

SEGMENTAÇÃO DE NÓDULOS EM IMAGENS DE MAMOGRAFIA¹

Éllen Grasiéli Vieira², Maikon Cismoski dos Santos³

¹ Trabalho de Conclusão de Curso de Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense, Câmpus Passo Fundo, 2022.

² Bacharel em Ciência da Ciência da Computação do IFSUL - Câmpus Passo Fundo

³ Docente do Curso de Bacharelado em Ciência da Computação do IFSUL - Câmpus Passo Fundo

RESUMO

Atualmente, na mortalidade proporcional por câncer em mulheres no Brasil, o Câncer de Mama é responsável pela maior taxa de mortalidade entre as mulheres, representando 16,1% do total. Diversos estudos ressaltam que um dos principais métodos para a detecção do Câncer de Mama é o diagnóstico por imagem através de exames de mamografia. As imagens geradas neste exame permitem identificar possíveis nódulos, sejam eles malignos ou benignos, além de, quando iniciado um tratamento, exames de imagem auxiliam no acompanhamento deste tratamento. Entretanto, nem sempre os exames geram imagens com alto grau de detalhamento, o que demanda tempo para a identificação de possíveis nódulos. Portanto, este projeto propôs o desenvolvimento de uma metodologia que, utilizando técnicas de processamento digital de imagens, auxilie na segmentação de nódulos em imagens de mamografia. Para isso, foram utilizados algoritmos de geração de superpixels, limiarização, detecção de bordas e operações morfológicas. O trabalho foi desenvolvido com o auxílio da biblioteca de visão computacional OpenCV aplicada à linguagem Python, empregando uma base de imagens pública para testes do protótipo. O protótipo desenvolvido se mostrou eficiente na identificação de nódulos, com um percentual 80% de acurácia em relação às 20 imagens utilizadas, das quais 15 possuíam nódulos malignos.

Palavras-chave: Processamento de imagens. Mamografia. Segmentação. OpenCV.

ABSTRACT

Currently, in the proportional mortality rate for cancer in women in Brazil, Breast Cancer is responsible for the highest mortality rate among women, representing 16.1% of the total. Several studies point out that one of the main methods for detecting breast cancer is diagnostic imaging through mammography. The images generated in this exam allow the identification of possible nodules, whether malignant or benign, in addition to, when treatment is started, imaging exams help in the follow-up of this treatment. However, the exams do not always generate images with a high degree of detail, which takes time to identify possible nodules. Therefore, this project proposed the development of a methodology that, using digital image processing techniques, would assist in the segmentation of nodules in mammography images. For this, algorithms for superpixel generation, thresholding, edge detection and morphological operations were used. The work was developed with the help of the OpenCV computer vision library applied to the Python language, using a public image base for prototype testing. The developed prototype proved to be efficient in identifying nodules, with an accuracy of 80% in relation to the 20 images used, of which 15 have malignant nodules.

Keywords: Image processing. Mammography. Segmentation. OpenCV.

INTRODUÇÃO

O Câncer de Mama é uma das doenças mais frequentes entre as mulheres globalmente, representando uma parcela significativa dos óbitos relacionados ao câncer feminino (INCA, 2019). No Brasil, ele é a segunda principal causa de morte e a primeira entre as neoplasias malignas (INCA, 2019). A Organização Mundial da Saúde estima que ocorram mais de um milhão de novos casos por ano em todo o mundo.

A mamografia é um método fundamental para diagnosticar o Câncer de Mama, gerando imagens que evidenciam áreas suspeitas. Imagens desempenham um papel essencial na medicina, fornecendo informações para diagnósticos mais precisos e decisões médicas informadas (NETO et al., 2004).

Apesar da redução na incidência, o Câncer de Mama ainda persiste como uma preocupação, com ênfase na detecção precoce (Assis, 2015). A otimização da identificação de nódulos em mamografias através do processamento de imagens pode acelerar o diagnóstico, beneficiando pacientes e profissionais de saúde.

Por esse motivo, a aplicação de técnicas de processamento e segmentação focadas em auxiliar no processo de reconhecimento destes nódulos poderá otimizar o tempo de identificação e auxiliar a equipe clínica no diagnóstico da doença. Além disso, o desenvolvimento de pesquisas com o propósito de aumentar a eficiência do diagnóstico de Câncer de Mama através do processamento de imagens em mamografias não só representa uma contribuição positiva para a sociedade, mas também enriquece o meio acadêmico com avanços tecnológicos relevantes.

A interseção entre medicina e tecnologia desempenha um papel cada vez mais crucial na busca por soluções inovadoras para desafios de saúde, e é nesse contexto que a aplicação de técnicas de processamento e segmentação para a detecção precoce de nódulos em mamografias se destaca. Além de contribuir para a agilidade do diagnóstico e para aprimorar a capacidade da equipe clínica, essa abordagem também está em perfeita sintonia com o Objetivo de Desenvolvimento Sustentável 3 - Saúde e bem-estar, estabelecido pelas Nações Unidas. Este objetivo visa assegurar uma vida saudável e promover o bem-estar para todas as idades, destacando a importância de intervenções inovadoras e tecnológicas que possam impactar positivamente a detecção e tratamento de doenças, como o Câncer de Mama.

Este trabalho tem como principal objetivo o desenvolvimento de uma metodologia que utilize técnicas de processamento digital de imagens para auxiliar na detecção das neoplasias malignas nas imagens de mamografia.

METODOLOGIA

Para atingir o objetivo proposto neste trabalho, foi criado um algoritmo para destacar nódulos nas imagens de mamografias. As imagens utilizadas para a Implementação do protótipo pertencem ao banco de dados Dataset of Breast mammography images with Masses, disponibilizado por Lin, Ting; Huang, Mei-Ling (2020). Essa base contém imagens de mamografias e ressonâncias de mama, já classificadas por médicos em imagens com achados malignos, benignos e imagens sem nódulos, além de ser uma base pública disponível para estudantes e pesquisadores.

Para o desenvolvimento do protótipo e realização dos testes de validação, foram selecionadas 20 imagens, sendo 15 mamografias contendo um nódulo maligno e 5 imagens sem nódulos identificados, a fim de avaliar se o algoritmo geraria falsos positivos. Essa quantidade de imagens foi escolhida com base na média de imagens utilizadas nos trabalhos relacionados existentes na literatura.

As imagens que contém neoplasias malignas foram escolhidas com o critério de possuírem apenas um nódulo maligno por imagem, a fim de simplificar a execução dos testes.

O protótipo foi desenvolvido na linguagem Python, em conjunto com as bibliotecas OpenCV e Scikit-Image, através da plataforma Colab, mantida pela Google. Essa plataforma permite a escrita e a execução de códigos em Python de forma online, via browser.

O funcionamento geral do algoritmo consiste na leitura da imagem; geração dos superpixels; limiarização; aplicação de operações morfológicas para eliminar possíveis imperfeições originadas no processo de limiarização; subtração da imagem original pela imagem limiarizada (para fins de visualização); detecção e desenho das bordas e por fim a plotagem de resultados. No diagrama da Figura 1 é apresentada a organização do protótipo desenvolvido.

Figura 1 - Diagrama de funcionamento do protótipo.



Fonte: do autor.

Cada uma das etapas é detalhada a seguir.

GERAÇÃO DE SUPERPIXELS

Na etapa de geração de *Superpixels* (aplicação do algoritmo *SLIC*), foi utilizada a função *slic*, da biblioteca *Scikit-image*. Essa função recebe por parâmetro a imagem de entrada, o número de segmentos em que a imagem será quebrada, utilizado o valor de 40 regiões por convenção e o nível de condensação de cada região (aqui utilizado o valor 20).

Para efetivamente gerar as regiões homogêneas na imagem foi utilizado a função *regionprops*, também da biblioteca *Scikit-image*, onde o objetivo principal é obter informações sobre determinada região da imagem, para isso, são passados como parâmetros os segmentos gerados na função *slic* e os valores de intensidade da imagem. A função *regionprops* devolve um array com as informações para cada região gerada na função *slic*.

Com as informações da função *regionprops*, é utilizada a função *color* para realizar a homogeneização da região em uma imagem de saída, considerando as propriedades *coords* e *mean_intensity* contidas em cada objeto do array de *regionprops*.

LIMIARIZAÇÃO

Para a etapa de limiarização, foi utilizada a função *threshold*, da biblioteca OpenCV, onde são passados como parâmetros a imagem obtida na etapa anterior, os valores de limite inferior e limite superior (aqui sendo 170 e 255, respectivamente) e o tipo de limiarização escolhida, sendo neste projeto utilizado o modelo de *THRESH_BINARY*, o qual obteve melhores resultados quando comparado aos outros modelos de binarização.

APLICAÇÃO DE FILTROS MORFOLÓGICOS

A aplicação de filtros morfológicos se fez necessária para a remoção de imperfeições originadas no processo de limiarização, através das operações de erosão e dilatação. Entretanto, essa é uma das etapas mais delicadas do processo, uma vez que os processos de erosão e dilatação podem afetar a extração das características de forma e margem, utilizadas para classificar o nódulo.

Para realizar estas operações, foram utilizadas as funções *erode* e *dilate* da biblioteca OpenCV, em sequência. A função *erode* é responsável por gerar a erosão da imagem, assim removendo pequenos agrupamentos de pixels que foram segmentados no processo anterior, mas que não se caracterizam como nódulo, entretanto, essa função acaba modificando a forma do nódulo. Para amenizar isso, é aplicada a função *dilate*, que irá dilatar o nódulo na mesma proporção em que foi corroído, a fim de retornar o objeto segmentado a uma forma muito próxima a obtida anteriormente.

Essas funções recebem por parâmetro a imagem limiarizada anteriormente, o elemento estruturante e o número de iterações que serão feitas, onde neste caso, foram duas iterações.

A criação do elemento estruturante é feita através da função *getStructuringElement(formato, tamanho_matriz_estruturante)*, onde o formato utilizado foi o de cruz (*MORPH_CROSS*). Para esta etapa, foram testados elementos estruturantes de tamanho 3x3, 4x4 e 5x5, onde o elemento de tamanho 3x3 obteve e melhores resultados, enquanto que os outros descaracterizaram as bordas dos nódulos.

DETECÇÃO E DESENHO DAS BORDAS

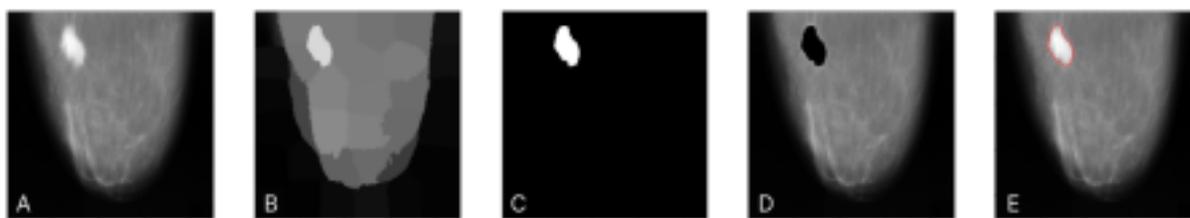
Para realizar o desenho do contorno do nódulo, foi utilizado o algoritmo de Canny, através da função *Canny(imagem, limite_inferior, limite_superior)*. Como a função foi aplicada sob a imagem limiarizada, foram utilizados os valores de 150 e 255 para limites inferior e superior.

Para localizar as coordenadas da borda do objeto segmentado, foi utilizada a função *findContours* sobre a imagem produto da função de Canny. Essa função retorna um array com todas coordenadas que, em seguida, são utilizadas como parâmetro na função *drawContours*, responsável por efetivamente desenhar as bordas do objeto localizado sobre a imagem original.

RESULTADOS

Para a análise, foram considerados dois cenários: Imagens com nódulos identificados e número de nódulos identificados por imagem. Conforme a Figura 2, para a exibição do resultado final, são geradas cinco imagens de saída para cada imagem de entrada, sendo A) imagem original, B) Imagem segmentada por SLIC, C) Imagem limiarizada, D) Imagem original mesclada a imagem limiarizada, E) Imagem original com a borda do nódulo destacada.

Figura 2 - Exemplo de imagem gerada como resultado do protótipo.



Fonte: do autor.

Neste trabalho, foi proposto um protótipo para processamento e segmentação de nódulos em imagens de mamografia que permite auxiliar no diagnóstico do Câncer de Mam

Para o cenário de testes em imagens com nódulos identificados, foram consideradas todas as imagens previamente selecionadas, já que o objetivo principal deste teste é validar se

o protótipo identificaria falsos positivos nas imagens sem nódulos. Aqui, o retorno esperado era se há ou não um objeto identificado na imagem.

No cenário de testes de quantidade de nódulos identificados por imagem, o objetivo principal é verificar se para cada imagem que continha um nódulo, foi identificado apenas um objeto, já que a proporção de nódulos por imagem era 1:1. Os resultados obtidos nos testes podem ser visualizados na Tabela 1.

Tabela 1 - Resultados obtidos.

| Cenário | Verdadeiros Positivos | Falsos Positivos | Verdadeiros Negativos | Falsos Negativos | Total | % Acurácia |
|-----------------------|-----------------------|------------------|-----------------------|------------------|-------|------------|
| Imagens | 11 | 1 | 5 | 3 | 20 | 80% |
| Nódulos Identificados | 11 | 4 | Não se aplica | 3 | 15 | 61.11% |

Fonte: do autor

A partir das informações contidas na Tabela 1, podemos verificar que das 15 imagens com nódulos utilizadas nos testes, o algoritmo se mostrou eficiente identificando objetos em 11 das imagens (para ambos os testes). Um exemplo de sucesso da identificação é apresentado na Figura 3, onde o nódulo foi corretamente segmentado e teve suas bordas contornadas.

Figura 3 - Exemplo de imagem em que o protótipo segmentou o nódulo com sucesso.

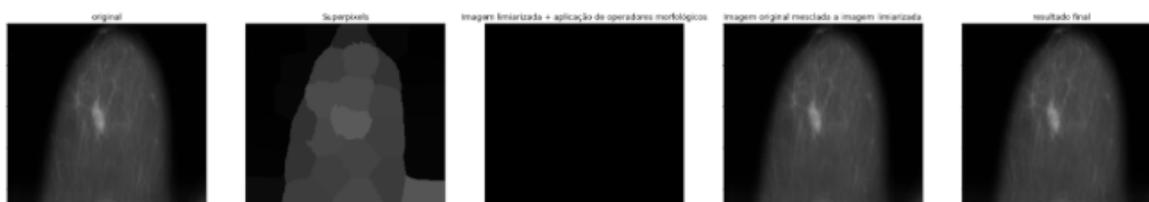


Fonte: do autor.

Nos testes de imagens com nódulos, uma das imagens que continha nódulo foi marcada como verdadeiro positivo, entretanto, a área segmentada se tratava de um reflexo existente na imagem, por esse motivo, ela foi contabilizada como falso positivo. No demais, o protótipo não gerou falsos positivos reais, mas gerou 3 falsos negativos em imagens que continham nódulos, mas que não tinham nenhum objeto segmentado.

Analisando as imagens que geram falsos negativos, foi possível identificar que o protótipo apresentou falha na etapa de geração dos superpixels para nódulos em imagens de baixo contraste, conforme a Figura 4, em que o nódulo não foi identificado na fase de limiarização, já que a imagem produto da etapa de geração dos superpixels possui baixo contraste.

Figura 4 - Exemplo de imagem com baixo contraste, onde o algoritmo não conseguiu localizar nenhum nódulo.



Fonte: do autor.

Figura 5 - Exemplo de imagens com mais de um objeto identificado, onde é possível ver reflexos e/ou estruturas que não oriundas da mama.



Fonte: do autor.

Também foi possível observar que nas imagens que obtiveram falsos positivos, o algoritmo acaba segmentando bordas e reflexos do exame, que por se tratarem de áreas muito grandes, acabam não sendo eliminadas na etapa de limpeza com os operadores morfológicos, como podemos visualizar na Figura 5.

CONSIDERAÇÕES FINAIS

Neste trabalho, foi proposto um protótipo para processamento e segmentação de nódulos em imagens de mamografia que permite auxiliar no diagnóstico do Câncer de Mama.

Para tal, foi desenvolvido um algoritmo que, fez o uso de Superpixels, limiarização, operadores morfológicos e segmentação com o operador de Canny.

O protótipo desenvolvido obteve sucesso identificando nódulos em 12 das 15 imagens que continham nódulos e não identificando falsos positivos nas imagens sem nódulos, alcançando um percentual de 80% de acurácia.

Nos testes de quantidade de nódulos identificada por imagem, o protótipo atingiu uma acurácia de 61,11% na detecção dos nódulos, entretanto, desconsiderando os valores de falsos positivos, os quais ocorreram devido a reflexos e/ou estruturas que não fazem parte da mama, o valor de acurácia passa para 78,57%. Nos testes de identificação de imagens que possuíam ou não o nódulo, o percentual de sucesso é 80%.

Como foi possível avaliar ao longo deste artigo, o algoritmo se mostra promissor, mas precisa de refinamentos, como o emprego de técnicas de correção de contraste a partir da equalização do histograma da imagem. Essa técnica é uma alternativa para corrigir o problema da geração de falsos negativos em imagens de baixo contraste.

Outra proposta para trabalhos futuros seria o uso de inteligência artificial e momentos invariantes para automatizar o processo de reconhecimento e caracterização dos nódulos, a fim de remover os falsos positivos para análises posteriores, além de que, o uso destes recursos pode ser utilizado para remover os falsos positivos obtidos no processo.

Além disso, em trabalhos futuros pode ser realizada uma comparação aprofundada da metodologia proposta neste trabalho com outros trabalhos relacionados, a fim de comparar os resultados obtidos.

REFERÊNCIAS BIBLIOGRÁFICAS

ACHANTA, Radhakrishna et al. **Slic superpixels**. 2010.

ALBUQUERQUE, Márcio Portes de; ALBUQUERQUE, Marcelo Portes de. **Processamento de imagens: métodos e análises**. Rio de Janeiro, Brasil, v. 12, 2000.

ASSIS, J. B. **Estudo do grau de risco de câncer de mama utilizando a dimensão fractal de imagens**. Dissertação (Conclusão de Curso). Universidade Tecnológica Federal do Paraná, 2015.

BRASIL. Instituto Nacional de Câncer, Ministério da Saúde. **Controle do Câncer de Mama**. Disponível em: <<https://www.inca.gov.br/controlado-cancer-de-mama>>. Acesso em 19 de abril de 2022.

BRASIL. Instituto Nacional de Câncer, Ministério da Saúde. **Parâmetros técnicos para rastreamento do câncer de mama**. Rio de Janeiro, RJ 2021. Disponível em: <<http://www.inca.gov.br>>. Acesso em 18 de abril de 2022.

CALVOSO, Beatriz De S. et Al., **Diagnóstico e Conduta Acerca De Nódulos Benignos De Mama: Uma Revisão Sistemática**. Revista de Medicina da Faculdade Atenas, p. 2236-9252, 2019.

GODINHO, Eduardo R; KOCH, Hilton A., **Breast imaging reporting and data system (BI-RADS™): como tem sido utilizado?** Radiologia Brasileira, São Paulo, v. 37, ed. 6, p. 413-417, 2004

GODINHO, Eduardo R; KOCH, Hilton A., **Rastreamento do câncer de mama: aspectos relacionados ao médico**. Radiologia Brasileira, São Paulo, v. 37, ed. 6, v. 37, p. 91-99, 2004.
GONZALEZ, R.C. e WOODS, R.E., **Processamento De Imagens Digitais**. 3ª ed. São Paulo (SP), Editora Pearson; 2011.

KANNAN, Subbu; GURUSAMY, Vairaprakash; NALINI, G. **Review on Image Segmentation Techniques**. Pattern recognition 26.9 (1993): 1277-1294.

LIN, T.; HUANG, M, **Dataset of Breast mammography images with Masse**. Mendeley Data, 2020.

NETO, G. H. RIBEIRO, G. C. VALERI, F. V. **Processamento e Segmentação de Mamogramas Digitais**. In Congresso Brasileiro de Informática em Saúde, 2004.

OLIVEIRA, Luiz E. S. **Processamento de Imagens: Morfologia Matemática Binária**. Anotações de aula, 2004.

OPENCV. Disponível em: <<https://opencv.org/>>. Acesso em: 8 de julho de 2022.

PAIVA, Fábio Augusto Procópio de. et al. **Introdução a python com aplicações de sistemas operacionais**. Editora IFRN, 2020.

PEDRINI, H; SCHWARTZ, W. R. **Análise de Imagens Digitais**. Editora Thomson Pioneira, 2008.

PYTHON. Disponível em: <<https://www.python.org/>>. Acesso em: 8 de julho de 2022.

REDLICH, Caroline R., **Segmentação de Imagens Baseada em Grafos de Superpixel**. Dissertação de Mestrado, PUC-Rio, 2018.

SALDANHA, M. F. S. FREITAS, C. C. **Segmentação de Imagens Digitais: Uma Revisão**. In: WORKSHOP DOS CURSOS DE COMPUTAÇÃO APLICADA DO INPE, 9. (WORCAP)., 2009, São José dos Campos. Anais... São José dos Campos: INPE, 2009 Disponível em: <<http://urlib.net/ibi/8JMKD3MGP8W/37NC86S>>. Acesso em: 23 de mai. de 2022.

SANTOS, Lyndainês A. dos. **Aplicação de Processamento de Imagem Digital para Detecção de Câncer de Mama**. In Anais da Mostra Nacional de Robótica - MNR 2018, 2018.

SHAPIRO, Linda; STOCKMAN, George. Computer vision. Vol. 3. New Jersey: Prentice Hall, 2001.

SILVA, Anna K. G. ; COSTA, Letícia F. de J. **Importância do diagnóstico por imagem no câncer de mama**. 2021. Trabalho de Conclusão de Curso (Bacharelado em Biomedicina) - UNA Câmpus Catalão, [S. l.], 2021. Disponível em: <https://repositorio.animaeducacao.com.br/handle/ANIMA/19213>. Acesso em: 21 jun. 2022.

SILVA, F. J. V.; ALVES, C. H. F. **Aplicação de técnicas de processamento de imagens digitais em imagens geradas por ultrassom**. 8º Encontro Regional de Matemática Aplicada e Computacional, 2008.

SILVA, T. C. et Al. **Detecção automática de tumores em mamografias utilizando técnicas de processamento digital de imagem**. In XXIV Congresso Brasileiro de Engenharia Biomédica – CBEB 2014, 2014.

VALE, G. M. D.; POZ, A. P. D. **Processo de detecção de bordas de canny**. Boletim de Ciências Geodésicas, v. 8, n. 2, 2002.