

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO**



PAULO GADIOLI DE MORAES

**UMA ANÁLISE DO USO DA MATRIZ DE COCORRÊNCIA
PARA DETECÇÃO E CLASSIFICAÇÃO DE CÂNCERES E
TECIDOS EM MAMOGRAFIAS DIGITAIS**

VITÓRIA – ES
DEZEMBRO/2016

PAULO GADIOLI DE MORAES

**UMA ANÁLISE DO USO DE MATRIZ DE COCORRÊNCIA PARA
DETECÇÃO E CLASSIFICAÇÃO DE CÂNCERES E TECIDOS EM
MAMOGRAFIAS DIGITAIS**

Parte manuscrita do Projeto de Graduação do aluno **Paulo Gadioli de Moraes**, apresentada ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Orientador: Prof. Dr. Klaus Fabian Coco

Coorientador: Prof. Dr. Patrick Marques Ciarelli

VITÓRIA – ES
DEZEMBRO/2016

PAULO GADIOLI DE MORAES

**UMA ANÁLISE DO USO DE MATRIZ DE COCORRÊNCIA PARA
DETECÇÃO E CLASSIFICAÇÃO DE CÂNCERES E TECIDOS EM
MAMOGRAFIAS DIGITAIS**

Parte manuscrita do Projeto de Graduação do aluno **Paulo Gadioli de Moraes**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

COMISSÃO EXAMINADORA:

Prof. Dr. Klaus Fabian Coco
Universidade Federal do Espírito Santo - UFES
Orientador

Prof. Dr. Patrick Marques Ciarelli
Universidade Federal do Espírito Santo - UFES
Coorientador

Prof. Dr. Jorge Leonid Aching Samatelo
Universidade Federal do Espírito Santo - UFES
Examinador

Prof. Dr. Wilian Hiroshi Hisatugu
Universidade Federal do Espírito Santo - UFES
Examinador

A todos os seres, materiais e espirituais, que me ajudaram a chegar até aqui.

Agradeço primeiramente ao Cosmos, no qual estamos todos ligados de alguma forma e que influencia em todas as minhas ações e decisões ao longo de minha vida.

Aos meus pais pela imensurável ajuda ao longo de todos esses anos, dedicando tempo e amor para o meu bem-estar.

A Pedro Henrique Neres dos Reis e Yuri da Rocha Sousa pela grandiosa amizade que se sustenta desde o ensino fundamental.

A Érika Mezabarba Riva pelo apoio, carinho e fonte de incentivo durante todo o tempo, principalmente nos últimos dias.

A Vinicius Belmuds Vasconcelos pela sua presença no dia da defesa.

Aos meus orientadores pela amizade que se firmou e enorme ajuda ao longo de todo o processo de realização desse trabalho.

A banca examinadora pela aceitação do convite e tempo investido para leitura e avaliação desse trabalho.

“A compaixão pelos animais está intimamente ligada à bondade de caráter, e quem é cruel com os animais não pode ser um bom homem.”
(Arthur Schopenhauer)

RESUMO

O câncer de mama é a doença que mais mata mulheres no Brasil e no mundo. Campanhas são realizadas regularmente a fim de combater essa doença, responsável por ceifar a vida de milhares de mulheres todos os anos. A mamografia digital, junto com o autoexame da mama, é um importante método para prevenção e detecção do câncer de mama, uma vez que o primeiro consegue detectar anomalias antes mesmo de ser perceptível pelo tato. Um diagnóstico auxiliado por computador favorece um melhor resultado final para o paciente e a diminuição de possíveis conclusões errôneas por parte dos radiologistas e médicos. Portanto, faz-se necessário o desenvolvimento de técnicas computacionais que auxiliem o exame de mamografia, a fim de ajudar o diagnóstico dos profissionais de saúde.

Este trabalho consiste numa análise do uso das medidas de textura extraídas da matriz de coocorrência de nível de cinza para a detecção e classificação automática de cânceres e tecidos em mamografias digitais de forma a auxiliar o diagnóstico das mamas. A escolha da análise da matriz de coocorrência se deve ao seu uso comum para detecção de câncer de mama em trabalhos da literatura e a sua fácil obtenção. O presente trabalho se divide em duas partes: 1) detecção e classificação de câncer e 2) classificação do tecido mamário. Em ambas as partes, calcularam-se as medidas de textura das regiões de interesse baseadas na matriz de coocorrência. Finalmente, foram usadas duas técnicas de classificação (kNN e SVM) para verificar a validade de detecção e classificação.

Palavras-chave: Diagnóstico auxiliado por computador, câncer de mama, tecido mamário, matriz de coocorrência de nível de cinza, processamento digital de imagem.

ABSTRACT

Breast cancer is the disease that most kill women in Brazil and world. Campaigns are realized regularly in order to fight against this disease, responsible to reap thousands of women's life every year. Digital mammography, together with the mamma's self-exam, is an important method to prevent and detect breast cancer, since the first can detect anomalies even before to be noticeable by touch. A computer-aided diagnosis favors a best final result to the patient and the decrease of possible erroneous conclusions by radiologists and physicians. Therefore, it is necessary the development of computer techniques that assist mammography exam, in order to help the diagnosis of health professionals.

This labor consists in a use analysis of texture measures extracted from gray level co-occurrence matrix to automatic detection and classification of cancers and tissues in digital mammograms in order to auxiliary mamma's diagnosis. The choice of co-occurrence matrix analysis should to your common use to detect breast cancer in literature's labors and your easy achievement. The present labor is divided in two parts: 1) detection and classification of cancer and 2) classification of breast tissue. In this way in both parts, were calculated the region of interest's textures measures based in the co-occurrence matrix. Finally, were used two classification techniques (kNN and SVM) were tested to verify the validity of detection and classification.

Keywords: Computer-aided diagnosis, breast cancer, breast tissue, gray level co-occurrence matrix, digital image processing.

LISTA DE FIGURAS

Figura 1 – Tecido com a presença de células cancerígenas.....	16
Figura 2 – Mamografia digital da mama.	21
Figura 3 – Diferença no formato de nódulos.	22
Figura 4 - Anatomia da mama e alguns aspectos representados em uma mamografia.	23
Figura 5 – Diferença de tecidos mamários.	24
Figura 6 – Etapas do processamento de imagens e classificação.	31
Figura 7 – Suavização usando filtro de média.	32
Figura 8 – Realce através da equalização do histograma.	33
Figura 9 – Binarização de uma imagem.	35
Figura 10 – Erosão em tons de cinza em uma matriz de pixels.....	36
Figura 11 – Erosão em uma imagem.	36
Figura 12 – Dilatação em tons de cinza em uma matriz de pixels.	37
Figura 13 – Dilatação em uma imagem.	37
Figura 14 – Abertura em uma imagem.	38
Figura 15 – Extração de fronteiras em uma imagem.	39
Figura 16 – Mapeamento de pontos no novo espaço de parametrização da curva.....	40
Figura 17 – Transformada de Hough em uma imagem.	41
Figura 18 – Diferentes tipos de texturas.	42
Figura 19 – Matriz de coocorrência de tons de cinza.	44
Figura 20 – Diferentes direções para a matriz de coocorrência.....	44
Figura 21 – Classificador supervisionado.	53
Figura 22 – Classificação através do algoritmo kNN.....	54
Figura 23 – Classificação através do algoritmo SVM.....	55
Figura 24 – Exemplos de mamografias utilizadas.	58
Figura 25 – Fluxograma do processo de detecção e classificação de câncer e tecido mamário.....	60
Figura 26 – Mamografia com câncer e sua respectiva ROI dada pelo banco de dados.	61
Figura 27 – Mamografia sem câncer e sua respectiva ROI manual.	61
Figura 28 – ROIs (com câncer) com e sem realce.....	62
Figura 29 – Histogramas para ROIs (com câncer) com e sem realce.....	62
Figura 30 – ROIs (sem câncer) com e sem realce.	63
Figura 31 – Histogramas para ROIs (sem câncer) com e sem realce.	63

Figura 32 – Mamografia para classificação do tecido mamário.....	64
Figura 33 – Etapa de espelhamento da mama para o lado esquerdo.	65
Figura 34 – Suavização da mamografia.....	65
Figura 35 – Limiarização da mamografia.....	66
Figura 36 – Objeto responsável pela abertura da mamografia.	66
Figura 37 – Contagem de objetos e permanência do maior objeto da mamografia.	67
Figura 38 – Preenchimento de buracos da mamografia.....	68
Figura 39 – Obtenção da mamografia sem ruídos e objetos e com músculo peitoral.	68
Figura 40 – Obtenção do contorno da mamografia.	69
Figura 41 – Colunas selecionadas para delimitar o polígono contendo o músculo peitoral...	70
Figura 42 – Polígono determinado.	70
Figura 43 – Obtenção da mamografia em forma do polígono.....	71
Figura 44 – Binarização e contorno da mamografia em forma do polígono.	71
Figura 45 – Obtenção da mamografia sem ruídos e objetos e sem músculo peitoral.....	72
Figura 46 – Comparativo da mamografia original espelhada e da mamografia obtida na última etapa.....	73
Figura 47 – Etapa de definição dos limites da ROI.	73
Figura 48 – Marcação da ROI para o tecido mamário.....	74
Figura 49 – ROIs (tecido mamário) com e sem realce.	75
Figura 50 – Histogramas para ROIs (tecido mamário) com e sem realce.	75
Figura 51 – Gráfico em radar para as melhores texturas no teste de detecção de câncer usando kNN.	79
Figura 52 – Gráfico em radar para as melhores texturas no teste de detecção de câncer usando SVM.	80
Figura 53 – Gráficos do SFS para as métricas na detecção de câncer usando kNN.....	81
Figura 54 – Gráfico em barra para diferentes ângulos da GLCM no teste de detecção de câncer usando kNN.....	83
Figura 55 – Gráfico em barra para diferentes ângulos da GLCM no teste de detecção de câncer usando SVM.....	83
Figura 56 – Gráfico em radar para as melhores texturas no teste de classificação de câncer usando kNN.	86
Figura 57 – Gráfico em radar para as melhores texturas no teste de classificação de câncer usando SVM.	87
Figura 58 – Gráficos do SFS para as métricas na classificação de câncer usando kNN.	88

Figura 59 – Relação entre as duas primeiras medidas de textura no teste de classificação de câncer usando SFS e kNN.	89
Figura 60 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação de câncer usando kNN.	91
Figura 61 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação de câncer usando SVM.	91
Figura 62 – Gráfico em radar para as melhores texturas no teste de classificação do tecido mamário usando kNN.	94
Figura 63 – Gráfico em radar para as melhores texturas no teste de classificação do tecido mamário usando SVM.	95
Figura 64 – Gráficos do SFS para as métricas na classificação do tecido mamário usando kNN.	96
Figura 65 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação do tecido mamário usando kNN.	98
Figura 66 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação do tecido mamário usando SVM.	98
Figura 67 – Três situações obtidas quando da remoção do músculo peitoral.	99

LISTA DE TABELAS

Tabela 1 – Estimativa de novos casos em 2016 para os tipos de câncer no Brasil.	18
Tabela 2 – Matriz de confusão.....	55
Tabela 3 – Média para as métricas no teste de detecção de câncer usando kNN e SVM.....	78
Tabela 4 – Média para as métricas em cada ângulo no teste de detecção de câncer usando kNN e SVM.	82
Tabela 5 – Média para as métricas no teste de classificação de câncer usando kNN e SVM.	85
Tabela 6 – Média para as métricas em cada ângulo no teste de classificação de câncer usando kNN e SVM.....	90
Tabela 7 – Média para as métricas no teste de classificação do tecido mamário usando kNN e SVM.....	93
Tabela 8 – Média para as métricas em cada ângulo no teste de classificação do tecido mamário usando kNN e SVM.	97

LISTA DE ABREVIATURAS E SIGLAS

UFES	Universidade Federal do Espírito Santo
MIAS	Mammographic Image Analysis Society
INCA	Instituto Nacional de Câncer
CAD	Computer-Aided Diagnosis
MATLAB	MATrix LABoratory
ROI	Region of interest
GLCM	Gray Level Co-occurrence Matrix
kNN	k-Nearest Neighbor
SVM	Support Vector Machine
VP	Verdadeiro Positivo
FN	Falso Negativo
FP	Falso Positivo
VN	Verdadeiro Negativo
SFS	Sequential Forward Selection

LISTA DE SÍMBOLOS

$f(x,y)$	Função bidimensional que representa uma imagem digital
$b(x,y)$	Função bidimensional que representa a imagem binarizada
μ	Média
σ	Variância
P	Matriz de Coocorrência de Nível de Cinza
ρ	Função de densidade de probabilidade

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Mamografia	20
1.2	Tipo de câncer de mama.....	22
1.3	Tecido mamário.....	23
1.4	Motivação do trabalho.....	24
1.5	Trabalhos relacionados.....	25
1.6	Objetivos geral e específico.....	28
1.7	Estrutura do trabalho	29
2	TÉCNICAS ABORDADAS NO TRABALHO	30
2.1	Aquisição e digitalização das imagens	31
2.2	Pré-processamento.....	31
2.2.1	Suavização.....	32
2.2.2	Realce de contraste.....	33
2.3	Segmentação.....	34
2.3.1	Binarização.....	34
2.4	Pós-processamento	35
2.4.1	Erosão.....	36
2.4.2	Dilatação.....	37
2.4.3	Abertura.....	38
2.4.4	Extração de fronteiras.....	39
2.4.5	Transformada de Hough.....	40
2.5	Extração de atributos	42
2.5.1	Análise de Texturas	42
2.5.2	Matriz de Coocorrência de Nível de Cinza (GLCM).....	43
2.5.3	Medidas de Texturas	45
2.6	Classificação e Reconhecimento	52
2.6.1	Algoritmos de Classificação.....	53
2.6.1.1	<i>k-Nearest Neighbor</i> (kNN).....	53
2.6.1.2	<i>Support Vector Machine</i> (SVM)	54
2.6.2	Validação dos Classificadores.....	55
2.7	Seleção de atributos.....	57
3	METODOLOGIA E DESENVOLVIMENTO	58

3.1 Detecção e Classificação de câncer	61
3.2 Classificação do tecido mamário	64
3.3 Classificação do Algoritmo	76
4 EXPERIMENTOS E RESULTADOS.....	77
4.1 Detecção de câncer	77
4.2 Classificação de câncer.....	84
4.3 Classificação do tecido mamário.....	92
4.4 Remoção do músculo peitoral	99
5 CONCLUSÃO	100
6 REFERÊNCIAS BIBLIOGRÁFICAS	102

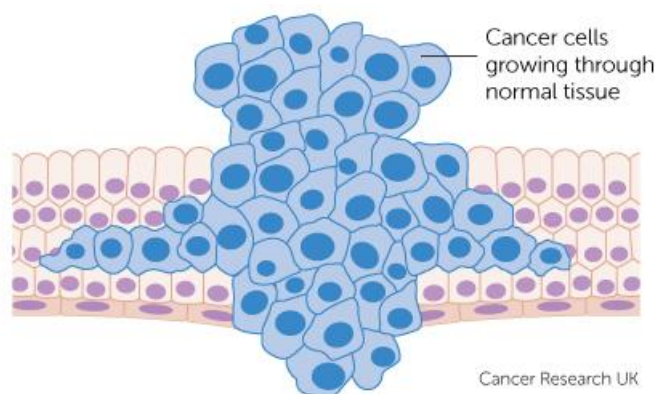
1 INTRODUÇÃO

Dentre as diversas doenças conhecidas pela medicina atual, o câncer tem crescido consideravelmente e gerado preocupação, incentivando o desenvolvimento de métodos de prevenção cada vez mais precoce.

Segundo o Instituto Nacional de Câncer (INCA, s.d.), câncer é o nome dado a um conjunto de mais de cem doenças que têm em comum o crescimento desordenado de células, que invadem tecidos e órgãos. Estas células, por sua vez, podem dividir-se rapidamente e se tornarem muito agressivas e incontroláveis, podendo ocasionar a formação de tumores malignos e se espalharem para outras regiões do corpo. A esse espalhamento dá-se o nome de metástase. Por outro lado, um tumor benigno significa apenas uma massa localizada de células que se multiplicam de forma menos agressiva, assemelhando-se ao tecido original e raramente oferecendo risco ao paciente (INCA, s.d.).

A figura 1 mostra uma representação de um tecido com a presença de células cancerígenas em um estágio avançado de desenvolvimento, rompendo os limites da estrutura e atingindo outros órgãos.

Figura 1 – Tecido com a presença de células cancerígenas.



Fonte: Cancer Research UK, 2014.

A diferença entre os vários tipos de câncer se dá pela grande variedade de células do corpo. Em um mesmo tipo de órgão como, por exemplo, a pele, pode existir diversos tipos de câncer, uma vez que esse órgão possui muitos tipos de células. O câncer que se origina em tecidos epiteliais como pele e mucosas é denominado carcinoma, enquanto o que se desenvolve em tecidos conjuntivos como ossos, músculos ou cartilagens é denominado sarcoma (INCA, s.d.).

Não existe uma causa principal para o surgimento e manifestação do câncer, e sim um inter-relacionamento de diversos fatores, sejam eles ambientais e/ou hereditários.

As causas ambientais são também chamadas de causas externas e se dá pelos hábitos ou costumes próprios do indivíduo em um ambiente social e cultural. Segundo (INCA, s.d.), dentre todos os casos, 80% a 90% dos cânceres estão associados a fatores ambientais, muitos deles velhos conhecidos: o uso excessivo de cigarro, a exposição excessiva ao sol, a ingestão de determinados tipos de alimentos, o alcoolismo, uso excessivo de medicamentos, dentre outros. Outras causas, porém, ainda são completamente desconhecidas. Sabe-se, no entanto, que o surgimento do câncer depende da intensidade e duração da exposição das células aos agentes causadores de câncer.

As causas hereditárias são chamadas de causas internas e são, em sua maioria, geneticamente pré-determinadas, estando ligado à capacidade do organismo de se defender de agressões externas. Raramente um câncer se dá exclusivamente por fatores hereditários, familiares e étnicos. Sabe-se, no entanto, que o fator genético tem um papel importante no surgimento e desenvolvimento do câncer. Em nível de curiosidade, alguns tipos de câncer tais como mama, estômago e intestino parecem ter uma influência genética alta, ao passo que certos tipos de câncer são raros em algumas etnias (leucemia linfocítica em orientais e sarcoma de Wing em negros).

Dentre os vários tipos de câncer existentes, podem-se destacar: câncer de próstata, câncer de cólon e reto, câncer de pulmão, câncer de estômago, câncer da cavidade oral, câncer do esôfago, câncer de pele, câncer de ovário, câncer do corpo de útero e câncer de mama.

O tipo de câncer mais frequente entre as mulheres no Brasil e no mundo é o câncer de mama. Segundo dados do INCA (2016), são esperados, somente no Brasil, 57.960 novos casos de câncer de mama para o ano de 2016, com um risco estimado de 56 casos a cada 100 mil mulheres.

A tabela 1 mostra a estimativa de novos casos em 2016 das taxas brutas de incidência por 100 mil habitantes e do número de novos casos de câncer, segundo sexo e localização primária. Pode-se perceber que, dentre os diversos tipos de câncer acometidos pela população, o câncer de próstata (mais presente em homens) e o câncer de mama (mais presente em mulheres) possuem um número muito maior de novos casos estimados quando comparado aos demais. Cerca de 58.000 novos casos de câncer de mama são esperados nos estados ao redor do Brasil, onde quase 19.000 são somente nas capitais. O câncer de mama é a doença com maior taxa bruta por 100 mil habitantes entre as capitais.

Tabela 1 – Estimativa de novos casos em 2016 para os tipos de câncer no Brasil.

Localização Primária da Neoplasia Maligna	Estimativa dos Casos Novos							
	Homens				Mulheres			
	Estados		Capitais		Estados		Capitais	
	Casos	Taxa Bruta	Casos	Taxa Bruta	Casos	Taxa Bruta	Casos	Taxa Bruta
Próstata	61.200	61,82	13.940	64,93	-	-	-	-
Mama Feminina	-	-	-	-	57.960	56,20	18.990	79,37
Colo do Útero	-	-	-	-	16.340	15,85	4.550	19,07
Traqueia, Brônquio e Pulmão	17.330	17,49	4.430	20,59	10.890	10,54	3.230	13,49
Cólon e Reto	16.660	16,84	5.560	25,80	17.620	17,10	6.210	25,95
Estômago	12.920	13,04	3.130	14,54	7.600	7,37	2.180	9,07
Cavidade Oral	11.140	11,27	2.780	12,95	4.350	4,21	1.230	5,04
Laringe	6.360	6,43	1.600	7,50	990	0,94	320	0,97
Bexiga	7.200	7,26	2.110	9,79	2.470	2,39	830	3,21
Esôfago	7.950	8,04	1.460	6,75	2.860	2,76	610	2,27
Ovário	-	-	-	-	6.150	5,95	2.170	8,92
Linfoma de Hodgkin	1.460	1,46	450	1,74	1.010	0,93	400	1,33
Linfoma não Hodgkin	5.210	5,27	1.550	7,15	5.030	4,88	1.670	7,02
Glândula Tireoide	1.090	1,08	350	1,27	5.870	5,70	1.800	7,46
Sistema Nervoso Central	5.440	5,50	1.290	5,86	4.830	4,68	1.250	5,20
Leucemias	5.540	5,63	1.370	6,38	4.530	4,38	1.180	4,88
Corpo do Útero	-	-	-	-	6.950	6,74	2.530	10,47
Pele Melanoma	3.000	3,03	840	3,86	2.670	2,59	740	2,96
Outras Localizações	51.850	52,38	11.890	55,45	47.840	46,36	11.820	49,33
Subtotal	214.350	216,48	52.750	245,63	205.960	199,57	61.710	257,55
Pele não Melanoma	80.850	81,66	17.370	80,90	94.910	91,98	21.910	91,65
Todas as Neoplasias	295.200	298,13	70.120	326,51	300.870	291,54	83.620	348,99

Fonte: INCA, 2016.

Segundo dados da Agência Internacional para a Pesquisa do Câncer, agência da Organização Mundial da Saúde, 25% dos casos de câncer em mulheres no mundo são de mama. Ou seja, um a cada quatro tipos de câncer que afetam as mulheres é de mama (IARC, 2016 apud SANTOS, 2016).

Em 2012, últimos dados mundiais publicados pela agência, a taxa de incidência desse tipo de câncer era de 1,6 milhão de casos entre mulheres. Ainda, de acordo com a Agência Internacional para a Pesquisa do Câncer, em 2012, 14,7% dos casos de morte por câncer eram de pacientes com câncer de mama. O número de mortes naquele ano foi de 521.907 (SANTOS, 2016).

A fim de diminuir os riscos provenientes do câncer de mama, no mês de outubro é celebrado o Outubro Rosa. Com origem em Nova York, nos Estados Unidos, em 1991, com uma corrida pela cura do câncer de mama, o referido mês é caracterizado por uma série de ações realizadas para conscientizar as mulheres (e também os homens) de que devem realizar o exame de toque, a fim de se detectar calcificações e se prevenir o mais cedo possível do câncer de mama. É comum, neste mês do ano, diversos prédios e monumentos serem iluminados por luzes rosa, a fim de lembrar a todos do risco da doença e da importância do diagnóstico precoce (Glamour, 2016).

A campanha conta com o apoio de diversos setores da saúde, ONGs, órgãos públicos, mídia e população em geral, e chama cada vez mais a atenção de marcas e celebridades em solidariedade à causa.

Após cinco anos de detecção e evolução do câncer de mama, a média de sobrevivência é de 61% na população mundial. Sua incidência aumenta de forma rápida e progressiva a partir dos 35 anos de idade (RIUL, 2012). Qualquer mulher pode desenvolver o câncer de mama, porém nos homens ele ocorre mais raramente. Vários fatores podem contribuir para o surgimento desse tipo de câncer, dentre eles podemos citar: ter tido filho após os 30 anos ou nunca ter tido filhos, ter histórico familiar (parentes de primeiro e segundo grau que possuíram a doença), consumo excessivo de álcool, ser fumante, ter feito tratamentos hormonais e obesidade (DANTAS; COURA; ALVARENGA, 2012).

Segundo o INCA (2016), 30% dos casos de câncer de mama podem ser evitados através de bons hábitos, que envolvem a prática regular de atividade física, uma alimentação saudável, evitar o consumo de bebidas alcoólicas e manter um peso adequado. Portanto, ter uma vida saudável é a receita ideal para a prevenção da doença.

O autoexame da mama, divulgado enormemente pelo Outubro Rosa, caracteriza-se pela apalpação das mamas, e deve ser realizado todo mês. Mulheres que ainda menstruam devem realizar o autoexame de quatro a seis dias após o fim da menstruação, pois nesse período as mamas estão menores e menos sensíveis. Já as mulheres que estão na menopausa devem escolher um dia no mês e realizar o autoexame todo mês nesse mesmo dia.

Em outubro de 2016, uma pesquisa do INCA (2016) apontou que 66,2% dos casos de câncer de mama são descobertos pelas próprias pacientes ao notarem alguma alteração na mama. Apesar do autoexame geralmente só detectar nódulos e malformações que já ultrapassaram um centímetro de extensão, esse dado reforça ainda mais sua importância para a prevenção e detecção precoce da doença. Completando os dados, segundo o estudo, o percentual de mulheres que identificaram a doença por meio da mamografia ou outro exame de imagem foi de 30,1%, ao passo que apenas 3,7% tiveram a suspeita vinda de um profissional da saúde.

1.1 Mamografia

Um método considerado padrão para a detecção precoce do câncer de mama é a mamografia digital. É um exame de imagem, em que uma dosagem baixa de raios-X é aplicada sobre a mama com o objetivo de examiná-la e auxiliar no diagnóstico das doenças mamárias.

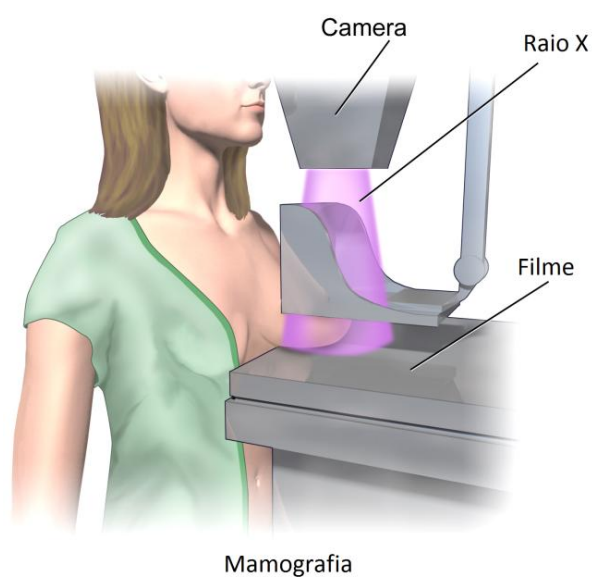
A mamografia é de vital importância, pois possibilita a detecção de anomalias na mama com até dois anos de antecedência de ser detectada pelo tato. Enquanto o exame clínico das mamas quando realizado por um médico ou uma enfermeira é capaz de detectar tumores com cerca de um centímetro, a mamografia consegue mostrar uma lesão, em fase inicial, muito menor (na escala de milímetros) (INCA, s.d.).

Vale lembrar que existem dois métodos de mamografia existentes: a mamografia digital e a mamografia convencional. No primeiro método, o detector da máquina de raios-X torna a radiação incidente em uma imagem digital que é transmitida imediatamente à tela de um computador. No segundo, também chamada de indireta, o detector possui uma substância especial que armazena os dados radiográficos, sendo essa substância direcionada a uma máquina, tornando as informações em uma imagem digital (Clínica da Mama, 2014).

Apesar de não haver consenso hoje entre os médicos e especialistas quanto ao início da prevenção, de um modo geral as mulheres com um risco médio de câncer de mama passam a realizar o exame de mamografia a cada 1-2 anos a partir dos 40 anos. Mulheres com um alto risco de câncer de mama começam antes dos 40 anos a se submeter ao exame de raios-X da mama.

A figura 2 mostra como se dá o posicionamento da paciente quando do exame. A paciente posiciona-se de pé e próxima ao aparelho de mamografia. O seio é fixado em uma espécie de prensa para que não haja movimento durante o intervalo de realização do exame. Uma câmera localizada acima do seio emite um feixe de raios-X em direção a um filme fotográfico que transmite os dados de imagem diretamente a um computador. Características como presença de nódulos cancerígenos e tipo de tecido mamário determinam o grau de absorção e penetração de raios-X pela mama.

Figura 2 – Mamografia digital da mama.



Fonte: Instituto Espaço da Vida, 2016.

Como mostrado na imagem acima, a compressão da mama é necessária devido a uma série de razões (SILVA, 2013):

- O tecido mamário deve ser espalhado para que pequenas anormalidades não sejam obscurecidas pelo tecido mamário superior;
- Radiografar uma camada mais fina do tecido mamário, de modo a permitir o uso de uma dose mais baixa de raios-X;
- Conter a mama, evitando assim possíveis borramentos na imagem causados pelo movimento da paciente;
- Reduzir a radiação espalhada e assim produzir imagens com maior nitidez;
- Igualar a espessura da mama de modo que todo o tecido possa ser visualizado.

É importante ressaltar que a mamografia não consegue detectar todos os tipos de câncer. Um determinado tipo de câncer pode estar em uma área que é difícil de se ver pelo exame de mamografia, como a axila. Em geral, mamografias podem perder 1 em cada 5 cânceres em mulheres (Saúde Melhor, s.d.). Portanto, podemos concluir que a mamografia é um método de auxílio ao diagnóstico, mas não é o exame principal e suficiente. O autoexame regular ainda é fortemente indicado para todas as pacientes.

1.2 Tipo de câncer de mama

Um tumor mamário é considerado do tipo benigno se é limitado a pequenas camadas de célula, não ultrapassando tecidos vizinhos ou órgãos. Porém, se o tumor se espalha para tecidos vizinhos ou órgãos, ele é considerado maligno.

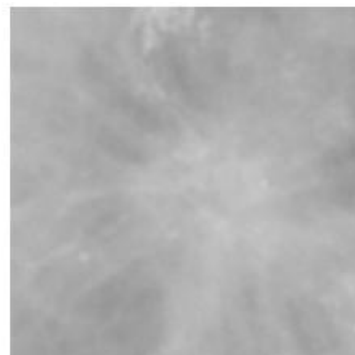
Um indicativo do grau de risco de um nódulo mamário é o seu formato. Aqueles que apresentam formato irregular ou “espiculado” têm maiores chances de serem malignos. No entanto, alguns casos malignos também podem apresentar contornos bem definidos. Portanto, a identificação do contorno dos nódulos é de suma importância, tarefa essa de caráter rotineiro para os radiologistas.

A figura 3 mostra a diferença no formato entre dois nódulos em uma mamografia digital: circunscrito (a), com maior possibilidade de ser benigno, e espiculado (b), com maior possibilidade de ser maligno. Outros formatos existem (calcificações, distorção arquitetural, assimétrico e mesmo os sem definição), portanto, torna-se um grande desafio classificar o tipo de câncer de acordo com o formato de seu nódulo.

Figura 3 – Diferença no formato de nódulos.



a) Nódulo circunscrito.



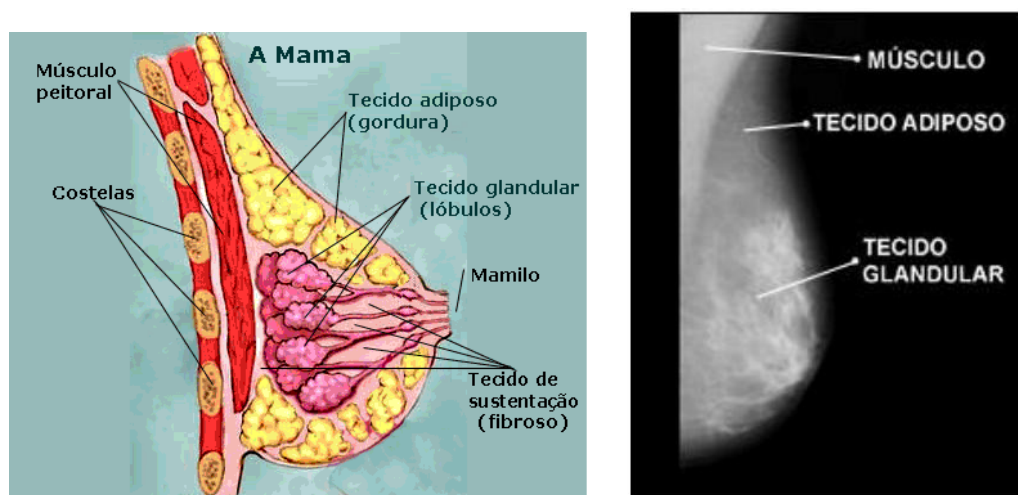
b) Nódulo espiculado.

Fonte: Produção do próprio autor.

1.3 Tecido mamário

De um modo geral, as mamas consistem de três tipos principais de tecidos: adiposo, glandular e fibroso. A quantidade de cada um desses tecidos na mama pode variar bastante dependendo da idade da paciente. A figura 4 abaixo mostra, em (a) um esboço dos principais aspectos da anatomia da mama e, em (b), alguns desses aspectos são mostrados em uma mamografia.

Figura 4 - Anatomia da mama e alguns aspectos representados em uma mamografia.



a) Esboço da anatomia da mama.

b) Mamografia.

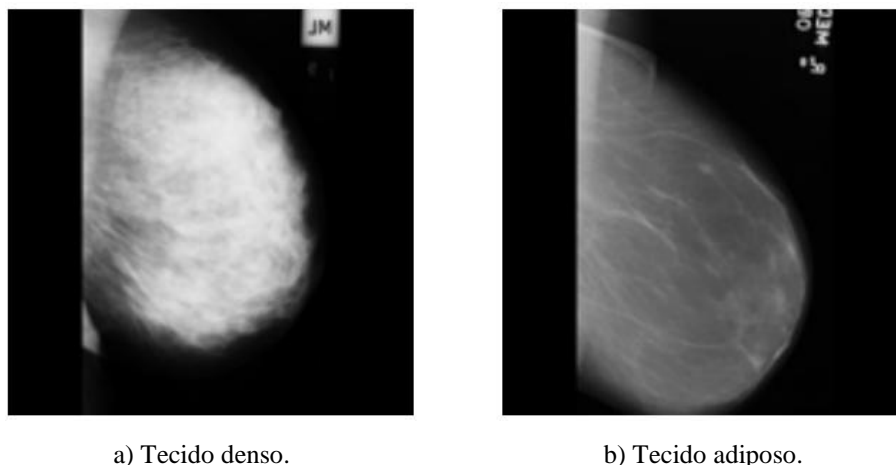
Fonte: REY, 2012 (a) e MASCARO, 2007 (b).

Devemos destacar que mamografias em mulheres mais jovens podem ser mais difíceis de serem interpretadas. Isso se deve ao fato de que os seios de jovens possuem mais glândulas e ligamentos do que os de mulheres mais velhas, resultando em tecido mamário mais denso e podendo obscurecer sinais de câncer. Com o passar dos anos, o tecido da mama torna-se menos glandular e mais adiposo, tornando mais fácil a interpretação e detecção de alterações na mamografia (SANTOS, 2002).

O tecido adiposo (gordura) aparece mais escuro na mamografia digital, ao passo que o tecido glandular (mais presente em mulheres mais jovens) aparece mais claro na imagem. Esse fato dificulta muito a identificação de doenças em jovens com menos de 35 anos (MASCARO, 2007).

A figura 5 mostra a diferença entre as mamografias de uma paciente jovem e de uma paciente mais idosa. Em (a) temos uma paciente mais jovem, com a predominância de tecido denso. Em (b) uma paciente mais idosa com mais tecido adiposo.

Figura 5 – Diferença de tecidos mamários.



Fonte: MIAS s.d.

1.4 Motivação do trabalho

A população mundial necessita, cada vez mais, de novas tecnologias na área da saúde. A expectativa de vida das pessoas está crescendo cada vez mais, novas curas para doenças estão sendo descobertas e novos equipamentos estão sendo projetados para auxiliar cirurgias e profissionais da saúde.

O câncer, porém, ainda é uma doença que agride uma parcela grande da população e muitos deles carecem de um método eficiente de cura. Muitas vezes, por demora na detecção da doença, a cura por cirurgia não é mais possível, sendo necessária a realização da quimioterapia para retardar o avanço da doença e prolongar o tempo de vida do paciente. Os efeitos colaterais desse processo são vários, podendo-se citar: queda de cabelo, enjojo, vômito, perda de apetite, fraqueza, diarreia, dor no corpo, febre e baixa imunidade. Não é, portanto, a melhor forma de se combater o câncer.

A melhor maneira de se combater o câncer é, e sempre será, a prevenção e a detecção precoce. Para isso, é necessário educar a população para que se realizem exames rotineiros de forma a se detectar o câncer em um estágio inicial e, dessa forma, poder ser retirado por meio de cirurgia.

Métodos computacionais podem auxiliar no diagnóstico dessa detecção de uma maneira mais eficaz do que os métodos tradicionais. Essa é a premissa do diagnóstico auxiliado por computador (CAD), em que algoritmos são utilizados para dar o diagnóstico. Dessa forma, o profissional de saúde pode comparar o seu próprio diagnóstico com aquele fornecido pelo sistema computacional e tirar melhores conclusões.

Na Seção 1.5 são apresentados alguns trabalhos que usam técnicas computacionais para a detecção e classificação de câncer e tecido mamário.

1.5 Trabalhos relacionados

Diversos trabalhos abordam o tema de detecção e classificação de câncer, além da classificação do tecido mamário. Dentre os diversos trabalhos lidos para a elaboração deste, pode-se citar:

- TAYLOR et al. (1994) também classificaram as mamas através de seus tecidos mamários adiposo e denso, selecionando manualmente uma janela de 128x128 pixels contendo o tecido a ser analisado e calculando medidas de textura. O teste foi aplicado em 104 imagens com um resultado de 67,4% de acurácia.
- MARTINS et al. (2006) analisaram a aplicação da matriz de coocorrência para a caracterização de tecidos mamário em normal, benigno ou maligno nas mamografias do banco de dados MIAS (*Mammographic Image Analysis Society*). O número total de mamografias utilizadas foi de 218, sendo 119 normais e 99 com anomalias (68 benignas e 31 malignas). As medidas de textura de (HARALICK, 1973) utilizadas foram: Contraste, Homogeneidade, Momento de Diferença Inversa, Entropia e Segundo Momento Angular. Para computar as matrizes de coocorrência, usaram-se as quatro direções de 0°, 45°, 90° e 135° e a distância de $d = 1$ entre cada pixel. Para as imagens com câncer, utilizou-se a ROI previamente fornecido pelo banco de dados, ao passo que para as imagens sem câncer, a ROI foi selecionada manualmente, evitando-se englobar o músculo peitoral. As ROIs foram redimensionadas para 50x50 pixels. Utilizou-se em cada interação seis diferentes quantidades de níveis de cinza (8, 16, 32, 128 e 256) no cálculo da matriz de coocorrência. Por fim, uma Rede Neural com Aproximação Bayesiana (BNN) se encarregou de classificar as medidas através de 180 amostras de treino e 38 imagens, fornecendo um resultado de acerto médio de 86,84% para a classificação entre normal, benigno e maligno; sensibilidade de 83,3% e especificidade de 95% para a classificação entre benigno e normal; uma taxa de acerto de 100% para a classificação entre normal e maligno; e sensibilidade de 90% e especificidade de 71,4% para a classificação entre benigno e maligno.

- PRATIWI et al. (2015) utilizaram a base de dados MIAS para classificar um conjunto de 262 imagens normais e com anomalias (benigno e maligno). Para isso, fez o uso da matriz de coocorrência calculando quatro medidas de textura de (HARALICK, 1973): Segundo Momento Angular, Correlação, Entropia da Soma e Variância da Soma. O trabalho se dividiu em duas etapas: classificação entre normal ou câncer; e classificação do câncer em benigno ou maligno. A ROI foi extraída de maneira semelhante ao realizado em (MARTINS et al., 2006), com a diferença de que houve um realce e um redimensionamento da mesma para 128x128 pixels. Por fim, para ambas as etapas de classificação utilizaram-se duas redes neurais para efeito de comparação: Rede Neural de Retro Propagação (BPNN) e Rede Neural de Base Radial (RBFNN). Utilizando a primeira rede neural, obteve-se uma sensibilidade de 88,9% e uma especificidade de 94,4% para a classificação entre normal e doença, ao passo que se obteve uma sensibilidade de 81,6% e especificidade de 76,9% para a classificação entre benigno e maligno. Para a segunda rede neural, obteve-se uma sensibilidade de 94,4% e uma especificidade de 93,6% para a classificação entre normal e doença, ao passo que se obteve uma sensibilidade de 100% e especificidade de 89,5% para a classificação entre benigno e maligno.
- ALCÂNTARA, JÚNIOR e RAMOS (2016) classificaram os tecidos mamários em normais e anormais (benignos e malignos) das imagens presentes no banco de dados *Digital Database for Screening Mammography* (DDSM, s.d.), baseados na matriz de coocorrência e utilizando um classificador *Support Vector Machine* (SVM). Propuseram um novo método de descritor de textura da ROI na imagem utilizando a decomposição multinível, a matriz de coocorrência, a decomposição de valor singular (*Singular Value Decomposition* – SVD) e a extração de entropia Tsallis. Todas as medidas de textura propostas por (HARALICK, 1973) foram extraídas. A melhor acurácia deu-se para a matriz de coocorrência na direção 0°, com um valor de 84,4%, ao passo que a sensibilidade e a especificidade tiveram valores iguais a 85,4% e 83,4%, respectivamente.
- OLIVEIRA et al. (2016) propuseram um sistema automático para a detecção de distorção arquitetural mamária através de extração de características de medidas de textura, sendo que dos seis métodos utilizados, quatro foram baseados nas matrizes de coocorrência. As medidas de textura extraídas das imagens do banco de dados DDSM foram: Contraste, Correlação, Energia e Homogeneidade. Cada imagem foi recortada manualmente em duas ROIs, uma contendo a distorção arquitetural e outra contendo tecido normal, sadio. As regiões selecionadas passaram por uma etapa de pré-processamento, onde foi efetuado o

realce do contraste e a sub-amostragem deixando as ROIs com dimensão 128x128 pixels. Extraídas as medidas de textura, foi utilizada uma rede neural artificial com arquitetura do tipo Perceptron Multi Camadas (PMC), contendo apenas uma camada escondida. Dentre os descritores de textura utilizados, os baseados na matriz de coocorrência apresentaram melhores resultados do que os descritores *Local Binary Pattern* (LBP) e *Local Mapped Pattern* (LMP) quando comparados em separado, com uma taxa de Verdadeiros Positivos (VP) de aproximadamente 90%.

- BOVIS e SINGH (s.d.) classificaram as mamografias presentes na base de dados MIAS entre os tecidos mamários de duas maneiras: 1) adiposo, glandular e denso; e adiposo e glandular/denso. Para isso, calcularam medidas de textura baseadas na matriz de coocorrência, dimensão fractal e valores estatísticos de níveis de cinza. As medidas de textura de coocorrência utilizadas foram: Segundo Momento Angular, Contraste, Correlação, Momento de Diferença Inversa, Média da Soma, Variância da Soma, Entropia da Soma, Entropia, Média da Diferença, Variância da Diferença, Entropia da Diferença, Medida de Informação da Correlação 1, Medida de Informação da Correlação 2 e Variância. Para a primeira classificação, utilizou-se uma Rede Neural em 3 Camadas, obtendo-se uma taxa de acerto de 70,4%, ao passo que na segunda classificação utilizou-se o classificador kNN (*k-Nearest Neighbour*) com uma análise de Característica de Operação do Receptor (ROC), obtendo uma área abaixo da mesma de 0,832.

Como pode ser observado, o uso de medidas de texturas extraídas a partir da matriz de coocorrência é aplicável para a detecção e classificação de câncer de mama e identificação de tipo de tecido, e resultados consideráveis são obtidos com o seu uso. Portanto, um trabalho que pudesse analisar as medidas extraídas da matriz de coocorrência poderia ser útil para melhorar o desempenho dos métodos de detecção e classificação de cânceres e tecidos mamários.

1.6 Objetivos geral e específico

Diversos trabalhos se basearam em selecionar algumas medidas de textura para serem utilizadas dos descritores de textura, não ficando muito claro o motivo da escolha inicial delas. Nesse trabalho, optou-se por medir a contribuição individual de várias medidas de textura extraídas da Matriz de Coocorrência de Nível de Cinza e fazer um comparativo entre as melhores. Também foi usado um algoritmo de seleção de atributos para verificar a combinação delas. A correta seleção das medidas pode trazer um menor custo computacional e tornar o sistema mais simples de ser analisado.

O objetivo geral deste trabalho é realizar uma análise das medidas extraídas das matrizes de coocorrência de nível de cinza na tarefa de detectar e classificar cânceres e tecidos mamários, utilizando para isso técnicas de processamento digital de imagens e técnicas de classificação.

Algumas tarefas são, portanto, requisitadas: analisar a bibliografia sobre processamento digital de imagens e imagens médicas de mamografia; estudar características de textura de uma imagem; estudar técnicas de classificação; realizar simulações e testes com o banco de dados MIAS – *Mammographic Image Analysis Society* (MIAS, s.d.), fornecido online e gratuitamente, utilizando o software MATLAB.

Os objetivos específicos deste trabalho são:

- Desenvolver uma técnica capaz de detectar a presença ou não de câncer em uma mamografia digital e classificar o câncer em benigno ou maligno com o auxílio das medidas de textura extraídas das matrizes de coocorrência;
- Desenvolver uma técnica capaz de classificar o tecido mamário em adiposo ou glandular (glandular adiposo e glandular denso), de acordo com a composição mais abundante presente em uma mamografia digital com o auxílio das medidas de textura extraídas das matrizes de coocorrência;
- Desenvolvimento de uma metodologia para a segmentação da mama de imagem de mamografia de forma a remover o músculo peitoral.

1.7 Estrutura do trabalho

O restante deste trabalho está organizado como segue.

O Capítulo 2 descreve os aspectos teóricos relacionados ao processamento digital de imagens, análise de texturas e técnicas de classificação, assuntos estudados e utilizados na elaboração desse trabalho.

O Capítulo 3 apresenta, de forma detalhada, a metodologia e o desenvolvimento do projeto, etapa por etapa, desde a aquisição da imagem até a definição da região de interesse (ROI), passando por toda a segmentação. São mostradas as dificuldades encontradas e como foram solucionadas, além de exemplos das imagens mostrando os resultados de cada técnica usada no procedimento.

O Capítulo 4 fornece os resultados obtidos em cada parte do trabalho, com valores comparativos e análises.

O Capítulo 5 descreve a conclusão final do trabalho, baseada em todo o contexto, aspecto teórico, desenvolvimento e, principalmente, dos resultados encontrados. Por fim, são descritas sugestões para trabalhos futuros.

2 TÉCNICAS ABORDADAS NO TRABALHO

O processamento digital de imagens está presente em diversas áreas profissionais e acadêmicas, como por exemplo: astronomia, medicina, geografia, arqueologia, física e indústria. Dentre suas aplicações podemos citar: realce de contraste em imagens de raios-X, caracterização de padrões de poluição ou relevo em imagens aéreas e de satélite e visão computacional industrial para montagem e inspeção de produtos.

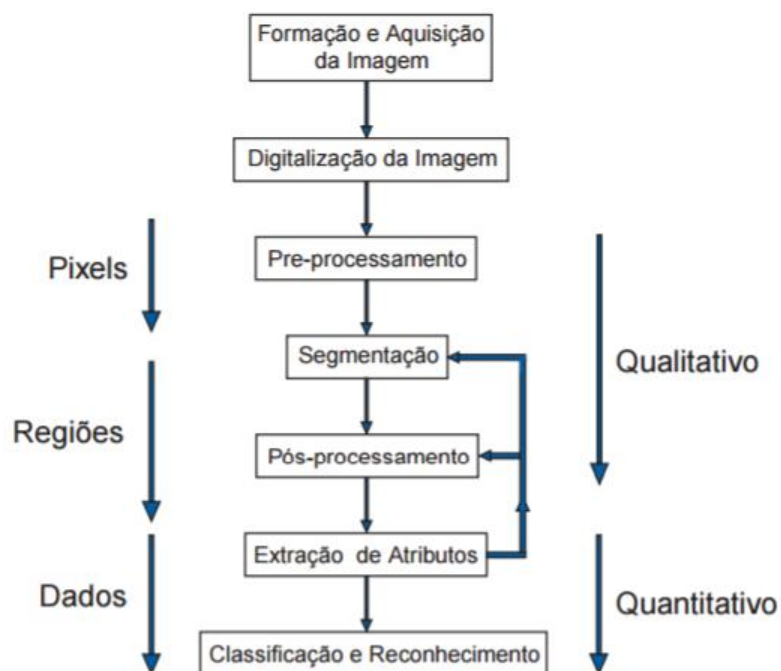
Na medicina, podemos destacar o processamento digital de imagens médicas, cujas pesquisas vêm crescendo exponencialmente devido à preocupação da população com a saúde, prevenção de doenças e melhor qualidade de vida. Exemplos de aplicações de processamento digital de imagens médicas são: radiografias, tomografias, ultrassonografias e ressonâncias.

Uma imagem pode ser definida como uma função bidimensional, $f(x,y)$, em que x e y são coordenadas espaciais (plano), e a amplitude de f em qualquer par de coordenadas (x,y) é chamada de intensidade ou nível de cinza da imagem nesse ponto. Quando x , y e os valores de intensidade de f são quantidades finitas e discretas, chamados de imagem digital. O campo do processamento digital de imagens se refere ao processamento de imagens digitais por um computador digital. Observe que uma imagem digital é composta de um número finito de elementos, cada um com localização e valor específicos. Esses elementos são chamados de *elementos pictóricos*, *elementos de imagem*, *pels* e *pixels*. *Pixel* é o termo mais utilizado para representar os elementos de uma imagem digital (GONZALEZ, 2011).

Uma imagem em nível de cinza, foco deste trabalho, pode ser definida como uma matriz f retangular de tamanho $N \times M$, onde cada elemento dessa matriz é chamado de pixel, e seu valor numérico costuma ser representado por 8 bits, permitindo assim até 256 valores diferentes (0 a 255).

Um sistema de processamento de imagens típico, que inclui classificação de suas partes segmentadas, é constituído de diversas etapas, tais como: formação e aquisição da imagem, digitalização, pré-processamento, segmentação, pós-processamento, extração de atributos, classificação e reconhecimento (ALBUQUERQUE et al., 2011). A figura 6 representa esquematicamente as etapas que serão utilizadas neste trabalho. A seguir, cada etapa será detalhada indicando quais foram as técnicas usadas em cada categoria neste trabalho.

Figura 6 – Etapas do processamento de imagens e classificação.



Fonte: ALBUQUERQUE et al., 2011.

2.1 Aquisição e digitalização das imagens

O processamento digital de imagens começa pela aquisição da imagem, que pode ser feita pela captura da cena com uma câmera digital, a digitalização de uma imagem através de scanners ou então pela recepção de uma imagem cujo formato já seja digital.

2.2 Pré-processamento

As técnicas de pré-processamento têm a função de melhorar a qualidade da imagem, corrigindo algum defeito proveniente de sua aquisição e/ou realçando detalhes importantes para análise. Essas técnicas estão divididas em duas categorias: técnicas que operam no domínio espacial e técnicas que operam no domínio da frequência (GONZALEZ, 2011).

Para que se realce determinadas características de uma imagem, é comum que se combine vários métodos que estejam baseados nestas duas categorias. Alguns exemplos de procedimentos utilizados no pré-processamento são: melhoras no brilho e no contraste, correção de iluminação irregular, redução de ruídos (suavização) e realce de bordas.

Vale lembrar que, para que a etapa de segmentação tenha resultados satisfatórios, é necessário que a imagem esteja com o mínimo de imperfeições, daí a importância da etapa de pré-processamento.

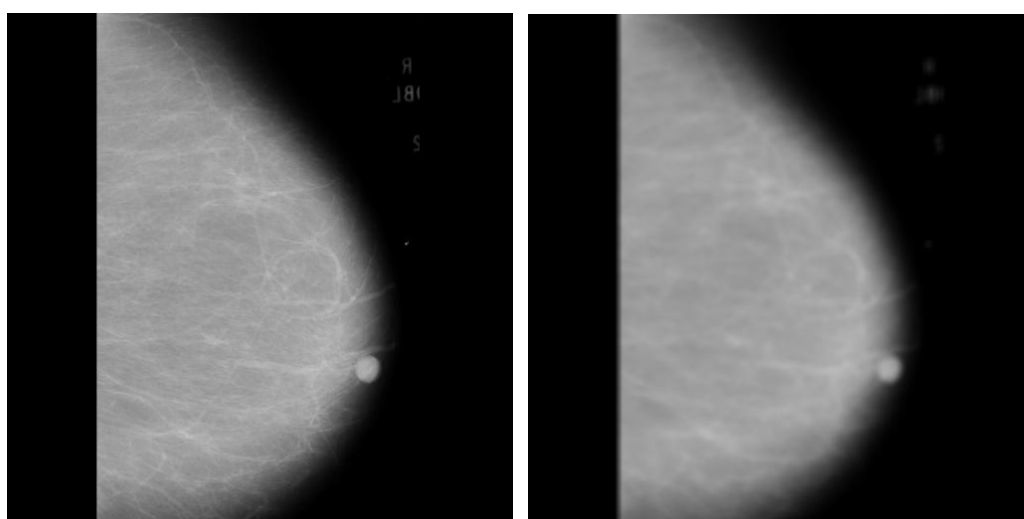
2.2.1 Suavização

Durante o processo de aquisição e digitalização de uma imagem, geralmente são introduzidos ruídos de alta frequência espacial. Filtros passa-baixa, também conhecidos como filtros de suavização, são empregados na remoção desses ruídos. Dentre os tipos de filtros passa-baixa existentes, destaca-se o filtro de média, uma técnica de suavização não conservativa, que atenua as frequências espaciais elevadas da imagem e elimina detalhes como linhas finas e curvas agudas, causando o efeito de *blurring* (borramento).

No filtro de média, uma janela (elemento estruturante) percorre toda a imagem, pixel a pixel, e cada pixel no centro dessa janela receberá a média de todos os elementos da janela. Quanto maior for o elemento estruturante, maior será o efeito de borramento da imagem resultante, visto que levaremos em consideração um número maior de pixels.

A figura 7 exemplifica o processo de suavização de uma imagem médica contendo uma mamografia através de um filtro de média. Perceba que a imagem sofreu um borramento e os ruídos presentes foram eliminados/reduzidos.

Figura 7 – Suavização usando filtro de média.



a) Sem suavização.

b) Com suavização.

Fonte: Produção do próprio autor.

Outros filtros de suavização bastante utilizados são os filtros de mediana e gaussiana. No entanto, neste presente trabalho, utilizou-se somente o filtro de média.

2.2.2 Realce de contraste

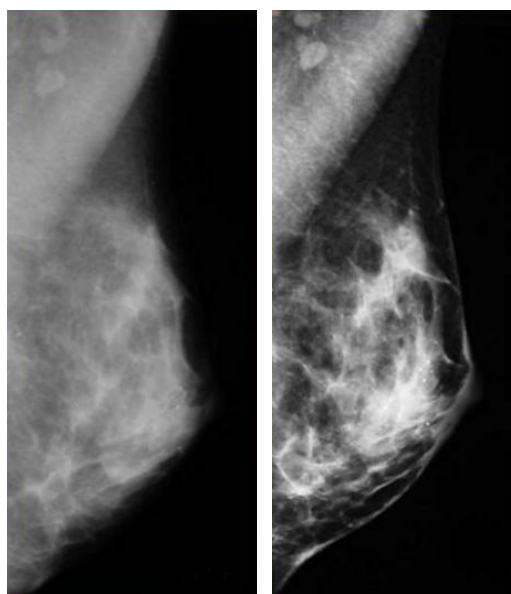
O realce de uma imagem digital consiste em melhorar sua qualidade visual geral, aumentando o contraste entre os elementos apresentados na imagem e facilitando processamentos adicionais, como detecção de bordas. Tem como objetivo realçar características específicas relacionadas aos alvos imageados (FELGUEIRAS, s.d.).

Existem várias técnicas de realce utilizadas no processamento digital de imagens, a saber: realce linear, mínimos e máximos, raiz quadrada, quadrado, logarítmico, negativo, por fatia, por edição e, por últimos, através da equalização do histograma e exponencial.

A equalização do histograma, utilizada no presente trabalho, procura redistribuir os valores dos níveis de cinza na imagem, de forma a se obter um histograma uniforme, ou seja, tenta obter uma distribuição uniforme dos valores de nível de cinza. Pode ser considerado um pré-processamento para outras técnicas, como a segmentação e a extração de atributos.

A figura 8 exemplifica a técnica de realce de imagem. Perceba que a imagem realçada possui características realçadas devido ao histograma mais uniforme e distribuído ao longo dos níveis de cinza.

Figura 8 – Realce através da equalização do histograma.



a) Sem realce.

b) Com realce.

Fonte: Adaptado de MARANHÃO, 2016.

2.3 Segmentação

É na etapa de segmentação que são definidas as regiões de interesse para processamento e análise posteriores. A determinação da região de interesse (ROI – *Region of Interest*) possibilita economizar recursos computacionais em casos onde apenas uma parte da imagem possui dados pertinentes à análise que estiver sendo realizada (NETO, 2014).

Segmentar a imagem significa, de um modo simplificado, separar a imagem como um todo nas partes que a constituem e que se diferenciam entre si. Deve-se ressaltar que não há um modelo formal e correto para a segmentação de uma imagem. A segmentação é um processo empírico e adaptativo, procurando sempre se adequar às características particulares de cada tipo de imagem e aos objetivos que se pretende alcançar. (ALBUQUERQUE et al., 2011).

2.3.1 Binarização

Também chamada de limiarização ou *Thresholding*, a binarização é a técnica que divide a imagem em duas ou mais partes considerando valores de limiares (NUNES, 2006). Na maioria das vezes, utiliza-se a técnica da binarização para tornar a imagem com somente dois níveis de cor: preto e branco.

O principal desafio dessa técnica é a definição do valor de limiar, que definirá quais pixels terão seu valor atribuído igual a 1 e quais terão valor atribuído igual a 0. Algumas técnicas são utilizadas para se definir esse limiar, podendo-se citar: pontos de mínimo e máximo do histograma e a binarização de Otsu (MONTEIRO, s.d.).

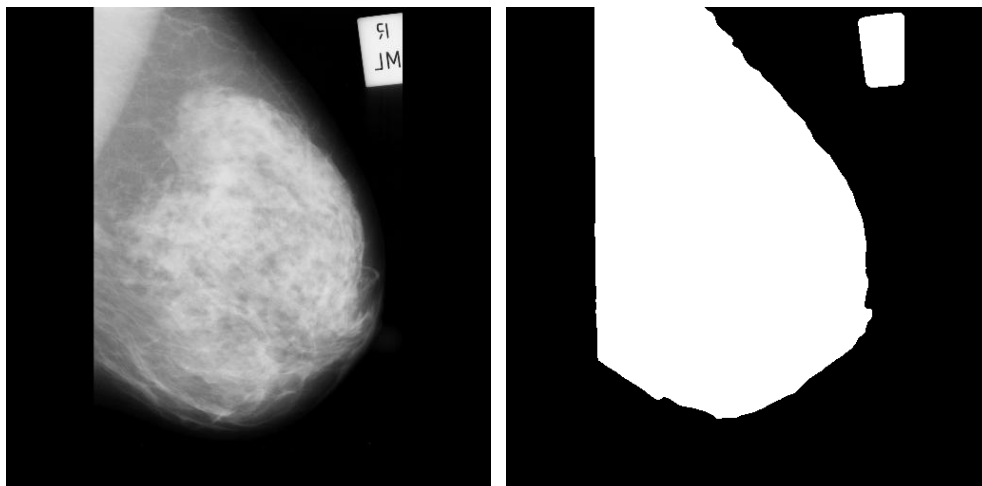
A binarização de Otsu, utilizada no presente trabalho, assume que a imagem está dividida em duas classes, objeto e fundo, e através de múltiplas iterações tenta encontrar o limiar (nível de cinza) que divide essas duas classes, com a máxima variância interclasse. Ele é um método automático que não necessita de parâmetros (FILHO et al., 2011).

Considere a figura 9, uma imagem médica contendo uma mamografia e que possui objetos claros sobre um fundo escuro. Todo ponto (x,y) da imagem f que possuir intensidade $f(x,y) \geq T$, onde T é o limiar de segmentação, será parte do objeto de interesse e o restante da imagem será pertencente ao fundo.

A imagem binarizada, representada por $b(x,y)$ é definida pela expressão (1):

$$b(x, y) = \begin{cases} 0, & \text{se } f(x, y) < T \\ 1, & \text{se } f(x, y) \geq T \end{cases} \quad (1)$$

Figura 9 – Binarização de uma imagem.



a) Sem binarização.

b) Com binarização.

Fonte: Produção do próprio autor.

2.4 Pós-processamento

O pós-processamento é a etapa realizada após a segmentação e que se destina a corrigir os principais defeitos ou imperfeições da etapa anterior.

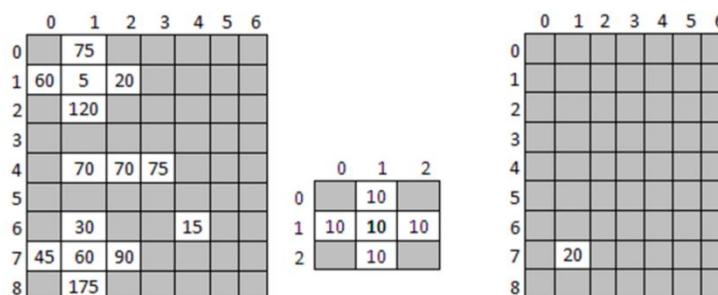
Normalmente, estes defeitos de segmentação são corrigidos por meio de técnicas de morfologia matemática, em que filtros morfológicos são aplicados em sequência para realizar uma análise quantitativa dos pixels da imagem. Dois operadores muito utilizados são os de erosão e dilatação.

2.4.1 Erosão

O filtro morfológico de erosão provoca efeitos de desgaste das partes claras da imagem (altos níveis de cinza), gerando imagens mais escuras. De forma geral, a técnica funciona fazendo uma comparação da subtração das intensidades dos pixels da imagem original com as intensidades dos pixels correspondentes no elemento estruturante, definido previamente. O pixel correspondente da imagem de saída terá a intensidade da menor subtração desta comparação. O menor valor para o pixel é 0. (MENEZES, 2010).

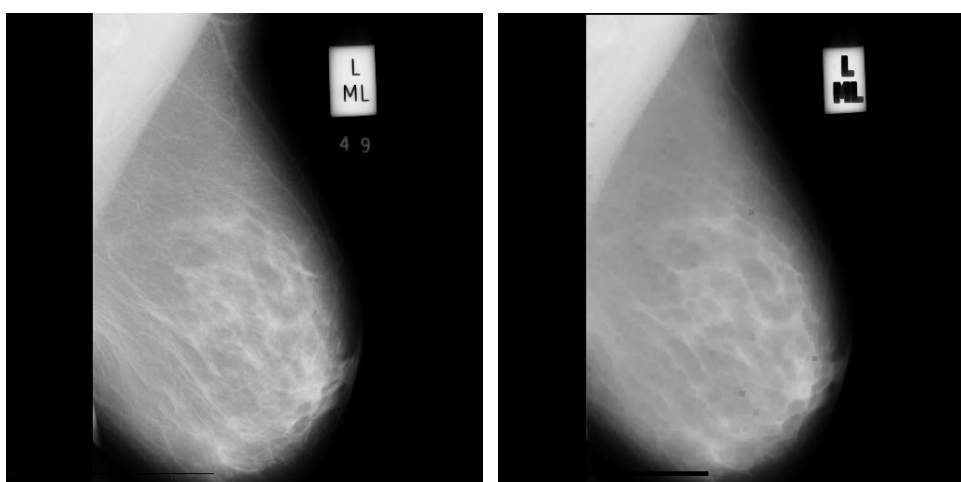
As figuras 10 e 11 mostram o processo de erosão. Na primeira imagem é apresentado o processo resultante na matriz de valores, cujo resultado é dado pelo menor valor da subtração do pixel da imagem original e dos pixels correspondentes do elemento estruturante. Na segunda imagem, pode-se perceber que houve um escurecimento da mesma, devido à erosão das partes claras.

Figura 10 – Erosão em tons de cinza em uma matriz de pixels.



Fonte: MENEZES, 2010.

Figura 11 – Erosão em uma imagem.



a) Sem erosão.

b) Com erosão.

Fonte: Produção do próprio autor.

2.4.2 Dilatação

O filtro morfológico de dilatação, ao contrário da erosão, provoca efeitos de expansão das partes claras (altos níveis de cinza), gerando imagens mais claras. De forma geral, a técnica funciona fazendo uma comparação da soma das intensidades dos pixels da imagem original com as intensidades dos pixels correspondentes no elemento estruturante, definido previamente. O pixel correspondente da imagem de saída terá a intensidade da maior soma desta comparação. O maior valor para o pixel é 255 (MENEZES, 2010).

As figuras 12 e 13 mostram o processo de dilatação acontecendo. Na primeira imagem é mostrado o processo resultante na matriz de valores, cujo resultado é dado pelo maior valor da soma do pixel da imagem original e dos pixels correspondentes do elemento estruturante. Na segunda imagem, pode-se perceber que houve um clareamento da mesma, devido à dilatação das partes claras.

Figura 12 – Dilatação em tons de cinza em uma matriz de pixels.

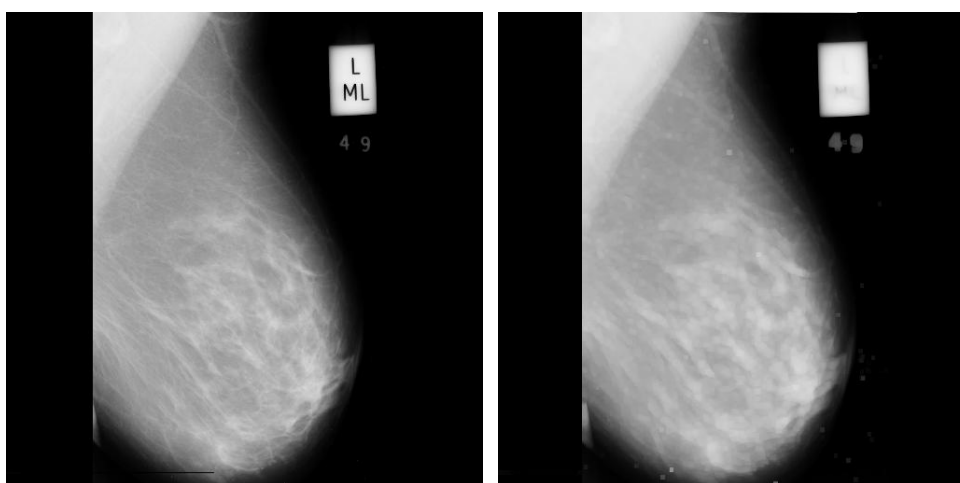
	0	1	2	3	4	5	6
0							
1		5	20				
2							
3							
4		70	70	75			
5							
6					15		
7		60					
8							

	0	1	2
0		10	
1	10	10	10
2		10	

	0	1	2	3	4	5	6
0		15	30				
1	15	30	30	30			
2		15	30				
3		80	80	85			
4	80	80	80	85	85		
5		80	80	85	25		
6		70		25	25	25	
7	70	60	70		25		
8		70					

Fonte: MENEZES, 2010.

Figura 13 – Dilatação em uma imagem.



a) Sem dilatação.

b) Com dilatação.

Fonte: Produção do próprio autor.

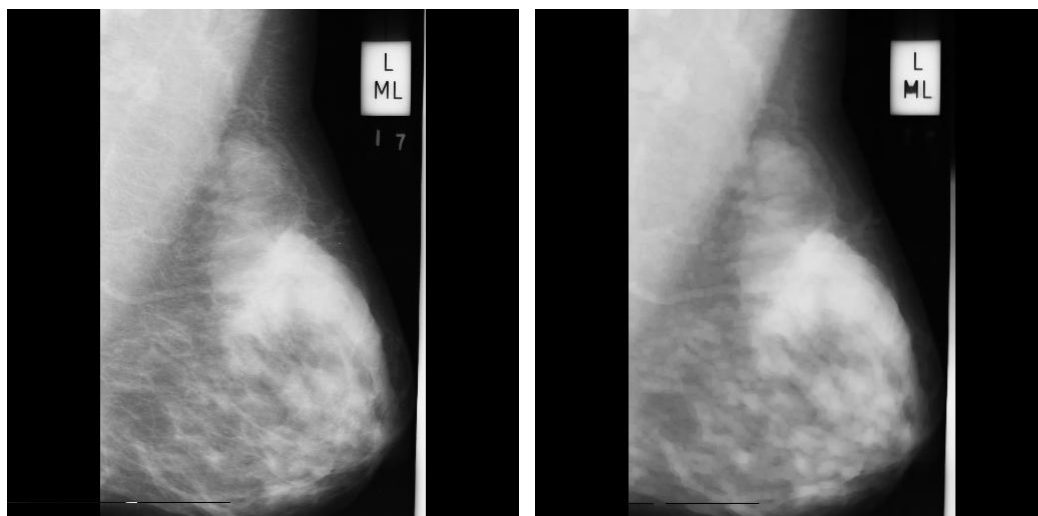
2.4.3 Abertura

A técnica de abertura é uma operação que suaviza o contorno de uma imagem e elimina protuberâncias finas. O processo consiste em uma operação de erosão seguida por uma dilatação, resultando na abertura de pequenos vazios ou espaços entre objetos próximos na imagem. Essa operação elimina as protuberâncias da imagem de forma mais suave que a erosão.

Interessante citar que nessa técnica, os pontos pretos aleatórios e isolados podem ser removidos e a forma dos objetos é recuperada pela dilatação sem restaurar o ruído anteriormente presente (SOARES, s.d.).

A figura 14 exemplifica a técnica de abertura em uma imagem. Percebe-se que na abertura os contornos foram suavizados e os ruídos eliminados, ao passo que pequenos espaços foram abertos ou deixaram de existir.

Figura 14 – Abertura em uma imagem.



a) Sem abertura.

b) Com abertura.

Fonte: Produção do próprio autor.

2.4.4 Extração de fronteiras

A fronteira de uma imagem pode ser obtida por meio da erosão da imagem binarizada por um elemento estruturante, e posterior subtração dessa erosão da própria imagem (GONZALEZ, 2011). Esse filtro busca extrair da imagem, componentes que sejam úteis na representação e na descrição de formas.

A figura 15 mostra a técnica de extração de fronteiras aplicada a uma imagem médica contendo uma mamografia.

Figura 15 – Extração de fronteiras em uma imagem.



a) Imagem binarizada.

b) Fronteiras da imagem.

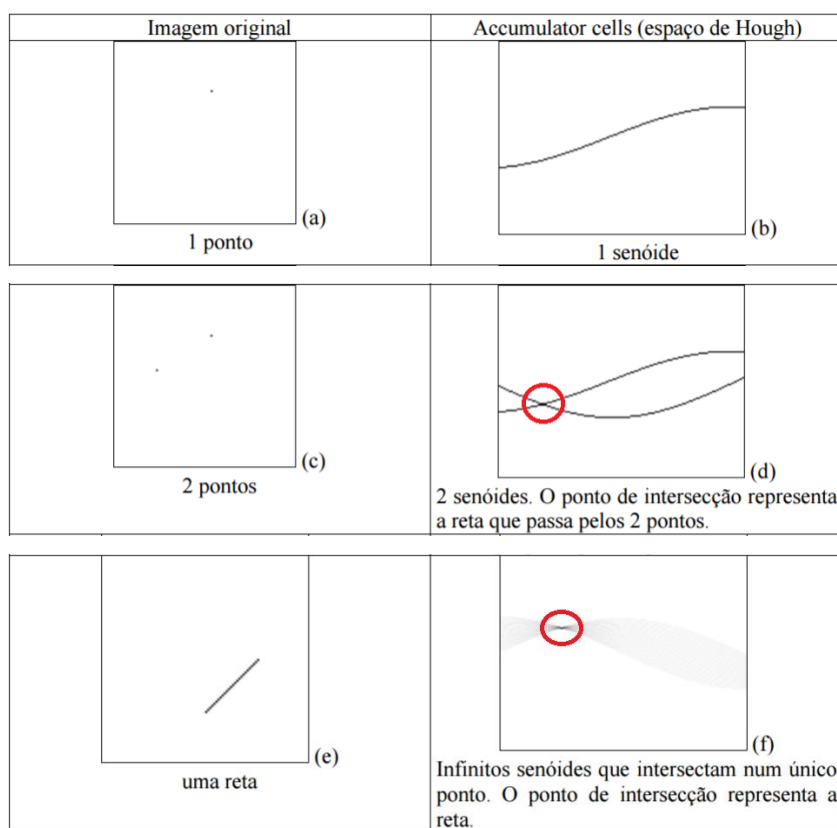
Fonte: Produção do próprio autor.

2.4.5 Transformada de Hough

A Transformada de Hough é uma técnica muito atrativa computacionalmente capaz de detectar formas que são facilmente parametrizadas (retas, círculos e elipses, por exemplo) em imagens digitalizadas. Particularmente, a detecção de retas, como feito no presente trabalho, é de suma importância para muitas aplicações na área de processamento digital de imagens. Para realizar esse procedimento, a ideia é aplicar na imagem uma transformação tal que os pontos pertencentes a uma mesma curva sejam mapeados num único ponto de um novo espaço de parametrização da curva procurada (PIVETTA et al., s.d.).

Conforme mostra a figura 16, a imagem original (a) é um único ponto. Infinitas retas passam por esse único ponto no plano. Todas as retas que passam por esse ponto formam uma senóide no plano de Hough (b). Dois pontos no plano da imagem (c) definem uma reta. Dois pontos correspondem a duas senóides no plano de Hough (d). A intersecção das duas senóides representa a reta que passa pelos dois pontos no plano da imagem. Uma reta no plano da imagem (e) corresponde a infinitas senóides no plano de Hough que se intersectam num único ponto (f). Este ponto do plano de Hough representa a reta (KIM, 2016).

Figura 16 – Mapeamento de pontos no novo espaço de parametrização da curva.



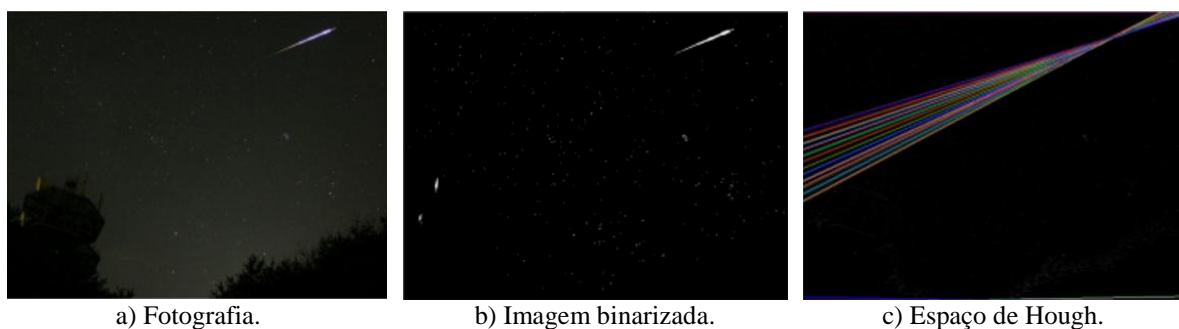
Fonte: Adaptado de KIM, 2016.

Em princípio, as retas do espaço de parâmetros que corresponde a todos os pontos (x,y) no plano xy podem ser traçadas e as retas principais nesse plano poderiam ser determinadas identificando os pontos no espaço de parâmetros nos quais uma grande quantidade de retas do espaço de parâmetros se intercepta (GONZALEZ, 2008).

Um passo necessário no processo de aplicação da Transformada de Hough é a limiarização (binarização) e detecção de bordas (extração de fronteiras) da imagem cujas retas desejam-se obter.

A figura 17 mostra o resultado da Transformada de Hough na detecção de um meteoro atravessando a atmosfera da Terra. A fotografia de um meteoro é registrada em (a). Pode-se perceber que o rastro que um meteoro deixa em nossa atmosfera é semelhante a uma reta. Essa reta diferencia-se dos demais pontos brilhantes do céu noturno que representam diferentes estrelas e corpos celestes. A imagem é então binarizada em (b) e a reta do rastro do meteoro é obtida no espaço de Hough em (c) pelo encontro das infinitas senóides que representam a reta na imagem original.

Figura 17 – Transformada de Hough em uma imagem.



Fonte: Adaptado de ANNA, 2011.

2.5 Extração de atributos

A etapa final de um sistema de processamento de imagens é aquela em que as informações úteis da imagem já processada são extraídas para posterior avaliação e classificação.

De uma maneira geral, a extração de atributos se divide em dois tipos de medidas: de campo e de região. As medidas de campo se referem ao campo como um todo, como na medição do número de objetos, área total dos objetos e medição de área, gerando como resultado apenas um valor por medida. As medidas de região se referem aos objetos individualmente, ou seja, é extraído um parâmetro de cada objeto na imagem, como por exemplo, tamanho, forma e posição das partículas (GOMES, 2001).

2.5.1 Análise de Texturas

Textura pode ser definida com um padrão de cores ou intensidade de nível de cinza usado pelo sistema visual humano para reconhecer a forma e a composição de um objeto, por exemplo. Dessa forma, podem-se reconhecer objetos numa imagem e se ter o perfeito entendimento dela (MASCARO, 2007).

A figura 18 exemplifica oito tipos diferentes de texturas encontrados no cotidiano.

Figura 18 – Diferentes tipos de texturas.



Fonte: CONCI, 2007.

A textura refere-se a um padrão visual que possui algumas propriedades de homogeneidade que não resultam simplesmente de uma cor ou intensidade. Ao contrário de outras características (brilho e cor), a textura não pode ser definida em um pixel, mas sim através de uma região ou conjunto de pixels (NUNES, 2013).

Fazer a análise de uma imagem a partir de suas texturas permite identificar regiões através de um padrão de cores ou intensidades de nível de cinza. Com a análise de texturas é possível se ter uma medida geral de como ocorre o relacionamento entre os pixels de uma dada região (MASCARO, 2007).

Entre as dificuldades encontradas na caracterização de uma textura, destacam-se:

- Irregularidades presentes nas texturas;
- Quantificação da textura para diversas resoluções de uma mesma imagem.

Diversos descritores de textura são usados quando da extração das características de textura de uma imagem. Neste trabalho foi analisada a Matriz de Coocorrência de Nível de Cinza (GLCM – *Gray Level Co-occurrence Matrix*).

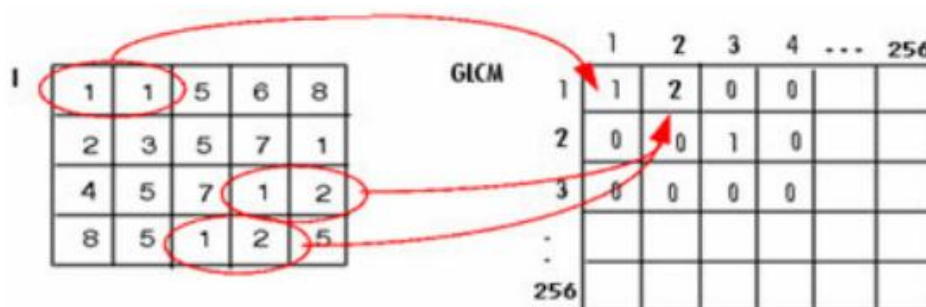
2.5.2 Matriz de Coocorrência de Nível de Cinza (GLCM)

Esse descritor de textura fornece uma medida de probabilidade da ocorrência de níveis de cinza separados por uma dada distância d em uma dada direção θ (0° , 45° , 90° e 135°).

A figura 19 fornece uma ideia de como a GLCM é preenchida. Para uma dada imagem, cada célula (i,j) da matriz de coocorrência funciona como um contador e armazena a frequência com que dois pixels ocorrem na imagem separados por uma distância d , um com o tom de cinza i , outro com o tom de cinza j , na direção de 0° . À esquerda dessa figura há uma imagem de exemplo onde aparecem em duas regiões da imagem os tons 1 e 2 separadas por um pixel de distância. Assim, se $d = 1$, a matriz GLCM tem sua posição $(1, 2)$ incrementada para 2, indicando a ocorrência de 2 pixels com tons de cinza 1 e 2 separados pela distância d (MASCARO, 2007).

Cada posição da matriz de coocorrência conterà a variação dos níveis de cinza presentes na imagem, considerando os índices da matriz. Então, por exemplo, a posição $(1, 2)$ da matriz de coocorrência conterà a quantidade de vezes que houve a ocorrência do nível de cinza 1 com o vizinho no nível de cinza 2 na horizontal à direita (para o ângulo 0°) (VASCONCELOS, 2015).

Figura 19 – Matriz de coocorrência de tons de cinza.



Fonte: MASCARO, 2007.

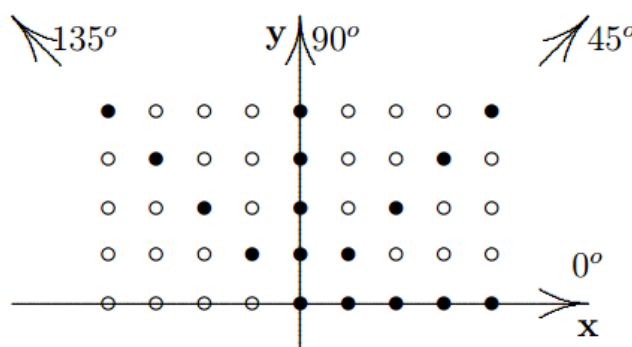
Para chegar-se a forma final da matriz de coocorrência precisa-se dividir cada elemento pelo número total, ou seja, a matriz de coocorrência é dada pela matriz quadrada de probabilidade $P(i, j)$, onde cada elemento (i, j) representa a probabilidade de um certo valor aparecer na matriz, sendo então calculado pela equação (2):

$$P(i, j) = \frac{M(i, j)}{\sum_{i, j=0}^{N-1} M(i, j)} \quad (2)$$

onde i é o número da linha e j é o número da coluna; $M(i, j)$ é o conteúdo da célula (i, j) ; $P(i, j)$ é a probabilidade da célula (i, j) ; N é o número de linhas ou colunas, uma vez que M é uma matriz quadrada.

Para o presente trabalho, utilizaram-se, para cada imagem, quatro diferentes GLCMs, cuja diferença se dá pelo ângulo de direção entre cada pixel, separados de uma distância $d = 1$ para todas elas. Os ângulos utilizados foram 0° , 45° , 90° e 135° , conforme mostra a figura 20.

Figura 20 – Diferentes direções para a matriz de coocorrência.



Fonte: ALBREGTSEN, 2008.

2.5.3 Medidas de Texturas

Um conjunto de 22 atributos foi definido para descrever uma textura a partir da matriz de coocorrência, segundo (HARALICK, 1973), (SOH, 1999), (CLAUSI, 2002) e (BOVIS, K. SINGH, S. s.d.). Esses atributos são largamente utilizados até hoje em diversos trabalhos:

- Segundo Momento Angular
- Contraste
- Correlação
- Entropia
- Entropia da Soma
- Entropia da Diferença
- Variância
- Variância da Soma
- Variância da Diferença
- Momento de Diferença Inversa Normalizada
- Diferença Inversa Normalizada
- Média da Soma
- Média da Diferença
- Medida de Informação da Correlação 1
- Medida de Informação da Correlação 2
- Coeficiente de Correlação Máximo
- Autocorrelação
- Dissimilaridade
- Sombra de Agrupamento
- Proeminência de Agrupamento
- Máxima Probabilidade
- Inércia

No caso específico de imagens médicas de mamografia, objeto de estudo do presente trabalho, medidas de textura podem fornecer informações importantes a respeito dos tecidos presentes na mama (adiposo e glandular) e de eventuais anormalidades (presença, formato e classificação de câncer).

Neste trabalho foram utilizados todos os atributos citados acima.

- **Segundo Momento Angular**

Também conhecida como energia, expressa a medida de uniformidade entre pares de pixels em uma imagem e detecta distúrbios em uma textura (GADKARI, 2000). Uma imagem uniforme conterá alguns níveis de cinza, de modo que a GLCM terá alguns, porém relativamente altos valores de $P(i,j)$. Logo, a soma de quadrados será alta. Se a imagem é uniforme, o segundo momento angular terá valor alto (ALBREGTSEN, 2008). Sua equação (3) é dada abaixo:

$$ASM = \sum_{i,j=0}^{n-1} \{P(i,j)\}^2 \quad (3)$$

onde i e j são as coordenadas da GLCM, n é o número de níveis de cinza da imagem e $P(i,j)$ é o valor da posição (i,j) da GLCM.

- **Contraste**

Expressa a medida das variações locais apresentadas em uma imagem. Esta medida de contraste favorece as contribuições da GLCM para longe da diagonal ($i \neq j$) (ALBREGTSEN, 2008). Quando i e j são iguais, a célula está na diagonal e $i - j = 0$. Estes valores representam pixels inteiramente semelhantes ao seu vizinho, então eles recebem um peso de 0. Se i e j diferem em 1, há um pequeno contraste e o peso é 1. Se i e j diferem em 2, o contraste está aumentando e o peso é 4. Os pesos continuam a aumentar exponencialmente à medida que $i - j$ aumenta (RAUT et al., 2016). Se houver uma grande quantidade de variação em uma imagem, o contraste será alto. Sua equação (4) é dada abaixo:

$$Contraste = \sum_{n=0}^{n-1} n^2 \left\{ \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i,j) \right\} \quad |i - j| = n \quad (4)$$

- **Correlação**

Expressa a medida da dependência linear dos níveis de cinza sobre os pixels vizinhos ou pontos especificados. Indica a dependência de nível de cinza local na imagem de textura. Podem ser obtidos valores mais elevados de correlação para uma região com níveis de cinza semelhantes (ALBREGTSEN, 2008). Sua equação (5) é dada abaixo:

$$Correlação = \sum_{i,j=0}^{n-1} \frac{(i-\mu_x)*(j-\mu_y)*P(i,j)}{\sigma_i\sigma_j} \quad (5)$$

onde μ_x e μ_y são as médias ao longo das linhas e colunas de $P(i,j)$ e σ_x e σ_y são os desvios padrões ao longo das linhas e colunas de $P(i,j)$ (ALBREGTSEN, 2008). Suas respectivas equações (6), (7), (8) e (9) são:

$$\mu_x = \sum_{i=0}^{n-1} i * \rho_x(i) \quad (6)$$

$$\mu_y = \sum_{j=0}^{n-1} j * \rho_y(j) \quad (7)$$

$$\sigma_x^2 = \sum_{i=0}^{n-1} (i - \mu_x)^2 * \sum_{j=0}^{n-1} P(i,j) \quad (8)$$

$$\sigma_y^2 = \sum_{j=0}^{n-1} (j - \mu_y)^2 * \sum_{i=0}^{n-1} P(i,j) \quad (9)$$

onde $P_x(i)$ e $P_y(j)$ são, respectivamente, as i -ésima e j -ésima entradas das matrizes probabilidades obtidas pela soma dos linhas e colunas de $P(i,j)$ (ALBREGTSEN, 2008). Suas equações (10) e (11) são:

$$\rho_x(i) = \sum_{j=0}^{n-1} P(i,j) \quad (10)$$

$$\rho_y(j) = \sum_{i=0}^{n-1} P(i,j) \quad (11)$$

- **Entropia**

Expressa a medida de aleatoriedade (distúrbio) da textura da imagem (distribuição de intensidade). A entropia é grande quando a imagem não é uniforme e muitos elementos da GLCM têm pequenos valores. Portanto, uma imagem homogênea resultará em um menor valor de entropia, enquanto que uma região não homogênea (heterogênea) e complexa resultará em um maior valor de entropia (GADKARI, 2000). Sua equação (12) é dada abaixo:

$$Entropia = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i,j) * \log(P(i,j)) \quad (12)$$

- **Entropia da Soma**

Temos que ρ_{x+y} é a matriz dada pela equação (13) abaixo (ALBREGTSEN, 2008):

$$\rho_{x+y}(k) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i, j) \quad i + j = k \quad (13)$$

onde $k = 0, 1, 2, \dots, 2(n-1)$.

A entropia da soma expressa a medida de aleatoriedade da textura da imagem ao longo das diagonais definidas pela condição $i + j = k$ (ALBREGTSEN, 2008). Sua equação (14) é dada abaixo:

$$\text{Entropia da Soma} = \sum_{i=0}^{2(n-1)} \rho_{x+y}(i) * \log\{\rho_{x+y}(i)\} \quad (14)$$

- **Entropia da Diferença**

Temos que ρ_{x-y} é a matriz dada pela equação (15) abaixo (HARALICK, 1973):

$$\rho_{x-y}(k) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i, j) \quad |i - j| = k \quad (15)$$

onde $k = 0, 1, 2, \dots, n-1$.

A entropia da diferença expressa a medida de aleatoriedade da textura da imagem ao longo das diagonais definidas pela condição $|i - j| = k$ (HARALICK, 1973). Sua equação (16) é dada abaixo:

$$\text{Entropia da Diferença} = \sum_{i=0}^{n-1} \rho_{x-y}(i) * \log\{\rho_{x-y}(i)\} \quad (16)$$

- **Variância**

Expressa a medida da dispersão dos valores da GLCM em torno da média das combinações de pixels de referência e de vizinhança. Uma alta variância indica maiores valores de distâncias superiores à média (GADKARI, 2000). Sua equação (17) é dada abaixo:

$$\text{Variância} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i - \mu)^2 * P(i, j) \quad (17)$$

onde μ é a média de $P(i, j)$, dada pela equação (18) abaixo (ALBREGTSEN, 2008):

$$\mu = \sum_{i=0}^{n-1} i * P(i, j) \quad (18)$$

- **Variância da Soma**

Sua equação (19) é dada abaixo (PRATIWI et al., 2015):

$$\text{Variância da Soma} = \sum_{i=2}^{2(n-1)} (i - \text{Entropia da Soma})^2 * \rho_{x+y} \quad (19)$$

onde *Entropia da Soma* é dada pela equação (14).

- **Variância da Diferença**

Sua equação (20) é dada abaixo (HARALICK, 1973):

$$\text{Variância da Diferença} = \text{variância de } \rho_{x-y} \quad (20)$$

- **Momento de Diferença Inversa Normalizada**

Expressa a medida da homogeneidade local de uma imagem. A incidência da coocorrência de pares de pixels é aumentada quando eles estão próximos do valor da escala de cinza e, assim, aumenta o valor do momento de diferença inversa. Por conta do fator de ponderação $(1 + (i - j)^2)^{-1}$, obterá pequenas contribuições de áreas não homogêneas $i \neq j$. (ALBREGTSEN, 2008). Sua equação (21) é dada abaixo:

$$\text{Momento de Diferença Inversa Normalizada} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{1}{1 + (i - j)^2} * P(i, j) \quad (21)$$

- **Diferença Inversa Normalizada**

Expressa outra medida de homogeneidade local de uma imagem. Ao contrário do momento de diferença inversa normalizada, a diferença inversa normalizada normaliza a diferença entre os valores dividindo pelo total de valores discretos (CLAUSI, 2002). Sua equação (22) é dada abaixo:

$$\text{Diferença Inversa Normalizada} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{1}{1 + |i - j|} * P(i, j) \quad (22)$$

- **Média da Soma**

Expressa a média dos valores das diagonais definidas pela condição $i + j = k$ (ALBREGTSEN, 2008) Sua equação (23) é dada abaixo:

$$\text{Média da Soma} = \sum_{i=0}^{2n-2} i * \rho_{x+y}(i) \quad (23)$$

- **Média da Diferença**

Expressa a média dos valores das diagonais definidas pela condição $|i - j| = k$ (BOVIS, K. SINGH, S. s.d.) Sua equação (24) é dada abaixo:

$$\text{Média da Diferença} = \sum_{i=0}^{2n-2} i * \rho_{x-y}(i) \quad (24)$$

- **Medida de Informação da Correlação 1**

Sua equação (25) é dada abaixo (HARALICK, 1973):

$$\text{Medida de Informação da Correlação 1} = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (25)$$

onde HXY, HXY1, HX e HY são dadas pelas equações (26), (27), (28) e (29) abaixo:

$$HXY = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i, j) * \log(P(i, j)) \quad (26)$$

$$HXY1 = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i, j) * \log(\rho_x(i) * \rho_x(j)) \quad (27)$$

$$HX = \text{Entropia de } \rho_x \quad (28)$$

$$HY = \text{Entropia de } \rho_y \quad (29)$$

- **Medida de Informação da Correlação 2**

Sua equação (30) é dada abaixo (HARALICK, 1973):

$$\text{Medida de Informação da Correlação 2} = \sqrt{(1 - e^{-2(HXY2 - HXY)})} \quad (30)$$

onde HXY2 é dado pela equação (31) abaixo:

$$HXY2 = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \rho_x(i) * \rho_x(j) * \log(\rho_x(i) * \rho_x(j)) \quad (31)$$

- **Coefficiente de Correlação Máximo**

Sua equação (32) é dada abaixo (HARALICK, 1973):

$$\text{Coefficiente de Correlação Máximo} = \sqrt{(\text{Segundo maior autovalor de } Q)} \quad (32)$$

onde Q é dado pela equação (33) abaixo:

$$Q = \sum_k \frac{P(i, k) * P(j, k)}{\rho_x(i) * \rho_y(k)} \quad (33)$$

- **Autocorrelação**

Expressa a medida da regularidade da finura e grossura de uma textura (LEW, 2001). Sua equação (34) é dada abaixo:

$$\text{Autocorrelação} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i * j) * P(i, j) \quad (34)$$

- **Dissimilaridade**

É uma medida semelhante ao contraste, apresentando alto valor se a região possuir alto contraste (SOH, 1999). Sua equação (35) é dada abaixo:

$$\text{Dissimilaridade} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} |i - j| * P(i, j) \quad (35)$$

- **Sombra de Agrupamento**

Expressa a medida de desigualdade e assimetria da GLCM. Uma sombra de agrupamento mais alta implica maior assimetria em relação ao valor médio (YANG et al., 2012). Sua equação (36) é dada abaixo:

$$\text{Sombra de Agrupamento} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i + j - \mu_x - \mu_y)^3 * P(i, j) \quad (36)$$

- **Proeminência de Agrupamento**

Expressa também um medida de assimetria da GLCM. Um valor mais elevado implica maior assimetria em relação ao valor médio, enquanto um valor inferior indica um pico em torno do valor médio e menos variação e relação à média (YANG et al., 2012). Sua equação (37) é dada abaixo:

$$\text{Proeminência de Agrupamento} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i + j - \mu_x - \mu_y)^4 * P(i, j) \quad (37)$$

- **Máxima Probabilidade**

Expressa a máxima probabilidade (SOH, 1999). Sua equação (38) é dada abaixo:

$$\text{Máxima Probabilidade} = \max\{P(i, j)\} \quad (38)$$

- **Inércia**

Sua equação (39) é dada abaixo (BOVIS, K. SINGH, S. s.d.):

$$\text{Inércia} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i - j)^2 * P(i, j) \quad (39)$$

2.6 Classificação e Reconhecimento

Nesta etapa os atributos de interesse já foram extraídos e deseja-se realizar, de forma automática, a identificação dos objetos segmentados na imagem. Há dois tipos de etapas no processo de classificação: o aprendizado e o reconhecimento. Os sistemas de aprendizado definem uma função que separa eficientemente todas as formas representadas no espaço de medida de N dimensões, onde N é o número de atributos extraídos (ALBUQUERQUE, 2011).

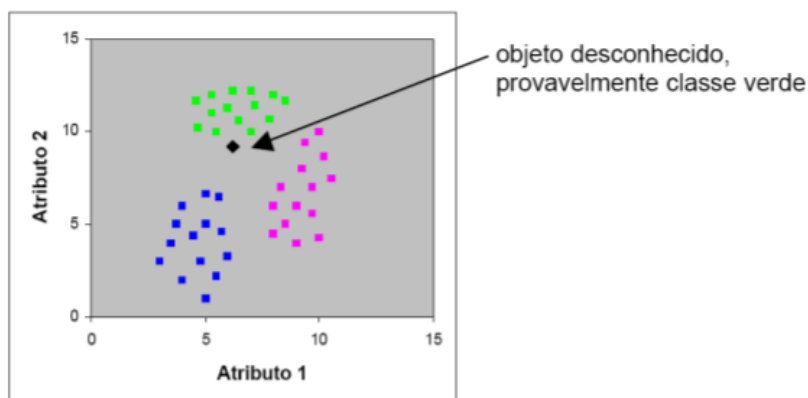
A escolha e a qualidade dos parâmetros para construir esse espaço de medidas são muito importantes para que o processo de reconhecimento tenha um bom desempenho. No entanto, essa etapa é delicada, uma vez que a escolha de um grande número de parâmetros leva a um espaço de grande dimensão e uma difícil etapa de aprendizado, enquanto que um pequeno espaço de medidas pode levar a uma baixa caracterização e erros no processo de reconhecimento.

Pode-se dividir o processo de aprendizado em dois tipos: supervisionado, o utilizado neste trabalho, e não supervisionado. Na classificação supervisionada, a cada objeto, ou seja, elemento caracterizado por um conjunto de atributos é associado uma classe previamente definida por um analista.

Na classificação não supervisionada, nenhuma informação a respeito das classes é fornecida, e os objetos são agrupados em distintos grupos através de suas proximidades (GOMES, 2001).

A figura 21 mostra um exemplo de um classificador do tipo supervisionado. O objeto desconhecido, com classe a ser determinada, encontra-se mais próximo, segundo seus atributos 1 e 2, da classe de cor verde. Tanto a classe de cor verde quanto as de cor azul e rosa foram definidas por um especialista e, portanto, seus atributos definem suas respectivas classes. Provavelmente o classificador irá atribuir a classe verde para o objeto.

Figura 21 – Classificador supervisionado.



Fonte: GOMES, 2001.

2.6.1 Algoritmos de Classificação

Neste trabalho foram usados dois algoritmos de classificação: o kNN e o SVM.

2.6.1.1 k-Nearest Neighbor (kNN)

O algoritmo kNN (*k-Nearest Neighbor*) é uma forma de aprendizado do tipo supervisionado e um dos mais simples e flexíveis quando se trata de classificação.

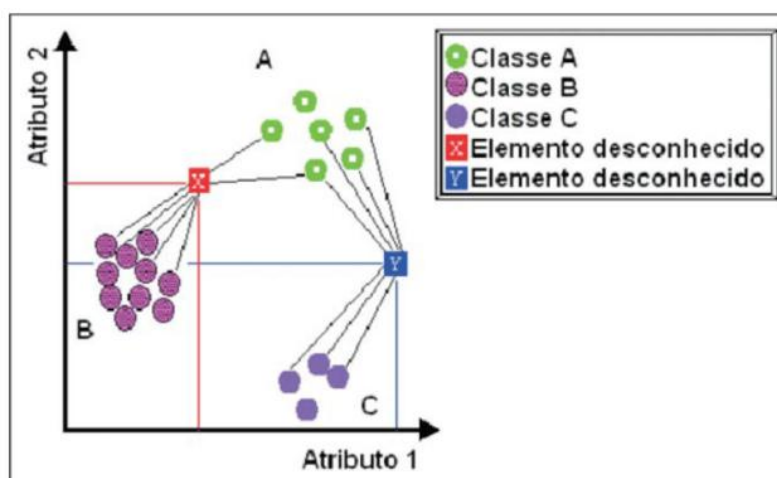
Basicamente, o algoritmo funciona com base em exemplos de treinamento que estão mais próximos no espaço de características. Para utilizá-lo é necessário (LIMA, 2014):

- Um conjunto de exemplos de treinamento;
- Definir uma métrica para calcular a distância entre os exemplos de treinamento;
- Definir o valor de k (o número de vizinhos mais próximos que serão considerados pelo algoritmo);
- Utilizar o rótulo da classe dos vizinhos mais próximos para determinar o rótulo de classe do exemplo desconhecido através de uma votação majoritária.

Devemos ressaltar que se k for muito pequeno, a classificação fica sensível a pontos de ruído, ao passo que se k for muito grande, a vizinhança pode incluir elementos de outras classes. Para se evitar empates na votação, usa-se um valor ímpar para k (LIMA, 2014).

A figura 22 abaixo exemplifica a classificação de dois elementos desconhecidos X e Y utilizando o algoritmo kNN. Observe que para $k = 7$, o elemento X será classificado como pertencente à classe B, enquanto que o elemento Y será enquadrado na classe A.

Figura 22 – Classificação através do algoritmo kNN.



Fonte: MARTINS, 2005.

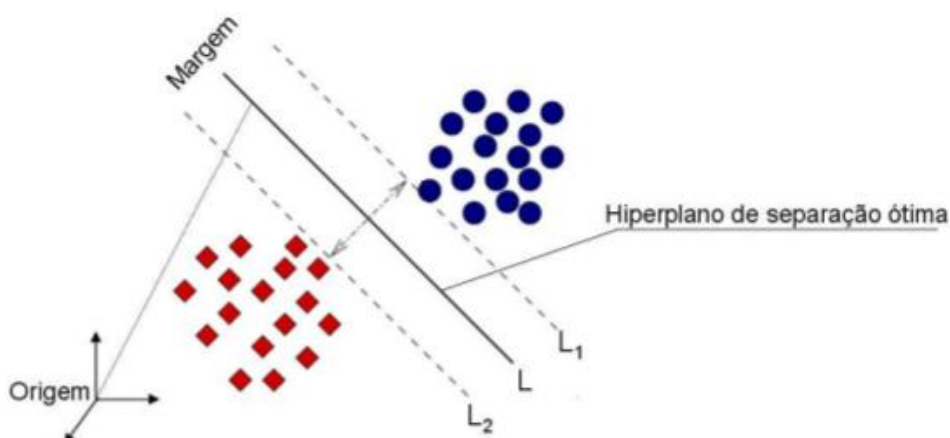
2.6.1.2 Support Vector Machine (SVM)

O algoritmo SVM muito se assemelha com o anteriormente citado algoritmo kNN, ou seja, é um analisador e classificador de dados para reconhecimento de padrões do tipo supervisionado, com a diferença que seu objetivo é a determinação de limites de decisão que produzam uma separação ótima entre classes por meio da minimização dos erros (NASCIMENTO et al., 2009). Ou seja, os novos exemplos são mapeados no mesmo espaço e preditos como pertencentes a uma categoria baseados em qual dos lados definidos por uma separação eles são colocados.

Os resultados da aplicação dessa técnica são comparáveis aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs). Exemplos de aplicações de sucesso podem ser encontrados em diversos domínios, como na categorização de textos, na análise de imagens e em Bioinformática (LORENA, 2007).

Conforme mostrado na figura 23, a separação ótima é definida por um hiperplano condicional (L) que separa as diferentes classes, maximizando a margem (distância entre as bordas, L_1 e L_2) entre os pontos extremos de cada classe.

Figura 23 – Classificação através do algoritmo SVM.



Fonte: NASCIMENTO, 2009.

Nem sempre é possível obter uma separação linear entre as classes e podem ocorrer erros de classificação. Nestes casos, é necessário considerar um parâmetro de custo c que indica o quão impactante o erro de classificação será no processo de obtenção do hiperplano de separação. Esse valor precisa ser previamente definido.

2.6.2 Validação dos Classificadores

Quatro formas de validar um classificador são utilizadas nesse trabalho: a acurácia (A), a sensibilidade (S), especificidade (E) e a medida F das duas últimas, chamada neste trabalho de média harmônica (MH). Essas medidas são definidas através da matriz de confusão, um tipo de matriz que relaciona os resultados finais dados pelo algoritmo (previsto) e pelo profissional (real). Um exemplo de matriz de confusão é ilustrado na tabela 2 abaixo:

Tabela 2 – Matriz de confusão.

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP Verdadeiro Positivo	FP Falso Positivo
	negativos	FN Falso Negativo	VN Verdadeiro Negativo

Fonte: SOUZA, 2009.

Analisando a matriz de confusão percebemos que são definidas quatro métricas diferentes de acordo com a interseção dos resultados previstos e reais. Suas definições estão listadas abaixo (BIANCA, 2004):

- Verdadeiro Positivo – VP:
Diz-se dos resultados considerados verdadeiros tanto no teste com o algoritmo quanto na classificação pelo banco de dados;
- Falso Negativo – FN:
Diz-se dos resultados considerados falsos no teste com o algoritmo e verdadeiros na classificação pelo banco de dados;
- Falso Positivo – FP:
Diz-se dos resultados considerados verdadeiros no teste com o algoritmo e falsos na classificação pelo banco de dados;
- Verdadeiro Negativo – VN:
Diz-se dos resultados considerados falsos tanto no teste com o algoritmo quanto na classificação pelo banco de dados.

De conhecimento das quatro métricas definidas a partir da matriz de confusão, podem-se definir as métricas de validação de um classificador pelas equações (38) (BHARATHI, 2013), (39) e (40) (BIANCA, 2004) e (41) abaixo:

$$A = \frac{VP + VN}{VP + VN + FP + FN} \quad (38)$$

$$S = \frac{VP}{VP + FN} \quad (39)$$

$$E = \frac{VN}{VN + FP} \quad (40)$$

$$MH = \frac{2SE}{S + E} \quad (41)$$

A acurácia é o percentual correspondente à assertividade do teste, ou seja, o teste deve informar de maneira correta a classe do objeto, de acordo com a classificação real.

Analisando ambas as equações (39) e (40), pode-se concluir que a sensibilidade nos dá a proporção de indivíduos identificados com a doença (tanto no classificador quanto no diagnóstico) dentre os indivíduos de fato com a doença. Por conseguinte, a especificidade nos dá a proporção de indivíduos identificados como normais (tanto no classificador quanto no diagnóstico) dentre os indivíduos com características normais.

2.7 Seleção de atributos

Nem todas as medidas extraídas da matriz de coocorrência devem ser úteis para realizar a tarefa de detecção e classificação de câncer e tecido mamário, inclusive, algumas podem até dificultar na realização da tarefa. Por isso, costuma-se realizar um passo adicional antes da etapa de classificação, que corresponde a parte de selecionar os atributos (medidas) que são os mais úteis para a classificação. Um dos métodos mais simples de seleção dos atributos mais proeminentes é o *Sequential Forward Selection* (SFS) (DUTRA, L. V. HUBER, R., 1999).

A seleção sequencial progressiva (SFS) é um método *bottom-up* de seleção de atributos estruturado de modo a identificar iterativamente o melhor subconjunto de atributos que possa ser obtido pela adição de um atributo por iteração, ao invés de selecionar um conjunto completo de atributos numa única iteração.

O algoritmo SFS seleciona os melhores atributos que representam um conjunto de classes a partir de um critério de seleção. O processo inicia-se com um conjunto nulo de atributos, de modo que, em cada iteração é selecionado o atributo que produza o melhor resultado segundo um determinado critério. Tradicionalmente são utilizadas como critério, medidas estatísticas de separabilidade. Assim, iterativamente, cada atributo é adicionado ao conjunto de atributos determinados pelas interações anteriores. Nesse processo não é permitido descartar um atributo que já tenha sido selecionado (BATISTA, 2006).

No presente trabalho, o critério de seleção utilizado pelo algoritmo SFS foi a média harmônica do classificador kNN entre as medidas de texturas utilizadas.

3 METODOLOGIA E DESENVOLVIMENTO

O presente trabalho consiste em detectar e classificar o tipo de câncer presente em uma mamografia digital, além de classificar o tecido mamário mais abundante da mama. Para isso, utilizou-se do banco de dados MIAS, fornecido gratuitamente e com um total de 322 imagens mamográficas de resolução 1024x1024 pixels (MIAS, s.d.). O banco de dados fornece diversas informações a respeito das mamografias, tais como: tecido mamário, presença ou não de câncer, tipo de câncer, classe do câncer e as ROIs do centro da anormalidade nas imagens com a presença do câncer.

Em relação à presença ou não de câncer, das 322 imagens presentes no banco de dados, 209 não possuem câncer e 113 possuem câncer. Das imagens que possuem câncer, 62 são do tipo benigno e 51 são do tipo maligno. Possuindo as coordenadas do centro da anormalidade são 109 imagens com câncer.

Em relação à classificação do tecido mamário, das 322 imagens presentes no banco de dados, 106 são adiposas e 216 são glandulares (glandular adiposa e glandular densa). Dentre as glandulares, 104 são do tipo glandular adiposa e 112 são do tipo glandular densa.

As mamas diferenciam-se de várias maneiras como formato, tamanho e sentido (direita ou esquerda), conforme mostra a figura 24.

Figura 24 – Exemplos de mamografias utilizadas.



Fonte: MIAS s.d.

Podemos observar nas mamografias do banco de dados a presença de marcações e números junto ao fundo da imagem, além do músculo peitoral na grande maioria delas. Essas características, muito comuns em mamografias, junto com a transição “nebulosa” entre o seio e o fundo, são indesejáveis e devem ser eliminados no processamento. Assim como visto em outros trabalhos, esses são detalhes que não trazem contribuição alguma para o resultado esperado e podem piorar os cálculos realizados pela introdução de valores irrelevantes.

A metodologia adotada neste trabalho está representada pelo fluxograma da figura 25. Temos dois tipos de procedimento: a detecção e classificação de câncer; e a classificação do tecido mamário.

No primeiro procedimento, o primeiro passo é a aquisição da imagem. Feito isso, faz-se a leitura das coordenadas da ROI fornecida pelo banco de dados (imagens com câncer). Para as imagens sem câncer, faz-se a leitura das coordenadas da ROI manual, cuja localização é o mais próximo possível do centro da mama.

De posse desses dados, o algoritmo obtém a ROI na imagem, que será realçada através da equalização do histograma e redimensionada para 128x128 pixels, etapa essa baseada em (PRATIWI et al., 2015). Por conseguinte, os atributos de textura são calculados e os classificadores utilizados para a verificação da veracidade do teste.

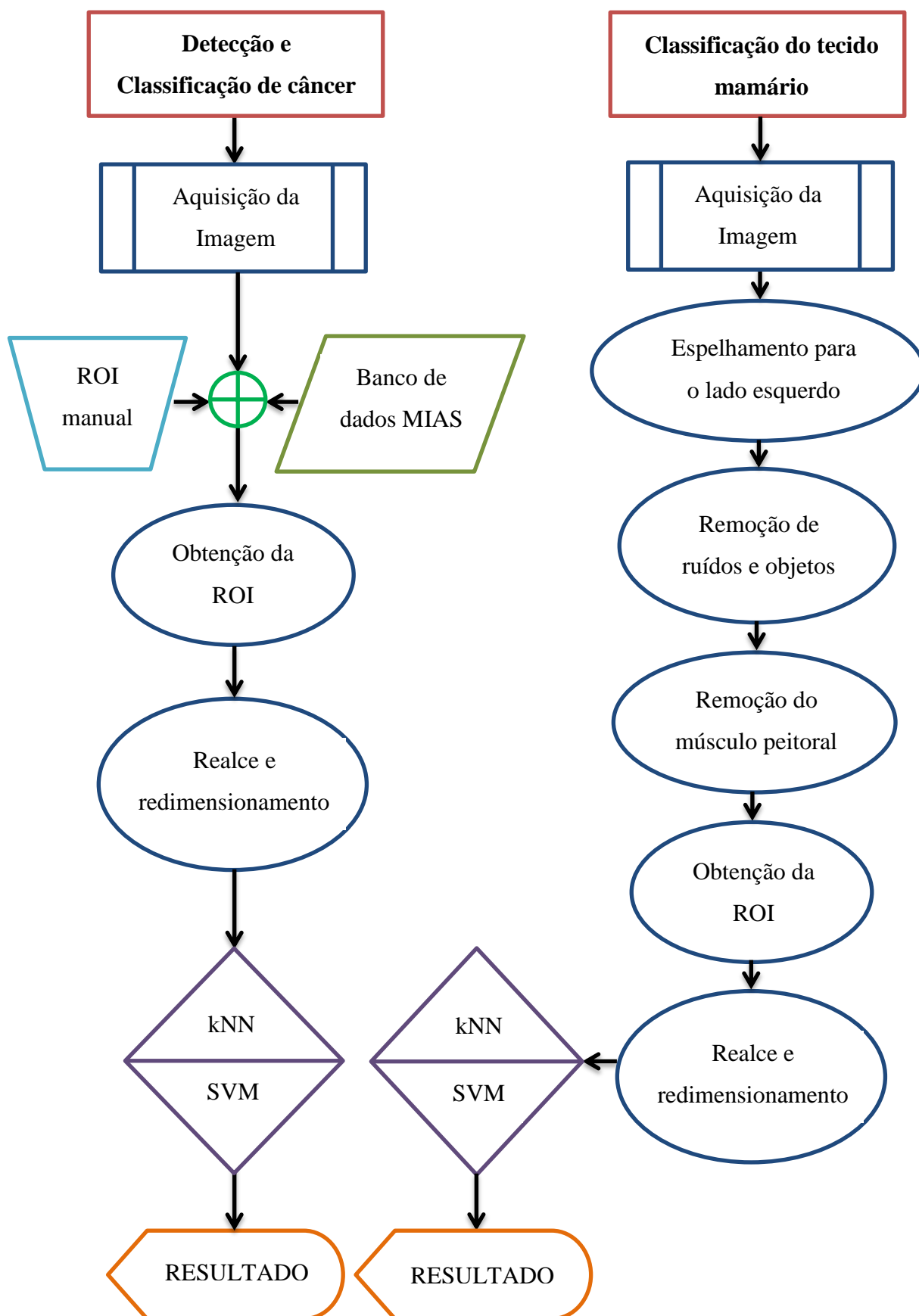
No segundo procedimento, o primeiro passo é também a aquisição da imagem. Posteriormente, deve se fazer diversos procedimentos de processamento de imagem na mesma, a começar pelo seu espelhamento para o lado esquerdo, etapa essa fundamental para que o algoritmo da Transformada de Hough funcione posteriormente.

A remoção de ruídos e objetos na imagem é a etapa seguinte, onde marcações, ruído de fundo e falhas de impressão na mamografia são eliminados. Técnicas como suavização, binarização, abertura e contagem de objetos são utilizadas nessa etapa.

Faz-se necessário, então, a retirada do músculo peitoral, etapa essa fundamental para que a ROI, definida posteriormente, não englobe esse músculo e forneça um resultado errado, uma vez que ele não pertence a nenhum dos tecidos considerados no teste. Para essa retirada utiliza-se a técnica da Transformada de Hough para a detecção de retas, uma vez que o limite do músculo peitoral, nas imagens, se assemelha a uma reta.

Feito todo o procedimento de retirada dos elementos não relevantes para o teste, faz-se a obtenção da ROI na imagem, que será realçada através da equalização do histograma e redimensionada para 256x256 pixels. Por conseguinte, os atributos de textura são calculados e os classificadores utilizados para a verificação da veracidade do teste.

Figura 25 – Fluxograma do processo de detecção e classificação de câncer e tecido mamário.



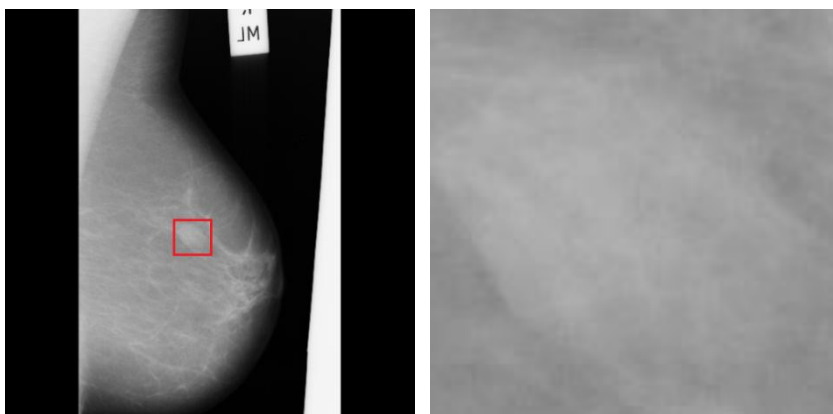
Fonte: Produção do próprio autor.

3.1 Detecção e Classificação de câncer

No primeiro procedimento, não é feito nenhum pré-processamento ou segmentação na imagem. Faz-se a leitura da região de interesse (ROI) pré-determinada pelo banco de dados MIAS para as imagens com câncer, procedimento esse semelhante ao realizado nos trabalhos citados na seção 1.7. Vale lembrar que para as imagens sem câncer foi feita uma escolha manual da ROI, posicionando-a o mais próxima possível do meio da mamografia e contendo um tamanho de janela igual a 128x128 pixels. Quatro mamografias contendo câncer foram descartadas nesse procedimento por não apresentarem ROI anteriormente definida, a saber: mdb059.pgm, mdb216.pgm, mdb233.pgm e mdb245.pgm.

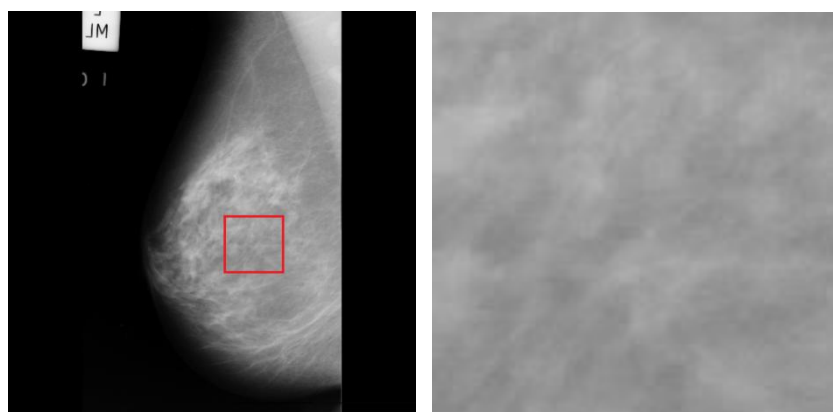
As mamografias mdb012.pgm (com câncer) e mdb007.pgm (sem câncer) serão usadas para ilustrar o passo-a-passo das técnicas utilizadas. As figuras 26 e 27 abaixo ilustram as mamografias citadas e sua respectiva ROI em destaque.

Figura 26 – Mamografia com câncer e sua respectiva ROI dada pelo banco de dados.



Fonte: Produção do próprio autor.

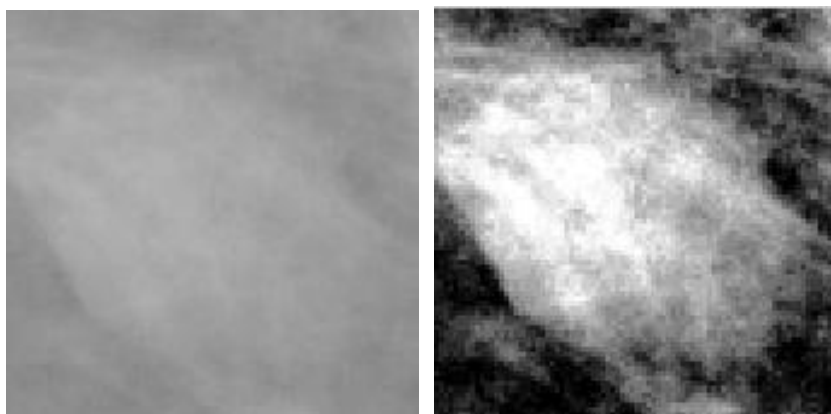
Figura 27 – Mamografia sem câncer e sua respectiva ROI manual.



Fonte: Produção do próprio autor.

Após a obtenção da ROI, realiza-se um realce de contraste da mesma através da técnica de equalização do histograma. As figuras 28 e 29 mostram o realce e o histograma equalizado para a ROI com câncer, ao passo que as figuras 30 e 31 mostram as mesmas características para a ROI sem câncer.

Figura 28 – ROIs (com câncer) com e sem realce.

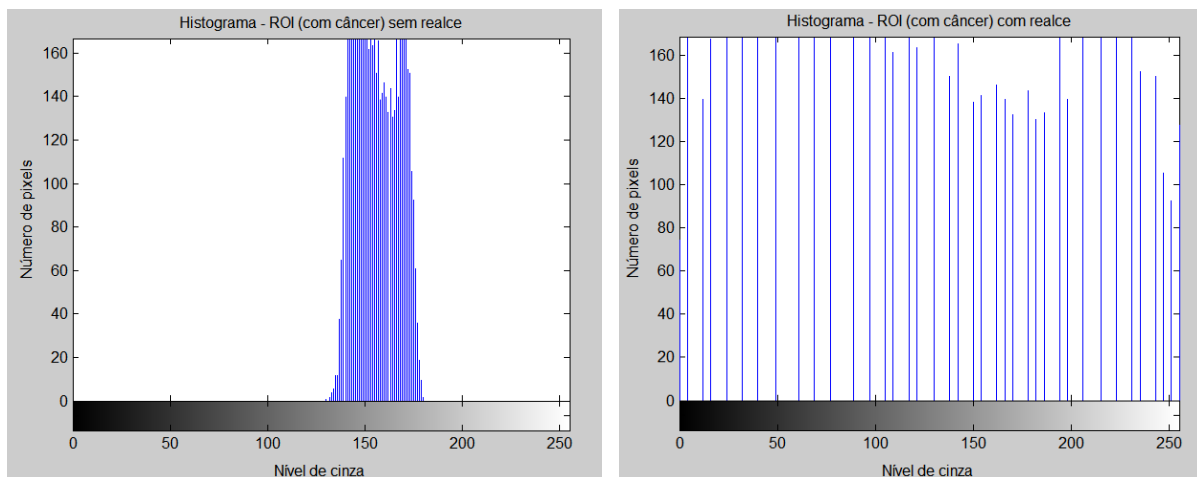


a) Sem realce.

b) Com realce.

Fonte: Produção do próprio autor.

Figura 29 – Histogramas para ROIs (com câncer) com e sem realce.

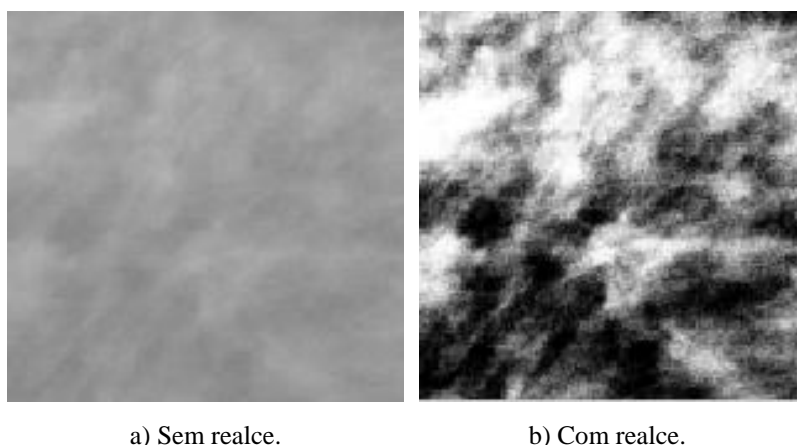


a) Histograma sem realce.

b) Histograma com realce.

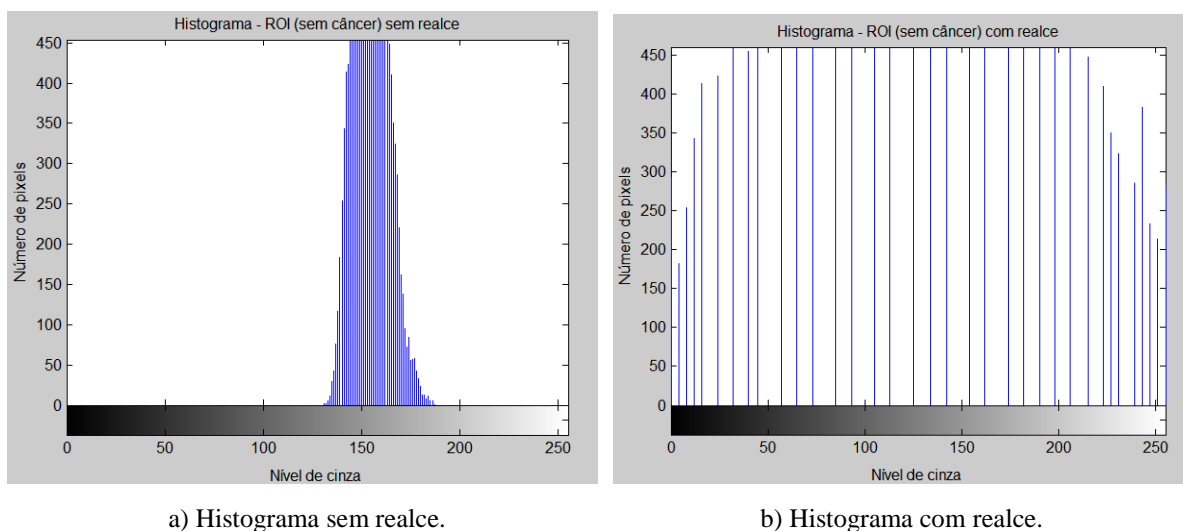
Fonte: Produção do próprio autor.

Figura 30 – ROIs (sem câncer) com e sem realce.



Fonte: Produção do próprio autor.

Figura 31 – Histogramas para ROIs (sem câncer) com e sem realce.



Fonte: Produção do próprio autor.

Obtida a ROI realçada, o próximo passo é redimensioná-la para 128x128 pixels. Esse redimensionamento foi inspirado nos testes realizados em (PRATIWI et al., 2015). Vale lembrar que as ROIs fornecidas pelo banco de dados possuem dimensões diferentes, enquanto todas as ROIs manuais possuem um tamanho de janela igual a 128x128 pixels, daí a necessidade de redimensioná-las de forma a ficarem com as mesmas dimensões.

Obtida a ROI dimensionada e realçada, calculam-se as medidas de textura usando as Matrizes de Coocorrência de Nível de Cinza para todos os ângulos de direção (0° , 45° , 90° e 135°) e distância $d = 1$ entre cada pixel para todas elas.

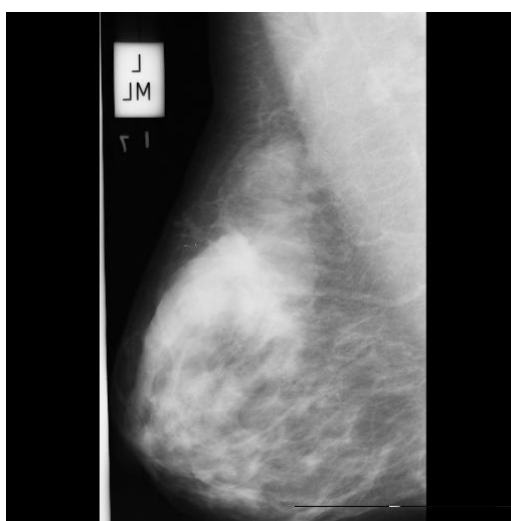
Finalmente, após a obtenção das medidas, utilizam-se os classificadores kNN e SVM para realizarem a detecção e classificação do câncer.

3.2 Classificação do tecido mamário

No segundo procedimento, é necessário um trabalho maior do que o realizado no primeiro procedimento, pois não são usadas as ROI pré-definidas, de forma a permitir que todas as 322 imagens fossem utilizadas, uma vez que a falta de uma ROI definida pelo banco de dados não é um empecilho para esta etapa.

A mamografia mdb115.pgm será usada para ilustrar o passo-a-passo das técnicas utilizadas. A figura 32 ilustra a mamografia após sua aquisição.

Figura 32 – Mamografia para classificação do tecido mamário.



Fonte: MIAS, s.d.

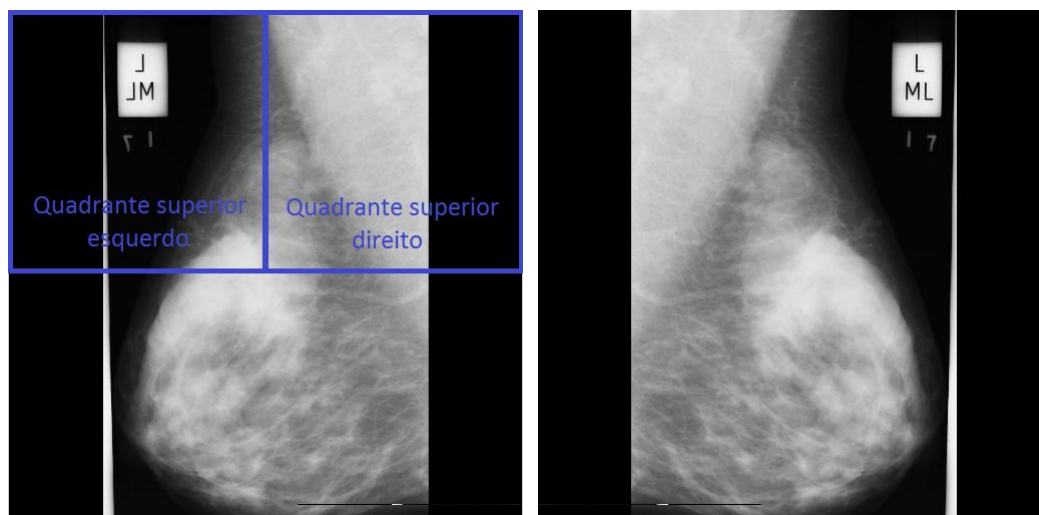
Feita a aquisição da imagem, o próximo passo é espelhá-la para o lado esquerdo. Esse procedimento é realizado para garantir que todas as mamas das imagens estejam do lado esquerdo, de forma a facilitar a aplicação das técnicas subsequentes. Importante ressaltar que na base de dados MIAS existem mamas tanto do lado direito quanto do lado esquerdo.

Para que todas as imagens adquiridas sejam espelhadas corretamente para o lado esquerdo, fez-se o seguinte procedimento:

- Para o quadrante superior esquerdo e o quadrante superior direito da imagem, calculou-se o número de zeros na imagem contidos em cada um deles;
- Se o número de zeros no quadrante superior esquerdo for maior, isso significa que a mama está do lado direito da imagem. Deve-se então espelhá-la.

A figura 33 exemplifica essa etapa. Em (a) a imagem é dividida nos quadrantes superior esquerdo e direito e a contagem é feita. Em (b), a imagem é espelhada para o lado esquerdo com seu devido resultado.

Figura 33 – Etapa de espelhamento da mama para o lado esquerdo.



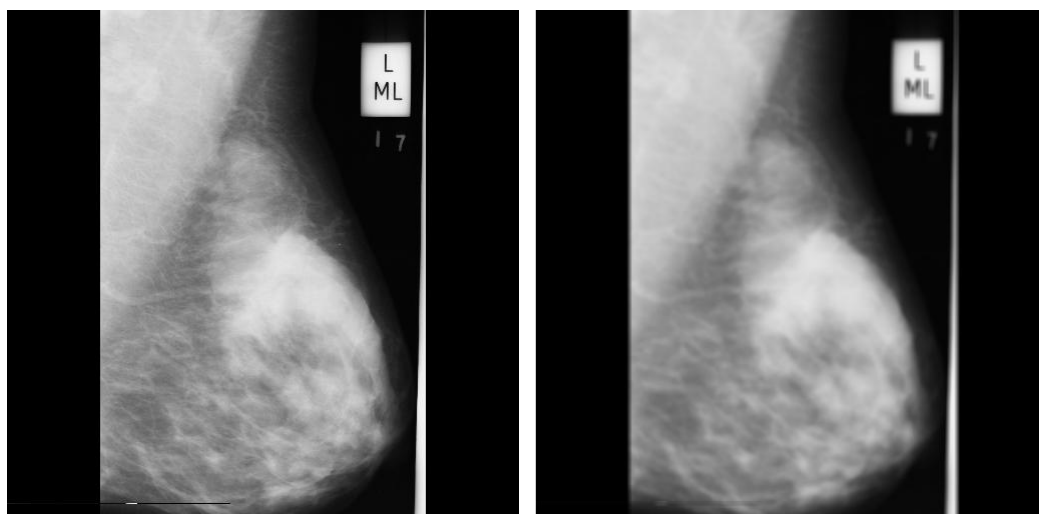
a) Imagem não espelhada.

b) Imagem espelhada.

Fonte: Produção do próprio autor.

Após a etapa de espelhamento, devem-se eliminar os ruídos e objetos presentes na imagem. Inicialmente, faz uma suavização da mesma utilizando um filtro de média de máscara 11x11. A figura 34 mostra o resultado, com a suavização ocorrendo de (a) para (b).

Figura 34 – Suavização da mamografia.



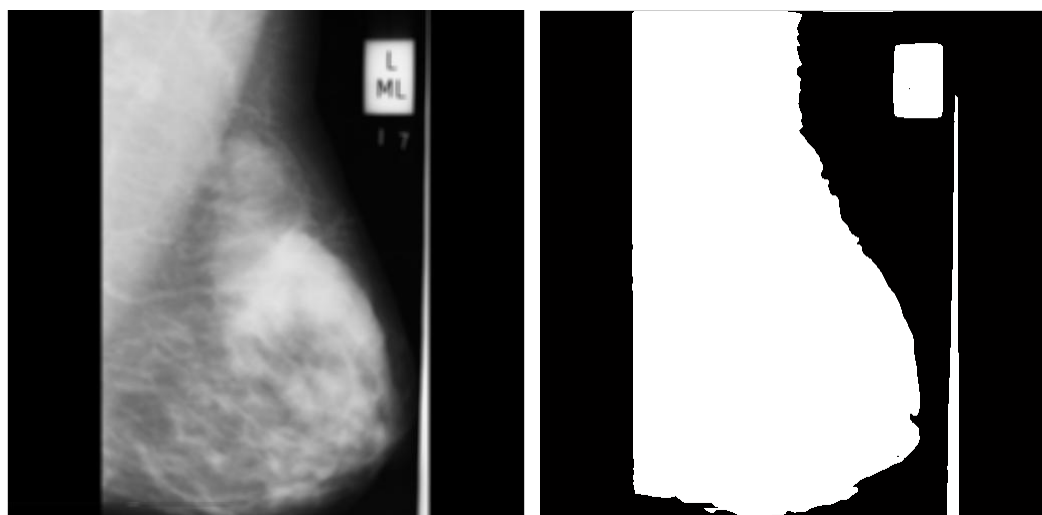
a) Imagem.

b) Imagem suavizada.

Fonte: Produção do próprio autor.

Feita a suavização, o próximo passo é a limiarização (binarização) da imagem. O limiar (*threshold*) foi escolhido a partir da binarização de Otsu. A figura 35 ilustra essa etapa, com a limiarização acontecendo de (a) para (b).

Figura 35 – Limiarização da mamografia.



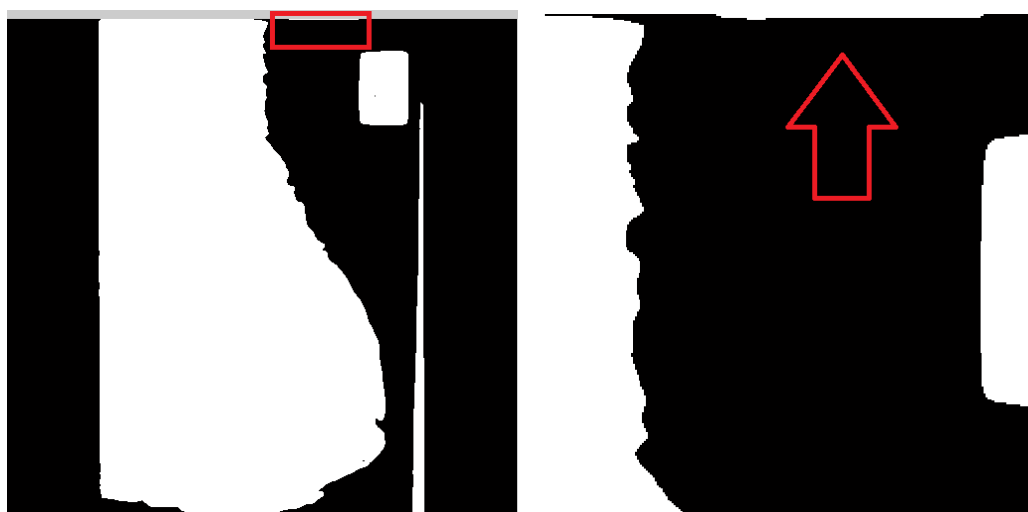
a) Antes da limiarização.

b) Após a limiarização.

Fonte: Produção do próprio autor.

Após a limiarização, pode-se perceber que a imagem possui quatro objetos de diferentes formas e tamanhos. A figura 36 mostra um objeto muito comum nas imagens do banco de dados. Em (a) temos a imagem binarizada e em (b) o objeto em destaque.

Figura 36 – Objeto responsável pela abertura da mamografia.



a) Objetos na imagem binarizada.

b) Destaque pra um objeto.

Fonte: Produção do próprio autor.

Esse objeto, em algumas imagens, acabava por se ligar à mama, dificultando a etapa seguinte de contagem de objetos e eliminação dos menores dentre esses. Não é o caso da imagem em destaque, mas em pelo menos dez imagens foi encontrado esse problema. Outro motivo para se eliminar esse objeto se dá pelo fato do mesmo se tornar um obstáculo quando da escolha dos limites do polígono envolvendo o músculo peitoral. Para eliminar esse empecilho faz-se uma operação de abertura com máscara 5x5.

Feita a operação de abertura, realiza-se a contagem de objetos na imagem e a permanência do maior dentre eles. A figura 37 ilustra o resultado dessa etapa. Em (a) a imagem com os quatro objetos e em (b) somente o maior objeto permanecendo.

Figura 37 – Contagem de objetos e permanência do maior objeto da mamografia.



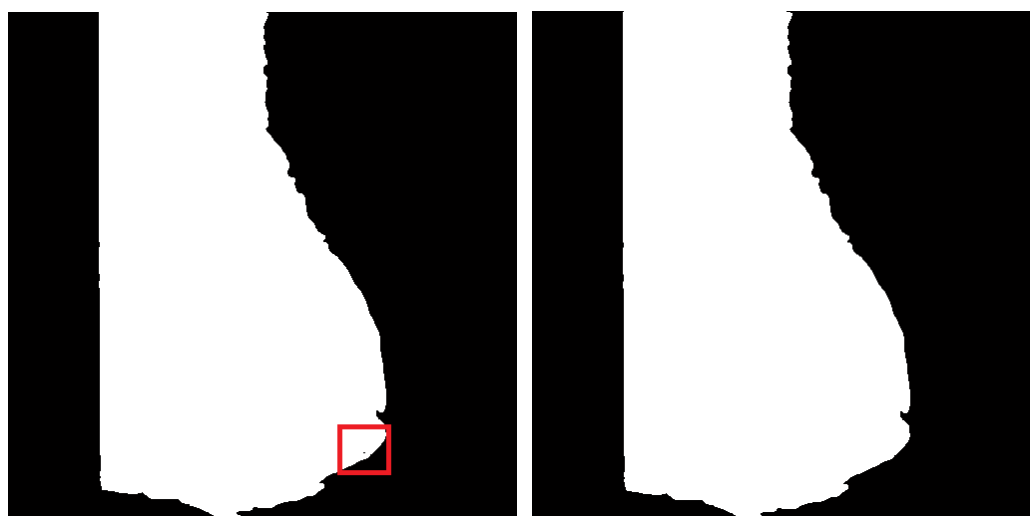
a) Imagem com objetos.

b) Seleção do maior objeto na imagem.

Fonte: Produção do próprio autor.

Após a etapa de eliminação dos menores objetos na imagem, é feito um preenchimento de buracos na mama, de forma a não se perder informações de seu interior. Na figura 38 é representada essa etapa que para a imagem em particular o resultado é quase imperceptível, porém necessário. Em (a) temos o buraco em destaque e em (b) a imagem preenchida.

Figura 38 – Preenchimento de buracos da mamografia.



a) Buraco no objeto.

b) Objeto preenchido.

Fonte: Produção do próprio autor.

Realizada a etapa de preenchimento de buracos, realiza-se a etapa de obtenção da mamografia sem ruídos e objetos e com músculo peitoral, como exemplificado na figura 39. É feita a multiplicação da imagem espelhada para o lado esquerdo (a) e a imagem binarizada preenchida (b), obtendo-se a imagem (c).

Figura 39 – Obtenção da mamografia sem ruídos e objetos e com músculo peitoral.



a) Imagem espelhada.

b) Imagem binarizada.

c) Multiplicação da imagem de a) por b).

Fonte: Produção do próprio autor.

Obtida a imagem sem ruído e objetos, o próximo passo é eliminar o músculo peitoral. Essa etapa é a mais extensa e trabalhosa, e consiste em diversas sub-etapas. A primeira delas é obter o polígono contendo o músculo peitoral, mas não necessariamente sua totalidade.

Para isso, obtém-se primeiramente o contorno da última imagem. A figura 40 ilustra essa etapa. Realiza-se a subtração entre a imagem binarizada (a) e sua imagem erodida com máscara 7x7 (b), resultando no contorno da imagem contendo o músculo peitoral (c).

Figura 40 – Obtenção do contorno da mamografia.



a) Imagem binarizada.

b) Imagem erodida.

c) Imagem com contorno.

Fonte: Produção do próprio autor.

Devemos agora definir os limites do polígono que contém o músculo peitoral. Para isso, fez-se o seguinte procedimento:

- Na metade superior da imagem binarizada com contorno, calculam-se os valores da coluna mais à esquerda e a mais à direita onde a imagem possui valor igual a 1;

A figura 41 ilustra esse procedimento, com as setas indicando as colunas escolhidas para delimitar o polígono contendo o músculo peitoral.

Figura 41 – Colunas selecionadas para delimitar o polígono contendo o músculo peitoral.



Fonte: Produção do próprio autor.

Com as medidas determinadas, delimita-se o polígono de forma que ele fique localizado acima da metade superior e entre os valores das colunas esquerda e direita. A figura 42 ilustra o polígono determinado.

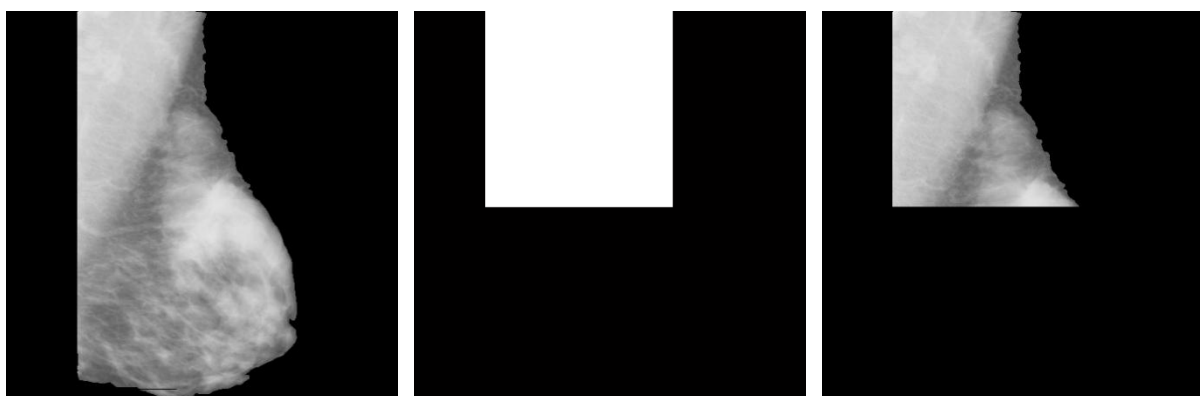
Figura 42 – Polígono determinado.



Fonte: Produção do próprio autor.

O próximo passo é obter a mamografia em forma do polígono. A figura 43 exemplifica esse passo, onde a imagem sem ruídos e objetos contendo o músculo peitoral (a) é multiplicada pelo polígono determinado (b), resultando na mamografia em forma do polígono (c).

Figura 43 – Obtenção da mamografia em forma do polígono.



a) Imagem.

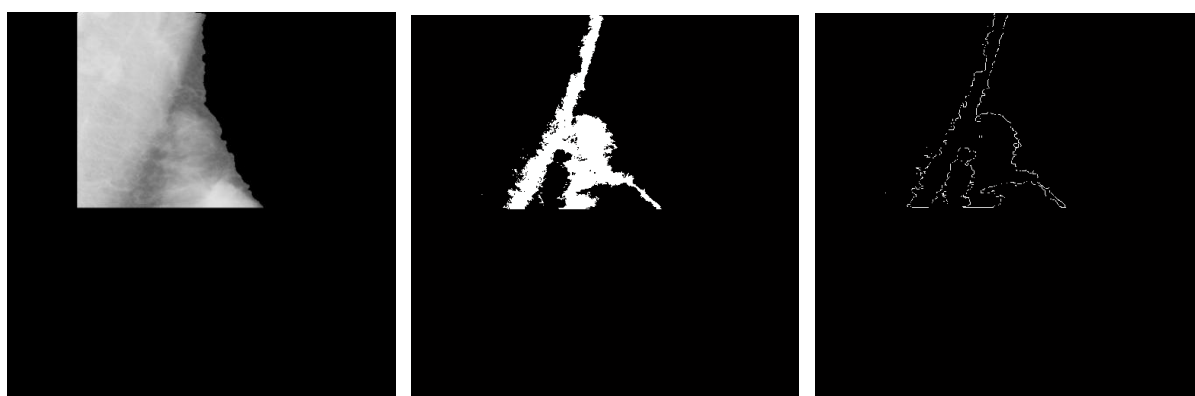
b) Polígono.

c) Multiplicação de a) por b).

Fonte: Produção do próprio autor.

O próximo passo é a aplicação da Transformada de Hough para encontrar a reta que contém o limite do músculo peitoral. Antes, porém, devemos binarizar (com a binarização de Otsu de 3 limiares) a mamografia em forma do polígono e obter o contorno da imagem resultante. Esses passos estão ilustrados na figura 44, de (a) para (b) e depois para (c).

Figura 44 – Binarização e contorno da mamografia em forma do polígono.



a) Imagem contendo o músculo peitoral.

b) Imagem binarizada.

c) Contorno da imagem.

Fonte: Produção do próprio autor.

Sabe-se que o músculo peitoral se assemelha a um triângulo retângulo na mamografia e que, portanto seu limite é uma reta inclinada. De forma a se obter essa reta, a Transformada de Hough foi usada nessa etapa para que se possa encontrar a reta mais bem definida. Para isso, ele foi limitado a procurar retas que façam ângulo de 15° a 80° com a reta horizontal. O maior valor acumulado no plano de Hough foi selecionado para representar a reta, pois isto representa a maior incidência de retas passando sobre os pontos em uma mesma direção. Com a equação da reta determinada, pode-se definir o triângulo retângulo contendo o músculo peitoral.

A figura 45 ilustra a obtenção da mamografia sem o músculo peitoral. Multiplica-se a imagem sem ruídos e objetos contendo o músculo peitoral (a) e a imagem negativada do triângulo retângulo cuja hipotenusa é o limite do músculo peitoral (b), resultando na imagem sem o músculo peitoral (c).

Figura 45 – Obtenção da mamografia sem ruídos e objetos e sem músculo peitoral.



a) Imagem da mama.

b) Divisão da imagem pela reta obtida.

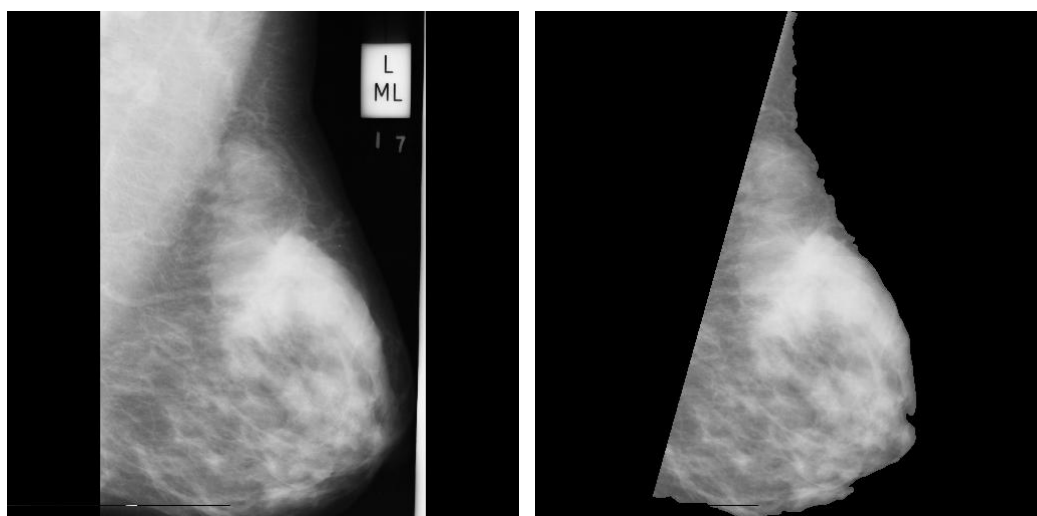
c) Imagem sem o músculo peitoral.

Fonte: Produção do próprio autor.

A figura 46 faz um comparativo entre a mamografia original (espelhada) (a) e a mamografia sem ruídos e objetos e sem músculo peitoral (b) obtida ao final da última etapa.

Podemos concluir, analisando ambas as mamografias, que os objetos e marcações foram eliminados, o ruído de fundo não existe mais, o músculo peitoral foi retirado e região “nebulosa” da transição entre o fundo da mamografia e a mama foi também desconsiderado. Todos esses elementos não são interessantes para a definição da ROI que será usada para a classificação do tecido mamário.

Figura 46 – Comparativo da mamografia original espelhada e da mamografia obtida na última etapa.



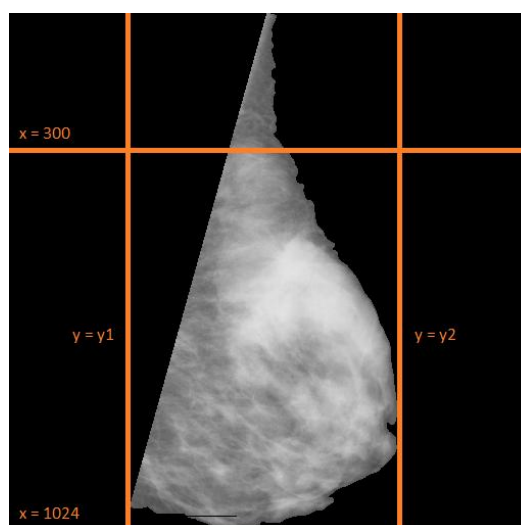
a) Imagem original espelhada.

b) Imagem após o procedimento.

Fonte: Produção do próprio autor.

Para podermos definir a ROI, procurou-se localizá-la no centro da mama. Sabendo-se que a grande parte das mamografias possui uma maior concentração de tecido na metade inferior da imagem, definiu-se o seguinte procedimento: enquanto não fosse encontrado um valor diferente de 0 (fundo da imagem), o limite superior da linha seria incrementado a partir da posição 300, enquanto que o limite inferior da linha seria decrementado a partir da posição 1024. Ao encontrar um valor diferente de 0, esses serão os limites superiores e inferiores utilizados pela ROI para sua localização na mama. As colunas são os limites à esquerda e à direita da imagem (fronteiras da mama y_1 e y_2). A figura 47 ilustra o procedimento adotado.

Figura 47 – Etapa de definição dos limites da ROI.



Fonte: Produção do próprio autor.

Definidos os valores tanto para as linhas quanto para as colunas, a ROI é centralizada no pixel de valor médio entre os limites das linhas e das colunas. Vale ressaltar que o tamanho da janela quadrada utilizada para a ROI foi de 128x128 pixels. A figura 48 ilustra essa etapa, onde (a) fornece a marcação da ROI e em (b) a ROI em destaque.

Figura 48 – Marcação da ROI para o tecido mamário.

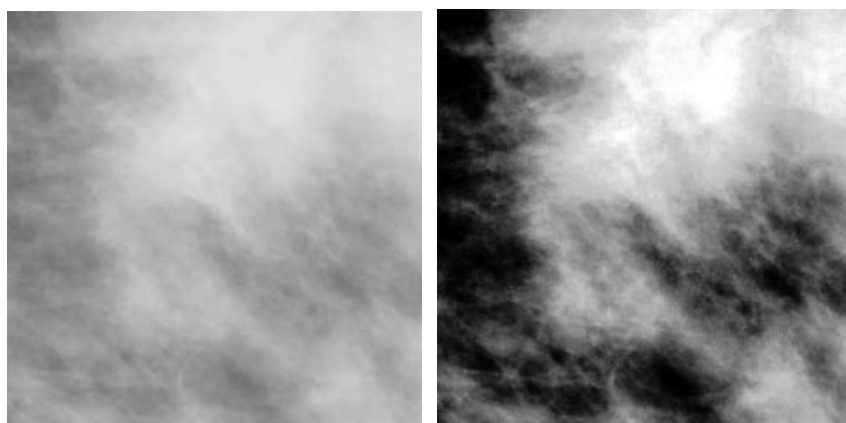


Fonte: Produção do próprio autor.

É importante ressaltar que, dependendo do tamanho da mamografia, a ROI acaba por selecionar fundo da imagem. Isso é prejudicial para a próxima etapa de realce do contraste pela equalização do histograma. Para evitar essa condição, faz-se uma contagem do número de pixels iguais a 0 ao redor da ROI selecionada. Caso o número exceda 5, o raio da ROI é decrementado de 1 pixel e uma nova contagem é feita, até que a contagem de zeros não exceda o valor estipulado. Para a imagem em destaque, não se precisou fazer a redução da dimensão da ROI.

Após a obtenção da ROI sem a presença de fundo da imagem, realiza-se um realce de contraste da mesma através da técnica de equalização do histograma. As figuras 49 e 50 mostram, respectivamente, o realce e o histograma equalizado para a ROI.

Figura 49 – ROIs (tecido mamário) com e sem realce.

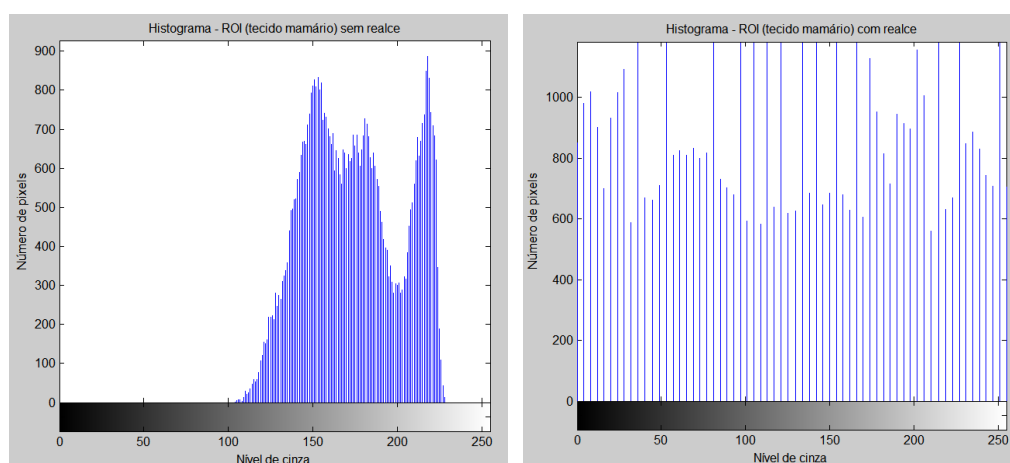


a) Sem realce.

b) Com realce.

Fonte: Produção do próprio autor.

Figura 50 – Histogramas para ROIs (tecido mamário) com e sem realce.



a) Histograma sem realce.

b) Histograma com realce.

Fonte: Produção do próprio autor.

Obtida a ROI realçada, o próximo passo é redimensioná-la para 128x128 pixels. Esse redimensionamento foi inspirado nos testes realizados em (TAYLOR et al., 1994). Vale lembrar que as ROIs que inicialmente selecionaram o fundo da imagem foram diminuídas através do decremento de seu tamanho, daí a necessidade de redimensioná-las de forma a ficarem com as mesmas dimensões.

Obtida a ROI dimensionada e realçada, calculam-se as medidas de textura usando as Matrizes de Coocorrência de Nível de Cinza para todos os ângulos de direção (0° , 45° , 90° e 135°) e distância $d = 1$ entre cada pixel para todas elas.

Finalmente, após a obtenção das medidas, utilizam-se os classificadores kNN e SVM para realizarem a classificação do tecido mamário.

3.3 Classificação do Algoritmo

Para realizar as tarefas de classificação neste trabalho foram usados os classificadores kNN e SVM, descritos na seção 2.6.1.

Ambas as técnicas precisam de dados de treinamento e ajustes de parâmetros antes de realizar a classificação de um conjunto de teste. Para isso, foi usado um método de validação cruzada chamada de *hold-out* (REFAEILZADEH, P. TANG, L. LIU, H., s.d). Neste método a base de dados é dividida em duas partes sem sobreposição de amostras: uma para treinamento e outra para teste, sendo que os dados de teste não são usados para o treinamento dos algoritmos, que corresponde à etapa em que os classificadores se adaptam para realizar uma tarefa, o mais correto possível, de acordo com uma métrica definida. As não sobreposições das partes permitem estimar de forma mais precisa o desempenho real do classificador (REFAEILZADEH, P. TANG, L. LIU, H., s.d).

Nos experimentos realizados, a parte de treinamento corresponde a 75% da base de dados, enquanto a parte de teste são os restantes 25%. Como ambos kNN e SVM têm parâmetros para ajustar, a base de dados de treinamento foi dividida ao meio para treinar o classificador e avaliar diferentes valores de vizinhos mais próximos k (1, 3, 5, 7, 9, 11 e 13, para o caso do kNN) e custo c (de 0,5 a 10, com acréscimos de 0,5, para o caso do SVM). O valor de k e de c que obter o maior valor de média harmônica entre sensibilidade e especificidade sobre a outra metade dos dados de treinamento é usado para classificar os dados de teste. Neste caso, os classificadores são treinados com todo o conjunto de treinamento. Antes de treinar os classificadores, de cada medida foi subtraída a média e o resultado foi dividido pelo desvio padrão da medida. Isso foi feito porque a ordem de grande parte dos valores varia muito.

Como a seleção das amostras é feita de forma aleatória, os resultados obtidos em diferentes execuções podem variar muito devido às amostras selecionadas para cada parte. Para reduzir esse efeito, dois procedimentos a mais foram realizados para tentar obter resultados mais confiáveis. O primeiro é a estratificação dos dados, que permite que cada parte da base de dados tenha uma proporção entre as classes similar à base de dados inteira (REFAEILZADEH, P. TANG, L. LIU, H., s.d). O segundo procedimento é realizar o procedimento de *hold-out* 10 vezes e obter a média dos resultados obtidos.

4 EXPERIMENTOS E RESULTADOS

4.1 Detecção de câncer

No primeiro experimento foram usadas 318 imagens com o objetivo de identificar se a mama tem alguma anormalidade ou se está normal. Isso é feito extraindo medidas de textura da matriz de coocorrência das ROI marcadas na base de dados, quando há anomalia, e ROI marcadas manualmente, quando a mama está normal. Neste primeiro experimento é avaliado cada característica individualmente, sendo que elas foram obtidas nas quatro orientações da matriz de coocorrência. Busca-se identificar também quais medidas são as mais relevantes.

A tabela 3 mostra os resultados de média μ das medidas de acurácia, sensibilidade, especificidade e média harmônica quando utilizados os classificadores kNN e SVM, e individualmente cada uma das medidas de textura. O melhor resultado em cada métrica está realçado em negrito. A medida de sensibilidade mede a proporção de vezes que a anormalidade foi corretamente identificada, enquanto especificidade mede a proporção de pessoas que receberam corretamente o diagnóstico de estarem saudáveis.

Tabela 3 – Média para as métricas no teste de detecção de câncer usando kNN e SVM.

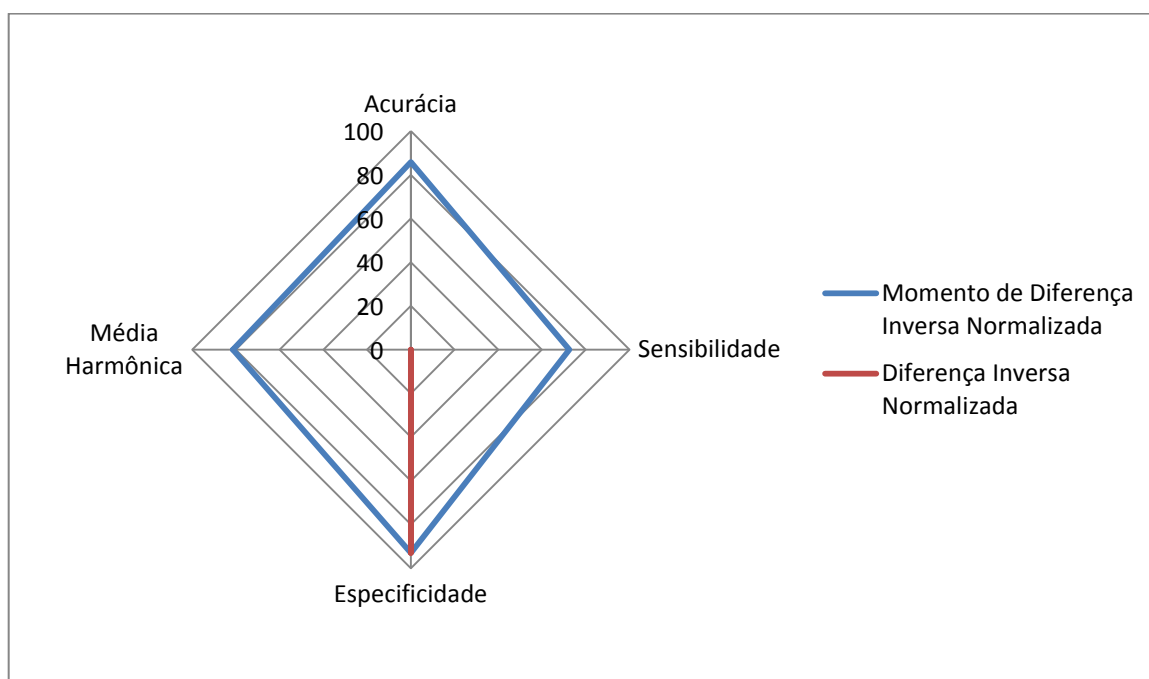
Medida de Textura	Média μ (%)							
	kNN				SVM			
	A	S	E	MH	A	S	E	MH
Segundo Momento Angular	72,5	52,5	83,3	64,1	66,0	3,6	99,4	6,3
Contraste	80,3	64,2	88,9	74,1	84,3	85,9	83,5	84,6
Correlação	80,8	66,2	88,5	75,5	84,3	85,5	83,7	84,5
Entropia	81,8	68,2	89,1	76,8	87,5	75,8	93,8	83,7
Entropia da Soma	62,2	41,6	73,2	52,6	66,2	4,3	99,4	7,7
Entropia da Diferença	81,1	65,2	89,7	75,2	89,8	76,1	97,1	85,2
Variância	70,4	50,4	81,2	61,8	80,2	57,7	92,4	70,7
Variância da Soma	74,2	58,4	82,7	68,1	84,4	62,7	96,1	75,8
Variância da Diferença	78,0	63,7	85,6	72,8	82,3	83,7	81,5	82,5
Momento de Diferença Inversa Normalizada	85,8	72,5	93,0	81,4	87,3	75,0	94,0	83,3
Diferença Inversa Normalizada	84,3	68,1	93,0	78,4	88,2	75,0	95,3	83,8
Média da Soma	74,2	56,9	83,5	66,8	80,5	46,4	98,8	62,8
Média da Diferença	80,8	65,2	89,1	75,1	91,5	85,1	94,9	89,6
Medida de Informação da Correlação 1	81,3	65,2	89,9	75,3	88,1	76,8	94,2	84,5
Medida de Informação da Correlação 2	80,5	63,8	89,5	74,1	87,6	87,0	88,0	87,3
Coefficiente de Correlação Máximo	69,4	56,4	76,2	63,5	69,1	19,1	95,9	28,3
Autocorrelação	76,3	62,7	83,7	71,4	85,4	67,8	94,9	79,0
Dissimilaridade	80,8	65,2	89,1	75,1	91,5	85,1	94,9	89,6
Sombra de Agrupamento	65,1	40,9	78,0	52,9	66,7	11,2	96,5	17,5
Proeminência de Agrupamento	74,2	63,0	80,2	69,9	73,5	45,8	88,6	56,9
Máxima Probabilidade	61,4	39,2	73,4	50,1	66,2	3,3	100,0	6,1
Inércia	80,3	64,2	88,9	74,1	84,3	85,9	83,5	84,6

Fonte: Produção do próprio autor.

Podemos concluir da tabela que o referido teste obteve boas médias de acerto, principalmente para algumas medidas de textura em particular: Dentre as melhores medidas de textura para cada métrica na detecção de câncer para o classificador kNN podemos citar: o Momento de Diferença Inversa Normalizada (85,8% de Acurácia, 72,5% de Sensibilidade,

93,0% de Especificidade e 81,4% de Média Harmônica) e a Diferença Inversa Normalizada (93,0% de Especificidade). Ambas as medidas estão relacionadas com a homogeneidade local das imagens, ou seja, quando os valores dos pares de pixels estão próximos na escala de cinza. As medidas de textura mais eficientes em cada métrica estão representadas na figura 51.

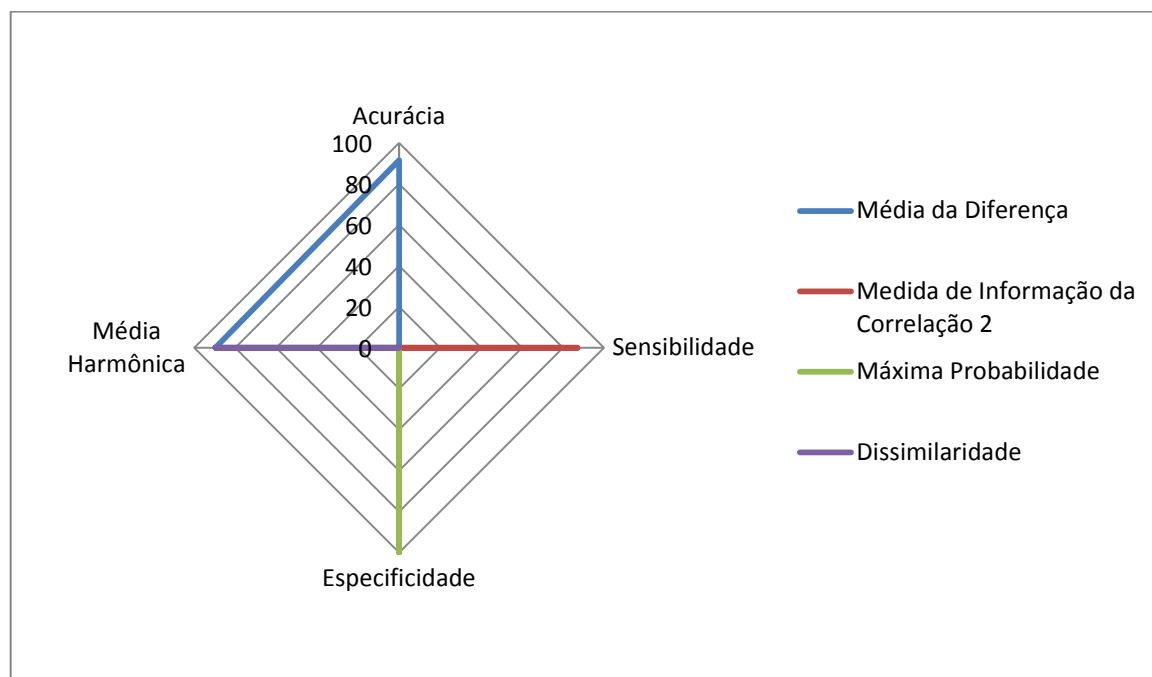
Figura 51 – Gráfico em radar para as melhores texturas no teste de detecção de câncer usando kNN.



Fonte: Produção do próprio autor.

Dentre as melhores medidas de textura para cada métrica na detecção de câncer para o classificador SVM, pode-se citar: a Média da Diferença (91,5% de Acurácia e 89,6% de Média Harmônica), a Medida de Informação da Correlação 2 (87% de Sensibilidade), a Máxima Probabilidade (100% de Especificidade) e a Dissimilaridade (89,6% de Média Harmônica). Para as medidas com maior valor de média harmônica, temos que a Média da Diferença mede os valores das diagonais da imagem determinadas por $|i - j| = k$, conforme equação (24), ao passo que a Dissimilaridade é uma medida das variações locais apresentadas em uma imagem de forma a favorecer as contribuições da GLCM para longe da diagonal principal, mas ao longo dos pixels cuja relação de coordenadas é dada por $|i - j| = n$, conforme equação (4). As medidas de textura mais eficientes em cada métrica estão representadas na figura 52.

Figura 52 – Gráfico em radar para as melhores texturas no teste de detecção de câncer usando SVM.

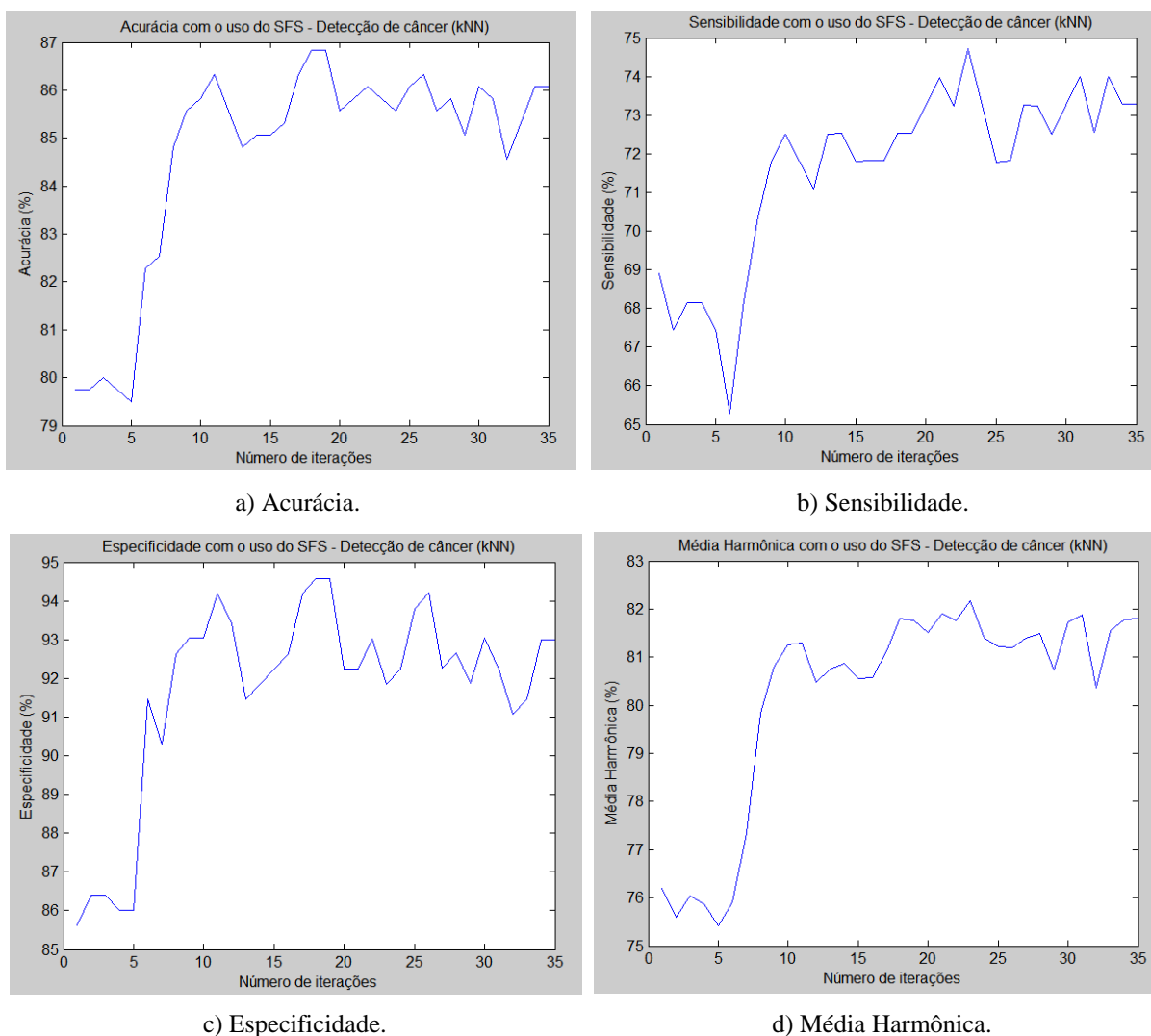


Fonte: Produção do próprio autor.

Os melhores resultados das métricas obtidos com o classificador SVM são todos superiores aqueles obtidos pelo classificador kNN.

O algoritmo SFS foi utilizado de forma a selecionar os melhores atributos (medidas). Os atributos avaliados foram as 22 medidas nas 4 orientações, totalizando 88 medidas. Foram realizadas 35 iterações, equivalente a selecionar 35 medidas, e suas métricas de acurácia, sensibilidade, especificidade e média harmônica foram medidas a cada nova iteração. Por motivos de custo computacional, foi usado somente o kNN e as métricas foram a média de 5 rodadas de *hold-out*. A figura 53 mostra os gráficos do SFS para as quatro métricas utilizando-se o classificador kNN.

Figura 53 – Gráficos do SFS para as métricas na detecção de câncer usando kNN.



Fonte: Produção do próprio autor.

Pode-se concluir, da análise dos gráficos do SFS, que a partir de 5 iterações, os testes passam a ter aumentos significativos na métrica considerada, e com um número ainda maior de iterações os valores das métricas oscilam em torno de um valor.

A acurácia teve seu maior valor de quase 87% em aproximadamente 17 iterações, mantendo-se em uma média de 85,5% ao longo das iterações posteriores. A sensibilidade atingiu seu maior valor de quase 75% em aproximadamente 24 iterações, ao passo que manteve uma média de 73,5% para as iterações posteriores. A especificidade, para cerca de 18 iterações, teve o maior valor atingido de 94,5%. Por fim, a média harmônica atingiu um pico de mais de 82% para cerca de 23 iterações. Observa-se, portanto, que o simples fato de adicionar novos atributos não garante que o desempenho dos classificadores irá aumentar. Além disso, o maior aumento das métricas ocorreu entre 5 e 10 iterações, e o ganho nas

métricas não foi muito alto com a adição dos atributos a partir da 11 iteração. Em outras palavras, os gráficos estão mostrando que somente um conjunto restrito de atributos é relevante para a tarefa.

Outro teste realizado foi feito com todas as medidas de texturas calculadas em cada direção da GLCM. Fez-se um comparativo das métricas entre cada ângulo utilizando ambos os classificadores kNN e SVM. A tabela 4 mostra o resultado atingido.

Tabela 4 – Média para as métricas em cada ângulo no teste de detecção de câncer usando kNN e SVM.

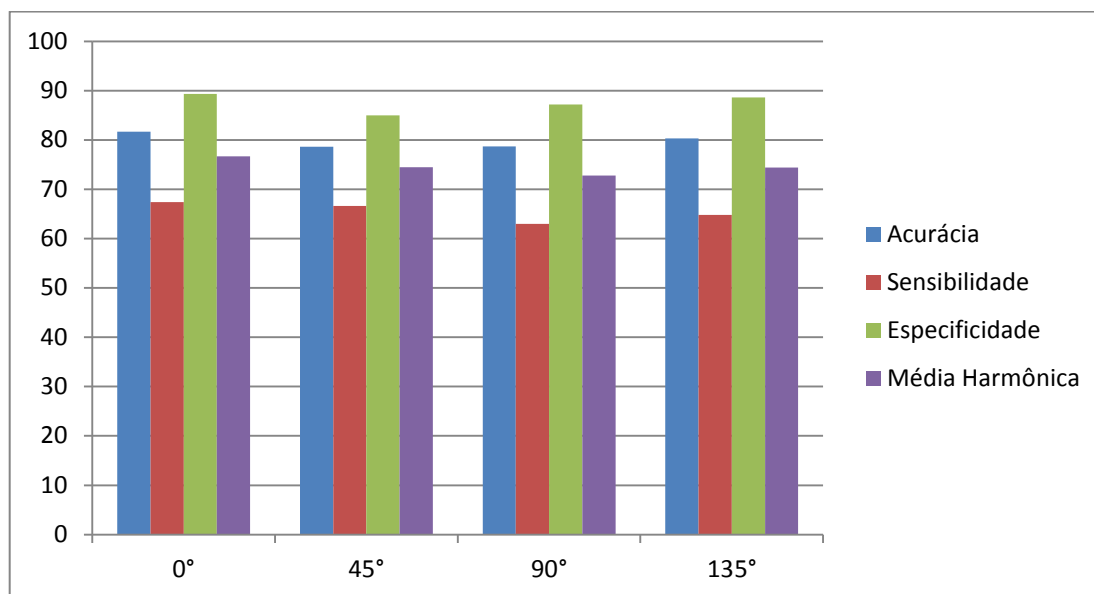
Ângulo Matriz de Coocorrência	Média μ (%)							
	kNN				SVM			
	A	S	E	MH	A	S	E	MH
0°	81,7	67,4	89,3	76,7	90,1	80,1	95,5	87,1
45°	78,6	66,6	85,0	74,5	84,4	76,4	88,8	81,8
90°	78,7	63,0	87,2	72,8	88,4	79,6	93,0	85,6
135°	80,3	64,8	88,6	74,4	83,8	72,1	90,1	79,7

Fonte: Produção do próprio autor.

Podemos concluir que para ambos os classificadores kNN e SVM, a direção de 0° ao longo da GLCM é a que melhor obtém resultados de acurácia, sensibilidade, especificidade e média harmônica. Vale ressaltar que para todas as métricas, o SVM obteve melhores resultados do que o kNN. ALCÂNTARA, JÚNIOR e RAMOS (2016) chegaram à mesma conclusão em seu trabalho de detecção de câncer, ao obterem os melhores valores de acurácia (84,4%), sensibilidade (85,4%) e especificidade (85,4%) para a direção de 0°.

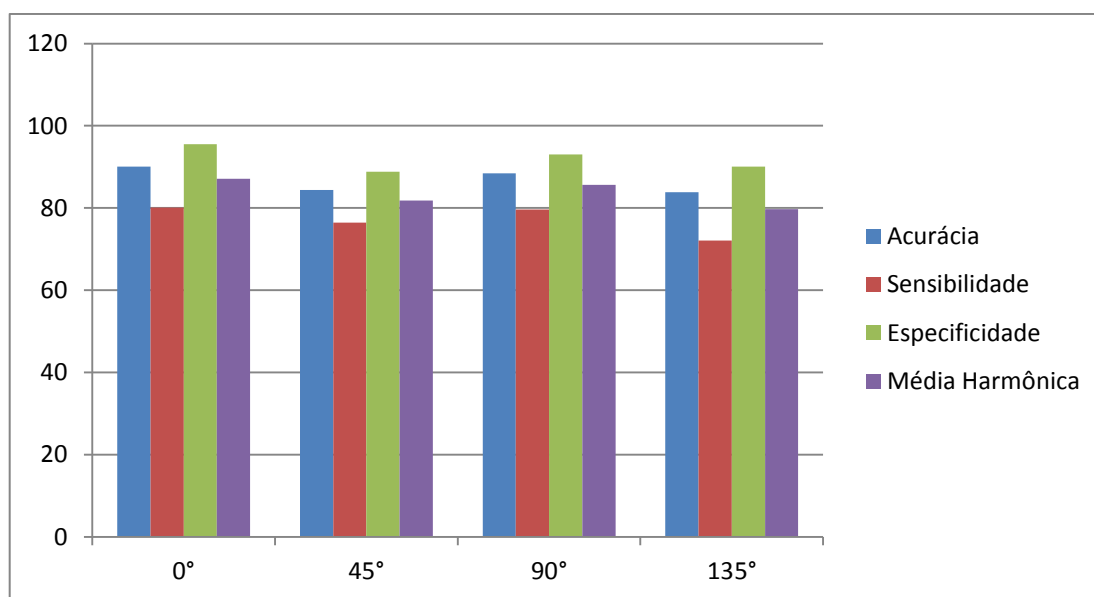
As figuras 54 e 55 mostram um comparativo das métricas entre cada uma das direções ao longo da GLCM tanto para o classificador kNN quanto para o classificador SVM.

Figura 54 – Gráfico em barra para diferentes ângulos da GLCM no teste de detecção de câncer usando kNN.



Fonte: Produção do próprio autor.

Figura 55 – Gráfico em barra para diferentes ângulos da GLCM no teste de detecção de câncer usando SVM.



Fonte: Produção do próprio autor.

4.2 Classificação de câncer

No segundo experimento foram usadas 109 imagens com câncer divididas em 61 do tipo benigno e 48 do tipo maligno. O objetivo é de classificar o câncer através das ROIs selecionadas através do banco de dados MIAS e das medidas de textura extraídas. Neste segundo experimento é avaliado cada característica individualmente, sendo que elas foram obtidas nas quatro orientações da matriz de coocorrência. Busca-se identificar também quais medidas são as mais relevantes.

A tabela 5 mostra os resultados de média μ das medidas de acurácia, sensibilidade, especificidade e média harmônica quando utilizados os classificadores kNN e SVM, e individualmente cada uma das medidas de textura. O melhor resultado em cada métrica está realçado em negrito. A medida de sensibilidade mede a proporção de vezes foi corretamente identificada com câncer maligno, enquanto especificidade mede a proporção de pessoas que receberam corretamente o diagnóstico de estarem com câncer benigno.

Tabela 5 – Média para as métricas no teste de classificação de câncer usando kNN e SVM.

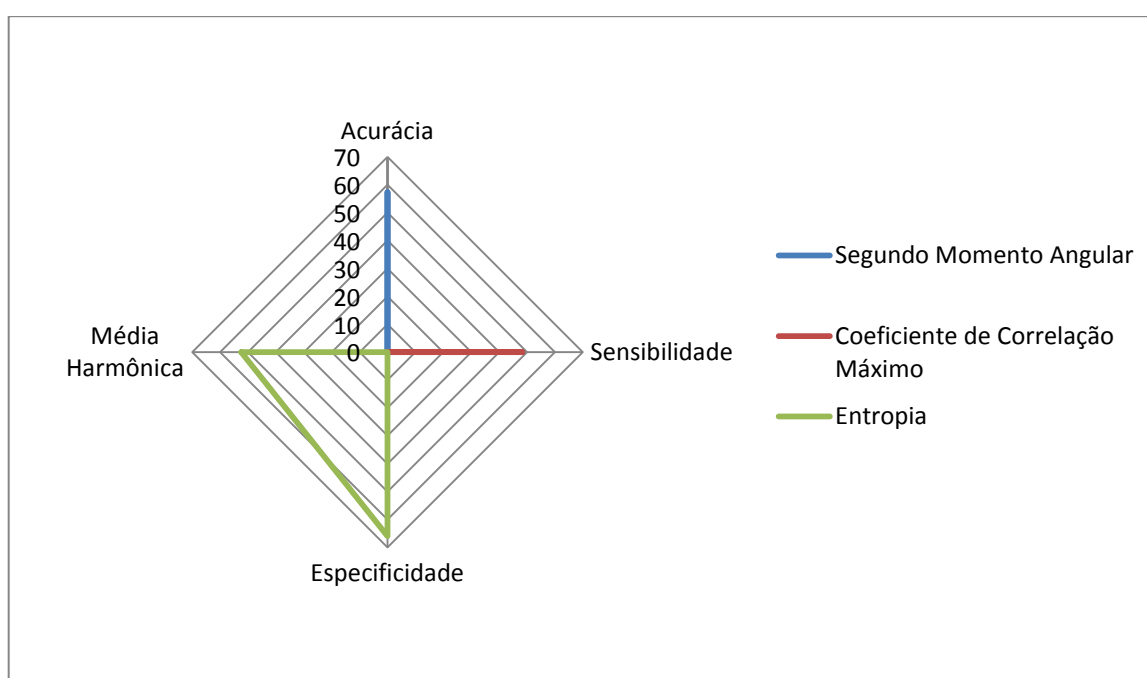
Medida de Textura	Média μ (%)							
	kNN				SVM			
	A	S	E	MH	A	S	E	MH
Segundo Momento Angular	57,4	47,5	65,33	51,5	55,2	0,8	98,7	1,5
Contraste	46,7	35,8	55,3	41,6	54,1	10,0	89,3	17,1
Correlação	44,1	38,3	48,7	41,6	54,8	13,3	88,0	22,3
Entropia	57,0	45,8	66,0	52,4	55,6	16,7	86,7	22,6
Entropia da Soma	48,2	45,0	50,7	45,6	55,6	1,7	98,7	3,1
Entropia da Diferença	38,9	27,5	48,0	32,8	58,9	49,2	66,7	52,8
Variância	52,6	43,3	60,0	47,2	55,6	0,8	99,3	1,5
Variância da Soma	43,0	36,7	48,0	39,9	55,6	0,8	99,3	1,5
Variância da Diferença	45,6	36,7	52,7	40,8	52,6	5,8	90,0	10,0
Momento de Diferença Inversa Normalizada	51,5	40,0	60,7	45,3	57,4	26,7	82,0	32,4
Diferença Inversa Normalizada	49,3	31,7	63,3	40,8	57,8	34,2	76,7	37,9
Média da Soma	45,9	35,8	54,0	38,7	55,9	1,7	99,3	3,1
Média da Diferença	41,5	29,2	51,3	35,2	59,6	37,5	77,3	48,7
Medida de Informação da Correlação 1	51,1	38,3	61,3	45,8	56,7	50,8	66,7	56,0
Medida de Informação da Correlação 2	48,9	39,2	56,7	44,9	52,2	27,5	80,0	37,4
Coefficiente de Correlação Máximo	48,9	48,3	49,3	44,0	55,6	7,5	88,0	10,1
Autocorrelação	53,7	46,7	59,3	50,7	59,6	0,8	99,3	1,5
Dissimilaridade	41,5	29,2	51,3	35,2	55,6	37,5	77,3	48,7
Sombra de Agrupamento	47,0	30,0	60,7	37,0	54,4	1,7	98,7	3,1
Proeminência de Agrupamento	46,7	31,7	58,7	38,5	55,6	1,7	96,7	3,1
Máxima Probabilidade	43,3	38,3	47,3	40,2	59,6	1,7	98,7	3,1
Inércia	46,7	35,8	55,3	41,6	54,1	10,0	89,3	17,1

Fonte: Produção do próprio autor.

Da análise dos resultados obtidos no teste de classificação de câncer, podemos concluir que houve uma taxa de acerto bem abaixo do esperado para todas as métricas das medidas de texturas individuais.

Pode-se concluir da tabela que o referido teste obteve médias de acertos de medianas, principalmente para algumas medidas de textura em particular. Dentre as melhores medidas de textura para cada métrica para o classificador kNN, podem ser citados: o Segundo Momento Angular (57,4% de Acurácia), o Coeficiente de Correlação Máximo (48,3% de Sensibilidade) e a Entropia (66% de Especificidade e 52,4% de Média Harmônica). A Entropia, com o maior valor de média harmônica, mede a aleatoriedade da textura da imagem. As medidas de textura mais eficientes em cada métrica estão representadas na figura 56.

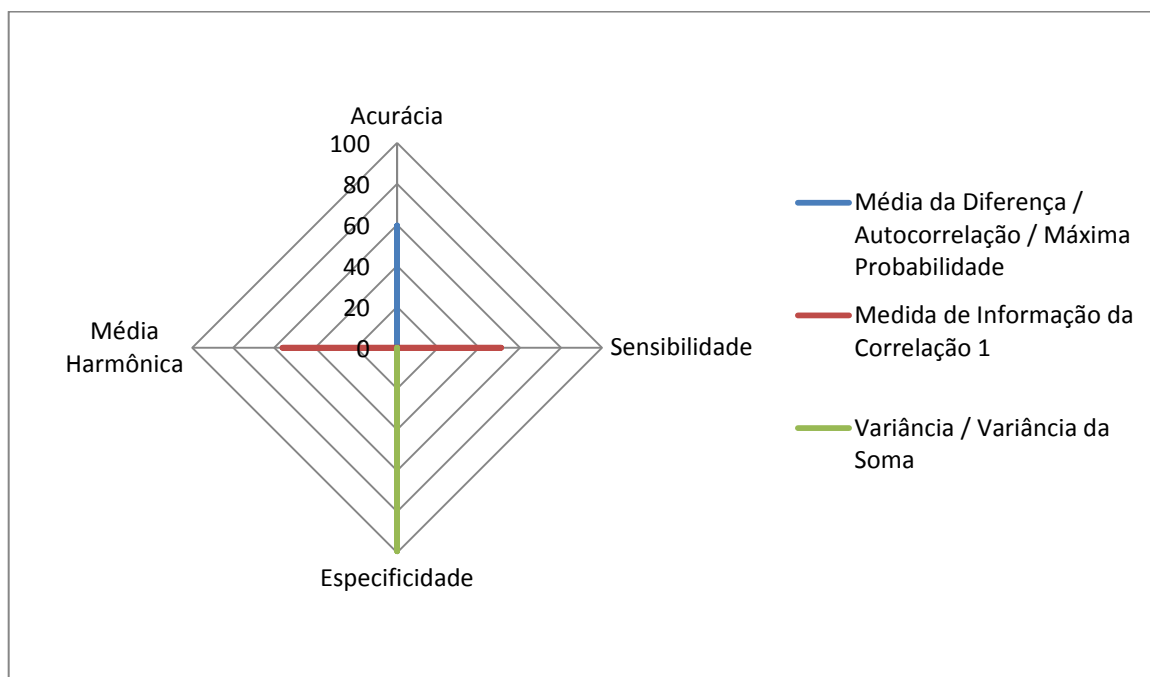
Figura 56 – Gráfico em radar para as melhores texturas no teste de classificação de câncer usando kNN.



Fonte: Produção do próprio autor.

Dentre as melhores medidas de textura para o classificador SVM, podem ser citados: a Média da Diferença (59,6% de Acurácia), a Autocorrelação (59,6% de Acurácia), a Máxima Probabilidade (59,6% de Acurácia), a Medida de Informação da Correlação 1 (50,8% de Sensibilidade e 56,0% de Média Harmônica), a Variância (99,3% de Especificidade) e a Variância da Soma (99,3% de Especificidade). A Medida de Informação da Correlação 1, com maior valor da média harmônica, é uma medida da correlação da imagem, ou seja, expressa a medida da dependência linear dos níveis de cinza sobre os pixels vizinhos. As medidas de textura mais eficientes em cada métrica estão representadas na figura 57.

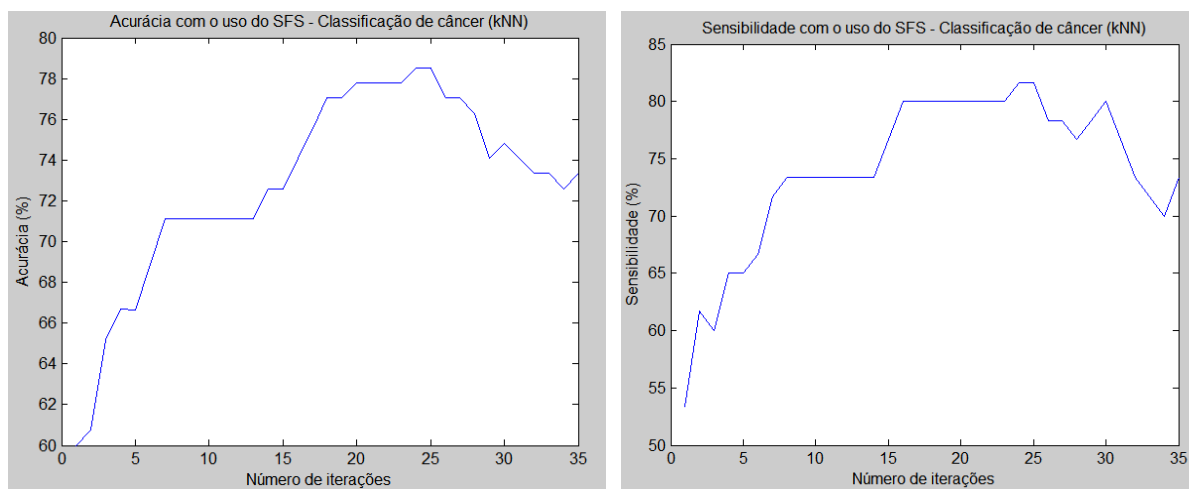
Figura 57 – Gráfico em radar para as melhores texturas no teste de classificação de câncer usando SVM.



Fonte: Produção do próprio autor.

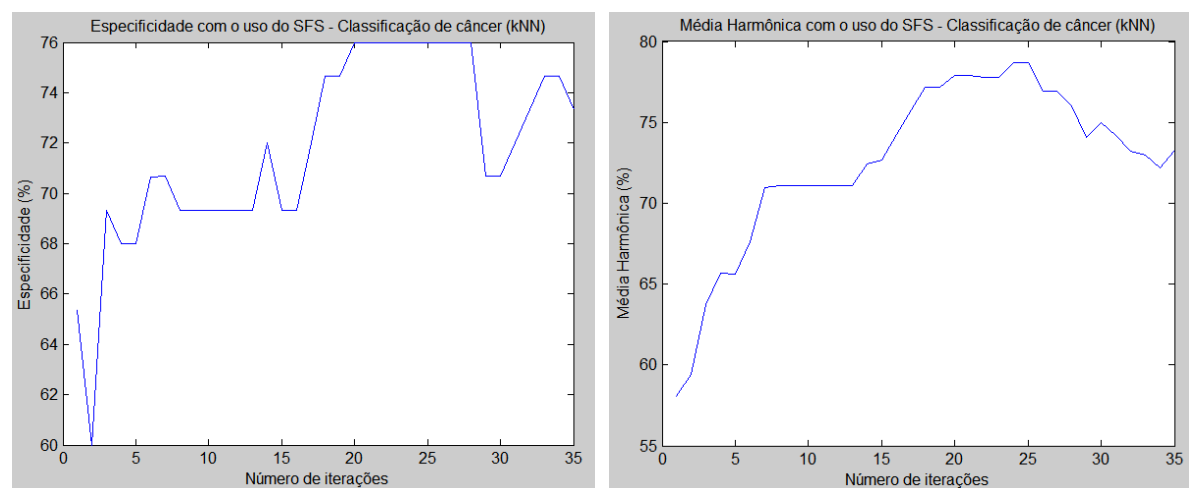
O algoritmo SFS foi utilizado de forma a selecionar os melhores atributos (medidas). Foram realizadas 35 iterações e suas métricas de acurácia, sensibilidade, especificidade e média harmônica foram medidas a cada nova iteração. Por motivos de custo computacional, foi usado somente o kNN e as métricas foram a média de 5 rodadas de *hold-out*. A figura 58 mostra os gráficos do SFS para as quatro métricas utilizando-se o classificador kNN.

Figura 58 – Gráficos do SFS para as métricas na classificação de câncer usando kNN.



a) Acurácia.

b) Sensibilidade.



c) Especificidade.

d) Média Harmônica.

Fonte: Produção do próprio autor.

Pode-se concluir, da análise dos gráficos do SFS, que a partir de 3 iterações, os testes passam a ter aumentos significativos na métrica considerada e que em cerca de 24 iterações, todas as métricas atingem seus maiores valores.

A acurácia teve seu maior valor em cerca de 79%, um valor considerável quando comparado àquele obtido pelo Segundo Momento Angular, melhor medida individual de textura para essa métrica. Para novas iterações, seu valor passa a diminuir, mas ainda mantendo-se acima dos valores das medidas individuais.

A sensibilidade atingiu seu maior valor de quase 82% em aproximadamente 24 iterações, valor acima daquele obtido pelo Coeficiente de Correlação Máximo. Para novas iterações, seu valor passa a diminuir, mas ainda mantendo-se acima dos valores das medidas individuais.

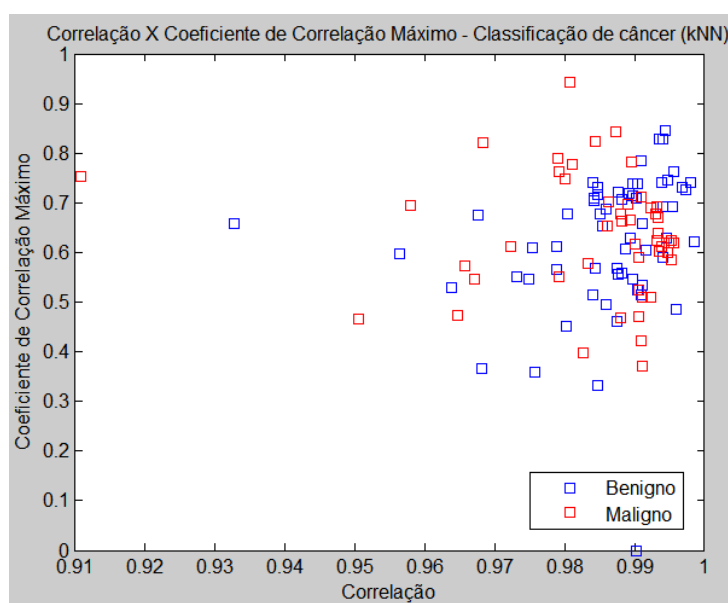
A especificidade, para cerca de 24 iterações, teve o maior valor atingido de 76%. Para iterações maiores apresentou um valor em torno de 73%, valor acima do apresentado pela Entropia, medida individual de textura com o melhor desempenho para essa métrica.

Por fim, a média harmônica obteve um valor máximo em 78%, ao passo que esse valor diminui para iterações maiores do que 24. Seu desempenho foi melhor do que a Entropia usada individualmente.

Observa-se da Figura 58 que a partir de um número de iterações os valores das métricas começam a cair. Novamente é notado o fato de que adicionar novos atributos não garante que o desempenho dos classificadores irá aumentar e, neste caso, é observado que as métricas até caíram de valor a partir de um número de iterações. Em outras palavras, os gráficos estão mostrando que somente um conjunto restrito de atributos é relevante para a tarefa.

Comparado com os experimentos anteriores para detecção, os valores das métricas de identificação foram abaixo. Para investigar esse problema, foi feito o gráfico da relação entre as duas primeiras medidas de textura consideradas pelo algoritmo SFS: a Correlação e o Coeficiente de Correlação Máximo, ambos na direção de 135°. A figura 59 ilustra o gráfico, onde cada quadrado corresponde a uma imagem representada pelas medidas de Correlação (eixo horizontal) e Coeficiente de Correlação Máximo (eixo vertical). Os quadrados azuis representam as imagens com câncer benigno e as vermelhas com câncer maligno.

Figura 59 – Relação entre as duas primeiras medidas de textura no teste de classificação de câncer usando SFS e kNN.



Fonte: Produção do próprio autor.

Na figura 59 é observado que os quadrados de ambas as classes estão muito sobrepostos, e que é muito difícil para um algoritmo como SVM conseguir traçar uma reta que separe bem as duas classes nesta situação. Embora essa análise tenha sido feita somente com duas medidas, ela pode ser uma possível explicação para os resultados não serem tão bons quanto nos experimentos de detecção de câncer.

Outro teste realizado foi feito com todas as medidas de texturas calculadas em cada direção da GLCM. Fez-se um comparativo das métricas entre cada ângulo utilizando ambos os classificadores kNN e SVM. A tabela 6 mostra o resultado atingido.

Tabela 6 – Média para as métricas em cada ângulo no teste de classificação de câncer usando kNN e SVM.

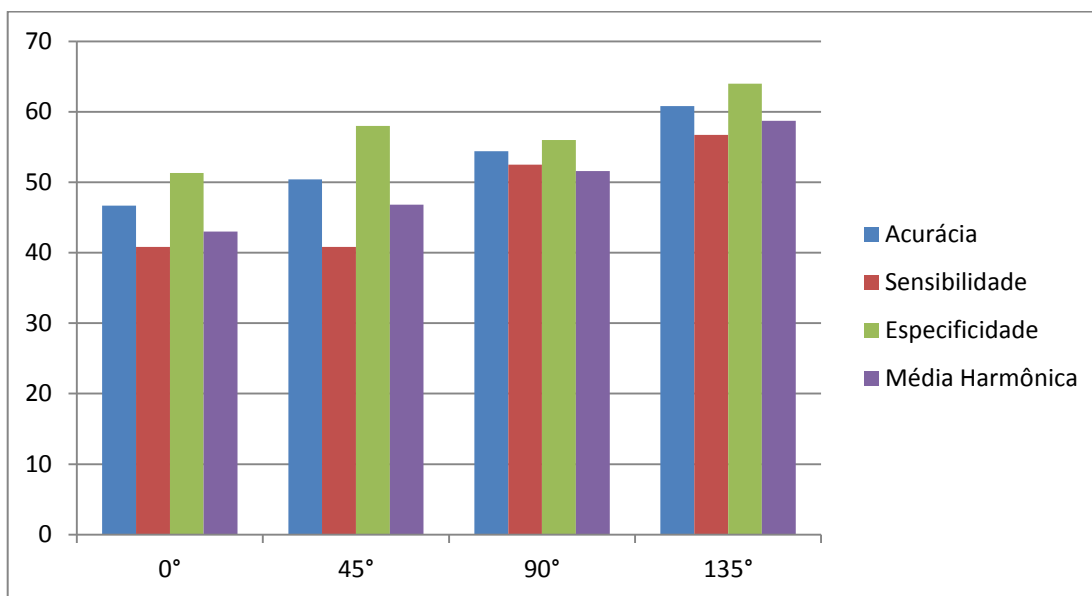
Ângulo Matriz de Coocorrência	Média μ (%)							
	kNN				SVM			
	A	S	E	MH	A	S	E	MH
0°	46,7	40,8	51,3	43,0	51,1	18,3	77,3	25,5
45°	50,4	40,8	58,0	46,8	53,3	26,7	74,7	33,3
90°	54,4	52,5	56,0	51,6	51,9	18,3	78,7	26,3
135°	60,8	56,7	64,0	58,7	53,3	6,7	90,7	11,7

Fonte: Produção do próprio autor.

Pode-se concluir que para o classificador kNN, a direção de 135° ao longo da GLCM é a que melhor obtém resultados de acurácia, sensibilidade, especificidade e média harmônica. No entanto, para o classificador SVM, a direção de 45° se destacou melhor, com exceção da especificidade que teve melhor resultado na direção de 135°. Vale ressaltar que para todas as métricas, à exceção de especificidade, o kNN obteve melhores resultados do que o SVM.

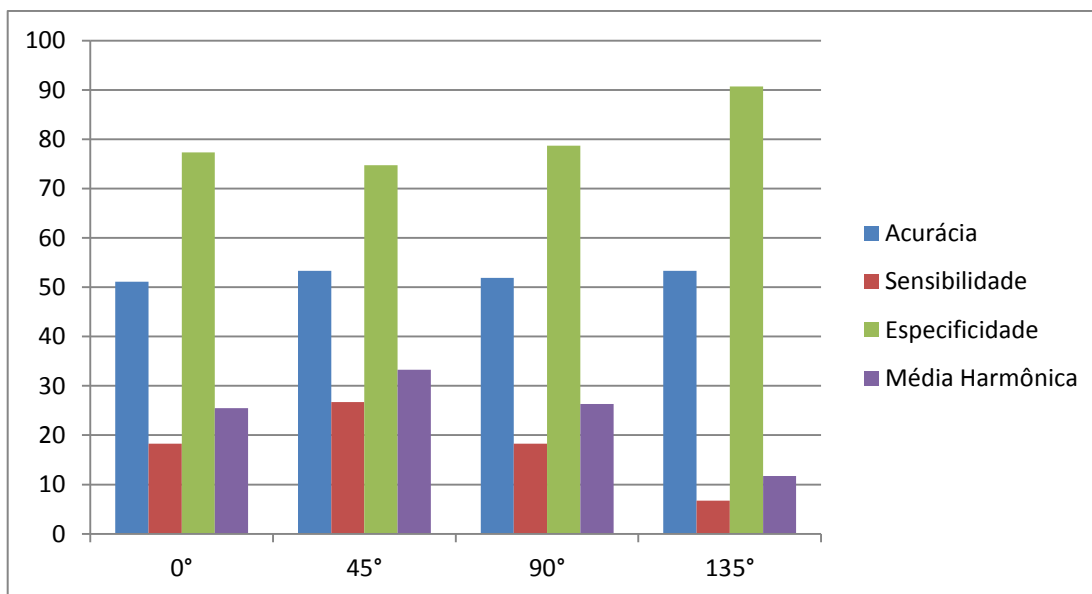
As figuras 60 e 61 mostram um comparativo das métricas entre cada uma das direções ao longo da GLCM tanto para o classificador kNN quanto para o classificador SVM.

Figura 60 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação de câncer usando kNN.



Fonte: Produção do próprio autor.

Figura 61 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação de câncer usando SVM.



Fonte: Produção do próprio autor.

4.3 Classificação do tecido mamário

No terceiro experimento foram usadas 322 imagens divididas em 106 adiposas e 216 glandulares (glandular adiposa e glandular densa). O objetivo é de classificar o tecido mamário através das ROI selecionadas automaticamente e das medidas de textura extraídas. Neste terceiro experimento é avaliado cada característica individualmente, sendo que elas foram obtidas nas quatro orientações da matriz de coocorrência. Busca-se identificar também quais medidas são as mais relevantes.

A tabela 7 mostra os resultados de média μ das medidas de acurácia, sensibilidade, especificidade e média harmônica quando utilizados os classificadores kNN e SVM, e individualmente cada uma das medidas de textura. O melhor resultado em cada métrica está realçado em negrito. A medida de sensibilidade mede a proporção de vezes que o tecido glandular foi corretamente identificado, enquanto especificidade mede a proporção de pessoas que receberam corretamente o diagnóstico de terem o tecido adiposo.

Tabela 7 – Média para as métricas no teste de classificação do tecido mamário usando kNN e SVM.

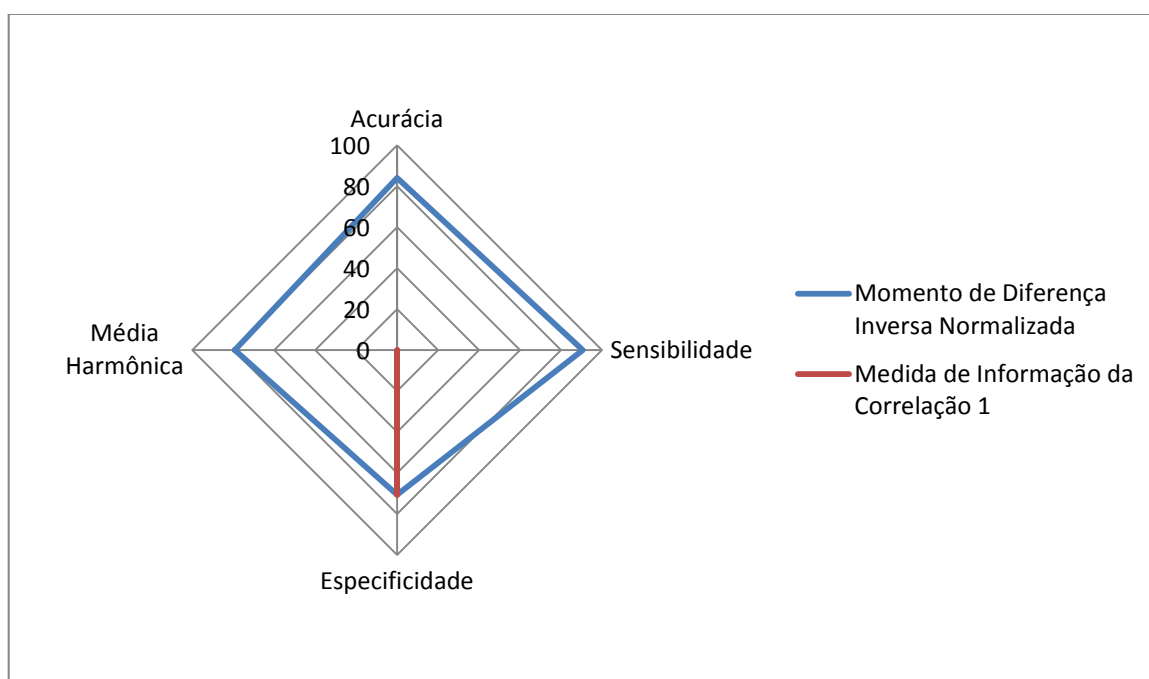
Medida de Textura	Média μ (%)							
	kNN				SVM			
	A	S	E	MH	A	S	E	MH
Segundo Momento Angular	77,1	84,1	63,1	71,2	80,8	87,3	67,6	75,7
Contraste	81,5	88,6	67,2	76,2	84,3	95,7	61,3	74,6
Correlação	81,1	88,6	66,1	75,2	84,1	95,7	60,9	74,3
Entropia	82,9	89,5	69,5	78,0	83,8	92,3	66,5	77,0
Entropia da Soma	69,5	81,9	44,7	57,0	74,0	88,8	44,3	58,6
Entropia da Diferença	81,4	88,6	66,9	75,9	84,1	93,3	65,8	77,0
Variância	63,9	73,0	45,5	55,8	71,5	93,1	28,2	42,8
Variância da Soma	65,4	79,7	36,4	49,4	71,8	93,3	28,5	42,9
Variância da Diferença	80,1	87,1	66,1	74,9	83,9	94,8	62,0	74,7
Momento de Diferença Inversa Normalizada	84,0	90,6	70,6	79,2	84,0	91,6	68,7	78,2
Diferença Inversa Normalizada	82,6	89,1	69,5	78,0	83,8	93,3	64,6	76,0
Média da Soma	61,6	71,7	41,4	51,7	68,3	94,8	15,0	24,2
Média da Diferença	81,6	88,8	67,3	76,2	84,4	94,9	63,1	75,6
Medida de Informação da Correlação 1	83,3	89,7	70,6	78,6	84,0	92,7	66,5	77,1
Medida de Informação da Correlação 2	82,3	89,1	68,4	77,1	84,4	94,8	63,5	75,8
Coefficiente de Correlação Máximo	71,6	85,0	44,7	58,1	73,9	88,2	45,1	59,0
Autocorrelação	74,5	83,3	56,7	66,9	80,5	92,5	56,4	69,9
Dissimilaridade	81,6	88,8	67,3	76,2	84,4	94,9	63,1	75,6
Sombra de Agrupamento	73,3	84,7	50,4	62,6	78,9	93,8	48,8	63,7
Proeminência de Agrupamento	79,8	89,1	60,9	71,9	83,6	94,4	62,0	74,5
Máxima Probabilidade	68,1	75,5	53,4	62,3	76,6	89,7	50,3	64,1
Inércia	81,5	88,6	67,2	76,2	84,3	95,7	61,3	74,6

Fonte: Produção do próprio autor.

Pode ser concluído da tabela que o referido teste obteve resultados razoáveis, principalmente para algumas medidas de textura em particular: Dentre as melhores medidas de textura para cada métrica na detecção de câncer para o classificador kNN, podem ser citados: o Momento de Diferença Inversa Normalizada (84,0% de Acurácia, 90,6% de

Sensibilidade, 70,6% de Especificidade e 79,2% de Média Harmônica) e a Medida de Informação da Correlação 1 (70,6% de Especificidade). O Momento de Diferença Inversa Normalizada, com melhor valor de média harmônica, mede a homogeneidade local da imagem, ao passo que a Medida de Informação da Correlação é uma medida da dependência linear dos níveis de cinza sobre os pixels vizinhos. As medidas de textura mais eficientes em cada métrica estão representadas na figura 62.

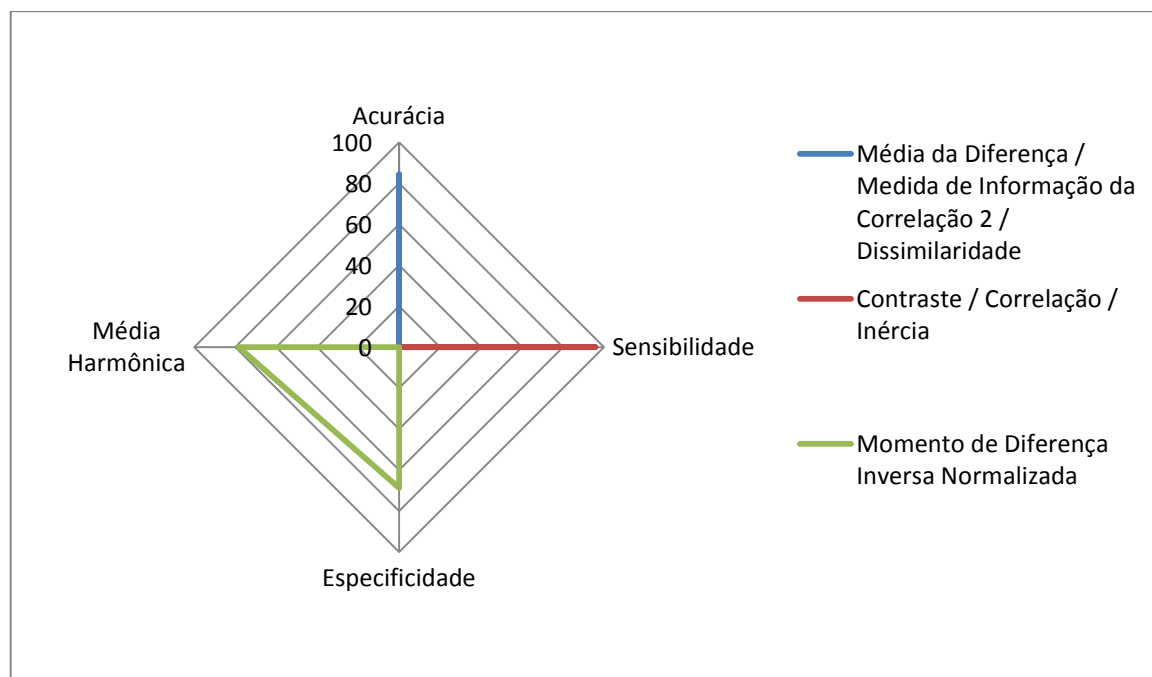
Figura 62 – Gráfico em radar para as melhores texturas no teste de classificação do tecido mamário usando kNN.



Fonte: Produção do próprio autor.

Dentre as melhores medidas de textura para cada métrica na classificação do tecido mamário para o classificador SVM, podem ser citados: a Média da Diferença (84,4% de Acurácia), a Medida de Informação da Correlação 2 (84,4% de Acurácia), a Dissimilaridade (84,4% de Acurácia), o Contraste (95,7% de Sensibilidade), a Correlação (95,7% de Sensibilidade), a Inércia (95,7% de Sensibilidade) e o Momento de Diferença Inversa Normalizada (68,7% de Especificidade e 78,2% de Média Harmônica). Mais uma vez, o Momento de Diferença Inversa Normalizada, com melhor valor de média harmônica, mede a homogeneidade local da imagem. As medidas de textura mais eficientes em cada métrica estão representadas na figura 63.

Figura 63 – Gráfico em radar para as melhores texturas no teste de classificação do tecido mamário usando SVM.

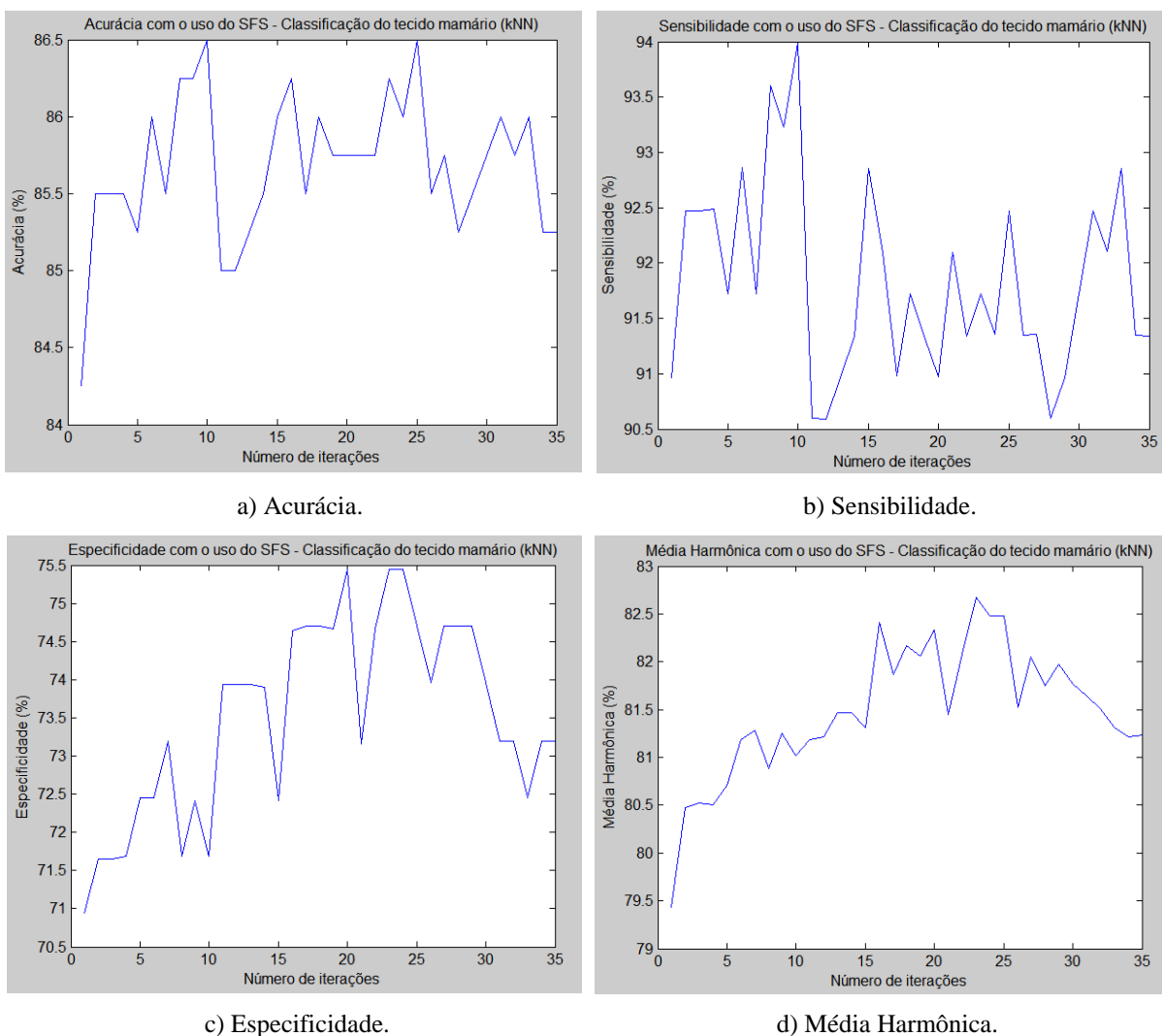


Fonte: Produção do próprio autor.

O algoritmo SFS foi utilizado de forma a selecionar os melhores atributos (medidas). Foram realizadas 35 iterações e suas métricas de acurácia, sensibilidade, especificidade e média harmônica foram medidas a cada nova iteração. Por motivos de custo computacional, foi usado somente o kNN e as métricas foram a média de 5 rodadas de *hold-out*. A figura 64 mostra os gráficos do SFS para as quatro métricas utilizando-se o classificador kNN.

Comparando-se ambos os classificadores kNN e SVM, pode-se verificar que o kNN obteve melhores resultados de especificidade e média harmônica, ao passo que o classificador SVM obteve melhores resultados de acurácia e sensibilidade.

Figura 64 – Gráficos do SFS para as métricas na classificação do tecido mamário usando kNN.



Fonte: Produção do próprio autor.

Pode-se concluir, da análise dos gráficos do SFS, que a partir da primeira iteração, os testes passam a ter aumentos significativos na métrica considerada.

A acurácia teve seu maior valor de 86,5% em dois momentos: com 9 iterações e 24 iterações. Ao longo das iterações sucessivas, a métrica teve seu valor em torno de 85,7%. A sensibilidade atingiu seu maior valor de 94% com 10 iterações. Posteriormente, o valor dessa métrica se mostrou em torno de 91,5%. A especificidade atingiu seu maior valor de 75,5% após cerca de 23 iterações. Por fim, a média harmônica atingiu seu maior valor de 82,7% após cerca de 23 iterações. É possível chegar às mesmas conclusões apresentadas anteriormente para os gráficos do SFS.

Outro teste realizado foi feito com todas as medidas de texturas calculadas em cada direção da GLCM. Fez-se um comparativo das métricas entre cada ângulo utilizando ambos os classificadores kNN e SVM. A tabela 8 mostra o resultado atingido.

Tabela 8 – Média para as métricas em cada ângulo no teste de classificação do tecido mamário usando kNN e SVM.

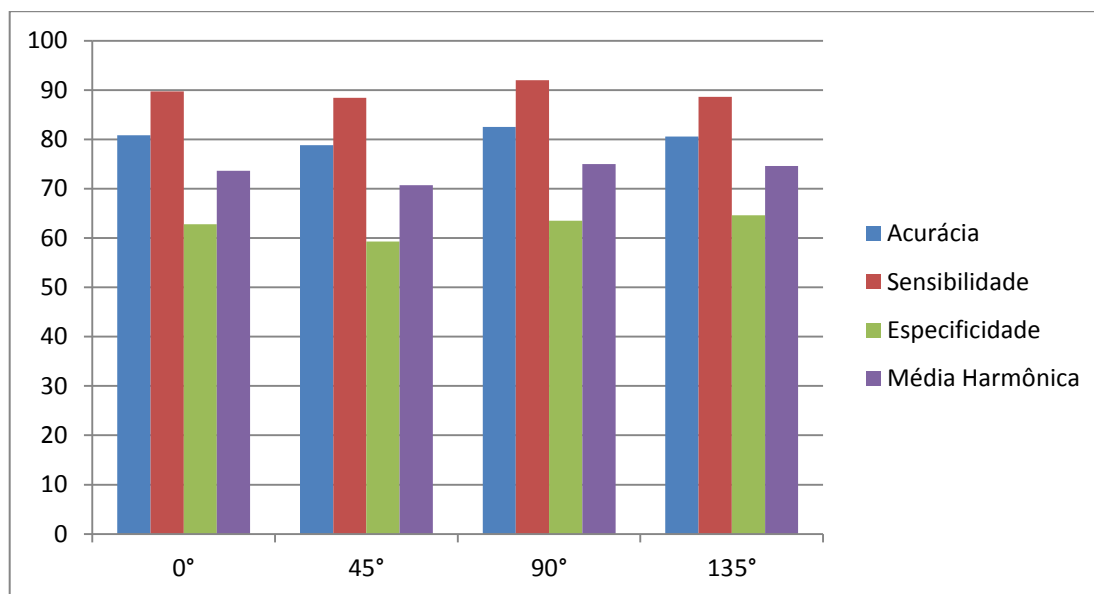
Ângulo Matriz de Coocorrência	Média μ (%)							
	kNN				SVM			
	A	S	E	MH	A	S	E	MH
0°	80,8	89,7	62,8	73,6	82,0	92,5	60,8	73,2
45°	78,8	88,4	59,3	70,7	82,9	93,8	60,9	73,6
90°	82,5	92,0	63,5	75,0	82,0	94,4	57,1	70,9
135°	80,6	88,6	64,6	74,6	80,5	91,9	57,5	70,6

Fonte: Produção do próprio autor.

Podemos concluir que para o classificador kNN, a direção de 90° ao longo da GLCM é a que melhor obtém resultados de acurácia, sensibilidade e média harmônica. No entanto, para o classificador SVM, a direção de 45° se destacou melhor, com exceção da sensibilidade que teve melhor resultado na direção de 90°.

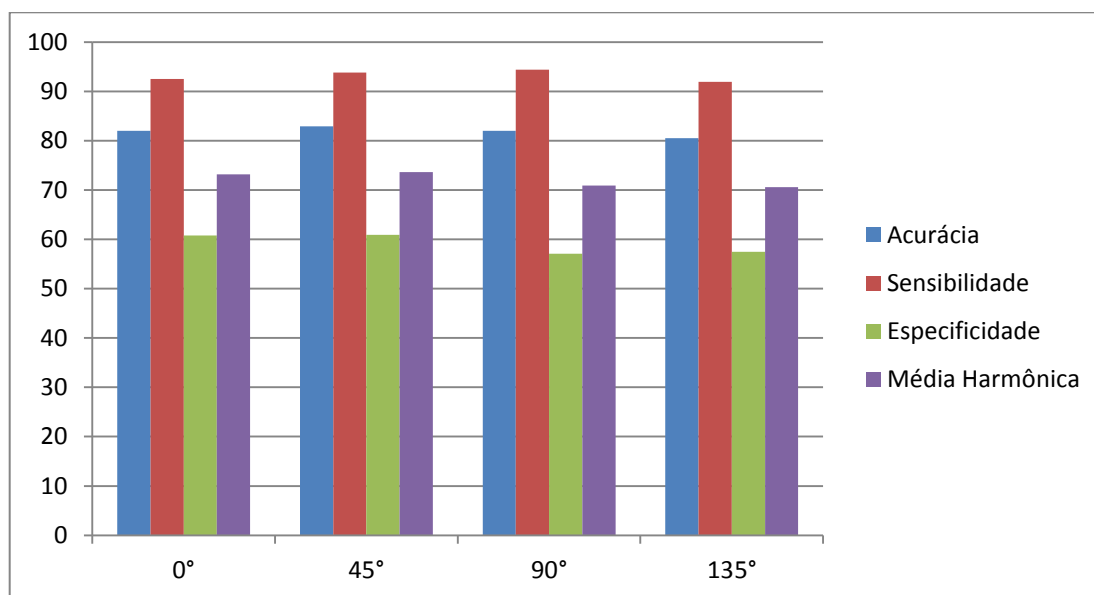
As figuras 65 e 66 mostram um comparativo das métricas entre cada uma das direções ao longo da GLCM tanto para o classificador kNN quanto pra o classificador SVM.

Figura 65 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação do tecido mamário usando kNN.



Fonte: Produção do próprio autor.

Figura 66 – Gráfico em barra para diferentes ângulos da GLCM no teste de classificação do tecido mamário usando SVM.



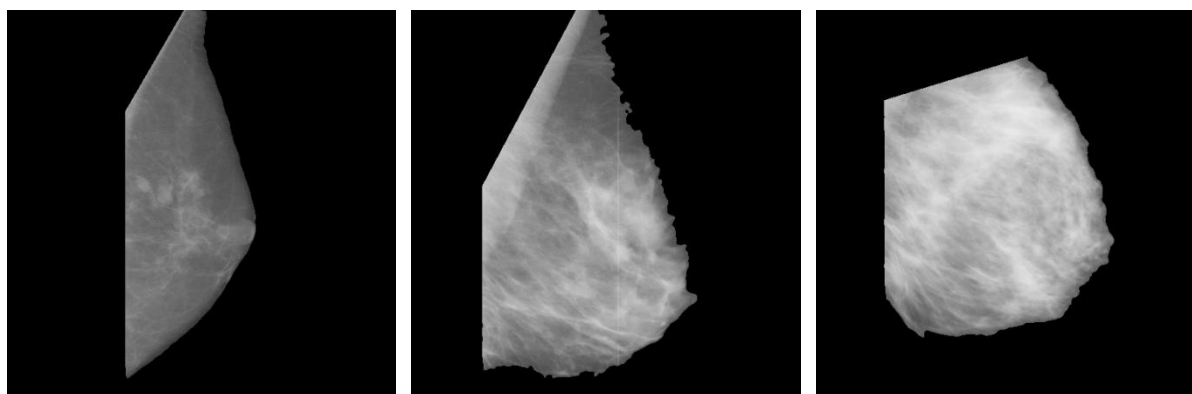
Fonte: Produção do próprio autor.

4.4 Remoção do músculo peitoral

A metodologia adotada para a segmentação da mama de forma a remover o músculo peitoral, remoção essa necessária para a correta definição da ROI envolvendo somente tecido mamário, foi bastante eficiente.

Dentre as 322 imagens utilizadas, a remoção do músculo peitoral obteve sucesso (ou seja, removeu totalmente o músculo e mais nenhuma outra parte da mama) em 295 imagens. O músculo não foi completamente removido em 13 imagens e parte da mama foi cortada em 14 imagens. A figura 67 mostra um exemplo de cada uma dessas situações.

Figura 67 – Três situações obtidas quando da remoção do músculo peitoral.



a) Completamente removido.

b) Parcialmente removido.

c) Parte da mama cortada.

Fonte: Produção do próprio autor.

Portanto, a metodologia adotada teve uma eficácia de 91,6%, se mostrando muito interessante para futuras aplicações que necessitem da remoção do músculo peitoral, uma vez que ele está presente nas mamografias em grande maioria das vezes e não é objeto de estudo para os trabalhos da área. Vale lembrar que mesmo para as imagens cuja remoção do músculo não se deu de forma esperada, a seleção da ROI não sofreu quaisquer complicações.

5 CONCLUSÃO

O diagnóstico auxiliado por computador é um recurso essencial para ajudar médicos e radiologistas na detecção e prevenção do câncer de mama, uma vez que consegue detectar nódulos e possíveis suspeitas antes mesmo de serem percebidos pelo tato (autoexame da mama).

Junto com as medidas extraídas matriz de coocorrência, neste trabalho foram usados os classificadores kNN e SVM para a tarefa de detecção e identificação de câncer e classificação de tecido mamário, como também foi usado o algoritmo SFS para a seleção das medidas. Além disso, foi proposto um método simples para remoção do músculo peitoral das mamografias.

A detecção de câncer apresentou resultados satisfatórios em ambos os classificadores para todas as métricas dependendo da medida de textura selecionada, com o Momento de Diferença Inversa Normalizada e a Diferença Inversa Normalizada se destacando dentre as outras medidas para o classificador kNN. Para o classificador SVM podem-se destacar as seguintes medidas individuais de textura: a Média da Diferença a Medida de Informação da Correlação 2, a Máxima Probabilidade e a Dissimilaridade.

A classificação de câncer não obteve bons resultados com os classificadores utilizados. Isso pode ter acontecido devido à proximidade dos valores de medidas de textura para as imagens com câncer benigno e maligno. As métricas obtidas com as medidas de forma individual foram baixas. As medidas com os maiores destaques para o classificador kNN foram: o Segundo Momento Angular, o Coeficiente de Correlação Máximo e a Entropia. Para o classificador SVM podem ser destacadas as seguintes medidas: a Média da Diferença, a Autocorrelação, a Máxima Probabilidade, a Medida de Informação da Correlação 1, a Variância e a Variância da Soma.

A classificação do tecido mamário em adiposo e glandular apresentou resultados satisfatórios em ambos os classificadores para todas as métricas dependendo da medida de textura selecionada. Para o classificador kNN, destacam as seguintes medidas de textura: o Momento de Diferença Inversa Normalizada e a Medida de Informação da Correlação 1. Para o classificador SVM podem-se destacar as seguintes medidas: a Média da Diferença, a Medida de Informação da Correlação 2, a Dissimilaridade, o Contraste, a Correlação, a Inércia e o Momento de Diferença Inversa Normalizada.

O SFS também foi usado nos experimentos, e foi observado que até certa iteração os resultados obtidos melhoravam, mas a partir de um determinado conjunto de medidas as métricas ou oscilavam ou reduziam. A partir dos resultados obtidos, chegaram-se às seguintes conclusões: 1) usar cada vez mais medidas não garante que os resultados irão melhorar, podendo inclusive piorar; 2) Para o caso das classificações realizadas, um número reduzido de medidas retornam resultados tão bons quanto os maiores valores alcançados no SFS.

O método proposto para remoção do músculo peitoral através de um polígono e do uso da Transformada de Hough obteve eficácia de 91,6%, ao passo que apenas 27 imagens, dentre as 322 utilizadas, apresentaram resultados não desejados.

Como sugestão para trabalhos futuros, destacam-se os seguintes pontos:

- Utilização de outros descritores de textura, como o Histograma da Soma/Diferença, o *Local Binary Pattern* (LBP) e o *Local Mapped Pattern* (LMP);
- Utilização de outros métodos de seleção de atributos tais como algoritmos genéticos, *Sequential Forward Floating Selection* (SFFS) e *Sequential Backward Floating Selection* (SBFS).

Por fim, o presente trabalho se mostrou muito satisfatório em nível de aprendizado em processamento digital de imagens, medidas de textura e classificadores, contribuindo para a formação do graduando e auxiliando futuros profissionais que se interessam pelo assunto.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, Marcio P. *et al.* Centro Brasileiro de Pesquisas Físicas. **Aplicação da Técnica de Momentos Invariantes no Reconhecimento de Padrões em Imagens Digitais**, out. 2011. Disponível em: <<http://www.cbpf.br/~chacon/mhu>>. Acesso em: 22 nov. 2016.

ALBREGTSEN, F. University of Oslo. **Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices**. 2008. 14 p. Disponível em: <<https://www.uio.no/studier/emner/matnat/ifi/INF4300/h08/undervisningsmateriale/g lcm.pdf>>. Acesso em: 03 dez. 2016.

ALCÂNTARA, R. S. JUNIOR, P. E. F. RAMOS, A. S. S. **SVM-based feature extraction for mammography classification**. *Proceedings of XII Workshop de Visão Computacional*, p. 21-26, 2016.

ANNA, K. E. Posterus. **Searching of meteors in astronomical images using Matlab GUI**, mai. 2011. Disponível em: <<http://www.posterus.sk/?p=10583>>. Acesso em: 02 dez. 2016.

BATISTA, M. H. **Classificação Hierárquica Orientada a Objeto em Imagens de Alta Resolução Espacial Empregando Atributos Espaciais e Espectrais**. 2006, 211 f. Dissertação (Mestrado em Sensoriamento Remoto). Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

BHARATHI, V. S. **Contourlet Based Texture Analysis and Classification of Mammogram Images**. *International Journal of Engineering Science and Technology*, v. 5, n. 06, p. 1228-1234, 2013.

BIANCA. Clínica Médica. **Sensibilidade, Especificidade, Valores Preditivos**, jun. 2004. Disponível em: <<https://clinicamedicarquivo.blogspot.com.br/2004/06/sensibilidade-especificidade-valores.html>>. Acesso em: 05 dez. 2016.

BOVIS, K. SINGH, S. CiteSeerX. **Classification of Breast Density in Digital Mammograms**, s.d. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.5702&rep=rep1&type=pdf>>. Acesso em: 05 nov. 2016.

Cancer Research UK. **What is câncer?**, out. 2014. Disponível em: <<http://www.cancerresearchuk.org/about-cancer/what-is-cancer>>. Acesso em: 03 nov. 2016.

CLAUSI, D. A. **An analysis of co-occurrence texture statistics as a function of grey level quantization**. *Can. J. Remote Sens.* **28**, p. 45–62, 2002.

Clínica da Mama. **Confira as diferenças entre a mamografia digital e a convencional**, jun. 2014. Disponível em: <<http://www.clinicadamama.com.br/confira-diferencas-entre-mamografia-digital-e-convencional>>. Acesso em: 15 nov. 2016.

CONCI, A. Azevedo E. Leta, F.R. **Computação Gráfica Vol. 2**. Campus, 2007.420 p.

DANTAS, W. COURA, A. ALVARENGA, D. RN Saúde - Assistência Médica. **Câncer de mama: a importância do diagnóstico precoce**, out. 2012. Disponível em: <<http://www.rnsaude.com.br/site/artigos/cancer-de-mama-a-importancia-do-diagnostico-precoce.html>>. Acesso em 10 nov. 2016.

DDSM, s.d. University of South Florida Digital Mammography Home Page. **DDSM: Digital Database for Screening Mammography**. Disponível em: <<http://marathon.csee.usf.edu/Mammography/Database.html>>. Acesso em: 06 nov. 2016.

DUTRA, L. V.; HUBER, R. **Feature extraction and selection for ERS-1/2 InSAR classification**. *International Journal of Remote Sensing*, v. 20, n. 5, p. 993-1016, 1999.

FELGUEIRAS, C.A. Instituto Nacional de Pesquisas Espaciais. **Processamento Digital de Imagens: Realce de Imagens**, s.d. Disponível em: <http://www.dpi.inpe.br/~carlos/Academicos/Cursos/Pdi/pdi_realce.html>. Acesso em: 23 nov. 2016.

FILHO, J. R. *et al.* **Segmentação do Músculo Peitoral em Imagens Mamográficas Através do Recorte Automático de ROI e da Binarização de Otsu**. *VII Workshop de Visão Computacional*, p. 198-203, 2011.

GADKARI, D. University of Pune. **Image Quality Analysis Using GLCM**, 2000. Disponível em: <http://etd.fcla.edu/CF/CFE0000273/Gadkari_Dhanashree_U_200412_MS.pdf>. Acesso em: 04 dez. 2016.

Glamour, Redação. Revista Glamour. **Outubro Rosa: conheça a campanha contra o câncer de mama**, out. 2016. Disponível em: <<http://revistaglamour.globo.com/Beleza/Saude/noticia/2016/10/outubro-rosa-conheca-campanha-contra-o-cancer-de-mama.html>>. Acesso em: 07 nov. 2016.

GOMES, O. F. M.. **Processamento e Análise de Imagens Aplicados à Caracterização Automática de Materiais**. 2001. 151 f. Dissertação (Mestrado em Ciências da Engenharia Metalúrgica). Pontifca Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2001.

GONZALEZ, Rafael C. Woods, Richard E. **Processamento de Imagens Digitais**. Pearson, 2011. 603 p.

HARALICK, R. et al. **Textural Features for Image Classification**. IEEE Transactions on Systems, Man and Cybernetics, p. 610–621, 1973.

HASSAN, I. et al. Scientific Reports. **Radiomic Texture Analysis Mapping Predicts Areas of True Functional MRI Activity**, mai. 2016. Disponível em: <<http://www.nature.com/articles/srep25295#t5>>. Acesso em: 05 nov. 2016.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. **O que é o câncer?**, s.d. Disponível em: <<http://www2.inca.gov.br/wps/wcm/connect/cancer/site/oquee>>. Acesso em: 01 nov. 2016.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. **Outubro Rosa: pesquisa do INCA revela importância de a mulher conhecer as próprias mamas**, out. 2016. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/agencianoticias/site/home/noticias/2016/pesquisa_do_inca_que_revela_importancia_do_conhecimento_das_proprias_mamas_chama_atencao_de_medicos_e_movimentos_sociais>. Acesso em: 12 nov. 2016.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. **Deteção precoce do câncer de mama**, s.d. Disponível em: <http://www.inca.gov.br/conteudo_view.asp?id=1932>. Acesso em: 12 nov. 2016.

Instituto Espaço de Vida. **Dia Nacional da Mamografia: Entenda o que é a Mamografia**, fev. 2016. Disponível em: <<http://www.espacodevida.org.br/noticias/dia-nacional-da-mamografia-entenda-o-que-e-a-mamografia-2388>>. Acesso em: 15 nov. 2016.

KIM, H. Y. Universidade de São Paulo. **Transformada de Hough para detectar retas**, 2016. Disponível em: <<http://www.lps.usp.br/hae/apostila/hough.pdf>>. Acesso em: 02 dez. 2016.

LADHA, L. LECTURER T. D. International Journal on Computer Science and Engineering. **Feature Selection Methods and Algorithms**, v. 3, n. 5, p. 1787-1797, 2011.

LEW, M.S. **Principles of Visual Information Retrieval**. Springer-Verlag London, v.1, 356 p., 2001.

LIMA, E. S. Edirlei 3DGB. **K-Nearest Neighbor (KNN)**, 2014. Disponível em: <http://edirlei.3dgb.com.br/aulas/ia_2012_1/IA_Aula_16_KNN.pdf>. Acesso em: 04 dez. 2016.

LORENA, A. C. Carvalho, A. C. P. L. F. **Uma Introdução às Support Vector Machines**. RITA, v. XIV, n. 2, p. 43-67, 2007.

MARANHÃO, N. Radiologia Clínica de Campinas. **Vantagens e Limitações do Diagnóstico Mamográfico**, jan. 2016. Disponível em: <<http://www.radiologiaclinicadecampinas.com.br/blog/?p=228>>. Acesso em: 25 nov. 2016.

MARTINS et al., Sistema Integrado de Gestão de Atividades Acadêmicas. **Classification of Normal, Benign and Malignant Tissues Using Co-occurrence Matrix and Bayesian Neural Network in Mammographic Images**, 2006. Disponível em: <<http://www.pgsc.ufma.br/arquivos/classificationofnormalbenignandmalignanttissues.pdf>>. Acesso em: 20 nov. 2016.

MASCARO, A. A. **Segmentação de Imagens de Mamografias Digitais**. 2007. 46 p. Trabalho de Conclusão de Curso, Universidade de Pernambuco, Recife, 2007.

MENEZES, G. Só falta testar. **Morfologia Matemática para imagens em tons de cinza**, jan. 2010. Disponível em: <<https://sofaltatestar.wordpress.com/2010/01/06/morfologia-matematica-para-imagens-em-tons-de-cinza>>. Acesso em: 29 nov. 2016.

MIAS, Database. **The mini-MIAS database of mammograms**, s.d. Disponível em: <<http://peipa.essex.ac.uk/info/mias.html>>. Acesso em: 20 nov. 2016.

MONTEIRO, L. H. Universidade Federal Fluminense. **Binarização por Otsu e Outras técnicas Usadas na Detecção de Placas**, s.d. Disponível em: <<http://www2.ic.uff.br/~aconci/OTSUeOutras.pdf>>. Acesso em: 27 nov. 2016.

NASCIMENTO, R. F. F. et al. Instituto Nacional de Pesquisas Espaciais. **O algoritmo Support Vector Machines (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2**, abr. 2009. Disponível em: <<http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2008/10.20.10.59/doc/2079-2086.pdf>>. Acesso em: 04 dez. 2016.

NETO, G. A. **Um Estudo sobre Detecção e Classificação de Defeitos em Tiras de Aço via Processamento Digital de Imagens**. 2014. 81 p. Trabalho de Conclusão de Curso, Universidade Federal do Espírito Santo, Vitória, 2014.

OLIVEIRA et al. **Detection of Architectural Distortion in Digitized Screen-film Mammograms Using Texture Descriptors**. *Proceedings of XII Workshop de Visão Computacional*, p. 260-265, 2016.

PIVETTA, C. et al. Universidade Estadual do Oeste do Paraná. **Transformada de Hough**, s.d. Disponível em <http://www.inf.unioeste.br/~adair/PID/Notas%20Aula/Transformada%20de%20Hough.pdf>. Acesso em: 01 dez. 2016.

PRATIWI, M. et al. Science Direct. **Mammograms Classification using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network**, 2015. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1877050915018694>. Acesso em: 05 nov. 2016.

RAUT, A. et al. **Texture parameters Extraction of Satellite Image**. *International Journal of Science Technology & Engineering*, v. 2, n. 11, p. 13-18, 2016.

REFAEILZADEH, P. TANG, L. LIU, H., s.d, Arizona State University. **Cross-Validaton**, s.d.

REY, C. A. Blog Dr. Carlos. **Mastalgia**, jan. 2012. Disponível em: <https://drcarlosrey.blogspot.com.br/2012/01/mastalgia.html>. Acesso em: 18 nov. 2016.

RIUL, S. S. *Revista Brasileira de Enfermagem*. **Câncer de mama: fatores de risco e detecção precoce**, jan. 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71672011000600005. Acesso em: 11 nov. 2016.

SANTOS, B. F. *Revista Exame*. **10 números preocupantes sobre câncer de mama no Brasil e no mundo**, out. 2016. Disponível em: <http://exame.abril.com.br/brasil/cancer-de-mama-brasil-mundo>. Acesso em: 03 nov. 2016.

SANTOS, V. T. **Segmentação de Imagens Mamográficas para Detecção de Nódulos em Mamas Densas**. 2002. 121 p. Dissertação (Mestrado em Engenharia Elétrica), Universidade de São Paulo, São Carlos, 2002.

Saúde Melhor. **Tem algum risco em fazer mamografia? Saiba tudo sobre este exame!**, s.d. Disponível em: <<https://www.saudemelhor.com/tem-algum-risco-fazer-mamografia-saiba-tudo-sobre-este-exame>>. Acesso em: 17 nov. 2016.

SILVA, T. C. **Deteção de Tumores em Mamografias via Processamento de Imagens**. 2013. 70 f. Trabalho de Conclusão de Curso, Instituto Federal de Santa Catarina, São José, 2013.

SOARES, F. A. A. M. N. Universidade Federal de Goiás. **Morfologia Matemática**, s.d. Disponível em: <<http://www.inf.ufg.br/~fabrizzio/mestrado/pdi/aulas/aula9.pdf>>. Acesso em: 01 dez. 2016.

SOH, L. K. Tsatsoulis, C. **Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices**. IEEE Trans. Geosci. Remote Sens. **37**, p. 780–795, 1999.

SOUZA, C. César Souza. **Análise de Poder Discriminativo Através de Curvas ROC**, jul. 2009. Disponível em: <<http://crsouza.com/2009/07/13/analise-de-poder-discriminativo-atraves-de-curvas-roc>>. Acesso em: 05 dez. 2016.

TAYLOR, P. et al. **Measuring image texture to separate *difficult* from *easy* mammograms**, The British Journal of Radiology, vol. 67, pp. 456-463, 1994.

VASCONCELOS, S. Universidade Federal Fluminense. **Matrizes de Co-Ocorrência**, 2015. Disponível em: <<http://www2.ic.uff.br/~aconci/co-ocorrencia.pdf>>. Acesso em: 03 dez. 2016.

YANG, X. et al. Nacional Center for Biotechnology Information. **Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An *in vivo* study of late toxicity**, set. 2012. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443195>>. Acesso em 05 dez. 2016.