

Anotação de imagens médicas assistida por IA: um estudo sobre segmentação de lesões de pele por não especialistas

Lorenzo M. Scaramussa¹, Andre G. C. Pacheco¹

¹Departamento de Informática e Programa de Pós-Graduação em Informática
Universidade Federal do Espírito Santo (UFES), Vitória, Espírito Santo, Brasil

lorenzo.scaramussa@edu.ufes.br, apacheco@inf.ufes.br

Abstract. *Image annotation is essential for building databases for training artificial intelligence (AI) algorithms. However, the dependence on specialists makes this process expensive and difficult to scale. This work proposes an interactive framework to assist non-specialists in segmenting skin lesions. Its effectiveness and usability were evaluated in an experiment with 50 volunteers, who segmented skin lesions in two modes: manual and assisted by the tool. The results indicate that the assisted approach improves efficiency without compromising accuracy, especially when combined with crowdsourcing. The proposed tool is available in open-source format and can be used in different medical areas beyond dermatology.*

Resumo. A anotação de imagens médicas é essencial para a construção de bases de dados destinadas ao treinamento de algoritmos de inteligência artificial (IA). No entanto, a dependência de especialistas torna esse processo caro e difícil de escalar. Este trabalho propõe um framework interativo para auxiliar não especialistas na segmentação de lesões de pele. Sua eficácia e usabilidade foram avaliadas em um experimento com 50 voluntários, que segmentaram lesões de pele em dois modos: manual e assistido pela ferramenta. Os resultados indicam que a abordagem assistida melhora a eficiência sem comprometer a precisão, especialmente quando combinada com crowdsourcing. A ferramenta proposta está disponível em formato open-source e pode ser utilizada em diversas áreas médicas além da dermatologia.

1. Introdução

Nos últimos anos, a inteligência artificial (IA) vem avançando significativamente na área da saúde. De acordo com o relatório *State of AI* de 2024 [Benaich 2024], a área de saúde foi a segunda que mais recebeu investimentos ativos na indústria de IA, sendo que, nos anos anteriores, esses valores superaram a casa de 6 bilhões de dólares [Clark et al. 2023]. Esses investimentos vêm impulsionando a criação de diversas ferramentas que auxiliam profissionais em diferentes tarefas, como monitoramento de pacientes [Shaik et al. 2023], descoberta e desenvolvimento de medicamentos [Abramson et al. 2024] e diagnóstico assistido [Costa et al. 2021].

Na área de imagens médicas, os Sistemas de Diagnóstico Assistido por Computador — do inglês *Computer-Aided Diagnosis* (CAD) — vêm sendo desenvolvidos para apoiar a tomada de decisão clínica em diversas condições. Exemplos incluem a identificação de invasão submucosa no câncer de Barrett [Souza Jr et al. 2024], avanços em radiologia, como a detecção de câncer de mama e doenças pulmonares

[Najjar 2023], a identificação de hemorragia cerebral [Kar et al. 2024], entre muitos outros [Barragán-Montero et al. 2021]. Especificamente na dermatologia, houve progressos significativos em técnicas de segmentação, identificação e classificação de lesões cutâneas. Isso inclui o uso de *deep learning* para segmentação de lesões de pele [Sarwar et al. 2024] e a aplicação de modelos com mecanismos de atenção para o diagnóstico de câncer de pele [Pacheco and Krohling 2021]. Independentemente da subárea de aplicação dos algoritmos de IA, um fator comum a todas elas é a necessidade de dados anotados e de alta qualidade. Sem essa premissa, o desenvolvimento desses modelos torna-se inviável.

Diferentes bases de dados de imagens médicas vêm se destacando pelo seu papel fundamental no avanço da IA aplicada à saúde. Especificamente na área de dermatologia, bases como o HAM10000 [Tschantl et al. 2018] e o PAD-UFES-20 [Pacheco et al. 2020] são amplamente utilizadas no treinamento de algoritmos de classificação e segmentação de lesões, contribuindo para o avanço na detecção de melanomas e outras condições cutâneas. No entanto, a construção de bases de dados, como as citadas, é um desafio amplamente reconhecido, visto que o processo demanda tempo e recursos significativos pois, na maioria dos casos depende da atuação intensa de especialistas para anotar os dados [Das et al. 2021]. Com intuito de enfrentar essa limitação, diversas iniciativas têm explorado a utilização de mão de obra não especialista na anotação de imagens. Dentre essas estratégias, o *crowdsourcing* emerge como uma alternativa interessante para expandir a disponibilidade de novas bases de dados [Cocos et al. 2017].

Diversos trabalhos vêm investigando métodos baseados em *crowdsourcing* para a criação de conjuntos de dados. Um exemplo notável é o estudo conduzido por pesquisadores da Google, que desenvolveu um banco de dados dermatológico por meio da colaboração de usuários online, utilizando anúncios para recrutar participantes [Ward et al. 2024]. Entretanto, essa abordagem resultou em um conjunto de dados com viés demográfico relacionado à idade e gênero, além de exigir investimento contínuo para a coleta de informações. Outro trabalho relevante, realizado por Veronika et al. [Cheplygina et al. 2021], demonstrou que indivíduos sem formação específica podem contribuir para a anotação de imagens médicas, particularmente na segmentação de vias aéreas em tomografias de tórax. Contudo, o método enfrentou obstáculos consideráveis, como a complexidade das orientações fornecidas, que resultou em um grande volume de anotações inadequadas, e a necessidade de agregar múltiplas contribuições para alcançar maior confiabilidade. Adicionalmente, observou-se que a precisão das segmentações está diretamente relacionada ao nível de treinamento oferecido aos colaboradores [Kentley et al. 2023].

A segmentação de imagens desempenha um papel importante na análise de imagens médicas, com aplicações relevantes em contextos clínicos, como na análise de função cardíaca para diagnóstico de doença cardíaca [Ouyang et al. 2020] e diagnóstico de doença de retina [De Fauw et al. 2018]. Contudo, a criação de máscaras de segmentação — representações que delimitam regiões de interesse em uma imagem — depende fortemente da contribuição de especialistas, o que torna o processo caro e limitante em escala. Assim como os trabalhos baseados em *crowdsourcing* já mencionados, outros estudos recentes exploram o uso de não especialistas como alternativa para anotação na área de segmentação de imagens [Damgaard et al. 2023]. Neste sentido, este trabalho propõe a criação de um *framework* para auxiliar não especialistas na geração de máscaras de

segmentação de lesões cutâneas. Este *framework* utiliza como base o *Segment Anything Model* (SAM) [Kirillov et al. 2023], um modelo no estado da arte que utiliza prompts de entrada para atuar na segmentação universal sem a necessidade de ajuste fino para tarefas específicas. A ideia da ferramenta proposta é combinar a automação com a intervenção humana, por meio de um sistema conhecido como *human-in-the-loop*, que neste caso, consiste em uma segmentação assistida, na qual um não especialista participa do processo de criação da máscara de segmentação através da indicação de áreas de região de interesse nas imagens médicas. O objetivo desta abordagem é reduzir o tempo e custos associados à anotação manual de bases de dados de imagens e, ao mesmo tempo, manter padrões de qualidade compatíveis com as exigências clínicas.

O *framework* proposto foi avaliado utilizando imagens clínicas de lesões de pele obtidas da base de dados pública PAD-UFES-20 [Pacheco et al. 2020]. Foi conduzido um experimento na qual 50 voluntários não especialistas utilizaram a ferramenta para gerar a segmentação das lesões de pele seguindo protocolo pré-estabelecido. As anotações obtidas foram comparadas com as máscaras de segmentação geradas por especialistas para mediar quantitativamente a eficácia da ferramenta. Os resultados preliminares indicam que o *framework* pode auxiliar efetivamente na geração de máscaras de segmentação, promovendo um equilíbrio entre eficiência e precisão. Por fim, o sistema está disponibilizado em formato *open-source* com licença CC-BY, possibilitando sua melhoria, expansão e utilização em outras áreas médicas além da dermatologia¹.

2. Métodos

Nesta seção, são apresentados os métodos e ferramentas utilizados no desenvolvimento do *framework* para segmentação de lesões cutâneas. Aborda-se o problema de segmentação de imagens, as métricas de avaliação, o *Segment Anything Model* (SAM) como base do sistema, e o *framework* proposto, que combina automação e intervenção humana para otimizar a anotação por não especialistas.

2.1. Segmentação de imagens e métricas

A segmentação de imagens consiste em dividir uma imagem em regiões distintas com base em características como cor, textura ou intensidade, visando facilitar a análise e interpretação de dados visuais. Seu objetivo é isolar objetos ou áreas de interesse para processamento independente, com diversas técnicas desenvolvidas para diferentes contextos [Tao Lei 2023, Yu et al. 2023].

Na área médica, a segmentação é crucial para diagnóstico, planejamento terapêutico e monitoramento de doenças. Imagens médicas, como as dermatológicas, apresentam desafios únicos, como estruturas complexas, variações de textura e bordas irregulares, exigindo precisão para distinguir regiões saudáveis e afetadas. A Figura 1 ilustra uma lesão de pele e sua máscara de segmentação anotada por um especialista.

Uma vez que existe uma máscara de segmentação anotada para uma imagem, é possível medir a qualidade de outras máscaras da mesma imagem – geradas por algoritmos ou manualmente – utilizando diferentes métricas. Neste trabalho, são utilizadas as seguintes métricas: *Intersection over Union* (IoU), *Dice Similarity Coefficient* (DSC) e *Normalized Surface Distance* (NSD). Elas foram escolhidas de acordo com análise de métricas ideais para o problema a ser abordado neste trabalho [Maier-Hein et al. 2024].

¹<https://github.com/life-ufes/segmentation-tool>



(a) Lesão de pele (b) Máscara de segmentação

Figura 1. Exemplo de uma lesão de pele e sua respectiva máscara de segmentação gerada por um especialista na área.

A combinação delas permite uma avaliação abrangente, equilibrando a análise de sobreposição global e a precisão de bordas. A seguir, cada uma das métricas é brevemente apresentada.

Intersection over Union (IoU): mede a sobreposição espacial entre a máscara predita (A) e a máscara de referência (B), sendo calculada como:

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (1)$$

onde $|A \cap B|$ é o número de pixels na interseção e $|A \cup B| = |A| + |B| - |A \cap B|$ representa a união das áreas. O IoU varia entre 0 (sem sobreposição) e 1 (sobreposição perfeita).

O *Dice Similarity Coefficient* (DSC) também quantifica a similaridade entre as regiões segmentadas, sendo definido por:

$$\text{DSC}(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}. \quad (2)$$

Assim como o IoU, o DSC varia entre 0 e 1, com 1 indicando segmentações idênticas. O DSC tende a ser mais tolerante a pequenas discrepâncias devido à sua ponderação linear da interseção em relação ao tamanho das regiões.

Normalized Surface Distance (NSD) avalia a precisão dos contornos das máscaras, comparando as distâncias entre as bordas da segmentação predita (∂A) e da referência (∂B). Seja $D_{\partial B}(p)$ a distância Euclidiana do ponto p na borda ∂A até a borda mais próxima em ∂B , e vice-versa para $D_{\partial A}(q)$ [Nikolov et al. 2021]. Para um limiar de tolerância τ , o NSD é calculado como:

$$\text{NSD}(A, B) = \frac{\sum_{p \in \partial A} \mathbb{I}(D_{\partial B}(p) \leq \tau) + \sum_{q \in \partial B} \mathbb{I}(D_{\partial A}(q) \leq \tau)}{|\partial A| + |\partial B|}, \quad (3)$$

onde $\mathbb{I}(\cdot)$ é a função indicadora (1 se a condição for verdadeira, 0 caso contrário). O NSD varia entre 0 e 1 e é especialmente útil para aplicações onde a precisão de contornos é crítica, como em diagnósticos médicos.

2.2. Segment Anything Model (SAM)

O *Segment Anything Model* (SAM) [Kirillov et al. 2023] é um modelo de segmentação universal baseado em *Vision Transformers* (ViT) [Dosovitskiy et al. 2021]. Desenvolvido

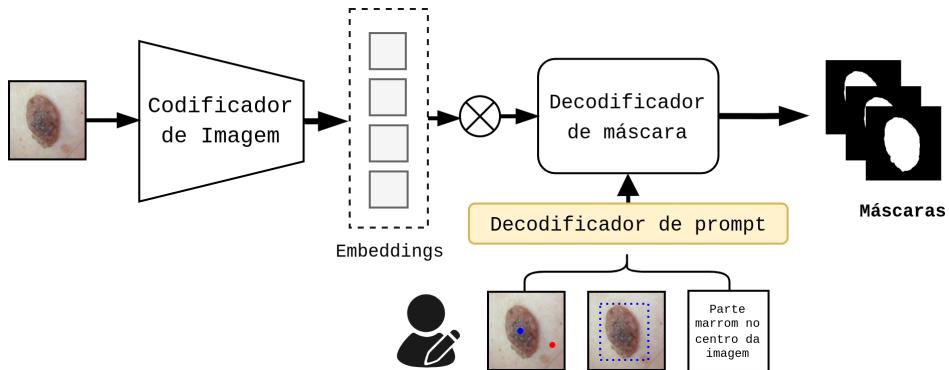


Figura 2. Diagrama ilustrativo do funcionamento do SAM adaptado para o contexto deste trabalho.

para lidar com qualquer tarefa de segmentação, o SAM gera máscaras de segmentação a partir de *prompts* simples, como pontos, caixas delimitadoras ou texto. Sua arquitetura é composta por três componentes principais: um codificador de imagem, um codificador de *prompt* e um decodificador. O codificador de imagem extrai características visuais (*embeddings*) de alta resolução da imagem de entrada. O codificador de *prompt* processa as entradas do usuário (pontos, caixas ou texto) e as mapeia para um espaço de características compatível com o codificador de imagem. O decodificador combina as características da imagem e do *prompt* para gerar máscaras de segmentação precisas.

Uma dos principais benefícios do SAM é sua capacidade de generalização: ele pode segmentar objetos em imagens de diversos domínios sem a necessidade de ajuste fino, tornando-o viável para aplicações assistidas na área médica. A Figura 2 ilustra um diagrama do SAM adaptado para o contexto deste trabalho. Nela, é possível observar os três componentes principais previamente descritos e como eles interagem entre si.

2.3. Framework Proposto

O framework proposto neste trabalho é uma plataforma web interativa desenvolvida para otimizar a segmentação de imagens, priorizando usabilidade e eficiência. O sistema foi concebido para facilitar sua replicação e implantação por meio de *containers Docker* e é totalmente *open source*, permitindo a personalização e substituição de componentes conforme necessário. A plataforma oferece dois modos de segmentação: manual e assistido por IA, ambos integrados em uma interface intuitiva que prioriza a experiência do usuário. Além disso, ele foi projetado para ser generalista, podendo ser aplicado a uma ampla gama de tarefas de segmentação, não se limitando a imagens médicas. A Figura 3 ilustra o funcionamento da plataforma em ambos os modos de segmentação.

No modo assistido por IA, o *Segment Anything Model* (SAM) é utilizado como núcleo do processo de segmentação. Durante o pré-processamento, as imagens são submetidas ao codificador de imagem do SAM para gerar seus respectivos *embeddings*, que são armazenados para uso posterior. Quando o usuário fornece um *prompt* – neste caso, uma caixa delimitadora –, o sistema utiliza o codificador de *prompt* e o decodificador de máscaras do SAM para prever a segmentação mais adequada com base no *embedding* da imagem e no *prompt* fornecido. A máscara com maior confiança é selecionada dentro da área indicada pelo usuário, permitindo uma experiência interativa e eficiente. Caso o resultado não seja satisfatório, o usuário pode ajustar a área selecionada e reprocessar a imagem, garantindo maior precisão na anotação.

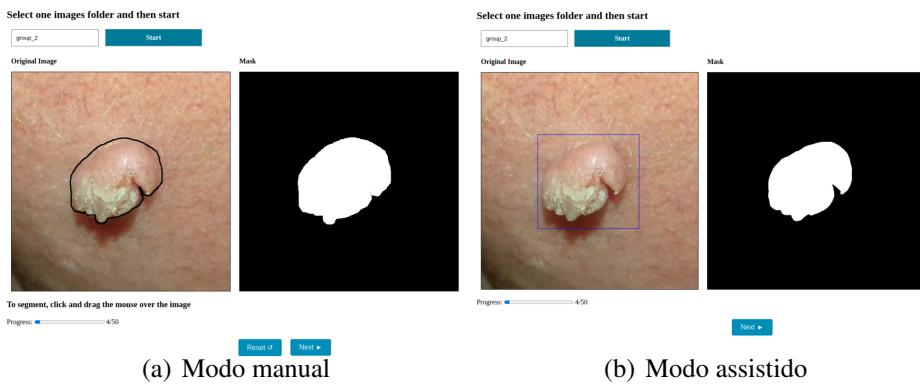


Figura 3. Ilustração dos dois modos de segmentação possível no framework proposto. No modo manual, o anotador é responsável por demarcar toda área da lesão. Já no modo assistido, ele indica a região através de uma caixa delimitadora.

No modo manual, o usuário pode desenhar livremente o contorno da lesão diretamente na interface, com a possibilidade de desfazer traços indesejados a qualquer momento. Ao concluir o desenho, um algoritmo de preenchimento automático completa a segmentação, gerando uma máscara a partir das bordas definidas pelo usuário. Esse modo é particularmente útil para casos em que a segmentação requer um alto grau de controle por parte do anotador. Porém, é uma tarefa mais demorada e exige maior habilidade do usuário para obter resultados precisos.

3. Experimentos e Resultados

Nesta seção, são apresentados os experimentos conduzidos para avaliar o *framework* proposto e os resultados obtidos. A intenção principal é investigar o desempenho de não especialistas na criação de máscaras de segmentação de lesões de pele com e sem o auxílio do *framework* descrito na seção anterior. O objetivo principal é avaliar se, na área de dermatologia, não especialistas podem ser úteis na anotação de dados.

3.1. Protocolo de avaliação

Um protocolo de avaliação foi desenvolvido para avaliar a eficácia e usabilidade da ferramenta de segmentação de lesões de pele por voluntários não especialistas, comparando os resultados obtidos nos modos manual e assistido por IA. Para isso, foram selecionados 50 voluntários sem formação em dermatologia, mas com habilidades básicas no uso de computadores, que utilizaram um desktop com mouse comum durante o estudo. O protocolo foi dividido em três etapas principais: treinamento, tarefas de segmentação e avaliação pós-tarefa, com duração média total de aproximadamente 30 minutos. Cada uma das etapas é descrita a seguir:

Etapa de treinamento: nesta etapa os voluntários foram introduzidos ao *framework* desenvolvido para entender como era o seu funcionamento geral. Porém, nenhuma orientação específica sobre como identificar e/ou delimitar as lesões foi fornecida.

Etapa de segmentação: nesta etapa cada voluntário utilizou a ferramenta para segmentar 50 lesões de pele em dois modos: manual e assistida. No modo manual, como o nome sugere, ele demarcava manualmente toda área da lesão. No modo assistido, ele indicava para o sistema, através de uma caixa de seleção, onde a lesão estava e o modelo era

responsável por realizar o restante da segmentação. O tempo gasto e as máscaras geradas foram armazenadas para análise posterior. As 50 imagens utilizadas nos modos manual e assistido foram as mesmas e apresentadas na mesma ordem para cada usuário.

Etapa de avaliação: após realizar a tarefa de segmentação, todos os voluntários preencheram um formulário de avaliação, fornecendo *feedback* sobre usabilidade, utilidade e percepção da ferramenta. Este formulário está disponível como material suplementar no repositório do projeto. Finalizadas as três etapas, o voluntário era liberado.

3.2. Base de dados

Para realização do protocolo previamente descrito, foi utilizada uma parte do conjunto de dados do dataset PAD-UFES-20 [Pacheco et al. 2020]. Esta base de dados é composta por imagens clínicas de lesões de pele e foi desenvolvido pelo Programa de Assistência Dermatológica e Cirúrgica (PAD) da Universidade Federal do Espírito Santo (UFES) para fomentar pesquisa e desenvolvimento tecnológico na área de dermatologia. A base possível amostras de três tipos de lesões benignas – nevo, ceratose actínica e ceratose seborreica – e três tipos malignas – carcinoma basocelular, carcinoma espinocelular e melanoma. Porém, na base original, não existem máscaras de segmentação para as lesões. Por conta disso, as máscaras utilizadas na avaliação advém do trabalho de [De Angelo et al. 2019], que obteve máscaras de segmentação por meio de especialistas para parte das lesões presentes no PAD-UFES-20.

Sendo assim, para a realização dos experimentos, foram selecionado quatro grupos de 50 imagens de lesões de pele, distribuídas entre os seis tipos de lesão presentes na base, para serem utilizadas na avaliação do *framework* conforme descrito no protocolo. Cada voluntário realizava a segmentação de um único grupo de imagens, uma vez no modo manual e outra no modo assistido. A escolha do grupo para o voluntário foi feita de forma aleatória, garantindo que todos os grupos fossem utilizados de forma proporcional. Todos os dados utilizados são públicos e disponibilizados juntamente do *framework* no repositório do projeto.

3.3. Resultados

As máscaras de segmentação geradas pelos voluntários, seguindo o protocolo previamente descrito, foram comparadas com as máscaras de referência fornecidas por especialistas, utilizando as métricas IoU (*Intersection over Union*), DSC (*Dice Similarity Coefficient*) e NSD (*Normalized Surface Distance*). Na Figura 4 é ilustrado um exemplo do processo para um único voluntário. Nela, é possível observar a imagem da lesão a ser anotada, a anotação manual do voluntário, a anotação assistida e anotação do especialista.

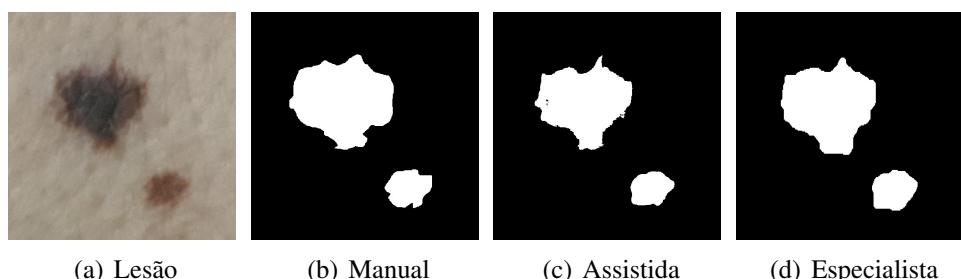


Figura 4. Exemplo de uma imagem de lesão (a) que foi anotada durante a execução do protocolo. A máscara (b) representa uma anotação manual do voluntário, (c) a assistida, e (d) a gerada por um especialista.

Na Tabela 1 são apresentados os resultados obtidos para cada métrica, em termos de média e desvio padrão, estratificados por modo de segmentação e por grupo de imagens, considerando a anotação individual de cada voluntário. Além das métricas, também é apresentado o tempo gasto para realizar a tarefa de segmentação. Para avaliar a significância estatística, foi utilizado o teste de Wilcoxon para amostras pareadas, adequado devido à não normalidade dos dados (verificado pelo teste de Shapiro-Wilk), com nível de significância de $p_{value} = 0,05$. Como pode ser observado, os resultados médios para ambos os modos de segmentação são similares. Essa observação é confirmada pelo teste estatístico, que não indicou diferença significativa entre os modos. Isso sugere que, no protocolo adotado, a ferramenta assistida por IA é equivalente ao modo manual, sem comprometer a precisão da segmentação.

Em relação ao tempo de segmentação, observa-se uma redução média de aproximadamente 75 segundos quando o modo assistido é utilizado. No entanto, o desvio padrão é elevado, sugerindo que, embora o modo assistido tenha acelerado a anotação para alguns voluntários, seu impacto foi menos significativo para outros. Neste caso, o teste estatístico apresentou diferença significativa ($p < 0,001$). Essa redução no tempo, aliada à manutenção da qualidade das máscaras, sugere que a abordagem assistida por IA pode ser uma opção viável para tarefas de segmentação de lesão de pele, sem comprometer a precisão dos resultados.

Os resultados apresentados na Tabela 1 agregam as métricas individuais obtidas por cada voluntário. Para explorar o potencial colaborativo de múltiplos anotadores, foi adotada uma estratégia de *crowdsourcing*, na qual as segmentações realizadas pelos voluntários no modo assistido por IA foram combinadas utilizando o método STAPLE (*Simultaneous Truth and Performance Level Estimation*) [Warfield et al. 2004]. O STAPLE estima uma máscara consenso a partir das anotações individuais, ponderando a confiabilidade de cada anotador com base na consistência entre as segmentações.

Na Tabela 2 é apresentado o resultado considerando as máscaras obtidas via STAPLE. De maneira geral, é possível observar uma melhora de aproximadamente 4% em todas as métricas, quando comparado com as obtidas por todos os grupos. Esse resultado indica que a combinação de crowdsourcing com a ferramenta proposta pode ser uma estratégia eficaz para melhorar a qualidade das máscaras de segmentação de lesão de pele, mesmo quando os anotadores não são especialistas.

3.4. Experiência do Usuário

Conforme descrito no protocolo, todos os voluntários responderam a um formulário de *feedback*, fornecendo informações sobre sua experiência com a ferramenta. A Figura 5 apresenta um gráfico com as principais percepções dos usuários em relação ao modo utilizado. No modo manual, os participantes destacaram três características principais: 72% descreveram a tarefa como trabalhosa, 50% como lenta e 47% como cansativa. Além disso, 34% enfatizaram o desgaste associado à anotação manual, classificando-a como exaustiva. Esses resultados estão alinhados com estudos anteriores que evidenciam os desafios da anotação manual de imagens médicas, especialmente para usuários sem experiência prévia.

Por outro lado, o modo assistido por IA impactou significativamente a experiência dos usuários. Os três aspectos mais mencionados foram: rápido (80% dos participantes), fácil (60%) e eficiente (45%). Além disso, emergiram respostas emocionais posi-

Modo	Grupo	NSD	IoU	DSC	Tempo (s)
Manual	1	0.86 ± 0.08	0.71 ± 0.16	0.82 ± 0.13	504 ± 233
	2	0.88 ± 0.08	0.73 ± 0.16	0.83 ± 0.12	449 ± 291
	3	0.90 ± 0.07	0.78 ± 0.13	0.87 ± 0.10	499 ± 177
	4	0.87 ± 0.08	0.73 ± 0.16	0.83 ± 0.12	470 ± 255
	Todos	0.88 ± 0.08	0.74 ± 0.15	0.84 ± 0.12	481 ± 242
Assistido	1	0.90 ± 0.13	0.77 ± 0.20	0.85 ± 0.17	428 ± 205
	2	0.82 ± 0.22	0.68 ± 0.20	0.76 ± 0.20	433 ± 243
	3	0.89 ± 0.13	0.76 ± 0.19	0.85 ± 0.16	341 ± 167
	4	0.89 ± 0.15	0.76 ± 0.20	0.85 ± 0.18	431 ± 287
	Todos	0.87 ± 0.16	0.75 ± 0.22	0.83 ± 0.21	406 ± 233

Tabela 1. Comparação de desempenho entre os modos manual e assistido considerando cada grupo de imagens e todos os grupos. Essa avaliação considera cada anotação individual do usuário e apresenta a média e o desvio padrão das métricas.

Modo	Grupo	NSD	IoU	DSC
Assistido	Todos	0.912 ± 0.078	0.793 ± 0.151	0.875 ± 0.112

Tabela 2. Métricas de avaliação das máscaras de segmentação utilizando o método de crowdsourcing STAPLE com as imagens geradas pelo modo assistido

tivas, com 25% dos voluntários descrevendo a experiência como divertida e 18% como agradável. Esses resultados indicam que a assistência por IA não apenas otimizou o processo de anotação, mas também tornou a tarefa mais envolvente e menos estressante.

Por fim, os voluntários foram questionados sobre a percepção da eficácia do modo assistido por IA, e mais de 90% relataram resultados positivos. Dentre eles, 50% consideraram a ferramenta muito útil, enquanto 42% a classificaram como útil. Apenas 4% permaneceram neutros quanto à sua eficácia, e a mesma porcentagem a considerou inútil. Essa alta taxa de satisfação reforça a viabilidade da ferramenta para a tarefa proposta.

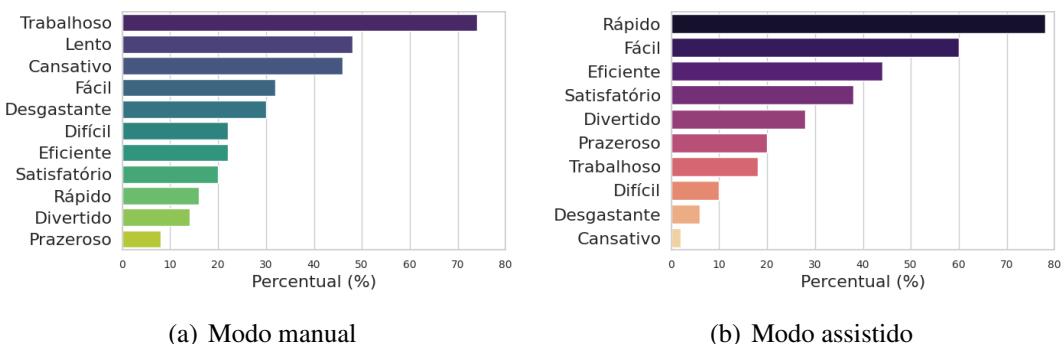


Figura 5. Avaliação de sentimento dos usuários com relação aos modos de segmentação do *framework*

4. Discussão

De maneira geral, os resultados deste estudo indicam que o *framework* proposto é uma alternativa viável para a geração de máscaras de segmentação de lesões de pele por não especialistas. Como ilustrado na Figura 4, um voluntário foi capaz de produzir máscaras semelhantes às de um especialista, cobrindo a maior parte da lesão. As principais diferenças observadas dizem respeito ao rigor na definição das bordas, um aspecto no qual até mesmo especialistas frequentemente divergem. Ainda assim, as métricas de avaliação indicaram a qualidade das máscaras geradas pelos voluntários. Esse resultado torna-se ainda mais relevante quando a técnica de *crowdsourcing* é aplicada para gerar uma máscara por consenso entre os voluntários. Nessa abordagem, obter um Dice ≈ 0.875 e um NSD ≈ 0.912 reforça a confiabilidade das máscaras geradas, validando a utilidade de estratégias colaborativas para a anotação de imagens médicas. Esse resultado é especialmente relevante para contextos onde a disponibilidade de especialistas é limitada.

Comparando os resultados obtidos nos modos manual e assistido por IA, observou-se que a abordagem assistida não comprometeu a qualidade das máscaras, mantendo métricas semelhantes às do modo manual. Além disso, houve um ganho interessante em termos de tempo de execução da tarefa. Essa equivalência em precisão, aliada à percepção positiva dos usuários e à redução no tempo de segmentação, reforça o potencial da ferramenta para auxiliar na criação de base de dados anotadas, tornando o processo mais acessível e menos oneroso.

Apesar do desvio padrão alto, é interessante notar que, no tempo de execução da segmentação em ambos os modos, a percepção dos voluntários foi amplamente favorável ao modo assistido, considerado mais rápido e intuitivo. Conforme relatado, 90% dos participantes avaliaram o modo assistido como mais rápido, e nenhum mencionou a sensação de lentidão. Durante o experimento, observou-se uma variação significativa no comportamento dos voluntários na abordagem da tarefa: enquanto alguns dedicavam mais tempo refinando a segmentação para obter contornos mais precisos, outros adotavam uma abordagem mais ampla e menos detalhista. Essa discrepância pode explicar tanto a variação no tempo médio entre os grupos – especialmente no Grupo 3, que lidou com lesões de contornos mais complexos – quanto o alto desvio padrão observado.

Por fim, uma possível melhoria para o *framework*, identificada durante a execução do protocolo, seria a implementação de um modo híbrido que permita aos usuários combinar as abordagens manual e assistida. Em casos de lesões com bordas mais complexas, o modelo assistido pode não capturar todos os detalhes com precisão, resultando em máscaras menos acuradas. Nesses cenários, a possibilidade de refinamento manual da máscara gerada pelo modelo poderia aprimorar a segmentação. Essa flexibilidade tem o potencial de aumentar a eficácia da ferramenta, permitindo que os usuários ajustem a abordagem conforme as características específicas de cada lesão.

5. Conclusão

Este trabalho apresentou um *framework* interativo para segmentação de lesões de pele, desenvolvido para otimizar a anotação de imagens por não especialistas. A ferramenta combina automação e intervenção humana para equilibrar eficiência e precisão, permitindo a geração de máscaras de segmentação satisfatórias. Os resultados obtidos indicam que voluntários não especialistas são capazes de produzir máscaras com métricas de avaliação satisfatórias, principalmente se combinadas por meio de um algoritmo de *crowdsourcing*.

A abordagem assistida se mostrou factível, acelerando o processo de segmentação sem comprometer a qualidade dos resultados. A percepção positiva dos usuários e a redução no tempo de execução reforçam a viabilidade da ferramenta para a criação de bases de dados anotadas, tornando o processo mais acessível e menos oneroso. Importante destacar que a ferramenta não busca substituir especialistas, mas sim auxiliá-los no processo de anotação. Futuras melhorias incluem a implementação de um modo híbrido que permita aos usuários combinar as abordagens manual e assistida, aumentando a flexibilidade e eficácia da ferramenta, além de um ajuste fino do modelo para melhorar a precisão da segmentação para tarefas específicas.

Agradecimentos

Os autores agradecem o apoio financeiro da Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES); o Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi); o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); o Ministério da Saúde (MS); e o Programa Nacional de Genômica e Saúde de Precisão (Genomas Brasil).

Referências

- Abramson, J., Adler, J., Dunger, et al. (2024). Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500.
- Barragán-Montero, A., Javaid, U., Valdés, G., and Others (2021). Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83.
- Benaich, N. (2024). State of ai report. Air Street Press.
- Cheplygina, V., Perez-Rovira, et al. (2021). Crowdsourcing airway annotations in chest computed tomography images. *PLoS One*, 16(4):e0249580.
- Clark, J. et al. (2023). Artificial intelligence index report. Stanford University. Último acesso: 9 de fevereiro 2025.
- Cocos, A., Qian, T., et al. (2017). Crowd control: Effectively utilizing unscreened crowd workers for biomedical data annotation. *JBI*, 69:86–92.
- Costa, G. S. S., Paiva, et al. (2021). Covid-19 automatic diagnosis with ct images using the novel transformer architecture. pages 293–301. SBC.
- Damgaard, C., Eriksen, T. N., Juodelyte, D., Cheplygina, V., and Jiménez-Sánchez, A. (2023). Augmenting chest x-ray datasets with non-expert annotations.
- Das, K., Cockerell, C. J., Patil, A., Pietkiewicz, et al. (2021). Machine learning and its application in skin cancer. *Int. J. Environ. Res. Public Health*, 18(24):13409.
- De Angelo, G. G., Pacheco, et al. (2019). Skin lesion segmentation using deep learning for images acquired from smartphones. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- De Fauw, J., Ledsam, et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, 24(9):1342–1350.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale.

- Kar, K. et al. (2024). Automated intracranial hemorrhage detection using deep learning in medical image analysis. In *2024 International Conference on Data Science and Network Security (ICDSNS)*, pages 1–6.
- Kentley, J., Weber, J., et al. (2023). Agreement between experts and an untrained crowd for identifying dermoscopic features using a gamified app: Reader feasibility study. *JMIR Medical Informatics*.
- Kirillov, A., Mintun, E., Ravi, et al. (2023). Segment anything. *arXiv:2304.02643*.
- Maier-Hein, L., Reinke, A., Godau, et al. (2024). Metrics reloaded: recommendations for image analysis validation. *Nature Methods*, 21(2):195–212.
- Najjar, R. (2023). Redefining radiology: a review of artificial intelligence integration in medical imaging. *Diagnostics*, 13(17):2760.
- Nikolov, S., Blackwell, et al. (2021). Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. *Journal of medical Internet research*, 23(7):e26151.
- Ouyang, D., He, et al. (2020). Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580(7802):252–256.
- Pacheco, A. G. C. et al. (2020). Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*.
- Pacheco, A. G. C. and Krohling, R. A. (2021). An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics*, 25(9):3554–3563.
- Sarwar, N., Irshad, et al. (2024). Skin lesion segmentation using deep learning algorithm with ant colony optimization. *BMC Medical Informatics and Decision Making*.
- Shaik, T., Tao, et al. (2023). Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2):e1485.
- Souza Jr, L. A., Pacheco, A. G., Passos, L. A., Santana, et al. (2024). Deepcraftfuse: visual and deeply-learnable features work better together for esophageal cancer detection in patients with barrett's esophagus. *Neural Computing and Applications*.
- Tao Lei, A. K. N. (2023). *Image Segmentation: Principles, Techniques, and Applications*. John Wiley & Sons.
- Tschandl, P., Rosendahl, et al. (2018). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*.
- Ward, A. et al. (2024). Creating an empirical dermatology dataset through crowdsourcing with web search advertisements. *JAMA Network Open*, 7(11):e2446615.
- Warfield, S. K., Zou, et al. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE(T-MI)*, 23.
- Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., Yang, T., and Gao, M. (2023). Techniques and challenges of image segmentation: A review. *Electronics*, 12(5).