

A Termografia infravermelha para detecção precoce de câncer de mama com aprendizado computacional.

Amanda C. Q. Leles
Faculdade de Engenharia Mecânica
Universidade Federal de Uberlândia,
Uberlândia, Brasil

Gilmar Guimarães
Faculdade de Engenharia Mecânica
Universidade Federal de Uberlândia,
Uberlândia, Brasil

Juliano R. Cunha
Hospital do Câncer de Uberlândia
Universidade Federal de Uberlândia
Uberlândia, Brasil

Caroline B Gonçalves
Faculdade de Ciência da Computação
Universidade Federal de Uberlândia
Uberlândia, Brasil

Lucimara E. Oliveira
Faculdade de Ciência da Computação
Universidade Federal de Uberlândia
Uberlândia, Brasil

Henrique Fernandes
Instituto Fraunhofer IZFP para testes
não destrutivos,
Campus E3 1, 66123 Saarbrücken,
Alemanha

Resumo: *As neoplasias benignas e principalmente a malignas, como o câncer de mama, é o segundo tipo de câncer mais comum no mundo e o mais comum entre as mulheres. Apesar de sua grande eficácia, o rastreamento também apresenta resultados falso-negativos/falso-positivos e sobrediagnóstico (quando o câncer é diagnosticado, mas, caso este não fosse, não traria danos ao paciente) que são consequência de técnicas de execução do exame inadequadas e da qualidade na mamografia sendo que não possuem estudos que comprovem sua real exatidão no pré-diagnóstico. A técnica de termografia (IRT) usa o princípio da medição da radiação térmica emitida por um objeto que por sua vez é medida através de uma câmera infravermelha para a determinação de sua temperatura. Neste trabalho propõe-se o uso da técnica de IRT para a detecção do câncer de mama precoce, além de uma abordagem para aquisição das imagens, o estudo com os pacientes em três diferentes classes: pacientes saudáveis, pacientes com alterações benignas e pacientes com alterações malignas (câncer), com a aprovação do comitê de ética. Um conjunto de recursos é extraído de cada imagem e duas abordagens são usadas no processo de classificação baseado em Redes Neurais Artificiais (ANN), e em Support Vector Machines. A abordagem mostra um grande potencial para ser usada como uma técnica de diagnóstico de rastreamento para detecção precoce do câncer de mama.*

Palavras-chave: *câncer de mama, diagnóstico, termografia infravermelha; análise computacional.*

I. INTRODUÇÃO

As neoplasias benignas e principalmente a malignas, como o câncer de mama, é o segundo tipo de câncer mais comum no mundo e o mais comum entre as mulheres [2]. Apesar de sua grande eficácia, o rastreamento também apresenta resultados falso-negativos/falso-positivos e sobrediagnóstico (quando o câncer é diagnosticado, mas, caso este não fosse, não traria danos ao paciente) que são consequência de técnicas de execução do exame inadequadas e da qualidade na mamografia [1].

Porém o mais recomendado é o exame clínico [1, 10, 11, 13, 19 e 22] que possui o mesmo princípio do autoexame mas é realizado por profissionais capacitados da área da saúde, sendo a base fundamental para a continuidade dos exames (como por exemplo a mamografia), que só são prescritos após algo anormal ser identificado.

A mamografia, segundo o Instituto Nacional de Câncer (INCA), é o principal meio de diagnóstico por imagem nos casos de neoplasia mamária e continua em avanço digital para o seu melhoramento [3]. Ela é recomendada em casos que se identifica algo estranho no exame clínico, é recomendado anualmente para mulheres que estão em grupos de risco e para mulheres com idade entre 50 e 69 anos sendo bial (2 em 2 anos), [3].

Mesmo com uma boa anamnese, exame clínico e exames de diagnósticos sendo o principal a mamografia, os resultados de falso-negativos/falso-positivos e sobrediagnóstico é muito grande e gera consequências ao paciente na sua qualidade de vida [1]. A detecção precoce da doença torna-se imprescindível acesso a melhores tratamentos e chances de cura do paciente, o que aumentando suas chances de sobrevida e qualidade de vida. Neste trabalho propõe-se o uso da técnica de IRT para a detecção do câncer de mama precocemente, além de uma abordagem para aquisição dessas imagens com o uso da técnica para a detecção do câncer de mama com a aprovação do comitê de ética [5]. A técnica de termografia (IRT) usa o princípio da medição da radiação térmica emitida por um objeto que por sua vez é medida através de uma câmera infravermelha para a determinação de sua temperatura [5]. Essa técnica possui um baixo custo e pode possibilitar a detecção principalmente em mulheres jovens, que geralmente têm mamas mais densas, quando comparadas à mamografia [6]. Algumas pesquisas confirmam que o IRT pode detectar o câncer de mama mais cedo do que outras técnicas, com grande potencial detecção precoce antes da mamografia [6].

II. METODOLOGIA

Devido a esses quesitos neste trabalho aborda a aquisição de imagens infravermelhas de pacientes saudáveis e pacientes com neoplasias benignas e malignas (câncer) com extração de características para a análise das imagens usando-se redes neurais. Um recurso de classificador será treinado automaticamente em que uma parte das imagens (pacientes) é usada para treinar os classificadores e a outra parte é usada para testar os classificadores. Depois que o classificador é treinado, ele pode ser usado para classificar qualquer novo paciente (imagem).

III.

BASE DE DADOS

O banco de dados utilizado contém dados de 70 pacientes com ou sem neoplasias benignas e malignas de mama. Essas imagens foram coletadas em um período de cinco meses. As imagens foram capturadas com uma câmera infravermelha de onda média e adquiridas utilizando o protocolo estático: paciente aguardou 15 minutos para aclimação do corpo temperatura e, em seguida, foi realizada a aquisição da imagem. Imagens, de cada paciente, foram adquiridas em quatro poses diferentes: frente com os braços levantados, frente com os braços para baixo, lado esquerdo (peito esquerdo apenas) e lado direito (mama direita apenas). Para as análises realizadas neste trabalho, apenas as imagens frontais com os braços levantados foram consideradas uma vez que as imagens nas outras posições não eram padronizadas (pacientes em poses diferentes).

Os pacientes foram classificados em três grupos: sem alterações, alterações benignas (nódulos benigno) e alterações malignas (câncer). Foi obtida uma classificação prévia de cada paciente através do diagnóstico do médico especialista com base nos exames: mamografia, ultrassonografia, ressonância magnética e biópsia. Porém, alguns pacientes não realizaram todos os exames. Eles foram submetidos apenas aos exames necessários para que o médico especialista tivesse o diagnóstico conclusivo e final. Os 70 pacientes considerados são divididos da seguinte forma: 21 pacientes sem alterações na mama (normal), 23 com alterações benignas e 26 com alterações malignas.

Cada imagem no banco de dados foi então pré-processada para extrair a região de interesse (ROI), ou seja, a região que contém apenas a mama do paciente. O ROI é a única região da imagem que contém informações sobre a presença ou não do câncer de mama. Este processo de segmentação foi realizada com uma abordagem de limiarização com base no método de Otsu [7]. Ajustes finos foram feitos para corrigir quaisquer valores discrepantes. As outras partes da imagem, como o fundo, os braços e a cabeça dos pacientes não foram considerados na fase de extração de características.

IV.

EXTRAÇÃO DE DADOS

Após a extração do ROI, um conjunto de recursos foi extraído desses ROIs. A extração de características busca obter a partir das imagens informações que possam descrevê-las de forma eficaz e que o especialista apenas olhando as imagens não perceberia.

Neste trabalho, uma compilação de características encontradas na literatura [5,6,7] foi utilizada para descrever cada mama.

Os recursos usados são: (1) Média, (2) Desvio padrão, (3) Mediana, (4) Temperatura mínima, (5) Temperatura máxima, (6) Amplitude térmica, (7) Assimetria, (8) Curtose, (9) Entropia, (10) Contraste, (11) Correlação, (12) Homogeneidade, (13) Energia, (14) Momento 2, (15) Momento 3, (16) Momento 4, (17) Dimensão de contagem de caixa. Esses recursos foram escolhidos com base em alguns trabalhos disponíveis na literatura [3,7,8]. Para cada mama (direita e esquerda) todas essas características foram calculadas. Detalhes de cada característica pode ser encontrada nos trabalhos citados anteriormente.

Várias combinações de diferentes recursos foram obtidas a partir dos recursos mencionados no parágrafo anterior. Cada combinação foi usada como entrada dos classificadores para encontrar o melhor conjunto de recursos apropriados para o banco de dados usado.

V.

BASE DE DADOS

O banco de dados utilizado contém dados de 70 pacientes com ou sem tumor de mama. Essas imagens foram coletadas em um período de cinco meses. As imagens foram capturadas com uma câmera infravermelha de onda média. As imagens, "Fig 1 e Fig 2", foram adquiridas pelo protocolo estático: o paciente aguardou 15 minutos para aclimação da temperatura corporal e, em seguida, foi realizada a aquisição das imagens. As imagens, de cada paciente, foram adquiridas em quatro poses diferentes: frente com os braços levantados, frente com os braços abaixados, lado esquerdo (mama esquerdo apenas) e lado direito (mama direito apenas). Para as análises realizadas neste trabalho, foram consideradas apenas as imagens frontais com os braços levantados, uma vez que as imagens nas demais posições não foram padronizadas (pacientes em poses diferentes). Os pacientes foram classificados em três grupos: normais (sem alterações), alterações benignas (nódulos benignos) e alterações malignas (tumores). A classificação prévia de cada paciente foi obtida através do diagnóstico do médico especialista com base nos exames: mamografia, ultrassonografia, ressonância magnética e biópsia. Porém, alguns pacientes não realizaram todos os exames. Os pacientes foram submetidos apenas aos exames necessários para que o médico especialista tivesse o diagnóstico conclusivo e definitivo. As 70 pacientes consideradas estão assim divididas: 21 pacientes sem alterações na mama (normal), 23 com alterações benignas e 26 com alterações malignas.

Cada imagem no banco de dados foi então pré-processada para extrair a região de interesse (ROI), ou seja, a região contendo apenas a mama do paciente. O ROI é a única região da imagem que contém informações sobre a presença ou não do câncer de mama. Este processo de segmentação foi realizado com uma abordagem de limiar, "Fig 1.2 e Fig 2", com base no método de Otsu [6]. Ajustes finos foram feitos para corrigir quaisquer discrepâncias. As demais partes da imagem, como fundo, braços e cabeça dos pacientes, não foram consideradas na etapa de extração de traços.

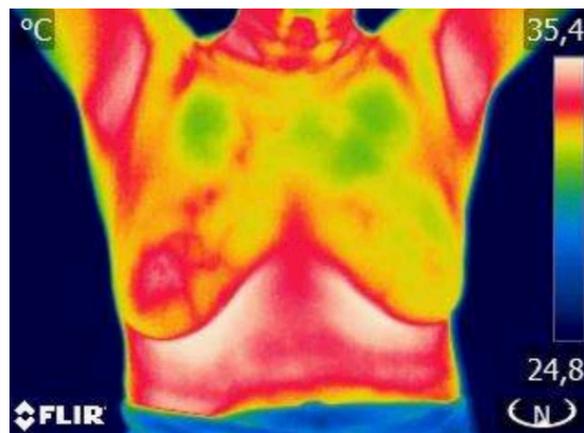


Figura 1: Imagem Infravermelha rainbow, diagnóstico de nódulo benigno. Acervo do autor.

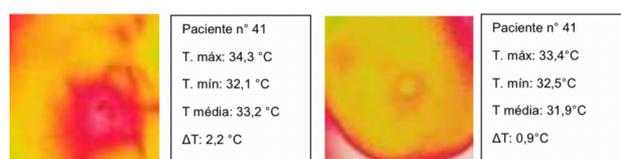


Figura 1.2 : ROIs , mama direita com nódulo benigno, mama esquerda normal e suas respectivas temperaturas. Fonte: Acervo do autor

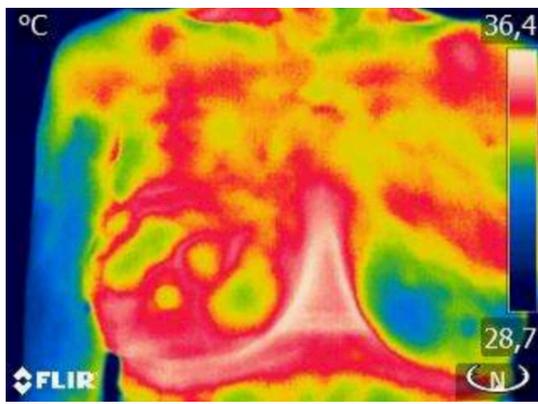


Figura 2: Imagem Infravermelha rainbow, diagnóstico de câncer. Fonte: Acervo do autor.

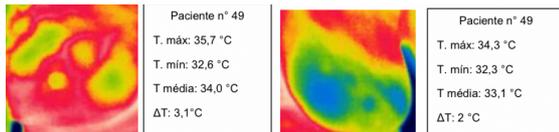


Figura 2.2 : ROIs, mama direita com câncer, mama esquerda normal e suas respectivas temperaturas. Fonte: Acervo do autor.

VI. CLASSIFICADORES

Neste trabalho, a Tabela Ia, que foi adquirida e analisada conforme descrito.

I. TABELA : HISTÓRICO DOS PACIENTES

TEMPERATURAS DAS IMAGENS TÉRMICAS									
CASO Nº	Exc	MMG	USG	Biopsia	T. MÁX (°C)	T. MIN (°C)	ΔT (°C)	Temperatura média(°C)	
								MD	ME
1	N	R	B	N	34,1	32,1	2,0	33,5	33,3
2	M	NR	B	M	35,5	32,2	3,3	33,1	34,5
3	N	X	N	N	34,8	33,7	1,1	30,5	32,1
4	M	X	M	M	33,8	31,0	2,8	31,2	32,1
5	M	M	M	M	35,7	32,7	3,0	34,7	35,1
6	M	X	M	M	34,0	31,1	2,9	31,9	32,5
7	B	B	B	B	34,1	31,6	2,5	32,6	32,2
8	N	X	N	N	33,8	32,2	1,6	32,8	32,5
9	M	NR	M	M	32,0	28,8	3,2	30,5	31,8
10	B	X	B	B	34,7	33,4	1,3	33,9	33,7
11	N	X	N	N	34,4	33,0	1,4	32,4	33,6
12	B	N	B	B	32,6	30,3	2,3	31,2	31,9
13	M	M	M	M	34,5	31,3	3,2	33,5	33,4
14	B	NR	B	B	32,5	30,5	2,0	31,0	31,4
15	B	X	B	B	33,8	31,2	2,6	32,4	30,4
16	B	NR	B	B	33,5	31,6	1,9	32,6	32,4
17	B	X	B	B	35,2	32,7	2,5	34,2	33,8
18	B	B	B	B	33,7	31,0	2,7	31,4	30,8
19	B	X	B	B	34,5	33,3	1,2	34,1	34,7
20	M	X	M	M	33,6	30,5	3,1	31,8	32,1
21	N	X	N	X	34,8	33,2	1,6	34,2	33,9
22	M	X	M	M	34,8	31,5	3,3	33,2	33,7
23	M	X	M	M	33,1	29,7	3,4	32,1	31,2
24	M	X	M	M	35,2	31,9	3,3	34,0	33,4
25	B	X	B	X	35,4	34,1	1,3	34,8	35,4
26	B	X	B	X	35,0	32,5	2,5	31,7	33,7
27	M	X	M	M	35,0	31,7	3,3	32,5	34,6
28	N	X	N	X	34,7	34,3	0,4	33,7	34,1
29	M	X	M	M	35,1	31,8	3,3	30,3	32,9
30	B	X	B	B	34,5	32,6	1,9	33,4	33,3
31	B	B	B	X	33,8	30,2	3,6	32,7	32,3
32	N	NR	N	N	33,3	31,8	1,5	32,1	32,8
33	B	X	B	X	35,1	32,7	2,4	34,0	34,3
34	N	X	N	N	33,4	32,4	1,0	32,8	33,9
35	B	X	B	B	35,6	33,7	1,9	34,4	34,7
36	B	X	B	B	34,3	31,9	2,4	31,8	32,8
37	B	X	B	B	34,6	32,3	2,3	32,6	34,3
38	N	X	N	N	35,3	33,0	2,3	34,2	34,1
39	M	M	M	M	33,8	30,3	3,5	31,4	32,5
40	B	B	B	B	35,2	33,4	1,8	34,5	34,3
41	B	X	B	B	34,3	32,1	2,2	33,4	32,5
42	B	X	B	X	33,2	30,6	2,6	31,0	31,9
43	M	X	M	M	34,7	31,5	3,2	33,7	32,6
44	N	X	N	X	34,6	33,9	0,7	34,6	34,1
45	N	X	N	X	36,4	33,7	2,7	35,6	35,3
46	M	X	M	M	34,8	31,5	3,3	33,9	32,9
47	N	NR	N	N	33,2	32,3	0,9	33,5	33,2
48	N	N	N	N	32,6	31,3	1,3	30,8	31,7
49	M	X	M	M	35,7	32,3	3,4	34,5	33,1
50	N	X	N	N	33,1	32,6	0,5	33,1	33,3
51	M	N	M	M	34,1	31,1	3,0	32,3	33,1
52	N	X	N	X	34,1	33,2	0,9	33,9	33,7
53	N	X	N	N	32,8	32,1	0,7	33,1	33,0
54	N	X	N	N	32,9	32,4	0,5	33,7	33,2
55	M	X	M	M	35,7	32,6	3,1	33,0	34,1
56	M	X	M	M	34,8	32,1	2,7	33,2	33,9
57	M	X	M	M	35,6	33,0	2,6	34,8	33,9
58	M	X	M	M	35,0	32,2	2,8	33,8	32,9
59	N	NR	N	N	32,8	32,2	0,6	32,4	32,8
60	M	X	M	M	34,9	32,1	2,8	32,9	33,7
61	M	X	M	M	34,0	31,1	2,9	31,7	32,6
62	M	X	M	M	33,9	31,6	2,3	32,5	31,9
63	M	X	M	M	34,0	31,1	2,9	32,3	33,8
64	M	X	M	M	35,3	32,2	3,1	33,4	34,6
65	B	X	B	X	33,1	30,9	2,2	31,9	31,1
66	N	X	N	X	32,8	31,7	1,1	33,3	33,2
67	B	NR	B	X	34,9	31,8	3,1	32,3	33,2
68	N	X	N	N	32,3	31,1	1,2	31,3	31,6
69	B	B	B	X	34,6	32,2	2,4	33,1	31,6
70	N	X	N	N	33,1	32,1	1,0	32,7	32,6

Exc=Exame Clínico; MMG=Mamografia; USG=Ultrassonografia; T. MÁX= Temperatura Máxima; T. MIN= Temperatura Mínima; ΔT=Diferença entre temperatura máxima e mínima; MD= Mama Direita; ME=Mama Esquerda; N= Normal, B= Benigno; M= Maligno.

a. Fonte: Acervo do autor

Foram considerados dois classificadores: Redes Neurais Artificiais (Artificial Neural Networks - ANN) e o Máquina de vetores de suporte (SVM). SVM foi escolhido porque é o classificador mais comum usado na literatura para imagens termográficas. Ambos foram implementados usando as caixas de ferramentas fornecidas por Matlab ©, *Neural Network Toolbox* (para implementação ANN) e *Machine Learning Toolbox* (para SVM implementação).

VII. EXPERIÊNCIAS

Os experimentos foram realizados com três imagens de entrada diferentes. Para o primeiro sem filtro espacial foi aplicado nas imagens. Na segunda, o filtro de mediana foi aplicado, enquanto na terceira foi aplicado um filtro gaussiano filtro com $\delta = 2$, foi aplicado. Com base nos recursos apresentados na Seção 2, sete recursos diferentes combinação foram testados. A escolha dessas combinações foi feita com base no que foi apresentado em [8,9,10]. A combinação de recursos será apresentada em detalhes na versão estendida do artigo.

VIII. RESULTADOS E DISCUSSÃO

Cada processo de treinamento foi repetido 50 vezes, sendo que foi escolhido o melhor resultado de cada teste. O melhor resultado obtido com a RNA foi alcançado com 15 neurônios na camada oculta, considerando a combinação de feições número 5 (cada uma inclui 14 feições: 1 a 11 e 14 a 16). Nessa análise cada mama deve ser analisada separadamente.

As RNAs consideradas são do tipo Rede de reconhecimento de padrões que contém uma camada oculta. A função de treinamento usada foi a retropropagação de Levenberg-Marquardt que foi escolhida empiricamente. Os critérios de parada da ANN foram 1000 iterações ou um gradiente mínimo. O número de neurônios na camada oculta foi um dos parâmetros considerados nos testes e.

O número de neurônios de entrada depende da combinação de características (recursos) que estão sendo considerados. Para o SVM, foi utilizado o modelo multiclasse de códigos de saída de correção de erros (ECOC), que permite a classificação em mais de duas classes; e a função fitcecoc do Matlab® que cria e ajusta o modelo para SVM. As funções do Kernel.

Esta configuração tinha uma precisão de 76,19%, uma especificidade normal de 57,1%, uma especificidade benigna de 83,3% e uma sensibilidade maligna de 87,5%. Para o SVM, o melhor resultado também foi alcançado usando a combinação de recursos número 5, com uma abordagem um-contratodos e uma função de kernel cúbico. Esta configuração teve uma precisão de 80,95%, uma normal especificidade de 83,33%, especificidade benigna de 85,71% e sensibilidade maligna de 75%.

A classificação (tem mudança benigna, tem alteração maligna ou nenhuma alteração) é para cada paciente. Assim, se houver uma mudança, não há informações se a alteração está localizada no seio esquerdo ou direito.

Experimentos adicionais foram realizados considerando apenas duas classes: pacientes sem alterações e pacientes com mudança. A primeira classe inclui pacientes sem alterações e com alterações benignas, enquanto a segunda

classe inclui apenas pacientes com alterações malignas. Todos os experimentos realizados antes eram repetidos e apenas as aulas (duas em vez de três) mudavam. Resultados de classificação obtidos eram melhores. Para a combinação número 5, aquela que obteve os melhores resultados nas três classes abordagem, o classificador ANN alcançou uma precisão de 90,5% e o SVM alcançou uma precisão de 85,7%.

IX. CONCLUSÃO

O uso de imagens infravermelhas para a detecção do câncer de mama é uma técnica de rastreamento promissora que pode auxiliar no diagnóstico da doença por ser uma técnica indolor, podendo identificar alterações em mamas densas (o que é difícil para a mamografia convencional), não há radiação nociva, pode identificar mudanças precoces e tem baixo custo.

Neste trabalho, foi adquirido e analisado um banco de dados com 70 pacientes. Primeiro, o ROI de cada imagem foi identificado e em segundo lugar, um conjunto de recursos foi extraído de cada ROI. Por último, um conjunto de diferentes combinações de recursos foi testado com dois classificadores diferentes (ANN e SVM), a fim de classificar os pacientes em três diferentes classes: pacientes sem alterações, pacientes com alterações benignas e pacientes com alterações malignas.

A grande maioria dos trabalhos disponíveis na literatura classifica os pacientes apenas em duas classes (sem alteração e com mudança). Assim, este trabalho foi a aquisição das imagens para desenvolver uma abordagem que pudesse diferenciar, utilizando apenas imagens infravermelhas, se possui uma alteração na mama e se ela é benigna ou maligna. Nossos resultados são promissores e confirmam que as imagens infravermelhas podem ser usadas para a detecção do câncer de mama.

O melhor resultado obtido teve uma precisão de 80,95%, uma especificidade normal de 83,33%, uma especificidade benigna de 85,71% e sensibilidade maligna de 75%. Trabalhos futuros, já em andamento, incluem (1) consideração de outras características como a assimetria térmica entre as mamas da paciente, (2) uso de genética algoritmos (GA) para selecionar automaticamente o melhor conjunto de recursos para descrever a imagem e (3) uso de profundidade aprender abordagens para classificar os pacientes.

REFERÊNCIAS

1. ASSIS, M; RAMOS, D.N; TOMAZELI, J.G. Detecção precoce do câncer de mama no Brasil: Um olhar a partir dos exames realizados no SUS. Rio de Janeiro; Instituto Nacional de Câncer José Alencar Gomes da Silva; 2014. y. Soc. London, vol. A247, pp. 529–551, April 1955.
2. BARBOSA, I.R; COSTA, I.C.C; PÉREZ, M.M.B; SOUZA, D.L.B. Mortalidade por câncer de mama nos estados do nordeste do Brasil: Tendências atuais e projeções até 2030. Revista Ciência Plural. 2015; 1(1): 4-14.
3. Instituto Nacional de Câncer (Brasil). Detecção precoce do câncer de mama. 2011. Rio de Janeiro; INCA.
4. TAVARES, T.R.P; ANDRADE, F.B; DANTAS, D.K.F; LUDOVICO, M.R.L; ARAUJO, D.V. Avaliação de indicadores para câncer de mama no período de 2009 a 2013. Revista Ciência Plural. 2016;2(1):30-41.
5. Kandlikar, S.G ; Perez-Raya, I ; Raghupathi, P.A ; Gonzalez-Hernandez, J.L ; Dabydeen, D ; Medeiros, L ;Phatak, P. Tecnologia de imagem infravermelha para detecção de câncer de mama - Status atual, protocolos e novosinstruções. Int. J. Heat Mass Transf. 2017, 108, 2303–2320.
6. Ng, E.K. A review of thermography as promising non-invasive detection modality for breast tumor. Int. J. Therm. Sci. 2009, 48, 849–859.
7. Otsu, N. A threshold selection method from gray-level histograms. In Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC), Bavaria, Germany, 8–10 October 1979; Volume 9, pp. 62–66.
8. LESSA, V.; Marengoni, M. Applying Artificial Neural Network for the Classification of Breast Cancer Using Infrared Thermographic Images. In International Conference on Computer Vision and Graphics; Springer: Berlin, Germany, 2016; pp. 429–438.
9. BORCHART, T. Análise de imagens termográficas para a classificação de alterações na mama; UFF: Niterói, Brazil, 2013.
10. ACHARYA, U.R.; Ng, E.Y.K.; Tan, J.H.; Sree, S.V. Thermography based breast cancer detection using texture features and support vector machine. J. Med. Syst. 2012, 36, 1503–1510.