

Metaheurísticas Aplicadas à Detecção, Localização e Classificação de Tumores Mamários Compostos por Múltiplos Focos via Termografia

Jan Pierre A. S. Rocha¹, Heder S. Bernardino¹, Helio J. C. Barbosa², Felipe S. Loureiro³

¹Programa de Pós Graduação em Modelagem Computacional
Universidade Federal de Juiz de Fora (UFJF)

²Universidade Federal de Juiz de Fora (UFJF)
Laboratório Nacional de Computação Científica (LNCC)

³Departamento de Ciências Térmicas e dos Fluidos
Universidade Federal de São João del Rei (UFJF)

jan.rocha@estudante.ufjf.br

Abstract. *Breast cancer is the most common neoplasm affecting women, and recently, auxiliary detection and monitoring techniques have been proposed, standing out the thermography due to its low cost. This study aims to compare three metaheuristics for the detection, localization, and classification of breast tumors with multiple foci through an inverse analysis using the skin surface temperature, simulated by a 2D nonlinear model of the Pennes' equation. The proposed method succeeded in distinguishing between breast with and without cancer, besides, it also classifies to the primary focus, considering the influence of external temperature.*

Resumo. *O câncer de mama é o neoplasma mais comum que afeta as mulheres, e recentemente, técnicas auxiliares de detecção e monitoramento foram propostas, destacando-se a termografia devido ao seu baixo custo. Este estudo tem como objetivo comparar três metaheurísticas para a detecção, localização e classificação de tumores de mama com múltiplos focos por meio de uma análise inversa utilizando a temperatura da superfície da pele, simulada por um modelo 2D não linear da equação de Pennes. O método proposto obteve sucesso em distinguir entre mamas com e sem tumor, além disso, também classifica o foco primário, considerando a influência da temperatura externa.*

1. Introdução

Com aproximadamente 2,3 milhões de novos casos e uma estimativa de 685.000 mortes em 2020 [Sung et al. 2021], o câncer de mama se tornou o câncer de maior incidência entre mulheres em todo o mundo, embora também possa ocorrer em homens, em casos raros. O diagnóstico do câncer de mama é geralmente realizado por meio de exames clínicos, como mamografia, ultrassom, biópsia ou ressonância magnética. Para classificar o tumor, a União Internacional Contra o Câncer (UICC) propõe o sistema TNM [Brierley et al. 2021], sendo T a extensão do tumor primário, N a ausência ou

presença de metástase em linfonodos regionais e M a ausência ou presença de metástase à distância.

Recentemente, pesquisadores têm se dedicado à exploração de abordagens alternativas e complementares mais acessíveis para o diagnóstico e detecção do câncer de mama. Entre essas abordagens, o exame de termografia tem se destacado devido ao seu custo reduzido e à natureza não invasiva em comparação com outras técnicas de diagnósticos. A proposta do uso da termografia consiste em medir a temperatura superficial da mama e a partir desta são utilizados algoritmos especializados para identificar áreas suspeitas, proporcionando apoio ao profissional de saúde no diagnóstico e acompanhamento. Além disso, a termografia apresenta a vantagem de não causar efeitos colaterais, tornando-a uma opção para o acompanhamento e evolução de casos diagnosticados.

Em outras palavras, uma possível abordagem para usar a termografia na detecção do câncer de mama envolve a medição da temperatura na superfície da pele e a aplicação de técnicas de modelagem para obter informações sobre o tumor [Figueiredo et al. 2018, Iljaž et al. 2020]. Esse processo é conhecido como um problema inverso, uma vez que envolve a determinação de variáveis desconhecidas, como as propriedades do tumor, sua forma e localização, sem medições diretas.

O problema inverso da detecção do câncer de mama por termografia pode ser dividido em duas etapas principais [Figueiredo et al. 2018, Iljaž et al. 2020]: i) inversão e ii) solução direta. O processo de inversão geralmente envolve a estimativa dos parâmetros usando algoritmos de otimização para minimizar a diferença entre as temperaturas simuladas e medidas. Já a solução direta leva em consideração um modelo de biotransferência de calor capaz de simular a distribuição da temperatura na mama com base nos parâmetros estimados na etapa de inversão.

Este trabalho tem como objetivo principal propor e comparar técnicas estocásticas de otimização aplicadas à detecção, localização e classificação TNM de tumores num modelo não-linear em regime estacionário da mama, composto por múltiplas camadas, 2D, e que apresenta diferentes números de focos tumorais, cujo número de focos também é objeto de estimativa [Rocha et al. 2024].

2. Problema Inverso

O problema inverso geométrico proposto consiste em estimar as propriedades do tumor através da minimização do erro entre a temperatura simulada e a temperatura medida ao longo da superfície da mama. Para o presente trabalho, a classe de algoritmos estocásticos baseados em população foi escolhida devido as seguintes características: i) a solução é encontrada a partir de uma população de soluções possíveis ao invés de uma única; ii) não há exigência de continuidade nas derivadas da função objetivo, sendo necessária apenas a avaliação da função objetivo e iii) a busca é adequada para implementações paralelas.

Com base nesses requisitos, três algoritmos foram escolhidos e comparados para estimar as variáveis de decisão: (i) Algoritmo Genético (GA) [Holland 1992], que é o método estocástico mais utilizado no contexto de problema inverso de biotransferência de calor; (ii) Evolução Diferencial (DE) [Storn and Price 1997], que é um algoritmo desenvolvido para otimizar funções com parâmetros reais; e (iii) Evolução Diferencial Autoadaptativa (SaDE) [Qin et al. 2009], que é um algoritmo de Evolução Diferencial cujos parâmetros são adaptados no decorrer da busca.

3. Problema direto

Neste trabalho, para o problema direto considera-se um modelo 2D em que a mama é representada como um semicírculo e é dividida em camadas distintas de tecido, incluindo epiderme, derme, gordura, glândula e músculo, conforme representado na Figura 1.

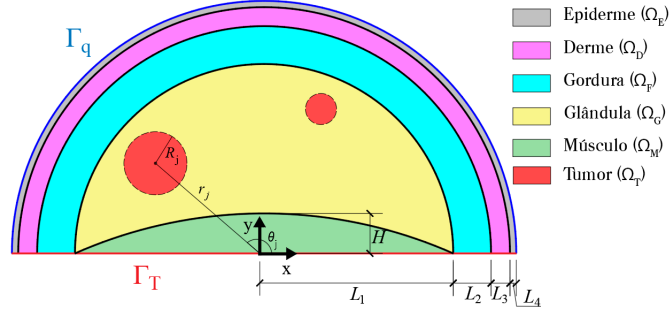


Figura 1. Geometria da mama utilizada com dois focos de exemplo.

O problema direto adotado aqui consiste em um modelo não linear governado pela equação de Pennes [Minkowycz et al. 2009]. Deste modo temos:

$$\nabla \cdot (k(T) \nabla T) + \omega_b(T) \rho_b c_b (T_a - T) + Q_m(T) = 0 \quad (1)$$

Sendo as condições de contorno (Figura 1):

$$q_n = -k(T) \nabla T \cdot \mathbf{n} = q_s, \quad \text{em } \Gamma_q \quad (2)$$

$$T = T_{a0}, \quad \text{em } \Gamma_T \quad (3)$$

onde ρ_b , c_b , T_{a0} e $\omega_b(T)$ representam, respectivamente, a densidade, calor específico, temperatura arterial e taxa de perfusão do sangue. Para o tecido, temos T , $k(T)$ e $Q_m(T)$ que correspondem, respectivamente, ao campo de temperatura, condutividade térmica e geração de calor metabólico.

Dado que a equação mencionada não possui uma solução analítica conhecida, optou-se por empregar o método dos elementos finitos para sua resolução através da utilização da biblioteca FEniCS. Para um detalhamento dos parâmetros e solução do modelo, consultar [Rocha et al. 2024].

4. Resultados e discussões

Inicialmente foi utilizado o pacote irace [López-Ibáñez et al. 2016] para buscar o melhor conjunto de parâmetros tanto para o Algoritmo Genético (GA) quanto para a Evolução Diferencial (DE). No caso da Evolução Diferencial Auto-adaptativa (SaDE) [Qin et al. 2009] consideram apenas o período de aprendizagem, o número de indivíduos e o critério de parada como variáveis de entrada. Além disso, os autores argumentam que o período de aprendizagem pode ter um baixo impacto na busca. O número de indivíduos e o critério de parada foram os mesmos para os três algoritmos, sendo 50 indivíduos e 2500 avaliações diretas do problema, respectivamente. Portanto, foi decidido não utilizar o irace para busca do valor do período de aprendizagem do algoritmo SaDE devido ao alto custo computacional da busca de parâmetros.

Aqui, os dados de temperatura na superfície da mama são simulados resolvendo o problema direto e obtendo em seguida a temperatura na superfície. Para avaliar a eficácia da busca quando dados reais são utilizados, um ruído é adicionado aos dados de referência para simular erros experimentais de medição inerentes a casos reais [Iljaž et al. 2020]. Para testes foram utilizados níveis de incerteza com valores de 25mK e 50mK [Iljaž et al. 2020]. Com objetivo de minimizar o efeito dos ruídos inseridos, foi utilizado o método proposto por [Figueiredo et al. 2018] que utiliza a média de 300 medições. Vale ressaltar que a comparação entre os algoritmos é feita utilizando uma referência fixada em dois focos.

Devido à natureza estocástica dos algoritmos, uma análise estatística é realizada considerando um total de 45 execuções para cada algoritmo. Os resultados para a busca de dois focos são apresentados na Tabela 1. É possível observar que o GA teve desempenho inferior às demais técnicas. Vale lembrar que o DE, e consequentemente o SaDE, foram originalmente projetados para resolver problemas de otimização em espaços de busca contínuos. Por exemplo, a mutação DE é baseada em diferenciação, tornando-a mais adequada do que a usada no GA para o problema aqui resolvido. O SaDE, além de apresentar melhores resultados, apresentou outra grande vantagem: a não necessidade de configurar manualmente os parâmetros CR e F.

Tabela 1. Estatísticas da função objetivo para 45 testes.

Algoritmo	Ruído	Média	Mediana	Desvio	Melhor	Pior
GA	0mK	2,545e-02	2,188e-02	1,427e-02	4,549e-03	6,556e-02
DE	0mK	1,644e-02	9,376e-03	1,459e-02	1,618e-03	6,058e-02
SaDE	0mK	1,183e-03	1,114e-04	2,107e-03	9,497e-06	8,593e-03
GA	25mK	2,497e-02	2,294e-02	1,523e-02	1,920e-03	8,209e-02
DE	25mK	1,495e-02	1,019e-02	1,193e-02	2,168e-03	5,552e-02
SaDE	25mK	3,239e-03	1,448e-03	4,718e-03	1,435e-03	2,598e-02
GA	50mK	2,214e-02	2,147e-02	1,051e-02	4,131e-03	4,429e-02
DE	50mK	1,097e-02	7,446e-03	9,722e-03	2,789e-03	4,113e-02
SaDE	50mK	4,143e-03	2,619e-03	4,255e-03	2,556e-03	2,883e-02

Visto que o algoritmo SaDE apresentou os melhores resultados na busca de dois focos, o mesmo foi escolhido para resolver o desafio de detecção, localização e classificação de n focos (círculos). Para a busca, foram considerados quatro tipos de referência: mama saudável e mamas com um, dois e três focos posicionados de formas arbitrárias.

Devido à uma maior complexidade dessa busca em relação à anterior, o número de indivíduos foi aumentado para 150 e o critério de parada foi definido como 6000 cálculos do problema direto. Para fins de teste estatístico, foram geradas 75 referências aleatórias, sendo 25 referências para cada número de foco (1, 2 e 3), além da referência saudável. Embora a UICC recomende a classificação com base no foco primário - mais especificamente aqui como sendo o maior foco - alguns autores recomendam o uso da soma de todos os focos; portanto, neste trabalho, foram consideradas as duas abordagens. É importante destacar que, devido à natureza estocástica do algoritmo, foram realizadas 25 rodadas in-

dependentes para cada uma das 76 referências (25 referências para cada número de foco mais uma referência sendo a mama saudável).

Na aplicação de conceitos de aprendizado de máquina na literatura, é comum encontrar critérios de avaliação para a qualidade das soluções propostas. Dentre os métodos mais frequentemente destacados estão quatro, conforme apontado por [Naser and Alavi 2021]: Acurácia, Precisão, Revocação e *F1-score*. Os resultados desses conceitos aplicados no problema proposto levando em consideração a capacidade do SaDE em separar mamas saudáveis de suspeitas são apresentados na Tabela 2. Ao analisar os resultados, torna-se evidente que o SaDE obteve êxito na maioria das classificações, demonstrando uma eficiente distinção entre mamas saudáveis e aquelas com tumor.

Tabela 2. Métricas de classificação na detecção do tumor.

Ruído	Acurácia	Precisão	Revocação	F1-score
0mK	0,9968	1,0000	0,9968	0,9984
25mK	0,9995	0,9995	1,0000	0,9997
50mK	0,9974	0,9995	0,9979	0,9987

Já a Tabela 3 mostra a acurácia do SaDE em classificar os focos com base no sistema TNM levando em consideração apenas o foco primário (que aqui é o maior foco) e a soma de todos os focos. É notável que a classificação com base no foco primário é mais robusta, uma vez que o foco primário, por ser o maior, tende a ter uma influência mais significativa na temperatura superficial. Mesmo no caso de três focos e 50mK de ruído, o SaDE conseguiu classificar corretamente a maioria dos tumores de referência quando o foco primário foi usado como base.

Tabela 3. Acurácia no acerto da classificação com base no sistema TNM.

Número de Focos	Método de classificação	0mK	25mK	50mK
1	Primário	0,9888	0,9584	0,9936
1	Soma dos Focos	0,9776	0,9472	0,9856
2	Primário	0,8592	0,8560	0,8464
2	Soma dos Focos	0,5632	0,5568	0,5632
3	Primário	0,8880	0,8928	0,8960
3	Soma dos Focos	0,7648	0,7616	0,7440

5. Conclusões

O objetivo deste estudo foi identificar casos de câncer de mama, localizar e classificar o tipo de câncer com base no sistema de classificação TNM utilizando algoritmos estocásticos e dados sintéticos simulados da temperatura superficial. Para isso, a equação não-linear de Pennes foi utilizada para modelar o problema direto em um modelo 2D da mama, composto por múltiplas camadas e focos.

Inicialmente uma comparação foi realizada entre Algoritmo Genético (GA), a Evolução Diferencial (DE) e a Evolução Diferencial Auto-adaptativa (SaDE). O SaDE obteve resultados superiores em comparação com o GA e o DE em um problema utilizando

uma referência com dois focos. Em seguida, utilizando o SaDE, uma busca foi realizada em um problema com uma quantidade arbitrária de focos. O algoritmo teve sucesso na distinção entre mamas com e sem tumor. Além disso, obteve resultados satisfatórios na classificação com base no foco primário, recomendado pela União Internacional Contra o Câncer, uma vez que o maior foco tende a exercer uma influência maior na temperatura externa.

6. Agradecimentos

Jan Rocha recebeu bolsas de pesquisa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES). Heder S. Bernardino recebeu apoio financeiro do CNPq e da FAPEMIG. Felipe S. Loureiro recebeu apoio financeiro do CNPq (402832/2021-3).

Referências

- [Brierley et al. 2021] Brierley, J. D., Asamura, H., van Eycken, E., and Rous, B. (2021). *TNM Atlas*, volume 7. John Wiley & Sons.
- [Figueiredo et al. 2018] Figueiredo, A. A. A., Fernandes, H. C., and Guimaraes, G. (2018). Experimental approach for breast cancer center estimation using infrared thermography. *Infrared Physics & Technology*, 95:100 – 112.
- [Holland 1992] Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge, MA, USA.
- [Iljaž et al. 2020] Iljaž, J., Wrobel, L. C., Gomboc, T., Hriberšek, M., and Marn, J. (2020). Solving inverse bioheat problems of skin tumour identification by dynamic thermography. *Inverse Problems*, 36(3):035002.
- [López-Ibáñez et al. 2016] López-Ibáñez, M., Dubois-Lacoste, J., Pérez Cáceres, L., Bittartari, M., and Stützle, T. (2016). The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives*, 3:43–58.
- [Minkowycz et al. 2009] Minkowycz, W., Sparrow, E., and Abraham, J. (2009). *Advances in Numerical Heat Transfer, Volume 3*, volume 3.
- [Naser and Alavi 2021] Naser, M. Z. and Alavi, A. H. (2021). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction*.
- [Qin et al. 2009] Qin, A. K., Huang, V. L., and Suganthan, P. N. (2009). Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Trans. on Evolutionary Computation*, 13(2):398–417.
- [Rocha et al. 2024] Rocha, J. P. A. S., Loureiro, F. S., Bernardino, H. S., Vieira, A. B., and Barbosa, H. J. (2024). Metaheuristics applied to the thermographic detection of multicentric breast tumor. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. Artigo Submetido.
- [Storn and Price 1997] Storn, R. and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.
- [Sung et al. 2021] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249.