

Auxílio ao Diagnóstico Precoce de Patologias da Mama Usando Imagens Térmicas e Técnicas de Mineração de Dados

Roger Resmini¹, Tiago B. Borchardt¹, Aura Conci¹, Rita C. F. Lima²

Instituto de Computação – Universidade Federal Fluminense (UFF)
24.210-240 – Niterói – RJ – Brazil

Departamento de Engenharia Mecânica – Universidade Federal de Pernambuco (UFPE)
52.071-530 – Recife – PE – Brasil

{rresmini,tbonini,aconci}@ic.uff.br,ritalima@ufpe.br

Abstract. *Cancer is the most common death cause among women in the world. Accord INCA (National Cancer Institute - Brazil), early detection and treatments are the main way of cure. Among developed countries the overlife rate after five years is 85%, while the overlife in developing countries is 60%. For 2012 in Brazil, are expected 52680 new cases. Cancer is a disease in which cells reproduce uncontrollably because they have lost their function in tissue, usually due to mutations. This reproduction increases the local temperature because new blood vessels, neo-angiogenesis, are required by cancer cells. The medical thermography is a way to register the skin temperature of volunteers, and after that, to make some analysis and find thermal patterns. The human body is a symmetric thermal system and some changes in the pattern between right and left breast could identify some pathology. This paper explores the feature extractions of breast thermal images and after that classify the volunteers among with or without pathology. We extract statistic features, fractal dimension and geo-statistic features. The results are very promises with accuracy near of 90% and area under ROC curve near of 0.900%. We test three classifiers: SVM, KNN and Naïve Bayes. And two feature reduction techniques: PCA and Information Gain Ratio.*

Resumo. *O câncer de mama é a maior causa de morte entre mulheres do mundo todo. Segundo o INCA (Instituto Nacional do Câncer), a detecção e tratamento precoce são as principais formas de cura. Dentre os países desenvolvidos a taxa de sobrevivência após cinco anos é de 85%, enquanto que em países em desenvolvimento é de 60%. Para 2012, no Brasil, são esperados 52680 novos casos. O câncer, caracteriza-se por uma proliferação (reprodução) descontrolada entre células que perderam sua função no tecido a que fazem parte, geralmente devido a mutações. O aumento da população celular mutante faz com que seja necessária uma maior nutrição da região, consequentemente uma maior vascularização para conduzir os nutrientes. Essa nova vascularização, neoangiogênese, ocasiona um aumento de temperatura local. A termografia médica é uma maneira de registrar a temperatura da pele das voluntárias para posterior análise e identificação de padrões térmicos. Tendo em vista que o corpo humano é um sistema que busca a simetria térmica, a presença de alteração no padrão térmico entre as mamas esquerda e direita, é um importante indício de presença de patologias. Este trabalho tem por objetivo explorar a extração de características de*

imagens térmicas das mamas e posterior classificação das voluntárias entre com e sem patologia. Para tal, utilizou-se características estatísticas simples, dimensão fractal e características de geoestatística. Os resultados se mostraram bastante promissores com acurácia próxima de 90% e área abaixo da curva ROC próxima de 0,9%. Foram testados três classificadores, SVM, KNN e Naïve Bayes e duas técnicas de redução de características: PCA e Ganho de Informação.

1. Introdução

Dados recentes do Instituto Nacional do Câncer [Inca 2011] apontam o câncer de mama como o segundo câncer com maior frequência no mundo, sendo este o mais comum entre as mulheres. Para reduzir a taxa de mortalidade por câncer é necessário o desenvolvimento de técnicas para detecção precoce da doença, o que permite que o tratamento seja iniciado ainda em estágio inicial [Inca 2011; Ng 2001]. Existem estudos indicando que no caso do câncer de mama, a detecção no início da doença eleva para 85% a possibilidade de cura, enquanto que quando detectada em estágio avançado esse percentual cai para 10% [Ng 2001].

Atualmente, os principais aliados da detecção do câncer de mama são o autoexame, que tem baixo custo, e a mamografia, que é o principal exame utilizado para o diagnóstico em estágios iniciais do câncer de mama (exame padrão-ouro). No entanto, uma limitação encontrada pela mamografia é a dificuldade de detectar tumores em pacientes mais jovens. Este fato deve-se as mamas de mulheres jovens serem constituídas de tecido glandular, tornando-as mais densas e firmes [Koay 2004]. Gradativamente, devido à diminuição dos níveis hormonais, esse tecido tende a ser substituído por tecido adiposo (menopausa). Esta nova constituição da mama a torna menos densa facilitando a detecção de regiões suspeitas através da mamografia [Bezerra 2007]. Na Figura 1, pode ser vista uma comparação entre duas mamografias, em (a) de uma paciente jovem e em (b) de uma paciente com mais idade. Pode-se perceber que na imagem (b) existe um contraste bem maior, facilitando a identificação de possíveis patologias.

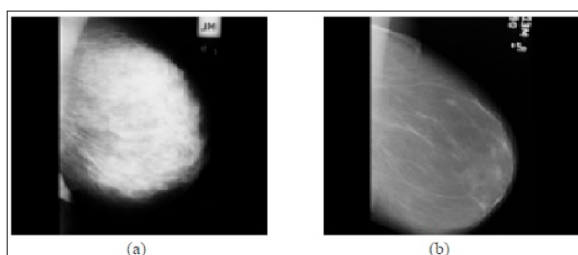


Figura 1. Mamografias: (a) mama jovem e (b) mama idosa.

Esta limitação da mamografia, juntamente com o crescente número de casos de câncer em pacientes mais jovens, motiva o desenvolvimento de novas técnicas para a detecção precoce de patologias na mama, dentre as quais, destaca-se a termografia.

A termografia começou a ser utilizada na mastologia em 1959, porém os equipamentos disponíveis na época fizeram com que o método fosse desacreditado e não recomendado para o diagnóstico de doenças na mama. Com o avanço tecnológico das novas gerações de câmeras termográficas, pesquisas envolvendo a termografia voltaram a ser

realizadas, agora utilizando novas técnicas tanto de diagnóstico como de processamento de imagens médicas [Moghbel 2011].

O objetivo deste trabalho é prover a melhoria de trabalho anterior, apresentado a seguir, através de uso de novos classificadores e métodos de seleção de atributos aplicados a três bases de dados diferentes. Estas bases de dados são constituídas de características extraídas a partir de texturas de imagens termográficas utilizando diversos extratores de características.

O trabalho está organizado da seguinte maneira: na Seção 2 é feita uma breve explicação do exame de termografia; na Seção 3 é explicado o trabalho no qual este trabalho foi baseado e expostos resultados da literatura; na Seção 4 a metodologia proposta e implementada é apresentada; na Seção 5 são apresentados os resultados obtidos e por fim, na Conclusão busca-se avaliar os resultados obtidos e discutir a viabilidade de trabalhos futuros. Este trabalho foi realizado com suporte da CAPES.

2. Termografia

A termografia é um exame fisiológico (funcional), diferente de outros exames tais como mamografia, ultrassonografia e ressonância magnética (sem uso de contraste), que são anatômicos (estruturais). As imagens térmicas, termogramas, são adquiridas por uma câmera termográfica que é sensível à radiação infravermelha. Todos os objetos que tenham temperatura acima do zero absoluto (-273°C), emitem radiação infravermelha na sua superfície [Ng 2008].

A medida da radiação infravermelha emitida pela pele pode ser convertida diretamente em um valor acurado de temperatura, devido ao alto valor de emissividade do corpo humano.

O sensor térmico de captura da radiação infravermelha varia de acordo com a câmera. Neste trabalho, são utilizadas imagens adquiridas por uma câmera termográfica modelo S45 da FLIR. O sensor presente nesta câmera tem sensibilidade de $0,08^{\circ}\text{C}$ em um intervalo de captura de -40°C à 2000°C , gerando imagens com uma resolução de 320 por 240 pixels [Flir Systems 2009]. A sensibilidade térmica desta câmera pode ser considerada dentro da faixa usada em aplicações médicas, conforme é relatado na literatura: $0,5^{\circ}\text{C}$ [Kuruganti 2002], $0,03^{\circ}\text{C}$ e $0,09^{\circ}\text{C}$ [Umadevi 2011]. Esta câmera estima a temperatura do objeto analisado baseando-se em três componentes de radiação: a radiação emitida pelo objeto de interesse, a radiação refletida pelos objetos adjacentes e a radiação emitida pelo meio.

Para que a câmera possa estimar a influência da radiação do meio em que estão sendo obtidas as imagens, é necessário configurar alguns parâmetros antes: a emissividade do objeto de interesse (no caso do corpo humano este valor é 0,98), a temperatura ambiente, a umidade relativa do ar e a distância entre a câmera e o objeto.

A partir de algoritmos internos, baseados na lei de Stefan-Boltzmann, a câmera converte a irradiação medida em uma imagem representativa de temperaturas pontuais da cena capturada. Esta imagem é apresentada em pseudocores que podem variar conforme a paleta de cores utilizada. As paletas são compostas por cento e vinte (120) cores discretas, com exceção da paleta Medical que apresenta somente dez (10) cores. Cada cor representa um intervalo de temperaturas. Ao lado de cada imagem é mostrada a paleta utilizada e os limites de temperatura representados na imagem.

