	86,90 ± 0,17	79,63 ± 0,19	84,97 ± 0,20	93,93 ± 0,10	96,52 ± 0,08	95,92 ± 0,07
	LM	83,31 ± 0,21	75,51 ± 0,23	81,92 ± 0,23	90,11 ± 0,18	96,99 ± 0,05	95,77 ± 0,06
	LB	84,79 ± 0,20	77,46 ± 0,22	82,48 ± 0,23	93,07 ± 0,14	95,71 ± 0,10	95,24 ± 0,09

padrão de 0,05, precisão de 88,19% e acurácia média de 96,92%. Analisando os demais casos para essa base, é perceptível que foram próximos do conjunto de teste da base HAM10000. Isso é justificável porque a quantidade de imagens é reduzida, sendo equiparada em tamanho com a base PH².

Tabela 5. Resultados de Segmentação dos Modelos com as Imagens da Base PH².

Dataset	Modelo	<i>Dice</i> (%)	<i>Jaccard</i> (%)	Precisão(%)	Sensitividade(%)	Especificidade(%)	Acurácia(%)
PH2 TL	TL	92,18 ± 0,07	86,20 ± 0,11	89,11 ± 0,08	96,83 ± 0,09	94,36 ± 0,06	94,87 ± 0,08
	LM	90,30 ± 0,08	83,18 ± 0,12	86,98 ± 0,10	95,76 ± 0,11	94,23 ± 0,05	93,67 ± 0,09
	LB	91,80 ± 0,08	85,71 ± 0,11	88,71 ± 0,09	96,71 ± 0,09	93,87 ± 0,07	94,46 ± 0,09
PH2 LM	TL	88,47 ± 0,13	81,29 ± 0,17	92,78 ± 0,10	87,93 ± 0,18	89,49 ± 0,08	86,64 ± 0,14
	LM	86,62 ± 0,13	78,25 ± 0,17	93,40 ± 0,09	84,31 ± 0,19	92,07 ± 0,07	83,89 ± 0,16
	LB	88,75 ± 0,12	81,50 ± 0,16	93,27 ± 0,08	87,59 ± 0,18	89,95 ± 0,07	86,61 ± 0,14
PH2 LB	TL	93,11 ± 0,05	87,43 ± 0,08	88,19 ± 0,08	99,05 ± 0,02	95,57 ± 0,04	96,92 ± 0,02
	LM	91,22 ± 0,06	84,41 ± 0,09	85,38 ± 0,09	98,62 ± 0,05	94,77 ± 0,04	96,11 ± 0,03
	LB	92,57 ± 0,07	86,76 ± 0,09	87,57 ± 0,09	98,99 ± 0,03	94,85 ± 0,07	96,43 ± 0,05

Os resultados obtidos apresentam algumas particularidades. Uma delas é que os modelos gerados *DeepLabV3+* são estáveis, mesmo com as bases desbalanceadas em todos os casos testados. Em alguns casos, é perceptível resultados de qualidade inferior, principalmente na base ISIC 2017. Isso é comprovado pelo *Dice* médio de 83,31%, desvio padrão de 0,21 e *Jaccard* de 75,51%. Alguns desses casos podem ser justificados pela qualidade das máscaras produzidas pelos modelos.

Para permitir uma melhor análise dos resultados, na base HAM10000, o método proposto foi comparado com nSknRSUNet [Karri et al. 2023], MFSNet [Basak et al. 2022], Rema-Net [Yang et al. 2023]. Em Rema-Net, alcançou-se um *Dice*

igual a 93,59%, a MFSNet, 90,60%, ambos com valores inferiores ao *Dice* de 94,42% do método proposto. Já a nSknRSUNet, mesmo bem próxima, obteve um resultado superior, alcançando um *Dice* de 94,60%.

5. Conclusões e Trabalhos Futuros

O sucesso de metodologias automáticas de classificação de lesões de pele depende da segmentação. Nossos experimentos demonstraram que essa tarefa é promissora com a arquitetura *DeepLabV3+*. Os resultados apresentados sugerem que o uso da segmentação semântica consegue atingir melhores definições entre bordas de lesões de pele, possibilitando uma separação mais próxima do especialista, já que os *pixels* são interpretados entre classes na imagem.

A utilização de um único conjunto de imagens para treino permitiu uma boa generalização para as demais bases de imagens sugeridas. Os modelos desenvolvidos conseguem produzir boas máscaras de segmentação, mesmo sem um pré-processamento prévio. Isso foi possível a partir de uma boa definição de hiper-parâmetros. Alguns deles são o número de épocas, o tamanho dos lotes, a taxa de aprendizado e a precisão no conjunto de validação. O monitoramento da *Dice loss* do conjunto de validação diminuiu as chances de ocorrência do *underfitting* e *overfitting* e pode possibilitar a parada antecipada.

Mesmo com as dificuldades na segmentação de lesões malignas como citado na Seção 1, o modelo treinado com todas as lesões proporcionou resultados superiores se comparado com os resultados individuais entre as classes maligna e benigna. Porém, vale ressaltar, que em alguns, o modelo treinado com apenas lesões benignas conseguiu se equiparar e até mesmo ultrapassar o modelo treinado com todas as lesões. Além disso, o modelo treinado com apenas lesões benignas foi superior ao treinado com apenas lesões malignas, mesmo em conjuntos das bases de imagens com apenas lesões malignas. Por fim, através do desvio padrão aceitável, os modelos demonstraram uma constância em seus resultados.

Para trabalhos futuros, é necessário a utilização de técnicas de aumento de dados dinâmica para superar o problema do desbalanceamento entre classes. Novos resultados mais precisos podem ser obtidos. Além disso, modelos de aprendizado profundo devem focar no tratamento aprimorado das imagens, para identificar características relevantes por canal com módulos de atenção. Nesse sentido, recomenda-se o treinamento com uma nova estrutura personalizada da arquitetura sugerida. Inclusive, uma variação dessa abordagem está sendo preparada para segmentar a base de imagens ISIC 2020.

Referências

- Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*.
- Azad, R., Asadi-Aghbolaghi, M., Fathy, M., and Escalera, S. (2020). Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation. In *Computer Vision—ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 251–266. Springer.

- Bagheri, F., Tarokh, M. J., and Ziaratban, M. (2021). Skin lesion segmentation from dermoscopic images by using mask r-cnn, retina-deeplab, and graph-based methods. *Biomedical Signal Processing and Control*, 67:102533.
- Basak, H., Kundu, R., and Sarkar, R. (2022). Mfsnet: A multi focus segmentation network for skin lesion segmentation. *Pattern Recognition*, 128:108673.
- Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., and Liao, W. (2020). Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatology and therapy*, 10:365–386.
- Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.
- Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kallou, A., Liopyris, K., Mishra, N., Kittler, H., et al. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE.
- de Oliveira Santos, M., de Lima, F. C. d. S., Martins, L. F. L., Oliveira, J. F. P., de Almeida, L. M., and de Camargo Cancela, M. (2023). Estimativa de incidência de câncer no brasil, 2023-2025. *Revista Brasileira de Cancerologia*, 69(1).
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., and Mahnashi, M. H. (2021). Skin cancer detection: a review using deep learning techniques. *International journal of environmental research and public health*, 18(10):5479.
- Fan, H., Xie, F., Li, Y., Jiang, Z., and Liu, J. (2017). Automatic segmentation of dermoscopy images using saliency combined with otsu threshold. *Computers in biology and medicine*, 85:75–85.
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., and Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International journal of cancer*, 149(4):778–789.
- Ge, Z., Demyanov, S., Bozorgtabar, B., Abedini, M., Chakravorty, R., Bowling, A., and Garnavi, R. (2017). Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, pages 986–990. IEEE.
- Goyal, M., Oakley, A., Bansal, P., Dancey, D., and Yap, M. H. (2019). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access*, 8:4171–4181.

- Gutman, D., Codella, N. C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., and Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397*.
- Hong, Y., Zhang, G., Wei, B., Cong, J., Xu, Y., and Zhang, K. (2022). Weakly supervised semantic segmentation for skin cancer via cnn superpixel region response. *Multimedia Tools and Applications*, pages 1–19.
- Isa, N. A. M., Salamah, S. A., and Ngah, U. K. (2009). Adaptive fuzzy moving k-means clustering algorithm for image segmentation. *IEEE Transactions on Consumer Electronics*, 55(4):2145–2153.
- Karri, M., Annavarapu, C. S. R., and Acharya, U. R. (2023). Skin lesion segmentation using two-phase cross-domain transfer learning framework. *Computer Methods and Programs in Biomedicine*, page 107408.
- Kaur, R., GholamHosseini, H., Sinha, R., and Lindén, M. (2022). Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images. *BMC Medical Imaging*, 22(1):1–13.
- MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297. University of California Los Angeles LA USA.
- Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., and Rozeira, J. (2013). Ph 2-a dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 5437–5440. IEEE.
- Monard, M. C. and Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.
- Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The ham10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9.
- Yang, L., Fan, C., Lin, H., and Qiu, Y. (2023). Rema-net: An efficient multi-attention convolutional neural network for rapid skin lesion segmentation. *Computers in Biology and Medicine*, 159:106952.
- Yuan, Y., Chao, M., and Lo, Y.-C. (2017). Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Transactions on Medical Imaging*, 36(9):1876–1886.