

Detecção do Câncer de Mama em Imagens Infravermelhas Utilizando Características Radiômicas

Elisson C Carvalho¹, Alessandra M Coelho¹, Matheus F O Baffa²

¹Departamento Acadêmico da Ciência da Computação – Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais (IF Sudeste MG)
Rio Pomba – MG – Brasil

²Departamento de Computação e Matemática – Universidade de São Paulo (USP)
Ribeirão Preto – SP – Brasil

elissoncarvalho314@gmail.com

Abstract. *Breast cancer is the second most common type of cancer around the world. When early diagnosed, the patient has a better prognosis and increases their chances of cure. This work proposes the development of an infrared image classification method to aid in the diagnosis of breast cancer. For this, radiomic-based features were used to represent the image along with a Deep Neural Network for pattern recognition and to build a classification model. We conducted an experiment following the Cross-Validation protocol and the proposed method reached an accuracy of 97.27% and a sensitivity of 96.33%. New medical imaging modalities have proven effective in the early detection of breast cancer.*

Resumo. *O câncer de mama é o segundo tipo mais frequente de câncer no mundo. Quando diagnosticado precocemente, o paciente pode apresentar um melhor prognóstico e aumentar suas chances de cura. Neste trabalho é proposto o desenvolvimento de um método de classificação de imagens infravermelhas para o auxílio no diagnóstico do câncer de mama. Para isso, utilizou-se características baseadas em radiômicas para representar o conteúdo da imagem em conjunto com uma Rede Neural Profunda para detecção de padrões e construção do modelo de classificação. Avaliado seguindo o protocolo de Validação Cruzada, o método proposto atingiu uma taxa de acerto de 97,27% e uma sensibilidade de 96,33%. Novas modalidades de imagens médicas têm se mostrado eficazes na detecção precoce do câncer de mama.*

1. Introdução

A Visão Computacional denota um campo na Inteligência Artificial, centrada em estudos teóricos em processamento de informação visual, tendo como um de seus objetivos principais desenvolver sistemas de interpretação de imagens [Poggio et al. 1987]. Com o avanço do conhecimento científico na área de Processamento de Imagens e da Inteligência Artificial para reconhecimento de padrões visuais e, consequentemente no desenvolvimento de novos métodos na área de Visão Computacional, surgiram aplicações em diferentes linhas de pesquisa, entre elas, na área da saúde. Podemos citar, por exemplo, trabalhos para detecção e classificação automática de doenças na

retina [Rasti et al. 2017] e trabalhos no auxílio ao diagnóstico do câncer de mama [Silva et al. 2016, Baffa and Lattari 2018].

A interpretação de imagens médicas tem sido amplamente utilizada para o diagnóstico do câncer de mama, e possuem um papel fundamental na detecção e no acompanhamento do tratamento do paciente. Para isso, além do exame clínico das mamas, podem ser recomendados exames de imagem, como a mamografia, a ultrassonografia ou a ressonância magnética [Instituto Nacional do Cancer. 2021].

Além das imagens comumente empregadas no diagnóstico do câncer de mama, a termografia é um método não invasivo, que permite obter uma imagem infravermelha a partir da temperatura da superfície da pele, de forma rápida, barata e indolor ao paciente. É uma modalidade de imagem relativamente simples que detecta a variação de temperatura na superfície da pele humana [Ng 2009].

Dada as limitações do sistema visual humano, torna-se difícil detectar as diferenças de temperatura [Roslidar et al. 2020]. Nesse sentido, um sistema de Visão Computacional capaz de detectar mudanças na distribuição de temperatura pode auxiliar no processo de identificação do câncer de mama em exames de termografia. Esses sistemas encontraram padrões, usando Redes Neurais Artificiais, que permitem distinguir, de forma precisa, entre exames de imagens de pacientes saudáveis e pacientes com tumor em alguma das estruturas que compõem as mamas.

Neste trabalho, apresentamos um método de classificação de imagens infravermelhas baseado na utilização de características radiômicas, extraídas da região de interesse das imagens infravermelhas. Posteriormente, um algoritmo baseado em Redes Neurais Profundas Totalmente Conectadas foi utilizado para detectar padrões no vetor descritor das imagens. A principal contribuição deste trabalho é uma metodologia de classificação de imagens baseada na utilização de características radiômicas e aprendizagem profunda.

2. Materiais e Métodos

2.1. Base de Imagens Infravermelhas da Mama

A base de dados *Database for Mastology Research* (DMR), desenvolvida pelo Grupo de Pesquisa *Visual Lab*, da Universidade Federal Fluminense (UFF), foi utilizada para realizar o treinamento e avaliação do método proposto [Silva et al. 2014]. Para o desenvolvimento do classificador, foram utilizadas 146 imagens, sendo 73 provenientes de pacientes doentes, com câncer de mama, e 73 de pacientes saudáveis.

2.2. Pré-Processamento

Antes de extrair as características descritivas das imagens, foi necessário realizar uma segmentação da região das mamas. Essa segmentação limita a quantidade de informações extraídas de forma a tornar a extração de características mais relevante para o processo de detecção de padrões. A segmentação foi realizada utilizando o método proposto em [Baffa et al. 2021], o qual utiliza uma Rede Neural Convolutiva (U-Net CNN) para segmentar imagens infravermelhas da mama e criar uma máscara binária de segmentação.

2.3. Extração de Características

Posteriormente, as imagens foram submetidas a um processo de extração de características, a fim de obter um vetor descritor, cujo objetivo foi representar o conteúdo

semântico das imagens em um conjunto de dados numéricos. Foi utilizada uma combinação entre diferentes métodos de extração de características, a saber, *Haralick* (13 características), *Local Binary Patterns* (LBP) (26 características), *Threshold Adjacency Statistics* (TAS) (54 características) e características baseadas na abordagem radiômica (93 características), tais como as estatísticas de primeira ordem das distribuição de níveis, características de forma, matrizes de repetição de nível, matrizes de quantificação de zonas de níveis, dentre outras. Os métodos de *Haralick*, o LBP e o TAS foram implementados utilizando a *framework* de Visão Computacional Mahotas¹, enquanto as características radiômicas foram implementadas utilizando a *framework* PyRadiomics². Ao todo, 186 características foram extraídas para representar as regiões de interesse das imagens infravermelhas.

2.4. Metodologia de Classificação

Após a extração de características, foi desenvolvido um algoritmo de detecção de padrões baseado nas Redes Neurais Totalmente Conectadas (*Fully-Connected Neural Network* – FCNN). A rede neural proposta foi composta por nove camadas, sendo uma camada de entrada, sete camadas ocultas e uma camada de saída. A camada de entrada é composta por 186 neurônios, referentes as 186 características extraídas na etapa anterior. A informação contida na camada de entrada é propagada para as camadas ocultas, que possuem 145 neurônios, com a função de ativação *Rectifier Linear Unit* (ReLU) e uma taxa de desligamento aleatório (*dropout*) de 20%. Por fim, a camada de saída é composta por um neurônio, referente a classificação binária, com a função de ativação *Sigmoid*. A estrutura da rede foi definida empiricamente.

3. Experimentos e Resultados

Para avaliar o modelo de classificação, um experimento foi proposto utilizando o protocolo de experimentação de validação cruzada (*K-Fold Cross-Validation*). A base de dados foi dividida em dez partes ($k = 10$), de maneira que os dados utilizados para o treinamento não fossem utilizados na validação do modelo.

Em cada iteração da validação cruzada, cinco métricas quantitativas foram calculadas, sendo elas a acurácia, a *Area Under the ROC Curve* – AUC, a precisão, a sensibilidade e a especificidade. Essas métricas são calculadas a partir da taxa de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos extraídas da classificação da base de validação. A Tabela 1 contém o resultado médio calculado a partir da validação cruzada.

Tabela 1. Resultados obtidos pela avaliação quantitativa.

| | Acurácia | AUC | Precisão | Sensibilidade | Especificidade |
|-----------------|----------|--------|----------|---------------|----------------|
| Resultado Médio | 97,27% | 98,28% | 98% | 96,33% | 98,57% |

Fonte: autor.

O método de classificação alcançou resultados satisfatórios, atingindo taxas de acerto elevado e competitivo na literatura. Analisando as taxas de sensibilidade e especificidade, foi possível observar que o modelo aprendeu padrões satisfatórios, uma vez que não realizou uma classificação tendenciosa a uma classe específica.

¹Mahotas. Disponível em: <https://mahotas.readthedocs.io/en/latest/>

²PyRadiomics. Disponível em: <https://pyradiomics.readthedocs.io/en/latest/>

4. Conclusão

Neste trabalho propomos o desenvolvimento de um método de classificação das mamas utilizando características radiômicas e imagens infravermelhas. Os resultados obtidos são promissores e podem auxiliar médicos a proverem um diagnóstico mais preciso, aumentando as chances de cura do câncer de mama. Em trabalhos futuros, pretende-se avaliar outras arquiteturas de aprendizagem profunda para maximizar as taxas de acerto, bem como desenvolver o *software* de auxílio ao diagnóstico.

Agradecimentos

Os autores agradecem ao IF SudesteMG e ao Programa de Educação Tutorial (PET - Conexões de Saberes Ciência da Computação) pelo apoio técnico e financeiro para o desenvolvimento do projeto.

Referências

- Baffa, M. F. O., Conci, A., and Coelho, A. M. (2021). Segmentação de imagens infravermelhas para detecção do câncer de mama utilizando u-net cnn. In *XXI Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. No Prelo.
- Baffa, M. F. O. and Lattari, L. G. (2018). Convolutional neural networks for static and dynamic breast infrared imaging classification. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 174–181.
- Instituto Nacional do Cancer. (2021). Câncer de mama. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>>. Acessado em: 25 mai. 2021.
- Ng, E.-K. (2009). A review of thermography as promising non-invasive detection modality for breast tumor. *International Journal of Thermal Sciences*, 48(5):849–859.
- Poggio, T., Torre, V., and Koch, C. (1987). Computational vision and regularization theory. *Readings in computer vision*, pages 638–643.
- Rasti, R., Rabbani, H., Mehridehnavi, A., and Hajizadeh, F. (2017). Macular oct classification using a multi-scale convolutional neural network ensemble. *IEEE transactions on medical imaging*, 37(4):1024–1034.
- Roslidar, R., Rahman, A., Muharar, R., Syahputra, M. R., Arnia, F., Syukri, M., Pradhan, B., and Munadi, K. (2020). A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access*, 8:116176–116194.
- Silva, L., Saade, D., Sequeiros, G., Silva, A., Paiva, A., Bravo, R., and Conci, A. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100.
- Silva, L. F., Santos, A. A. S., Bravo, R. S., Silva, A. C., Muchaluat-Saade, D. C., and Conci, A. (2016). Hybrid analysis for indicating patients with breast cancer using temperature time series. *Computer methods and programs in biomedicine*, 130:142–153.