

Análise de Imagens de Termografias Dinâmicas para Classificação de Alterações na Mama Usando Séries Temporais

Felipe Jordão P. de Andrade¹, Aristófanés Correa da Silva¹

¹Núcleo de Computação Aplicada – Universidade Federal do Maranhão (UFMA)
Av. dos Portugueses, 1966 – Bacanga – CEP 65080-805 – São Luís - MA – Brasil

felipejordaopinheiro@gmail.com, ari@dee.ufma.br

Abstract. *With the increase in the number of cases of breast cancer in recent years, there is a need for auxiliary techniques for detection of the disease. Dynamic thermography can be used as an alternative detection method based on the fact that the lesions have a higher temperature than the neighboring healthy tissues. In this work we propose a methodology for the transformation of the thermal signals into time series, from which are extracted features for a classification task. We used the K-NN and Deep Belief Network obtaining average results of 75% accuracy.*

Resumo. *Com o aumento do número de casos de câncer de mama nos últimos anos, percebe-se a necessidade de técnicas auxiliares para a detecção da doença. A termografia dinâmica pode ser usada como método alternativo de detecção com base no fato que as lesões apresentam uma temperatura mais elevada do que os tecidos saudáveis vizinhos. Nesse trabalho iremos propor uma metodologia para transformação dos sinais térmicos em séries temporais, da onde serão extraídas características para a tarefa de classificação. No trabalho usamos o K-NN e a Deep Belief Network obtendo resultados na média de 75% de acurácia.*

1. Introdução

Segundo os dados do Instituto Nacional do Câncer, o câncer de mama foi o responsável por cerca de 28% dos novos casos no último ano, sendo o segundo tipo de câncer que mais atinge as mulheres no Brasil. A detecção do câncer de mama é comumente feita pela mamografia de rastreamento, porém o exame apresenta seus pontos negativos, como alto índice de falsos positivos e a exposição da paciente à radiação ionizante [INCA 2017]. Dessa forma percebe-se a necessidade de outras metodologias que deem suporte ao diagnóstico da doença.

A termografia é uma técnica de triagem não invasiva, barata e rápida que pretende detectar as variações de temperatura da superfície da pele. As variações de temperatura do tecido canceroso em relação ao tecido vizinho saudável se dá pelo processo de angiogênese. Por esse processo o tumor estimula a criação de novos vasos sanguíneos para sua alimentação. Com mais vasos alimentando o tumor a temperatura da região se mostra superior. Kayserlingk et al. reportou em seu trabalho que a sensibilidade para a detecção de câncer do tipo carcinoma ductal é significativamente melhorada quando combinado o

exame de mamografia com a imagem térmica. A sensibilidade ao combinar essas duas técnicas foi de 95% [Keyserlingk et al. 1998].

O trabalho tem por objetivo a classificação, do conjunto de termogramas, em saudáveis ou anômalos. Dessa forma, apresentaremos uma metodologia envolvendo o processamento dos termogramas para criação de séries temporais, a extração de características dessas séries e classificação dessas características.

2. Materiais e Métodos

2.1. Aquisição

Os termogramas usados nesse trabalho foram obtidos a partir da técnica de Termografia Infravermelha Dinâmica. Esses exames pertencem a Database for Mastology Research with Infrared Image (DMR-IR) e foram adquiridas no Hospital Antônio Pedro da Universidade Federal Fluminense (HUAP-UFF) ¹. Durante o protocolo de captura, as pacientes são posicionadas em pé em uma sala a 22°C com os braços posicionados acima da cabeça e um fluxo de ar constante direcionado a região das mamas. Quando a temperatura central da paciente atingir 30,5°C, o fluxo de ar é desligado e 20 termogramas são capturados com um intervalo de 15 segundos entre eles [Silva 2015]. Esse protocolo resultou em 32 exames saudáveis e 33 anômalos.

2.2. Pré-processamento

Na etapa de pré-processamento pretendemos preparar os dados para a construção da série temporal a ser analisada.

2.2.1. Registro das Imagens

A etapa de registro tem a finalidade de alinhar as imagens de forma a compensar os movimentos causados pela respiração da paciente. A primeira imagem capturada será a imagem de referência enquanto as outras 19 passarão pelo processo de registro. Foi usado o registro deformável para o alinhamento das imagens.[Silva 2015]

2.3. Segmentação das Mamas

Também foi feita uma segmentação manual da região das mamas com objetivo de excluir os ombros e barriga do termograma. Após essa etapa foi feita a separação em mama em esquerda e direita para que seja feita a análise individual. Ao final da separação temos 130 casos, sendo 86 mamas saudáveis e 44 anômalas.

2.3.1. Conversão das Imagens para Matrizes de Temperatura

Devido ao fato das imagens originais possuírem informações adicionais inclusas pelo software da câmera, como logomarca e marcações, as imagens são convertidas para matrizes de temperatura, contendo apenas as informações de temperatura da paciente. Essa etapa é feita pelo software TermoCad desenvolvido com base no kit de desenvolvimento de software(SDK) fornecido pelo fabricante da câmera[Borchartt 2013].

¹Projetos aprovados no Comitê de Ética, registrados da Plataforma Brasil do Ministério da Saúde para pesquisa sob o CAE 01042812.0.0000.5243, aprovado pelo parecer numero 31646.

2.4. Construção das Séries

Na etapa da construção da série, temos a intenção de analisar o comportamento temporal da temperatura. A série é uma representação no plano cartesiano da evolução da temperatura da paciente enquanto seu corpo reaquece. Na série teremos um padrão que irá repetir vinte vezes (número de termogramas de um exame) variando a temperatura. Em séries saudáveis não veremos uma diferença significativa na variação da temperatura com o passar do tempo, enquanto em séries anômalas teremos uma variação notável em relação a temperaturas dos tempos anteriores. Podemos ver dois exemplos de séries criadas na Figura 1.

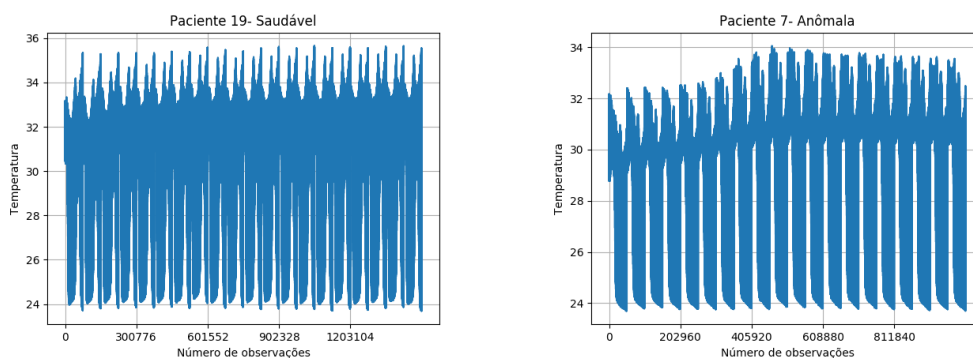


Figura 1. Exemplos das séries temporais criadas

2.5. Extração de Características

Nessa etapa extrairemos características das séries que representem nossos dados da melhor forma. As características extraídas foram: comprimento da linha, área sob a curva, energia média, amplitude média dos picos e vales e número de picos normalizados. As características escolhidas foram usadas anteriormente na análise de uma série temporal [Wulsin et al. 2011].

- Comprimento da Linha: Descreve a soma das diferenças entre consecutivos pontos da série.

$$CL = \sum_{i=2}^N |x_i - x_{i-1}| \quad (1)$$

- Área sob a curva: Descreve a área positiva normalizada sob a curva.

$$A = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (2)$$

Nessa equação o x é a lista de temperaturas para série, o i é o índice dessa lista e o N é a dimensão do vetor de temperatura.

- Energia Média: Descreve a energia média em todos os pontos da série.

$$EM = \frac{1}{N} \sum_{i=1}^N |x_i^2| \quad (3)$$

- Amplitude média dos vales: Mede a amplitude média dos vales pelo logaritmo do quadrado da amplitude para cada vale:

$$AMV = \log_{10} \left(\frac{1}{k} \sum_{i=1}^k x_{v(i)}^2 \right) \quad (4)$$

O k é o numero de vales e $v(i)$ são as posições destes.

- Amplitude média dos picos: Mede a amplitude média dos picos pelo logaritmo do quadrado da amplitude para cada pico:

$$AMP = \log_{10} \left(\frac{1}{k} \sum_{i=1}^k x_{p(i)}^2 \right) \quad (5)$$

O k é o numero de picos e $p(i)$ são as posições destes.

- Número de Picos Normalizado: Descreve o número de picos (K) normalizado pela diferença média entre pontos adjacentes da série:

$$NPN = K \left(\frac{1}{N-1} \sum_{i=1}^{N-1} x_{i+1} - x_i \right)^{-1} \quad (6)$$

2.6. Classificação das características

Para a tarefa de classificação usamos 2 classificadores de famílias diferentes. Usamos o K-Nearest Neighbors e a Deep Belief Network. O primeiro representa um classificador consistente na literatura e o usaremos como base de resultados para a comparação. O segundo classificador em comparação apresenta uma arquitetura profunda e pretendemos investigar a eficácia desse em relação ao primeiro.

Também fizemos a seleção de características usando a ferramenta Weka com o método de busca BeFirst e diversos classificadores disponíveis [Hall et al. 2009]. A seleção retornou, para os vários testes, a área e a amplitude média dos picos como os atributos que melhor separam esses dados.

2.6.1. K-Nearest Neighbors (K-NN)

O K-NN é um classificador que utiliza a distancia da instancia de teste para as instancias de treino para tomar uma decisão de classe. No trabalho usamos distancia euclidiana como função de distância e número de vizinhos igual a 3.[Cover and Hart 2006]

2.6.2. Deep Belief Network (DBN)

A DBN é um modelo generativo probabilístico, composta de uma camada visível e quantas camadas ocultas forem necessárias. Cada camada oculta pretende aprender uma relação probabilística entre as unidades da camada anterior. Esse modelo em específico tem como neurônio uma Máquina de Boltzmann Reestruturada(RBM) e é treinado gulosamente, onde cada camada oculta é treinada por vez, numa abordagem bottom-up[Hinton et al. 2006]. Na rede, usamos uma camada oculta de 500 unidades, taxa de aprendizado de 0.01, 50 iterações e função de ativação linear.

3. Resultados e Discussões

Durante os testes fizemos a divisão da base na proporção de 60% para treino e 40% para teste. Também fizemos o balanceamento de forma randômica das classes para evitar que os classificadores tenham mais instancias de uma classe para aprender. Dessa forma teremos 88 instancias, 44 instancias para cada classe. Para validar os resultados usamos a validação cruzada com 10 folds para os dois classificadores. Os resultados estão organizados na tabela a seguir:

Tabela 1. Resultados dos testes com validação cruzada

	K-NN		DBN	
	Saudável	Anômala	Saudável	Anômala
Acurácia	67%		75%	
Precisão	72%	55%	74%	76%
Sensibilidade	82%	40%	92%	40%

Com base nos resultados, percebemos que os classificadores acertam mais casos saudáveis. Vemos que a sensibilidade para os casos saudáveis é superior ao a casos com anomalias. Isso nos indica que as características conseguem mapear o padrão saudável das séries melhor que o padrão de anomalia.

Na primeira imagem da figura 2 temos um caso anômalo na mama direita da paciente no tempo 10 e a série temporal a qual ela corresponde na segunda imagem. Percebemos que o padrão de anomalia que esperávamos ver, apresentado na Figura 1, não pode ser encontrado nessa série. Podemos observar que a série é constante e com quase imperceptíveis variações de temperatura para todos os tempos observados.

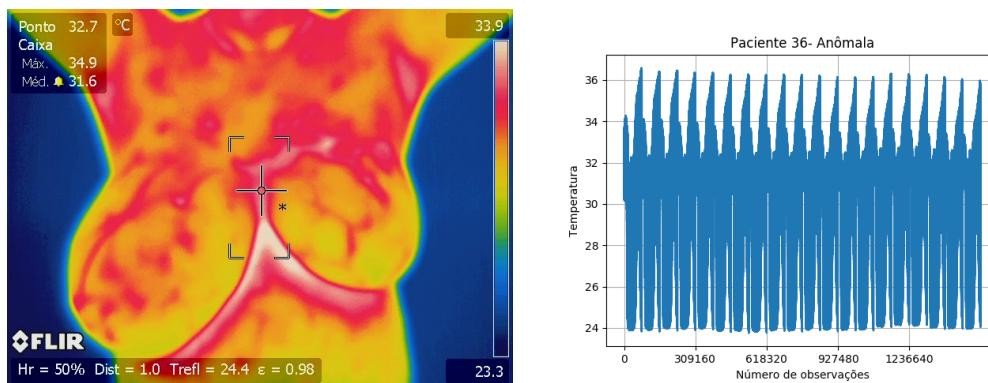


Figura 2. Exemplo de série anômala que foi classificada erroneamente

4. Conclusão

Fazendo uma análise do trabalho feito até esse ponto, concluímos que a metodologia apresenta resultados promissores, ao mesmo tempo que existe espaço para melhorias. Nos próximos passos pretendemos usar novas características que mapeiem melhor as diferenças entre as séries anômalas e saudáveis. Outra oportunidade de melhoria seria usar funções que melhorem o ruído das séries para que os padrões sejam mais evidentes, facilitando assim o processo de extração de características.

5. Agradecimentos

Os autores agradecem o apoio financeiro da Fundação de Amparo a Pesquisa e Desenvolvimento Científico do estado do Maranhão (FAPEMA).

Referências

- Borchardt, T. (2013). *Uma Análise de Imagens Termográficas para Classificação de Alteração na Mama*. PhD thesis, Universidade Federal Fluminense.
- Cover, T. and Hart, P. (2006). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.*, 13(1):21–27.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.
- Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554.
- INCA (2017). Câncer de Mama. http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/cancer_mama.
- Keyserlingk, J., Ahlgren, P., Yu, E., and Belliveau, N. (1998). Infrared imaging of the breast: Initial reappraisal using high-resolution digital technology in 100 successive cases of stage i and ii breast cancer. *The Breast Journal*, 4(4):245–251.
- Silva, L. (2015). *Uma Análise Híbrida para Detecção de Anomalias da Mama usando Séries Temporais de Temperatura*. PhD thesis, Universidade Federal Fluminense.
- Wulsin, D., Gupta, J., Mani, R., Blanco, J., and Litt, B. (2011). Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. *Journal of neural engineering*, 8(3):036015.