

CENTRO UNIVERSITÁRIO DA FEI

ALBERT DA COSTA XAVIER

**ANÁLISE DISCRIMINANTE E CLASSIFICAÇÃO DE
IMAGENS 2D DE ULTRASSONOGRAFIA DA MAMA**

São Bernardo do Campo
2011

ALBERT DA COSTA XAVIER

**ANÁLISE DISCRIMINANTE E CLASSIFICAÇÃO DE
IMAGENS 2D DE ULTRASSONOGRAFIA DA MAMA**

Dissertação de Mestrado apresentada ao Centro
Universitário da FEI, como parte dos requisitos
necessários para obtenção do título de Mestre em
Engenharia Elétrica, sob orientação do Prof. Dr.
Carlos Eduardo Thomaz.

São Bernardo do Campo
2011

Xavier, Albert da Costa

Análise discriminante e classificação de imagens 2D de ultrasonografia da mama / Albert da Costa Xavier. São Bernardo do Campo, 2011.

109 f. : il.

Dissertação - Centro Universitário da FEI.

Orientador: Prof. Dr. Carlos Eduardo Thomaz

1. Análise Estatística Univariada. 2. Análise Estatística Multivariada. 3. Imagem de Ultrassom da Mama. 4. Extração de Características Discriminantes. I. Thomaz, Carlos Eduardo, orient. II. Título.

CDU 681.32:616



Centro Universitário da FEI

APRESENTAÇÃO DE DISSERTAÇÃO ATA DA BANCA JULGADORA

PGE- 10

Programa de Mestrado de Engenharia Elétrica

Aluno: Albert da Costa Xavier

Matrícula: 108106-6

Título do Trabalho: ANÁLISE DISCRIMINANTE E CLASSIFICAÇÃO DE IMAGENS 2D DE ULTRASSONOGRAFIA DA MAMA.

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Carlos Eduardo Thomaz

Data da realização da defesa: 24/ fevereiro/ 2011

A Banca Julgadora abaixo-assinada atribuiu ao candidato o seguinte:

APROVADO ☒

REPROVADO ☐

São Bernardo do Campo, 24, 02, 2011.

MEMBROS DA BANCA JULGADORA

Prof. Dr. Carlos Eduardo Thomaz

Ass.: _____

Prof. Dr. Paulo Sérgio Silva Rodrigues

Ass.: _____

Prof.^a Dr.^a Annie France Frère Slaets

Ass.: _____

VERSÃO FINAL DA DISSERTAÇÃO

ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS
RECOMENDAÇÕES DA BANCA EXAMINADORA

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

À minha namorada e aos meus pais.

Agradecimentos

Gostaria de agradecer ao meu orientador prof. Dr. Carlos Eduardo Thomaz pela excelente orientação prestada e também por me incentivar a continuar o trabalho face às adversidades. Agradeço também aos professores Dr. Paulo Sérgio e Dr. Ivandro Sanchez pelas sugestões propostas durante a qualificação. Sugestões estas que certamente contribuíram para o meu aprendizado. Também agradeço ao prof. Dr. Gilson Antônio Giraldi pelo apoio na fase de pesquisa no Laboratório Nacional de Computação Científica - LNCC/RJ. Aproveito também para agradecer o prof. Dr. João Ricardo Sato, que gentilmente cedeu o programa Darth Brain, cuja utilização foi essencial no desenvolvimento do estudo. Aos demais professores do programa de mestrado da FEI também expresso meus sinceros agradecimentos, prof. Dr. Flávio Tonidandel e prof. Dr. Paulo Eduardo Santos.

Aos meus pais, Hélio e Anita, e a minha namorada Karina, pelo apoio incondicional e incentivo aos estudos. Aos meus colegas do curso de mestrado, pelos momentos de estudo em grupo e pelo companheirismo.

Finalmente, agradeço a CAPES pelo financiamento dos meus estudos em Petrópolis, por meio do Programa Nacional de Cooperação Acadêmica (PROCAD) entre a FEI e o LNCC.

*"Não se pode esperar resultados diferentes
fazendo as coisas da mesma forma."
Albert Einstein*

Resumo

Os estudos relacionados à detecção e ao tratamento de cânceres por meio de imagens, incluindo o câncer de mama, vêm se aperfeiçoando ano após ano. Segundo especialistas, o câncer de mama se posiciona, na verdade, como o maior causador de morte de câncer na população feminina mundial. E, particularmente, a ultrassonografia mamária, por ter menor custo e evitar o contato prejudicial dos envolvidos com radiação, tem aumentado significativamente nos últimos anos. É com este enfoque que o presente trabalho aborda a análise discriminante de tumores mamários em imagens 2D ultrassonográficas. O conjunto de imagens utilizado, com dados pré-processados e segmentados, contém tumores mamários benignos e malignos. No processo de análise dessas imagens, métodos estatísticos univariado e multivariado foram investigados com objetivo de extrair informações discriminantes para fins de classificação. Os resultados da aplicação desses métodos estatísticos favorecem a adoção do método multivariado na análise das imagens de ultrassom. Em caráter complementar, este tipo de metodologia pode também evidenciar as diferenças estatísticas mais significativas entre os tumores, indicando no espaço original das imagens os casos mais simples e difíceis de classificação.

Palavras-chave: análise estatística univariada; análise estatística multivariada; imagem de ultrassom da mama; extração de características discriminantes.

Abstract

Research on detection and treatment of cancer using images, including breast cancer, has improved in the last years. Actually, experts say that breast cancer is the most common cause of cancer death among women worldwide. And, particularly, the breast ultrasound exam, due to its minor cost and no contact with radiation, has increased significantly in recent years. This thesis approaches this problem by carrying out a statistical discriminant analysis of mammary tumors in 2D ultrasound images. The set of images used in this work is composed of benign and malignant breast tumors only. To analyse these images, univariate and multivariate statistical methods have been investigated with the aim of extracting discriminant information for classification purposes. Our results show a good classification performance of the multivariate statistical method when using preprocessed and segmented 2D ultrasound images as a whole. Additionally, this method highlights the most statistically significant differences between the tumors, ranking in the original image space the simplest and most difficult cases for classification.

Keywords: univariate statistical analysis; multivariate statistical analysis; breast ultrasound image (BUS); discriminant features extraction.

Sumário

1	Introdução	1
1.1	Objetivo	2
1.2	Organização do Trabalho	2
2	Contextualização do Problema	3
2.1	Contexto e Estatísticas	3
2.2	Anatomia e Patologia da Mama	5
2.3	Fatores de Risco e Prevenção	10
2.4	Técnicas de Rastreamento	11
2.5	Diagnóstico e Prognóstico do Câncer de Mama	13
2.6	Diagnóstico do Câncer de Mama por Ultrassom	14
3	Métodos de Detecção e Classificação	16
3.1	Detecção e Classificação de Câncer de Mama em Imagens de Ultrassom . .	16
3.2	Pré-processamento de Imagens	17
3.2.1	Técnicas de filtragem	18
3.2.2	Técnicas de domínio Wavelet	22
3.2.3	Técnicas de composição	23
3.3	Segmentação de Imagens	24
3.3.1	Operadores morfológicos	25
3.3.2	Campos aleatórios de Markov	26
3.3.3	Histograma de limiarização	27
3.3.4	Modelo de contorno ativo	28
3.3.5	Outros modelos	29
3.4	Extração e Seleção de Características	31
3.5	Classificação	34
3.5.1	Análise estatística univariada	38
3.5.2	Análise estatística multivariada	40
3.6	Considerações complementares	47
4	Experimentos	48
4.1	Dados Experimentais	48
4.2	Pré-processamento de Imagens	50
4.3	Análise Estatística Univariada	51
4.4	Análise Estatística Multivariada	53
4.4.1	Desempenho dos classificadores SVM e MLDA	54
4.4.2	Comparação dos resultados das análises estatísticas univariada e multivariada	54

5	Análise dos Resultados	56
5.1	Representação do Conjunto de Características	56
5.2	Análise Univariada das Imagens Médicas	58
5.3	Análise Multivariada das Imagens Médicas	61
5.3.1	Desempenho dos classificadores SVM e MLDA	62
5.3.2	Análise visual de resultados dos classificadores SVM e MLDA . . .	75
5.3.3	Comparação visual dos resultados entre as análises estatísticas uni- varidada e multivariada	77
5.3.4	Navegação nos hiperplanos	79
6	Conclusão	81
6.1	Trabalhos Futuros	82
	Anexo I - Amostras investigadas	84
	Anexo II - Valores dos pesos discriminantes das imagens na classificação por meio dos métodos SVM e MLDA	87
	Glossário	88
	Referências Bibliográficas	96

Lista de Figuras

2.1	Distribuição das estimativas de câncer de mama no Brasil para o ano de 2010	5
2.2	Representação esquemática da anatomia básica da mama	6
2.3	Diagrama esquemático da mama	6
2.4	Esquema com corte transversal no ducto mostrando as células epiteliais e mioepiteliais	7
2.5	Imagens do carcinoma lobular invasivo detectado numa paciente de 66 anos de idade	9
2.6	Gânglios linfáticos axilares e o músculo esternal	10
2.7	Gráfico representativo das taxas de crescimento dos cânceres em geral	13
2.8	Anatomia da mama em imagem de ultrassom	15
3.1	Resultado da aplicação do filtro da média em imagem com manchas <i>speckles</i> e em imagem artificial com pequenos ruídos aditivos Gaussianos	19
3.2	Aplicação das transformações logarítmica e gamma em imagens	21
3.3	Aquisição de imagens ultrassonográficas com uso e sem uso de técnicas de composição	24
3.4	Síntese das técnicas aplicadas na redução de manchas e melhoramento de imagem	24
3.5	Resultado comparativo da segmentação manual e da segmentação semi-automática com Operadores Morfológicos	26
3.6	Imagens de ultrassom com seus respectivos histogramas e tumores segmentados	28
3.7	Resultado da aplicação da técnica de corte normalizado - Ncut	31
3.8	Esquema das regiões de interesse utilizadas na definição da característica sombra acústica posterior	33
3.9	Região delimitada pelo algoritmo <i>convex hull</i> e <i>concave polygon</i>	34
3.10	Projeção das amostras no espaço do PCA	41
3.11	Idéia geométrica da LDA proposta por Fisher	43
3.12	Idéia geométrica dos classificador MLDA	44
3.13	Idéia geométrica do classificador SVM	46
4.1	Exemplos de imagens do mesmo tumor mamário	49
4.2	Imagens ultrassonográficas de tumores benignos	50
4.3	Imagens ultrassonográficas de tumores malignos	51
4.4	Imagens redimensionadas e recortadas	51
4.5	Representação do cálculo da imagem média	52
5.1	Histograma dos valores de circularidade	57
5.2	Histograma dos valores de heterogeneidade	57

5.3	Histograma dos valores de sombra acústica	57
5.4	Imagem média	58
5.5	Subtração das imagens médias	59
5.6	Projeção das regiões discriminantes com nível de significância de 95%	60
5.7	Projeção das regiões discriminantes com nível de significância de 99%	60
5.8	Projeção das regiões discriminantes com nível de significância de 99.9%	60
5.9	Projeção de regiões transparentes representando nível de significância de 99.9%	61
5.10	Projeção simultânea de regiões representando todos os níveis de significância experimentados	61
5.11	Hiperplano de separação obtido pelo classificador SVM na análise das imagens como um todo	64
5.12	Hiperplano de separação obtido pelo classificador MLDA na análise das imagens como um todo	65
5.13	Hiperplanos de separação obtidos pelos classificadores SVM e MLDA na análise das imagens como um todo	66
5.14	Imagens analisadas nos hiperplanos dos classificadores SVM e MLDA	66
5.15	Hiperplano de separação do classificador SVM - Circularidade	68
5.16	Hiperplano de separação do classificador SVM - Heterogeneidade	69
5.17	Hiperplano de separação do classificador SVM - Sombra acústica	70
5.18	Hiperplano de separação do classificador MLDA - Circularidade	71
5.19	Hiperplano de separação do classificador MLDA - Heterogeneidade	72
5.20	Hiperplano de separação do classificador MLDA - Sombra acústica	73
5.21	Projeção das regiões mais discriminantes segundo o classificador SVM considerando a mesma quantidade de pixels projetada na análise univariada com graus de significância de 95%, 99% e 99.9%	76
5.22	Projeção das regiões mais discriminantes segundo o classificador MLDA considerando a mesma quantidade de pixels projetada na análise univariada com graus de significância de 95%, 99% e 99.9%	76
5.23	Projeção transparente das regiões mais discriminantes segundo os classificadores SVM e MLDA considerando a mesma quantidade de pixels projetada na análise univariada com grau de significância de 99.9%	77
5.24	Comparação visual entre as regiões mais discriminantes obtidas na análise univariada com teste t-student e multivariada segundo os classificadores SVM e MLDA	78
5.25	Formas dos tumores mamários propostas pelo BI-RADS	79
5.26	Transição de tumores benignos para malignos por meio de protótipos segundo classificador SVM	80
5.27	Transição de tumores benignos para malignos por meio de protótipos segundo classificador MLDA	80

Lista de Tabelas

2.1	Locais mais comuns de incidência de lesões na mama	8
4.1	Características das imagens ultrassonográficas analisadas nos experimentos	50
5.1	Desempenho dos classificadores SVM e MLDA na análise de imagens ultrassonográficas da mama	74
5.2	Correlação entre os resultados dos classificadores por ordenação dos valores discriminantes de cada imagem analisada	75
6.1	Pesos discriminantes das imagens projetadas no hiperplano de separação do classificador SVM	87
6.2	Pesos discriminantes das imagens projetadas no hiperplano de separação do classificador MLDA	88

Capítulo 1

Introdução

Na área médica a classificação de imagens está presente no controle de vários tipos de doença, incluindo o câncer de mama. Segundo Martins et al. (2010), o câncer de mama tem registrado crescimento muito preocupante nos últimos anos e se posiciona como maior causador de morte por câncer na população feminina. Como as causas dessa doença ainda são desconhecidas, a detecção precoce é a principal esperança para conter tal crescimento.

Compreender e controlar tal doença requer conhecimentos científicos e experiências que vão desde o conhecimento dos complexos mecanismos de regulação molecular às escolhas individuais do estilo de vida. A prevenção e o controle de câncer estão entre os mais importantes desafios científicos e de saúde pública da nossa época (INSTITUTO NACIONAL DE CÂNCER, 2009).

Neste contexto, os recursos computacionais atuais e o processamento de imagens, aliados à estatística, podem contribuir muito para o esclarecimento e o diagnóstico preciso do câncer de mama.

No entanto, a análise estatística de imagens, particularmente aquela baseada em ultrassom, pode apresentar sérios problemas em decorrência do processo de aquisição. A presença de manchas é comum e a calibração do aparelho está extremamente relacionada à qualidade da imagem obtida. Portanto, este trabalho visa compreender os mecanismos envolvidos na geração das alterações visíveis nas imagens de tumor de mama. Tal compreensão é essencial para explicar a classificação de possíveis tumores benignos ou potencialmente malignos. Este trabalho visa também comparar análises geométricas e de textura com análises da imagem como um todo. Em alguns momentos os testes serão isolados mas o resultado final envolve todas as análises com propósito único de atingir um nível mais significativo de detecção e classificação de tumores da mama.

1.1 Objetivo

O objetivo do presente trabalho é, mais especificamente, analisar o resultado da investigação de métodos estatísticos univariado e multivariado na extração de características discriminantes de um conjunto de imagens de ultrassom 2D com tumores mamários e, também, analisar e quantificar a classificação dessas imagens. As imagens analisadas contêm tumores mamários benignos e malignos cujos diagnósticos foram atribuídos por radiologistas experientes.

1.2 Organização do Trabalho

Este trabalho está dividido em 6 capítulos. O próximo capítulo, de número 2, contextualiza o problema abordando assuntos relacionados à patologia da mama e ao diagnóstico do câncer de mama com imagens ultrassonográficas. O capítulo 3 apresenta a base teórica que apoia o desenvolvimento de todo o trabalho, descrevendo os métodos estatísticos univariado e multivariado. Os experimentos realizados e os dados das imagens são apresentados no capítulo 4 por meio da descrição dos recursos utilizados e das técnicas adotadas na análise das imagens ultrassonográficas de tumores mamários. O capítulo 5 contempla a análise dos resultados e, por fim, o capítulo 6 conclui essa dissertação resumindo os resultados e apresentando sugestões de trabalhos futuros.

Capítulo 2

Contextualização do Problema

Este capítulo tem por objetivo apresentar uma revisão sobre conceitos e informações relacionados ao câncer de mama. Inicialmente são apresentadas estatísticas de incidência deste tipo de câncer no Brasil seguidas da discussão dos possíveis métodos de prevenção e diagnóstico. Na sequência, é apresentada a estrutura da mama, sob aspectos anatômicos e patológicos, abordando a importância da observação de determinadas características diretamente relacionadas ao desenvolvimento e diagnóstico. Por fim, são apresentados os tipos de técnicas de rastreamento do câncer destacando a utilização do ultrassom para diagnóstico precoce da doença.

2.1 Contexto e Estatísticas

O câncer de mama é a principal causa de morte não previsível por câncer. A estimativa para novos casos é preocupante. Segundo o INCA (*Instituto Nacional de Câncer*¹), a previsão para 2010 foi de 49.240 novos casos representando aproximadamente 25,5% de incidência perante todas as ocorrências de câncer, estimadas em 192.590. Na Figura 2.1 pode-se observar a distribuição dos casos estimados, por estado, para o ano de 2010. Os maiores índices, destacados na distribuição, se concentram nos estados de São Paulo, Rio de Janeiro, Minas Gerais e Rio Grande do Sul, respectivamente. Ainda segundo o INCA, 33 mulheres morrem por dia em decorrência do câncer de mama.

Anualmente nos Estados Unidos, mais de 180.000 mulheres são diagnosticadas com câncer de mama invasivo. Em complemento são detectados aproximadamente 25.000

¹<http://www.inca.gov.br/estimativa/2010/estimativa20091201.pdf>

novos casos de carcinoma ductal e lobular *in situ* (AMERICAN CANCER SOCIETY, 2010). Notadamente os 25.000 casos de carcinoma representaram percentual significativo perante a totalidade de casos diagnosticados de cânceres de mama invasivos. Segundo o INCA, a denominação *carcinoma* é atribuída ao câncer que tem início em tecidos epiteliais como pele ou mucosas². O carcinoma mamário caracteriza-se pela proliferação anormal das células do ducto e do lóbulo (Figura 2.2).

Nenhum avanço terapêutico pode ser comparado, em termos de benefícios globais, à contribuição que o diagnóstico precoce, por meio dos métodos radiológicos, pode trazer para pacientes portadores do câncer de mama (KOPANS, 2000).

Atualmente, não existe nenhuma forma de prevenção da doença. Os dados mostram que o melhor resultado obtido, mesmo distante do ideal, é a detecção e tratamento precoce. Tais medidas implicam diretamente no retardo da morte e na melhor qualidade de vida da paciente que poderá ser poupada do trauma físico e psicológico da mastectomia. Definitivamente, o rastreamento mamográfico não é a solução para o problema do câncer de mama, mas é a melhor solução disponível no presente e no futuro próximo. Historicamente, a detecção precoce proporcionou o uso mais adequado da terapia conservadora com predomínio da excisão e da utilização da radiação.

Como as razões pelas quais a mulher desenvolve o câncer ainda são desconhecidas e obscuras, o entendimento da epidemiologia e da etiologia desta malignidade é o alicerce para qualquer estudo desse assunto. Mas, diante desses dados e considerações brevemente descritos aqui, as condições econômicas do sistema de saúde pública podem influenciar as possíveis escolhas de diagnóstico e tratamento dessa doença.

²http://www.inca.gov.br/conteudo_view.asp?id=322 (acessado em 10/07/2010)

Estimativas para o ano 2010 de número de casos novos de câncer, por Estado*

Estados	Mama Feminina	Traqueia, Brônquio e Pulmão	Estômago	Colo do Útero	Próstata	Côlon e Reto	Esôfago
Acre	40	50	40	50	90	20	20
Alagoas	350	200	140	270	450	120	60
Amapá	40	50	70	60	60	20	20
Amazonas	310	290	320	560	430	160	40
Bahia	1.970	850	990	1.030	2.970	820	410
Ceará	1.660	860	1.170	860	2.240	530	350
Distrito Federal	670	300	240	230	560	370	90
Espírito Santo	820	500	520	440	1.040	480	290
Goiás	1.070	770	480	540	1.360	640	230
Maranhão	390	270	310	730	690	170	60
Mato Grosso	400	330	260	330	710	220	120
Mato Grosso do Sul	550	360	290	310	800	350	140
Minas Gerais	4.250	2.310	2.000	1.330	5.350	2.250	1.420
Pará	640	390	650	790	700	290	80
Paraíba	550	220	320	300	790	180	130
Paraná	2.990	1.950	1.700	1.250	3.620	2.060	1.030
Pernambuco	2.120	880	760	1.020	2.470	720	320
Piauí	350	210	150	350	680	140	60
Rio de Janeiro	7.470	3.300	2.060	2.080	6.020	3.920	990
Rio Grande do Norte	540	290	310	230	760	230	100
Rio Grande do Sul	4.750	3.960	1.490	1.250	4.510	3.130	1.460
Rondônia	160	160	120	120	270	60	50
Roraima	40	30	30	50	70	20	20
Santa Catarina	1.570	1.320	900	610	1.690	960	550
São Paulo	15.080	7.500	5.980	3.190	13.160	10.070	2.520
Sergipe	340	170	130	260	520	130	40
Tocantins	120	110	70	190	340	50	30
Brasil	49.240	27.630	21.500	18.430	52.350	28.110	10.630

*Números arredondados para 10 ou múltiplos de 10

Figura 2.1: Distribuição das estimativas de câncer de mama no Brasil para o ano de 2010. Adaptado de (INSTITUTO NACIONAL DE CÂNCER, 2009, p 42)

2.2 Anatomia e Patologia da Mama

A mama é uma glândula modificada da pele. Ela desenvolve-se na parede torácica, entre a clavícula e as costelas (da sexta a oitava). A Figura 2.2 apresenta a anatomia básica da mama destacando o ducto terminal e suas unidades lobulares. Um ducto terminal juntamente com seu lóbulo recebem o nome de unidade lobular do ducto terminal (ULDT), termo muito comum na literatura sobre esse assunto. As ULDT ainda serão mencionadas nas próximas seções pois acredita-se que a maioria dos tumores se desenvolve nestas. Especula-se que a rápida proliferação celular e a replicação do DNA que ocorre nas ULDT tem relação direta com a incidência da maioria dos cânceres nessa área. Segundo estudos, a camada de gordura vista na Figura 2.2 também tem relação com a

incidência do câncer (KOPANS, 2000).

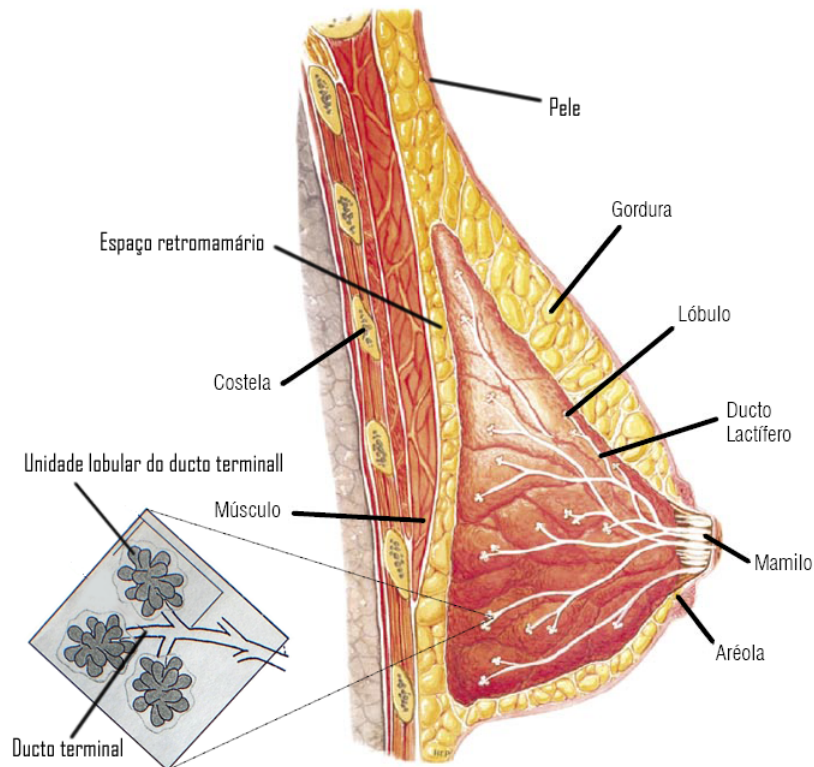


Figura 2.2: Representação esquemática da anatomia básica da mama destacando o ducto terminal e a unidade lobular do ducto terminal (ULD). Adaptado de (KOPANS, 2000)

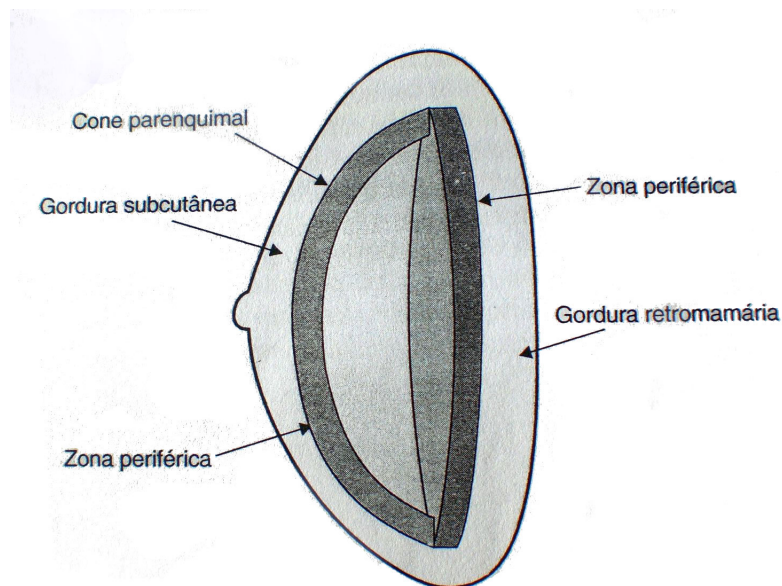


Figura 2.3: Diagrama esquemático da mama. Fonte: (KOPANS, 2000, p 14).

Existe também incidência de câncer de mama fora da região principal da mama. Denomina-se estadiamento o agrupamento dos tumores de acordo com a região envolvida.

Tal estadiamento contempla três níveis gerais de desenvolvimento tumoral. No primeiro nível o tumor cresce somente dentro da mama (local), no segundo nível o tumor é encontrado nos linfonodos axilares e por fim, no terceiro nível, o tumor se espalha por outros órgãos em função da habilidade metastática.

Estatísticas de incidência de lesões na mama, tanto benignas quanto malignas, indicaram que a maioria dessas desenvolveu-se na zona periférica do parênquima (Figura 2.3). Os ductos da mama apresentam em sua formação uma camada interna de células epiteliais envolvidas por outra camada mais fina de células mioepiteliais (Figura 2.4). Presume-se que essas células mioepiteliais exerçam papel na propulsão e expressão do leite durante o período de lactação. As células epiteliais são de suma importância para o diagnóstico patológico do câncer uma vez que estão envolvidas na formação do tumor maligno. Na análise tumoral, concorda-se que a presença de células epiteliais indica uma possível lesão maligna (KOPANS, 2000).

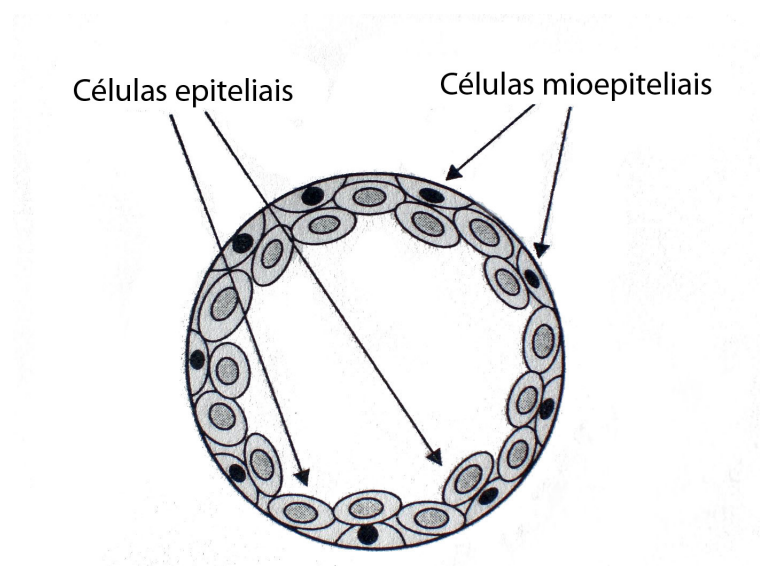


Figura 2.4: Esquema com corte transversal no ducto mostrando as células epiteliais e mioepiteliais. Fonte: (KOPANS, 2000, p 16).

Hoje, positivamente, o diagnóstico de lesões mamárias benignas e malignas tem sido facilitado pela divisão da mama em locais mais comuns de incidência relacionados na Tabela 2.1. Esses locais, exceto o tecido conjuntivo interlobular, podem ser observados na Figura 2.2 ilustrada anteriormente. Nesses locais ou componentes ocorrem muitas alterações histológicas e patológicas da mama. Dentre essas alterações podemos citar a *ectasia ductal* e o *papiloma ductal grande* nos ductos principais. Nos ductos menores e

terminais temos ocorrências da hiperplasia, do papiloma periférico, do carcinoma ductal *in situ* (CDIS), do carcinoma ductal invasivo, do carcinoma lobular *in situ* (CLIS) e do carcinoma lobular invasivo. Na imagem de mamografia da Figura 2.5(a) pode-se observar, na ponta da seta, um carcinoma lobular invasivo diagnosticado numa paciente de 66 anos de idade. Na Figura 2.5(b), o mesmo carcinoma pode ser observado na imagem de ultrassom. Alguns acreditam que muitos, senão todos, cânceres invasivos surgem de um CDIS. Por fim, destaca-se também a presença do cisto e do tumor filóide no lóbulo e no tecido conjuntivo interlobular, respectivamente. Apesar da natural preocupação, os cistos não representam grande ameaça, a menos que ocorram em mulheres com histórico de câncer familiar (ver seção 2.3). As lesões benignas mais comuns têm origem no lóbulo. O lóbulo pode ser visto também na Figura 2.2 anterior. Geralmente os tumores de mama malignos são invasivos, mas os tumores benignos também podem apresentar essa característica.

Locais mais comuns das lesões
Lesões dos ductos principais
Lesões dos ductos menores e terminais
Lesões do lóbulo
Lesões do tecido conjuntivo interlobular

Tabela 2.1: Locais mais comuns de incidência de lesões na mama. Adaptado de (KOPANS, 2000, p 21)

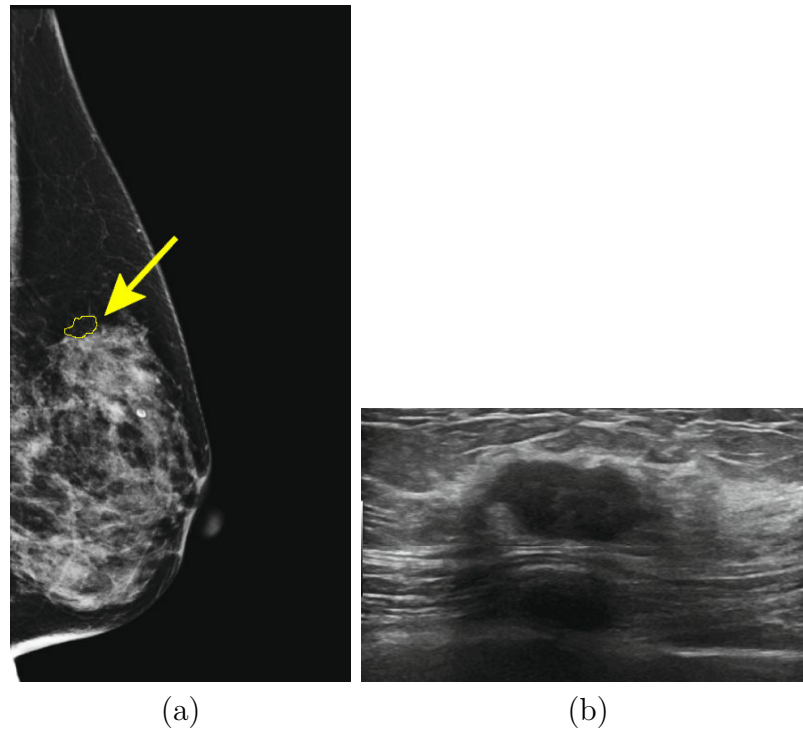


Figura 2.5: Imagem do mesmo carcinoma lobular invasivo detectado numa paciente de 66 anos de idade. Imagem de mamografia (a) e imagem de ultrassom (b). Fonte: (LEE et al., 2010, p 51-53).

Da mesma maneira que outros cânceres, quando o tumor originário na mama migra para outros órgãos, diz-se que as células deste adquiriram habilidade metástica. O processo de metástase consiste em habilitar a célula a adquirir alterações além do crescimento restrito (KOPANS, 2000). Esta célula metástica tem a capacidade de abandonar o local de origem (ducto ou linfonodos axilares) e, por meio da corrente sanguínea, hospedar-se em outro órgão. O ducto pode ser observado na Figura 2.2. Felizmente todo esse processo é muito complexo e tais células precisam sobrepor (vencer) muitas dificuldades, incluindo o sistema imunológico. O envolvimento dos gânglios linfáticos axilares (Figura 2.6) na lesão tumoral pode ser um sinal de metástase.

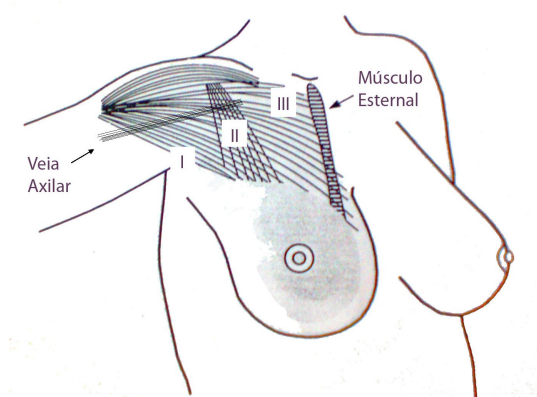


Figura 2.6: Gânglios linfáticos axilares (I, II e III) e o músculo esternal. Adaptado de (KOPANS, 2000, p 12)

2.3 Fatores de Risco e Prevenção

Embora sem comprovação científica, investigações indicam que a incidência do câncer tem relação direta com alguns fatores. Dentre estes podemos destacar a idade da mulher, a primeira gravidez associada com o período de lactação, o uso de anticoncepcionais e a reposição hormonal. Infelizmente, trata-se apenas de suposições pois até o presente momento não existe nenhuma comprovação concreta que justifique o aparecimento e desenvolvimento das lesões. Além desses fatores, existe uma forte crença que as células-tronco sejam progenitoras do câncer, ou seja, o câncer tem início nessas células. As células-tronco são especiais e têm a responsabilidade de substituir as células comuns que morrem frequentemente. Acredita-se que tais células-tronco, com defeito, possam se transformar em "*fábricas*" de células defeituosas que originam as lesões malignas. Mas, independentemente das afirmações anteriores, toda mulher apresenta grau de risco. Em número muito menor também são diagnosticados cânceres de mama em indivíduos do sexo masculino.

A hereditariedade do câncer de mama é outro fator amplamente discutido mas também ainda pouco consolidado. Algumas mulheres herdaram um gene anormal e apresentam risco muito elevado de desenvolvimento do câncer. Mesmo sem comprovação do resultado, recomenda-se iniciar o rastreamento 10 anos antes da idade que o familiar foi diagnosticado como portador do câncer de mama (ROBSON; OFFIT, 2002; COHN et al., 2003).

Ainda sem explicação também, Dupont e Page (1985) identificaram que a associ-

ação da presença de um cisto com o histórico de câncer familiar provoca um aumento da ordem de 2 a 3 vezes na possibilidade de desenvolvimento da doença.

A prevenção ainda é questionável. Pesquisas recentes têm demonstrado avanços significantes mas estes ainda estão distantes do desejado. Para complicar ainda mais, o próprio processo de detecção, utilizando raio-x, pode aumentar o risco de desenvolvimento da doença. É sabido que a energia da radiação produz radicais livres (partículas carregadas) que podem reagir e causar danos ao DNA. Ainda sem os devidos cuidados a exposição à radiação pode afetar também o próprio técnico radiologista que executa o rastreamento.

2.4 Técnicas de Rastreamento

O rastreamento nada mais é do que a avaliação de mulheres assintomáticas, ou seja, mulheres que não apresentam sintomas do câncer de mama. O objetivo é detectar a doença num estágio favorável à cura. Trata-se puramente de um processo que visa interromper o curso da doença por meio da detecção precoce. Dentre as técnicas, temos a mamografia, o exame clínico da mama e o auto-exame clínico da mama.

Infelizmente, nem todos os envolvidos no rastreamento têm noções claras e conhecimentos adequados do processo. O resultado positivo seria muito mais abrangente se os radiologistas e médicos trabalhassem em absoluta sincronia. Esta sincronia é comumente chamada de *rastreamento de alta qualidade*. Em muitos casos, por falta de conhecimento e de treinamento, ocorrem detecções tardias do câncer e, principalmente, recomendações de biopsias desnecessárias. Estima-se que 70% a 80% das biópsias executadas produzam resultados negativos para o tumor maligno (KOPANS, 2000). A biopsia não pode ser descartada mas deve ser indicada em situações realmente pertinentes. Outro fato inerente é a alteração da mama. Muitos fatores, como a lactação e idade, incidem em alterações mamárias que muitas vezes são mal interpretadas pelos responsáveis pelo rastreamento.

Outro assunto frequente é a relação do custo/benefício do rastreamento em termos de frequência e abrangência. Enquanto um modelo de rastreamento mais moderado pode descartar casos positivos da doença ou detectá-los tardiamente, um modelo mais agressivo gera mais custos monetários e pode gerar também preocupações oriundas da detecção de lesões que depois são classificadas como benignas. O modelo de rastreamento mais agressivo pode provocar efeitos físicos e psicológicos nas pacientes, que em muitos casos são

assintomáticas. Como descrito anteriormente a dosagem correta envolve muitos fatores incluindo os recursos financeiros da saúde pública. Geralmente os custos monetários ditam as regras para a adoção da política de rastreamento da população. Toda essa discussão financeira abre espaço para outra técnica de rastreamento denominada *Auto-Exame da Mama* (AEM). Nessa modalidade a própria paciente executa o exame, visual e de palpação, com intuito de detectar alguma anormalidade. Mesmo sem resultados comprovados, recomenda-se o AEM mensal. Este mesmo exame também pode ser feito em clínicas médicas e recebe a denominação de *Exame Clínico da Mama* (ECM).

O câncer de mama em estágio inicial não pode ser detectado por nenhum meio conhecido. Após um tempo, este pode ser detectado pela mamografia mas ainda é invisível ao ECM. Esse fato vai exatamente de encontro à política de rastreamento adotada, pois exerce influência direta no intervalo de rastreamento. Como mencionado na seção 2.1, a mamografia não é a solução definitiva para o problema do câncer de mama, mas ainda é a melhor solução. É sabido que a mamografia não detecta todos os cânceres e nem todos cânceres detectados, mesmo que precocemente, são curáveis.

Acredita-se que a taxa de crescimento dos cânceres, incluindo o câncer de mama, possa ser representada por uma curva em forma de sino em distribuição poissoniana. A distribuição de *Poisson* é uma distribuição de probabilidade discreta utilizada, neste cenário, para expressar a probabilidade de um certo número de eventos ocorrer num dado período de tempo (MAGALHÃES; LIMA, 2010). Assim, de forma geral, pode-se categorizar os cânceres em três grupos. Alguns crescem de forma muito rápida, outros têm crescimento moderado e outros apresentam crescimento muito lento (Figura 2.7). Essa categorização deve ser destacada por ser útil na definição do intervalo de rastreamento discutido nos dois parágrafos anteriores.

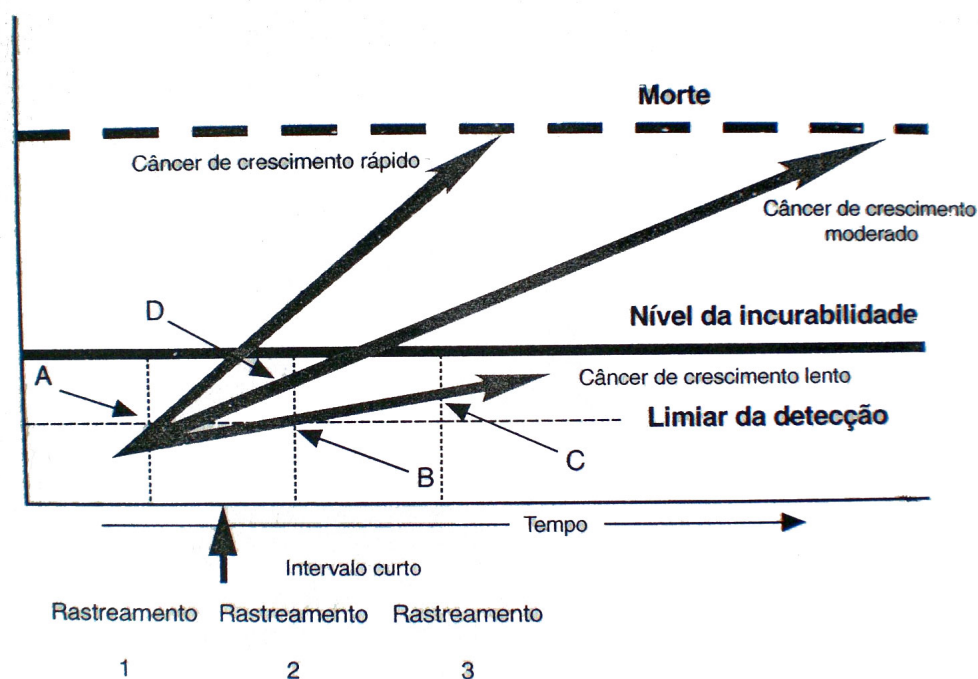


Figura 2.7: Gráfico representativo das taxas de crescimento dos cânceres em geral. Fonte: (KOPANS, 2000, p 64).

Na Figura 2.7, que ilustra um gráfico representando as taxas de crescimento dos cânceres em geral, a linha horizontal mais grossa representa o nível de incurabilidade. Se o tumor for detectado acima dela provavelmente as células do mesmo adquiriram potencial metástático e as chances de cura são mínimas. As linhas verticais finas e pontilhadas representam a execução do rastreamento. Nota-se que nos casos de crescimento tumoral moderado e lento a detecção pode ser feita a tempo, num segundo rastreamento, aumentando de modo significativo a chance de cura.

2.5 Diagnóstico e Prognóstico do Câncer de Mama

Deve ser bem clara a diferença dos termos detecção e diagnóstico. Detecção é a habilidade de encontrar anomalias enquanto o diagnóstico tem como responsabilidade classificar estas anomalias. A mamografia é uma ótima técnica de detecção mas na maioria das vezes, não se caracteriza como técnica diagnóstica.

Segundo Kopans (2000), o prognóstico tem relação direta com o tamanho do tumor, com o seu grau histológico e com a disseminação das células nos linfonodos axilares. Prognóstico é o processo de validação do diagnóstico.

Após a detecção de anomalias, na seção de mamografia, o radiologista nunca deve

recomendar uma intervenção, seja ela de qualquer gênero. Antes de mais nada, deve-se extrair a localização tridimensional exata da possível lesão na mama. Neste caso, a anomalia detectada na imagem de mamografia (raio-x) deverá ser corroborada com outros tipos de imagens, tais como ultrassom, tomografia computadorizada (TC) e ressonância magnética (RM). Até o momento a mamografia continua sendo o exame mais utilizado ((LEE et al., 2010)). A partir desta é que são indicados outros tipos de exame nos casos de suspeita de lesão na mama. Exemplificando, a RM é adequada para melhor visualizar a região do músculo esternal mostrado anteriormente na Figura 2.6. Em muitos casos a massa do músculo é confundida com tumores. Neste exemplo de análise do músculo esternal, a não utilização da RM poderia resultar num diagnóstico pouco preciso. Outra técnica também utilizada é a citologia por aspiração com agulha. Este método pode evitar a biópsia excisional e consequentemente proporcionar mais conforto e agilidade no tratamento da paciente.

2.6 Diagnóstico do Câncer de Mama por Ultrassom

Há um crescente avanço nos métodos de diagnóstico por imagem incluindo a utilização de ultrassom. Na imagiologia mamária, a ultrassonografia representa uma modalidade de diagnóstico muito significativa. Trata-se de um método baseado na reflexão do som. Os transdutores são dispositivos existentes no aparelho capazes de transmitir e receber ondas sonoras. As ondas produzidas alcançam os tecidos e, de acordo com os tipos dos mesmos, são refletidas de volta ao transdutor para ampliação em monitor. Este método é comumente chamado de ecografia pelo fato da reflexão do som também ser chamada de *eco*.

Para lesões de tamanho maior ou igual a 1(cm) a utilização do ultrassom é fortemente indicada para diferenciar um cisto de uma possível lesão sólida (KOPANS, 2000; SHAH et al., 2010). Na Figura 2.8 pode-se observar uma típica imagem de ultrassom descrevendo as principais partes da mama.

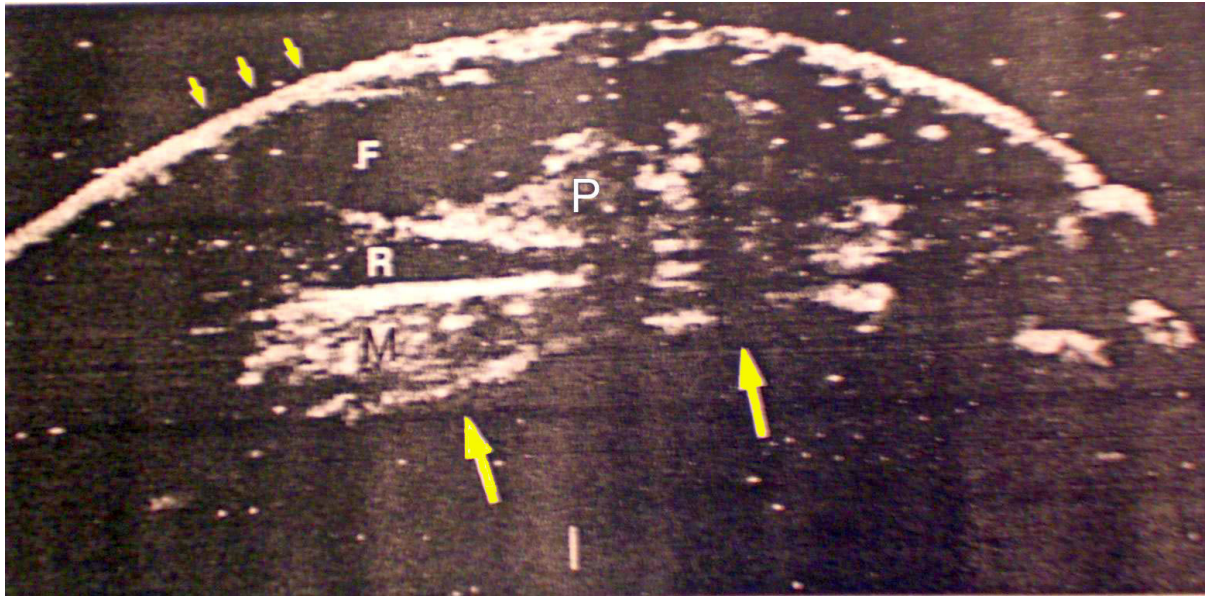


Figura 2.8: Anatomia da mama em imagem de ultrassom. Pode-se observar a pele (setas pequenas superiores), a gordura subcutânea (F), o parênquima (P), a gordura retromamária (R), o músculo peitoral (M) e a parede torácica (setas grandes inferiores). A mama está na posição horizontal. Fonte: (KOPANS, 2000, p 410).

Capítulo 3

Métodos de Detecção e Classificação

Este capítulo tem por finalidade descrever os métodos utilizados na análise de imagens ultrassonográficas. São apresentadas técnicas adotadas nas fases de pré-processamento, segmentação, extração de características e seleção, e, por fim, classificação.

Na fase de classificação, destaca-se a análise entre os grupos de amostras visando extrair as diferenças relevantes entre estes. Existem dois tipos básicos de análises estatísticas que podem ser utilizadas para extrair as diferenças relevantes entre grupos. Estes tipos de análises estatísticas são: a análise estatística univariada, onde cada variável de cada amostra é comparada isoladamente com a mesma variável correspondente das outras amostras e a análise estatística multivariada, que compara todas as variáveis da amostra simultaneamente com as demais amostras.

3.1 Detecção e Classificação de Câncer de Mama em Imagens de Ultrassom

Conforme mencionado anteriormente, o câncer de mama se posiciona como maior causador de morte por câncer na população feminina. Até os dias atuais os melhores resultados foram obtidos com a detecção precoce do tumor. A técnica de detecção e diagnóstico por imagens de ultrassom tem aplicação complementar à mamografia, mas também pode substituí-la. No caso da mamografia, a dependência do operador, considerando os fatos apresentados na Seção 2.4 do capítulo anterior, podem comprometer o processo e gerar consequências negativas. Além desses fatores, a ultrassonografia mamária tem menor custo perante a mamografia e reduz o contato prejudicial dos envolvidos com radiação (raio-x). Mesmo considerando algumas desvantagens relacionadas a seguir, nos últimos anos a proporção de utilização de imagens de ultrassom tem aumentado de forma

muito significativa. Segundo Chang et al. (2003a), no ano de 2003, uma em cada quatro pesquisas, na área médica, utilizou imagem de ultrassom visando melhorar o resultado da detecção e classificação automática do câncer de mama. Sistemas desse gênero visam, pela automatização do processo, aumentar a precisão diagnóstica e diminuir a alta dependência do operador humano. Tais sistemas são denominados CAD (*Computer-Aided Diagnosis*). De forma genérica, os sistemas CAD são compostos de quatro macro processos. O pré-processamento seguido da segmentação, o processo de extração de características e seleção e, por fim, a classificação. Nas próximas seções são detalhados cada macro processo destacando a classificação, foco deste trabalho.

3.2 Pré-processamento de Imagens

As maiores limitações das imagens de ultrassom são o baixo contraste e a presença de pequenas manchas conhecidas pelo termo *speckles*. Segundo Abd-Elmoniem et al. (2002), o *speckle* é uma forma de ruído multiplicativo resultante dos espalhadores que ficam dentro da célula de resolução do feixe de ultrassom. É um artefato introduzido pelas interferências construtivas e destrutivas das pequenas ondas de ultrassom que formam a frente de onda dos ecos que retornam dos tecidos. Assim, a qualidade da imagem é afetada diretamente por fatores inerentes ao instrumento utilizado no exame de ultrassom. Isso quer dizer que a calibração do aparelho está extremamente relacionada à qualidade da imagem obtida. Em decorrência desse problema de qualidade, nessa etapa destacam-se as técnicas aplicadas na redução de manchas e melhoramento da imagem (CHANG et al., 2003a). Tais técnicas podem ser aplicadas isoladamente ou em conjunto proporcionando melhoramento da imagem e remoção de manchas ao mesmo tempo. Pode-se relacionar as técnicas de filtragem linear e não-linear, as técnicas de domínio wavelet e as técnicas de composição.

Em seu trabalho, Abd-Elmoniem et al. (2002) propôs um método, em tempo real, para reforçar o contraste e reduzir as manchas em imagens ultrassonográficas, incluindo imagens de mama. O método, que aborda a técnica de difusão anisotrópica não linear, consiste num modelo matemático baseado em Equação Diferencial Parcial (EDP). Trata-se de uma alternativa aos filtros do tipo *passa-baixa* convencionais. Tal método consiste num processo seletivo, onde regiões homogêneas da imagem são suavizadas de acordo com a evolução temporal aplicada ao modelo. Em outras palavras, o nível de suavização está

relacionado à quantidade de informações indesejadas (manchas) que deverão ser eliminadas da imagem, ou seja, o modelo está diretamente relacionado com o nível desejado de suavização, eliminando assim informações irrelevantes e mantendo seletivamente regiões de interesse nas imagens analisadas.

As pequenas manchas presentes nas imagens de ultrassom não são facilmente visualizadas e muito menos interpretadas. Por tal fato, remover tais manchas sem afetar características que possibilitam o diagnóstico de uma possível lesão é um trabalho muito crítico e delicado. Muitas vezes o resultado da aplicação das técnicas de remoção de *speckles* é parcial ou totalmente dependente do tipo de ruído. Algumas técnicas de redução de manchas somente oferecem resultados satisfatórios com ruídos do tipo aditivo. Assim se faz necessário recorrer às técnicas que utilizam compressão logarítmica para converter ruídos multiplicativos, muito comuns em imagens de ultrassom, em ruídos aditivos (ABD-ELMONIEM et al., 2002).

3.2.1 Técnicas de filtragem

As técnicas de filtragem visam a redução e/ou remoção de ruídos em imagens. Trata-se de transformações aplicadas pixel a pixel na imagem. O objetivo é reconstruir a imagem degradada da maneira mais precisa possível, constituindo uma fase fundamental para a obtenção de resultados satisfatórios nas próximas etapas do sistema CAD. As técnicas de filtragem são categorizadas em termos de linear e não-linear.

Três requisitos são fundamentais na aplicação de filtros em imagens médicas: preservação das bordas da lesão mantendo detalhes da estrutura, redução dos ruídos (*speckles*) em regiões homogêneas e por fim, melhoramento das bordas (GERIG et al., 1992).

O filtro da média, do tipo linear, é uma técnica comumente utilizada na redução de ruídos. Trata-se de um método que calcula a média de todos os pontos (pixels) vizinhos de cada pixel da imagem. A implementação desse filtro é muito simples por envolver basicamente somas e divisões. O resultado da aplicação é o esmaecimento do ruído, porém com perda da nitidez da imagem (GONZALEZ; WOODS, 2002; CALÓOPE et al., 2004). Como geralmente as imagens de ultrassom apresentam manchas maiores (*speckles*) e não pequenos ruídos, a aplicação desse filtro não apresenta resultado tão satisfatório. As manchas presentes nas imagens de ultrassom são modelos multiplicativos com ruídos não-Gaussianos. O filtro da média é eficiente na redução de ruídos Gaussianos

aditivos orientados (no sentido) pelo erro quadrático médio. Na Figura 3.1 observa-se o resultado da aplicação do filtro da média em dois tipos de imagens. Conforme mencionado anteriormente o resultado da aplicação do filtro da média na imagem de ultrassom 3.1(b) não apresenta resultado satisfatório pois podem degradar regiões influentes no diagnóstico.

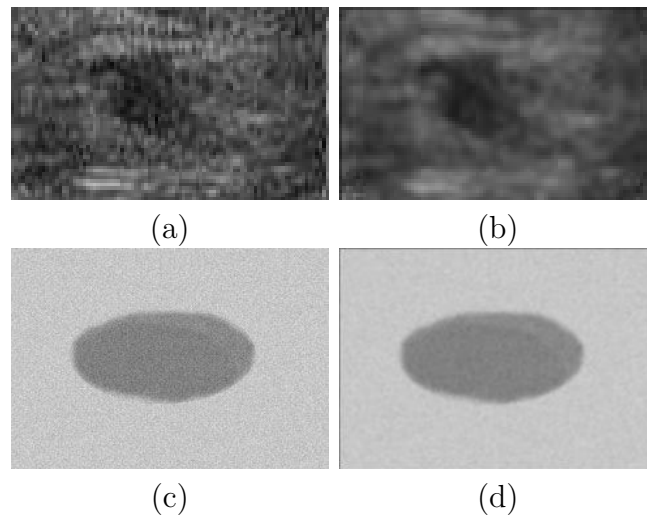


Figura 3.1: Imagem (b) resultante da aplicação do filtro da média em imagem ultrassono-gráfica de tumor com manchas *speckles* (a) e imagem (d) resultante da aplicação do filtro da média em imagem (c) artificial com pequenos ruídos aditivos Gaussianos artificiais.

Os filtros de média adaptativos, também do tipo linear, são empregados na atenuação do problema de nitidez mencionado anteriormente na descrição do filtro de média. A proposta desses é buscar um equilíbrio na aplicação dos filtros, utilizados na remoção de *speckles*, considerando as regiões homogêneas e as bordas das imagens. Tais filtros recorrem a artifícios estatísticos como média, variância e correlação para remover manchas preservando características importantes das imagens (WOODS; BIEMOND, 1984). Dentre os filtros de média adaptativos destacam-se os filtros de Lee, Kuan e Frost (KUAN et al., 1985; LEE, 1980; FROST et al., 1982; DUTT; GREENLEAF, 1996; DOMG et al., 2001). Estes tem como base o modelo de mancha multiplicativo representado por:

$$I(t) = R(t)u(t),$$

onde t é a coordenada (x,y) do pixel, $R(t)$ é a intensidade da imagem sem mancha, $I(t)$ é a intensidade observada e $u(t)$ é a mancha com média $\bar{u}$ e variância σ_u^2 .

Já na categorização de filtros não lineares destacam-se os filtros de ordem estatística (*Order-statistic filter*), os filtros MAP (*Maximum a Posteriori*) e os filtros de difusão não-

linear.

Os filtros de ordem estatística são eficientes na remoção de manchas representadas por uma função de densidade com variância média (CHENG et al., 2010). O filtro da mediana é um tipo de filtro de ordem estatística. Em geral, esse filtro é aplicado em imagens ruidosas mas apresenta resultados mais eficazes no tratamento de imagens com ruído impulsivo (GONZALEZ; WOODS, 2002). Há mais de 20 anos Loupas et al. (1989) desenvolveu um algoritmo denominado Filtro da Mediana Ponderado Adaptativo (*Adaptive Weighted Median Filter - AWMF*). Tal algoritmo proporcionou alcançar a redução máxima de manchas em regiões uniformes preservando as bordas e características das imagens. O AWMF é um filtro de mediana reforçado onde a mediana ponderada da região de pixels W_{ij} é definida como a mediana de uma sequência aumentada, formada pela replicação dos pixels em W_{ij} , por um montante calculado da sua distância para (i,j) e uma estimativa da homogeneidade h_{ij} da imagem local. Em um trabalho mais recente, Prager et al. (2002) demonstrou que os filtros adaptativos alcançaram melhores resultados perante o filtro da mediana.

O filtro MAP é outro tipo de filtro de ordem estatística. O filtro MAP estima um sinal não observado x por meio da maximização do conhecido teorema de Bayes (CALÓOPE et al., 2004). Para utilizar essa técnica de filtragem é necessário ter prévio conhecimento da função de densidade de probabilidade que representa a imagem a ser analisada.

Por fim, os filtros de difusão não linear também são adotados na redução de pequenas manchas em imagens. Esses são um tipo de filtro adaptativo onde uma função de detecção de borda controla a direção e força da difusão da imagem. Esses filtros adotam um modelo matemático baseado em Equação Diferencial Parcial (EDP) e, há anos, vem sendo estudados e melhorados. A difusão não linear anisotrópica foi primeiramente introduzida no trabalho de Perona e Malik (1990). Neste trabalho foram obtidos bons resultados no tratamento de imagens com ruídos Gaussianos aditivos. Porém esta técnica não apresentou eficácia no tratamento de imagens com ruídos multiplicativos pela dificuldade de tratamento das bordas da imagem.

Então, para tratamento de um tipo particular de imagem de ultrassom, foi proposto um novo método denominado *Speckle Reducing Anisotropic Diffusion (SRAD)*. Este método se mostrou eficaz no tratamento de imagens de ultrassom sem compressão logarítmica (YU; ACTON, 2002; ZHI; WANG, 2008). A compressão logarítmica, um tipo de

transformação na imagem, é adotada na expansão de valores dos pixels mais escuros e, ao mesmo tempo, compactação dos pixels de maior valor. A transformação *gamma* seria mais adequada a essa proposta, porém a função logarítmica apresenta uma característica muito importante de comprimir um intervalo dinâmico de imagens com grande variação dos valores dos pixels (CAVINATO, 2009). Consequentemente as regiões mais escuras serão realçadas enquanto as regiões mais claras serão degradadas. A expressão compressão logarítmica é também conhecida por variação dinâmica ou faixa dinâmica. A Figura 3.2 exhibe a comparação das imagens obtidas após a aplicação das transformações logarítmica e gamma.

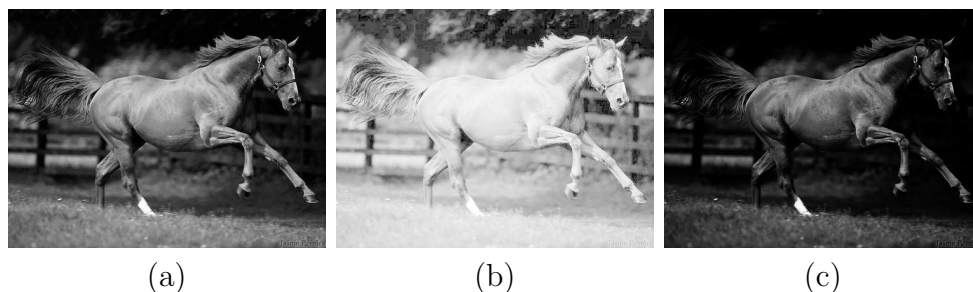


Figura 3.2: Imagem (a) original. Imagem (b) resultante da aplicação da transformação logarítmica e imagem (c) após a transformação gamma. Fonte: (CAVINATO, 2009, p 8).

Para manipular imagens de ultrassom com compressão logarítmica, foi proposto no trabalho de Abd-Elmoniem et al. (2002) o método denominado Difusão Reforçada Coerente Não Linear (*Nonlinear Coherence Enhancement Diffusion - NCD*). O NCD combina três diferentes modelos. De acordo com o tamanho da mancha e a anisotropia da imagem, o modelo NCD altera a imagem progressivamente da difusão isotrópica por meio da difusão anisotrópica para, finalmente, a curvatura média do movimento. O NCD é eficaz para encontrar pixels da borda e sua direção.

Outros filtros não lineares, como o filtro geométrico (*Geometric Filter - GF*) e o filtro de região de crescimento agressivo (*Aggressive Region Growing Filtering - ARGF*), também são empregados na redução eficiente de manchas e melhoramento das imagens. O filtro geométrico adota um algoritmo iterativo que altera os valores dos pixels em uma vizinhança especificada com base nos valores relativos desses pixels (CRIMMINS, 1985). Já o filtro ARGF aplica um algoritmo para selecionar uma região de filtragem na imagem usando, de forma apropriada, um limiar (valor) de homogeneidade estimado para a região de crescimento. As regiões homogêneas são suavizadas aplicando um filtro de média

aritmética e os pixels da borda são tratados por um filtro de mediana não linear (CHEN et al., 2003). Dentre outros, o trabalho de (CHEN et al., 2003) apresentou um comparativo dos resultados obtidos na aplicação do filtro ARGF e do filtro AWMF descrito anteriormente. Nesse trabalho o filtro ARGF se mostrou mais eficiente no tratamento de imagens.

As técnicas de filtragem são implementadas de maneira simples e rápida, porém ainda são limitadas ao tamanho e forma da janela de filtragem.

3.2.2 Técnicas de domínio Wavelet

Com o mesmo objetivo de outras técnicas comumente aplicadas, como a difusão anisotrópica, as técnicas de domínio wavelet também têm sido empregadas na redução de manchas e melhoramento de imagens. O modelo NCD descrito anteriormente, baseado na difusão anisotrópica, é eficaz para encontrar pixels da borda e sua direção. Porém, este modelo apresenta dificuldade para encontrar pixels formadores de bordas de tamanhos variados. E ainda a orientação da borda pode não representar características globais da imagem em decorrência da abordagem de escala simples presente nesse modelo. Segundo Quistgaard (1997), as imagens de ultrassom podem apresentar bordas de vários tamanhos.

Neste contexto são apresentadas as técnicas de domínio wavelet. Tais técnicas, pertencentes à abordagem multiescala, são aplicadas em imagens por meio de transformação discreta. A motivação da adoção dessas técnicas é a possibilidade de filtrar bordas de vários tamanhos por meio da utilização de informações de espaço e frequência (KIM; RA, 2005). Os filtros baseados na transformação wavelet são compostos de três etapas principais. Primeiramente a imagem é decomposta pela transformação discreta wavelet (*Discrete Wavelet Transform - DWT*). Na sequência os coeficientes wavelet são alterados para remover ruídos e manchas e, finalmente, a imagem é reconstruída por meio da transformação wavelet inversa (*Inverse Discrete Wavelet Transform - IDWT*). A transformação wavelet é adequada para compactação de energia. Possivelmente os coeficientes menores representam ruídos enquanto que os maiores representam características importantes da imagem.

Segundo Cheng et al. (2010), as técnicas de transformação wavelet são divididas em três grupos: encolhimento wavelet (*wavelet shrinkage*), remoção de manchas wavelet sob estrutura Bayesiana (*wavelet despeckling under Bayesian framework*) e difusão e filtragem wavelet (*wavelet filtering and diffusion*). O encolhimento wavelet é baseado no limiar do

coeficiente wavelet. Ele reduz os coeficientes que representam ruídos enquanto mantém os coeficientes que possivelmente representam características da imagem (HAO et al., 1999). O método wavelet de remoção de manchas sob estrutura Bayesiana é uma alternativa à técnica de limiar padrão. Este tem como base o conhecimento da estatística do coeficiente wavelet, obtido pela aplicação de regras Bayesianas. A desvantagem desse método é a dependência, a priori, da distribuição da imagem sem ruído. No contexto de imagens ultrassonográficas, essa dependência ora citada pode inviabilizar a aplicação do método dado que as pequenas manchas são inerentes à natureza das imagens de ultrassom (ACHIM et al., 2001).

Por fim, também podem ser utilizadas técnicas de difusão e filtragem no domínio wavelet com o propósito de reduzir manchas *speckles* nas imagens. No trabalho de Rangsanteri e Prasongsook (2002) foi proposto um filtro denominado Wiener. Em comparação com o método de limiar simples (encolhimento wavelet), os resultados da aplicação deste filtro se mostraram mais eficientes nos aspectos visual e quantitativo.

3.2.3 Técnicas de composição

Diferentemente das técnicas de filtragem as técnicas de composição são aplicadas no processo de aquisição da imagem. Trata-se da aquisição simultânea da imagem resultante de vários cortes em diferentes ângulos de insonação por meio de um transdutor especial. A imagem final é resultante da composição, em tempo real, da média dos vários cortes independentes em diferentes ângulos. Consequentemente, observa-se uma melhora significativa na detecção de lesões com baixo contraste quando comparado com imagens convencionais, obtendo-se bons resultados na avaliação de tumores de mama (FORSBERG, 2004). A Figura 3.3 apresenta duas imagens de um mesmo carcinoma mamário diagnosticado na imagem de ultrassom. A imagem 3.3(a) foi obtida sem a utilização da técnica de composição e a imagem 3.3(b) foi resultado do processo de aquisição utilizando-se a técnica de composição. É nítida a melhora alcançada na qualidade da imagem 3.3(b).

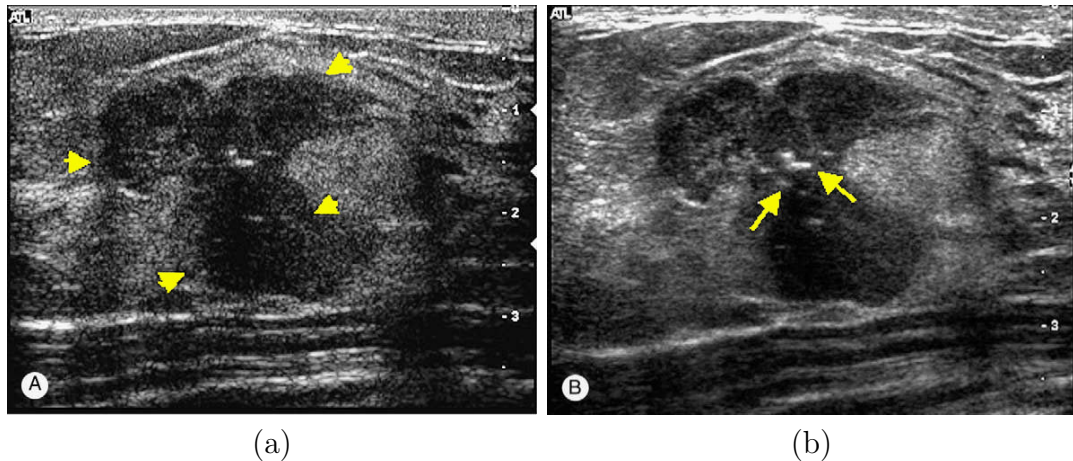


Figura 3.3: Aquisição de imagens ultrassonográficas com uso (a) e sem uso (b) de técnicas de composição. Fonte: (FORSBERG, 2004, p 25).

A Figura 3.4 apresenta uma síntese das técnicas de pré-processamento apresentadas, destacando as vantagens e desvantagens da aplicação das mesmas.

Técnica	Vantagem	Desvantagem
Filtros	Simple e rápido.	É sensível ao tamanho e forma da janela de filtragem. A representação de escala única dificulta a discriminação entre ruídos e sinais.
Wavelet	No domínio Wavelet a estatística dos sinais é simplificada. Ruídos e sinais são processados em escalas e orientações diferentes.	Cálculos da DWT / IDWT são computacionalmente complexos.
Composição	Simple.	Requer suporte de hardware.

Figura 3.4: Síntese das técnicas aplicadas na redução de manchas e melhoramento de imagem. Adaptado de Cheng et al. (2010, p 304).

3.3 Segmentação de Imagens

De forma genérica, este macro processo visa a extração de regiões de interesse da imagem. A segmentação é uma das etapas mais críticas e difíceis do processo de classificação de imagens. Um procedimento adequado favorece o alcance de um bom resultado final enquanto que uma segmentação mal realizada sempre determina falhas no processo final de classificação. No campo da imagiologia ultrassonográfica de tumores mamários, o objetivo da fase de segmentação é tentar extrair regiões da imagem que possam contribuir para um resultado de classificação mais eficiente.

Os algoritmos de segmentação são baseados em duas propriedades básicas: valores de intensidade de descontinuidade e valores de intensidade de similaridade. Na primeira categoria, a segmentação é realizada observando alterações abruptas de intensidade dos

pixels. As bordas de um objeto na imagem podem ser consideradas alterações abruptas e nessa categoria pode-se destacar as técnicas de segmentação que fazem uso de Operadores Morfológicos e do modelo de contorno ativo (*snake*). Já na segunda categoria, as imagens são segmentadas com base na semelhança de regiões em acordo com critérios pré-definidos. Nesta categoria pode-se destacar as técnicas de segmentação que fazem uso, por exemplo, de histograma de limiarização. Existem muitas técnicas de segmentação e, a seguir, são apresentadas as técnicas mais pertinentes a este trabalho.

3.3.1 Operadores morfológicos

Em Gomez et al. (2009) foi proposto um método de segmentação que evidencia as diferenças morfológicas das imagens ultrassonográficas de mama. São utilizados operadores morfológicos para segmentar as lesões nas imagens. Neste estudo foram aplicadas três técnicas conhecidas: filtragem morfológica, transformação morfológica *Watershed* e função Derivada Radial Média (*Average Radial Derivative function*). A proposta contempla uma segmentação semi-automática, combinando as diferentes técnicas morfológicas citadas, para pré-processamento das imagens e maximização da função Derivada Radial Média objetivando selecionar o limiar de nível de cinza ótimo correspondente ao contorno da lesão.

A Morfologia Matemática é uma ramificação da área de processamento de imagem aplicada na análise de formas. “Os operadores morfológicos podem simplificar os dados de informação de uma imagem, preservando as características essenciais e removendo estruturas irrelevantes” (HUANG; CHEN, 2004). A Figura 3.5 exhibe o resultado comparativo da segmentação manual e da segmentação semi-automática com operadores morfológicos. As linhas de cor vermelha foram traçadas manualmente por radiologistas experientes e as linhas de cor amarela foram resultantes do processo de segmentação semi-automático utilizando operadores morfológicos. Pelo resultado observado, constatou-se que os operadores morfológicos podem oferecer bons resultados na segmentação de lesões mamárias em imagens de ultrassom.

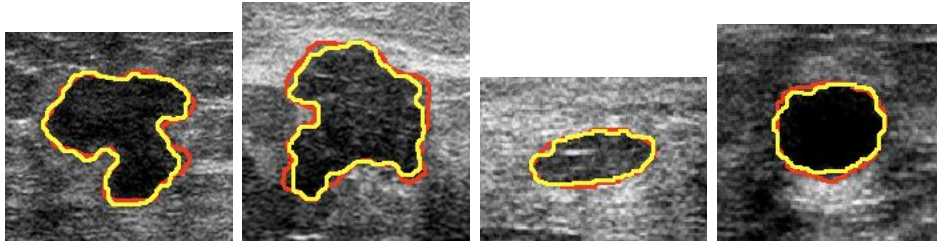


Figura 3.5: Resultado comparativo da segmentação manual e da segmentação semi-automática utilizando Operadores Matemáticos. Fonte: (GOMEZ et al., 2009).

3.3.2 Campos aleatórios de Markov

Na área de processamento de imagens, frequentemente, os modelos Markovianos são utilizados como conhecimento *a priori* no tratamento de diversos tipos de problemas como filtragem de ruídos e segmentação de objetos.

O processo de segmentação de imagens de ultrassom pode ser considerado um problema de rotulagem, cuja solução é atribuir um conjunto de rótulos para os pixels. Tal abordagem, baseada em textura, pode ser representada naturalmente pelos Campos Aleatórios de Markov (Markov Random Field - MRF) (CHANG et al., 2003b; PANJWANI; HEALEY, 1995). O processo de análise de imagens, que adota o modelo MRF, é formulado como um problema de estimação do *Maximum a Posteriori (MAP)*. A resolução deste problema de contorno envolve a minimização de energia.

Os modelos MRF têm sido empregados com bons resultados na segmentação de imagens de ultrassom, incluindo imagens de mama (BOUKERROUI et al., 1998, 2003; NOBLE; BOUKERROUI, 2006). Em um trabalho de revisão, detalhado e abrangente, (NOBLE; BOUKERROUI, 2006) analisaram os métodos de segmentação de imagens ultrassonográficas propostos até a publicação do artigo. Nesse trabalho foram destacados dez estudos cujos resultados demonstraram maior utilidade clínica ou segmentação eficiente das imagens de ultrassom.

Em razão da heterogeneidade de intensidade dos tecidos e da presença de ruídos nas imagens de ultrassom, o processo de segmentação torna-se mais complexo. Métodos de segmentação convencionais baseados em intensidade não oferecem resultados tão atraentes. Nesse cenário, Lihua et al. (2007) propôs um novo método de segmentação que contempla a combinação do modelo MRF com processamento morfológico da imagem de ultrassom. O objetivo é minimizar as deficiências dos métodos convencionais. Neste, o modelo MRF foi utilizado para estimar o rótulo da imagem e a fase de processamento

morfológico foi adotada para eliminação de artefatos e suavização das bordas segmentadas.

3.3.3 Histograma de limiarização

A técnica de histograma de limiarização é também utilizada para separar objetos do fundo da imagem. Para exemplificar, considere uma imagem, em nível de cinza, com um objeto escuro em fundo claro de tal forma que os pixels estejam agrupados em duas categorias dominantes (a mais clara e a mais escura). Uma maneira de separar o objeto do fundo é a adoção de um valor limiar de intensidade T . Se o valor do pixel (x, y) for maior ou igual a T então este pixel pertence ao objeto e no caso contrário pertence ao fundo da imagem.

Em seu trabalho, Horsch et al. (2001) propôs a utilização do histograma de limiarização em imagens de ultrassom da mama. Neste trabalho foi apresentado um algoritmo para segmentação de massas nas imagens de ultrassom. A implementação de tal algoritmo necessita de tratamento prévio da imagem por meio de filtros de mediana e, ainda, seleção manual ou semi-manual do centro, da largura e da altura da lesão na imagem de ultrassom.

O histograma de limiarização é uma técnica simples que proporciona bons resultados na segmentação de imagem cujo histograma é bimodal. Em imagem com histograma unimodal o resultado não é muito eficaz (CHENG et al., 2010). As características da imagem são determinantes no resultado do processo de segmentação. Nem sempre é possível separar o objeto do fundo da imagem, especialmente se o fundo tiver cores semelhantes ao objeto que se deseja isolar. Na Figura 3.6, pode-se observar duas imagens ultrassonográficas de tumor mamário com seus respectivos histogramas unimodal e bimodal utilizadas neste trabalho. Como já esperado, utilizando um limiar de intensidade (cor do pixel) de valor 50, a segmentação do tumor na imagem de ultrassom 3.6(b), cujo histograma é bimodal, foi bem eficiente. No entanto, na imagem de ultrassom com histograma unimodal o resultado não foi bom. Neste caso, a região do tumor acabou se misturando com outras partes da imagem que fazem parte do fundo. Como as imagens de ultrassom testadas estão em escala de cinza, o valor de intensidade do pixel pode variar entre 0 e 255. Ainda vale ressaltar que, nesse experimento, o limiar de intensidade foi escolhido de forma manual. Existem também técnicas para seleção automática desse valor de limiar.

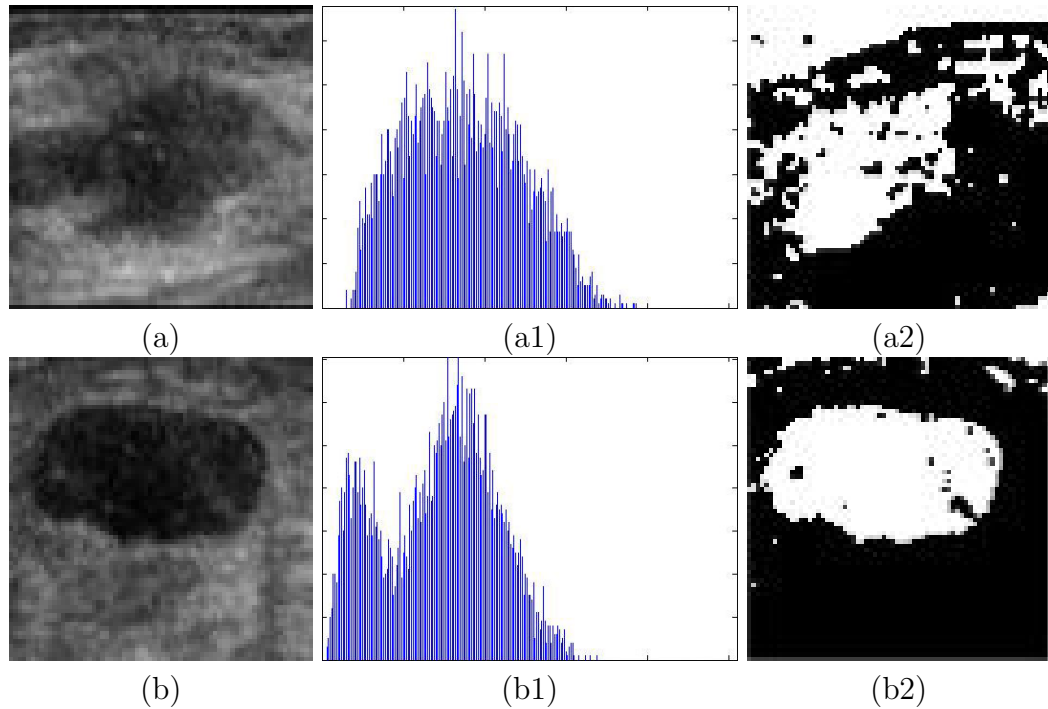


Figura 3.6: Imagens de ultrassom com tumor mamário (a, b) com seus respectivos histogramas unimodal e bimodal (a1, b1) e imagens binárias (a2, b2) obtidas no processo de segmentação com limiar de intensidade de pixel de valor 50.

3.3.4 Modelo de contorno ativo

O modelo de contorno ativo (*Active Contour Model - ACM*) é também conhecido como *snake*. Trata-se de um modelo onde curvas se movem dentro da imagem com objetivo de encontrar as bordas do objeto desejado. Trata-se de um processo de deformação onde o modelo inicial tem sua forma modificada ativamente aproximando-se do contorno desejado por meio da minimização de uma função de energia formulada. A força, aplicada na deformação, é calculada a partir da energia interna e externa (KASS et al., 1987). De forma geral nesse modelo, a segmentação segue três etapas: formulação da função de energia, dedução de uma equação diferencial de evolução para encontrar o mínimo da função de energia e solução da equação diferencial de evolução para a obtenção da segmentação (LIU et al., 2010). Uma desvantagem desse modelo é a necessidade de inicialização, seja manual ou semi-automática.

No trabalho de Huang e Chen (2005), o ACM foi aplicado na segmentação precisa de imagens 2D de ultrassom com tumor mamário. Neste trabalho, o ACM foi adotado em conjunto com a técnica de transformação *watershed*. Primeiramente a imagem passa pela transformação *watershed* visando a obtenção automática do contorno inicial da lesão

mantendo a forma e borda do tumor. Na sequência, o método ACM é aplicado para determinar, de forma automática, o contorno fino do tumor. Como resultado, as regiões segmentadas ficaram muito semelhantes às regiões delimitadas manualmente por radiologistas experientes, indicando se tratar de uma abordagem eficiente. Chang et al. (2005) também aplicou o modelo ACM em imagens ultrassonográficas de tumor mamário e obteve resultados bem satisfatórios.

Em outro trabalho, Jumaat et al. (2010) comparou o resultado da aplicação de dois métodos, baseados no modelo ACM, na segmentação de massas em imagens de ultrassom da mama. Os métodos, denominados *Balloon Snake* e *Gradient Vector Flow (GVF)*, foram aplicados em imagens artificiais e também em imagens reais destacando melhor resultado do método *Balloon Snake*.

Já em Liu et al. (2010) foi proposto um novo modelo de contorno ativo baseado em deformação de níveis. No método clássico de deformação de níveis uma equação diferencial deve ser formulada explicitamente para evolução da função auxiliar de deformação de níveis. Mas a formulação de tal equação diferencial nem sempre é conveniente e simples para modelar o problema de segmentação. Para minimizar este problema, foi então proposto um novo método, cuja principal diferença do modelo clássico apresentado no início desta seção, é a introdução de funções auxiliares de deformação de níveis nas funções de energia. Com essa abordagem os resultados dos experimentos foram mais eficazes perante os gerados na aplicação do modelo ACM clássico. De modo semelhante ao modelo de Campos Aleatórios de Markov, a resolução deste problema envolve a minimização de energia.

A geração automática de um contorno inicial adequado é uma tarefa difícil e os processos de deformação baseados no modelo ACM são computacionalmente custosos (CHENG et al., 2010).

3.3.5 Outros modelos

Liu et al. (2005) propôs um sistema automático de segmentação de lesões em imagens ultrassonográficas de mama. Tal sistema é composto por duas fases principais, o pré-processamento da imagem e a segmentação em si. Primeiramente, na fase de pré-processamento, é aplicada a técnica de filtragem não linear por difusão anisotrópica. Na sequência é aplicada uma máscara de ofuscamento (*Unsharp masking*) para melhorar a

aparência das bordas da possível lesão. Na Figura 3.7, pode-se observar as etapas do processo de segmentação proposto por Liu et al. (2005).

A fase de segmentação também é dividida em dois passos principais. No primeiro passo a imagem é segmentada utilizando a técnica de corte normalizado (*Normalized Cut Technique - Ncut*) proposta no trabalho de Shi e Malik (2000). A Ncut trata a segmentação como um problema de particionamento de grafos. Essa abordagem maximiza a dissimilaridade total entre os diferentes grupos e também a similaridade total dentro dos grupos. O resultado da aplicação da técnica Ncut é a segmentação da imagem em pequenas partes como observado na Figura 3.7(c). A Ncut é considerada uma técnica de segmentação não supervisionada. Por fim, é feita a junção das regiões segmentadas adjacentes em partes maiores. Tal junção, vista na Figura 3.7(d) foi baseada em regras empíricas de tamanho, forma e características de textura utilizadas por radiologistas no diagnóstico de lesões mamárias. Ainda nessa junção foi considerado que a lesão estava localizada na região central da imagem.

No trabalho de Rodrigues et al. (2006a) foi apresentado um método de segmentação da imagem que adota um algoritmo denominado *Algoritmo Recursivo de Segmentação Não Extensiva (Non-Extensive Segmentation Recursive Algorithm - NESRA)*, derivado da bem conhecida entropia de Boltzman-Gibbs-Shannon. O algoritmo foi aplicado na segmentação de imagens mamárias de ultrassom e apresentou ótimos resultados na separação da lesão do fundo da imagem.

Não menos importantes, mas em número menor, outros trabalhos utilizando métodos alternativos, como K-Means e Lógica Fuzzy, também tem sido apresentados na área de segmentação de imagens de ultrassom (LIU et al., 2006; KOTROPOULOS; PITAS, 2003).

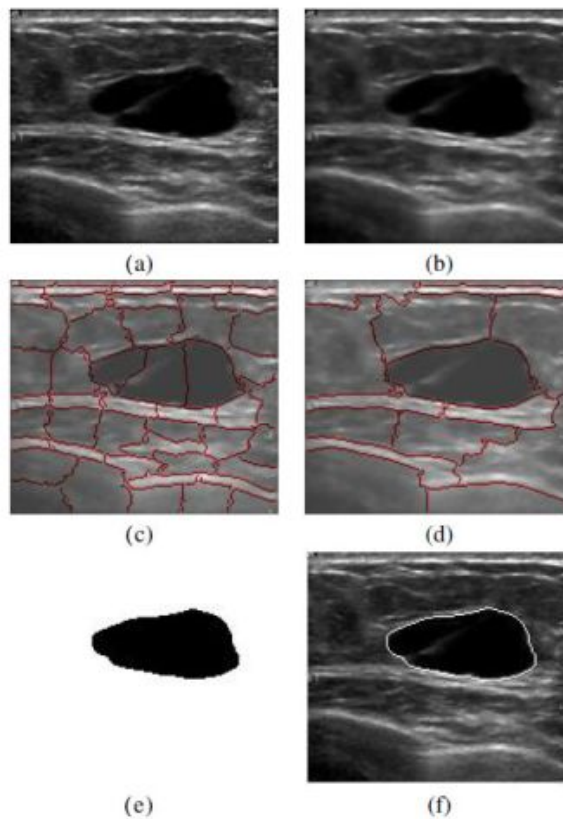


Figura 3.7: Imagem original (a). Imagem resultante da fase de pré-processamento (b). Segmentação da imagem utilizando a técnica Ncut (c). Junção das regiões adjacentes (d). Imagem binária com a região da lesão identificada (e). Resultado final da segmentação (f). Fonte: (LIU et al., 2005).

3.4 Extração e Seleção de Características

Este macro processo trata da extração e seleção de características discriminantes da imagem. Uma análise adequada pode extrair, de forma eficiente, características e, ainda, reduzir o espaço de dimensionalidade dessas características evitando o clássico problema da *maldição da dimensionalidade* (BELLMAN, 1961). A dimensionalidade é atribuída a quantidade de características que descrevem as amostras, ou seja, a dimensão do espaço de características. Análises, sem critérios bem definidos, podem gerar resultados com baixa significância e ainda consumir recursos desnecessários.

O problema da *maldição da dimensionalidade* é associado à situação onde o número de amostras de treinamento é dado por uma função exponencial da dimensão do espaço de características. Nesse cenário exponencial torna-se extremamente necessário a redução de dimensionalidade onde os padrões medidos possuem um número elevado de características e apenas um número limitado de amostras de treinamento (JAIN et al., 2000). A

redução de dimensionalidade tem como objetivo analisar somente as características mais salientes proporcionando melhor desempenho e eficácia. Técnicas para reduzir tal dimensionalidade, como a Análise das Componentes Principais (*Principal Component Analysis* - *PCA*), são fortemente recomendadas (JOHNSON; WICHERN, 2007; HASTIE et al., 2009). Geralmente a fase de extração de características é precedida pela fase de seleção. Após a extração das características dos dados, um algoritmo de seleção elimina as características mais irrelevantes em acordo com um critério determinado e, conseqüentemente reduz a dimensionalidade.

A escolha entre seleção e extração de características depende do domínio da aplicação e do conjunto específico de dados de treinamento disponíveis. Em geral a seleção de características reduz o custo de medição de dados, e as características selecionadas mantêm sua interpretação física original, mantendo as propriedades que possuíam quando foram criadas. Já as características transformadas obtidas na extração de características podem prover uma habilidade de discriminação mais eficaz que o melhor subconjunto das características originais. Mas as novas características (combinações lineares ou não lineares das características originais) podem não possuir um significado físico.

Segundo o estudo de Cheng et al. (2010), as características das imagens de ultrassom mamárias são divididas em quatro grupos: características de textura, morfológicas, baseadas em modelo e descritivas. Ainda nesse estudo foram citadas e comentadas, de forma resumida, as características típicas e comprovadamente eficazes. A seleção das características utilizadas é totalmente dependente do tipo de imagem. A adoção de características inadequadas certamente comprometerá o resultado final da classificação.

Na categoria de características de textura pode-se destacar os coeficientes de auto covariância e a sombra acústica posterior. Os coeficientes de auto covariância são tradicionalmente adotados para identificar a correlação interna dos pixels dentro da imagem (KUO et al., 2002; HUANG et al., 2006). Já a característica de sombra acústica posterior, também conhecida como comportamento acústico posterior, é quantificada por meio da comparação dos valores de nível de cinza posteriores à lesão com os valores de nível de cinza em tecidos adjacentes de mesma profundidade (DRUKKER et al., 2005). A fim de evitar sombras da borda define-se uma região de interesse que é a própria região da lesão sem uma parte de suas laterais. As regiões da esquerda, posterior e direita tem forma de retângulo e também possuem a mesma largura da região de interesse definida anteri-

ormente (HORSCH et al., 2002). O esquema dessas regiões de interesse pode ser visto na Figura 3.8. A característica de sombra acústica considerada é o mínimo da diferença dos lados (*Minimum Side Difference - MSD*) e sua definição é dada por:

$$MSD = \min(A_{post} - A_{left}, A_{post} - A_{right}), \quad (3.1)$$

onde A_{left} , A_{post} e A_{right} são as médias dos valores de nível de cinza sobre a região de interesse.

O mínimo da diferença dos lados é adotado por causa da presença frequente de sombra acústica nas lesões malignas.

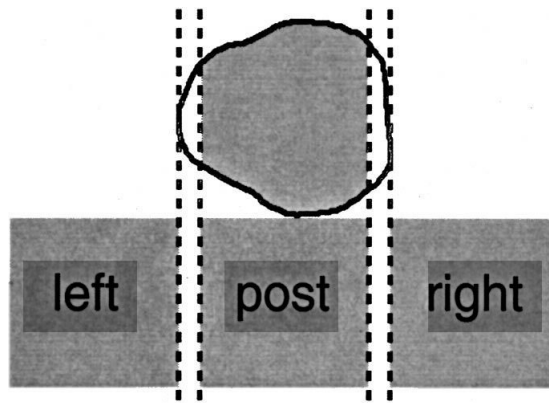


Figura 3.8: Esquema das regiões de interesse utilizadas na definição da característica sombra acústica posterior. Adaptado de (HORSCH et al., 2002, p 159)

Por sua vez, na categoria de características morfológicas, o foco é direcionado para os atributos de forma e borda da imagem. Nesse contexto, dentre outros métodos, são adotadas técnicas conhecidas como *convex hull* e *concave polygon* (CHEN et al., 2003). Para ilustrar, a Figura 3.9 demonstra o funcionamento das técnicas *convex hull* e *concave polygon* respectivamente. Na Figura 3.9 o tamanho da área das regiões A_1 , A_2 , A_3 e A_4 pode indicar uma possível lesão maligna dado que esta, na maioria da vezes, apresenta formato muito irregular.

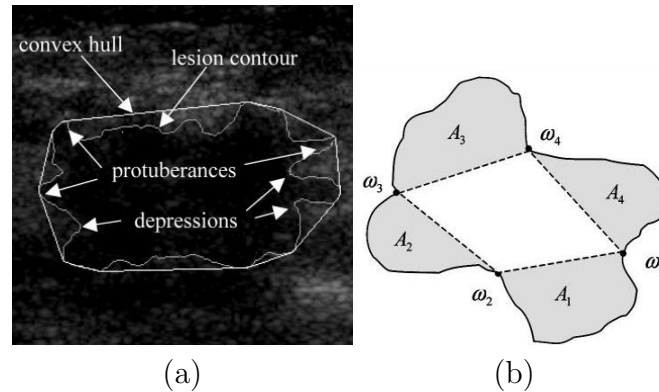


Figura 3.9: Região delimitada pelo algoritmo *convex hull* (a) e regiões A_1 , A_2 , A_3 e A_4 obtidas na aplicação da técnica *concave polygon* (b). Fonte: (CHEN et al., 2003, p 506).

Ao contrário dos outros tipos, as características baseadas em modelo são especiais e complexas por tratarem atributos particulares das imagens mamárias de ultrassom. Nesse tipo de imagem são focados os ecos de retroespalhamento do tecido mamário. Modelos orientados pelas distribuições *Nakagami* e *K de Jakeman* têm sido experimentados mas a estimativa dos parâmetros destes é de alta complexidade (CHENG et al., 2010).

Adicionalmente, na categoria de características descritivas são considerados critérios empíricos dos profissionais da área de saúde (médicos e radiologistas). Características como formato da lesão e presença de calcificações são os parâmetros utilizados para distinguir uma lesão maligna ou benigna. Também são aplicados métodos computacionais e estatísticos mas a observação empírica de atributos da imagem ainda indica o resultado do diagnóstico.

3.5 Classificação

Aplicada após a etapa de extração e seleção de características e dependente da mesma, a fase de classificação visa a categorização das imagens de ultrassom em termos de lesão e não lesão e, no caso positivo de lesão, categorização em termos de benigna e maligna. A proposta deste trabalho é direcionada à classificação de imagens ultrassono-gráficas de mama com presença de lesão (tumor).

Técnicas estatísticas multivariadas como LDA (*Linear Discriminant Analysis*) e SVM (*Support Vector Machine*) podem ser utilizadas nessa etapa de classificação (VAPNIK, 1998; FISHER, 1936; HASTIE et al., 2009). A classificação de imagens serve-se cada vez mais das técnicas de estatística multivariada para, entre outros propósitos, discriminação

de grupos. A denominação multivariada corresponde às técnicas que utilizam, de forma simultânea, todas as características (variáveis) na interpretação do conjunto de dados. Por sua vez, a análise estatística univariada, também abordada na seção de experimentos deste trabalho, propõe a análise de cada característica separadamente. Para alguns outros métodos de classificação avançados, tais como redes neurais artificiais, a dimensão dos vetores de características não só afeta o desempenho da classificação, mas também determina o tempo de treinamento do algoritmo. Assim, extrair recursos úteis e fazer uma boa seleção das características é uma tarefa crucial para os sistemas de classificação em geral.

LDA é um conhecido modelo probabilístico de classificação utilizado na extração de características. Este tem sido aplicado, com resultados satisfatórios, em muitos problemas de reconhecimento estatístico de padrões. O SVM é, essencialmente, um classificador de duas classes que maximiza a área vazia na proximidade da fronteira de decisão, isto é, maximiza a distância para os padrões de treinamento mais próximos entre as classes. Segundo Sato et al. (2009), o SVM detém a capacidade de minimizar o erro de classificação empírica considerando a complexidade do modelo. Essa propriedade, muito significativa, proporciona um bom poder de generalização e também evita problemas de *overfitting*. O termo *overfitting* é atribuído ao problema onde o modelo estatístico descreve erros e ruídos ao invés de relações pertinentes. Nesse problema, o modelo descreve bem os dados de treinamento mas não apresenta bom desempenho na avaliação de outros dados. A técnica SVM também pode ser estendida para classificação de mais de duas classes. Ao contrário da LDA, o método SVM é não-paramétrico, ou seja, não considera distribuição de probabilidade.

No estudo de (RODRIGUES et al., 2006a) foi proposto uma metodologia de classificação automática de lesão mamária em imagens de ultrassom. Este estudo contemplou a execução de cinco etapas. Na primeira delas foi feita a segmentação das imagens por meio de um algoritmo denominado Algoritmo Recursivo de Segmentação Não Extensiva (*Non-Extensive Segmentation Recursive Algorithm - NESRA*), derivado da bem conhecida entropia de *Boltzman-Gibbs-Shannon (BGS)*. A segunda etapa tratou da melhoria da segmentação e da extração da região de interesse (lesão) por meio da abordagem de cadeia morfológica e, em seguida na terceira etapa, foi feita a extração do contorno da lesão. Na quarta etapa foi destacada a extração de características da região de interesse. Nesta,

cinco características foram definidas, por radiologistas, como discriminantes entre lesões benignas e malignas. Então essas cinco características (área, circularidade, protuberância, heterogeneidade e sombra acústica) foram combinadas visando alcançar melhores resultados no processo de classificação realizado na quinta etapa. Para separar as lesões malignas e benignas foi aplicado o classificador SVM cujas entradas foram as características combinadas testadas na quarta etapa. Neste estudo ainda foram testados três tipos de kernel do classificador SVM: *BSpline*, polinomial e exponencial.

Já Giraldi et al. (2008b) apresentaram um método, denominado *Discriminant Feature Analysis (DFA)*, para seleção e reconstrução das características mais discriminantes das amostras. Tal método propõe a utilização de pesos discriminantes dados por hiperplanos de separação obtidos na aplicação das técnicas lineares de classificação SVM e MLDA. Esta última é uma extensão do LDA para problemas de poucas amostras (THOMAZ et al., 2006). Os hiperplanos obtidos são então utilizados para selecionar as características mais discriminantes visando a separação dos grupos de amostras. A DFA foi aplicada na classificação de imagens de face e também em imagens mamárias ultrassonográficas com presença de lesões. Neste estudo ainda foram analisadas, em imagens de face, o desempenho da extração de características utilizando a técnica PCA. Em imagens de face o estudo confirmou que a técnica PCA capturou características que têm uma variação considerável entre as amostras, como gênero e formato da cabeça. No entanto, PCA não apresentou eficiência no tratamento de imagens com variações específicas como as alterações de face (expressão séria e expressão de sorriso).

Nas imagens mamárias de ultrassom, foco do trabalho apresentado aqui, destacou-se neste estudo a classificação dos tumores em termos de benigno e maligno por meio da combinação das mesmas características das imagens apresentadas no trabalho de Rodrigues et al. (2006a). Após a segmentação, as características extraídas foram utilizadas pelo classificador SVM para discriminação dos grupos benigno e maligno. Os resultados foram satisfatórios e eficientes. Em outro estudo, os mesmos autores apresentaram e discutiram aspectos avançados da aplicação de diferentes metodologias de aprendizado estatístico supervisionado em imagens faciais e médicas. Neste estudo de caso foram analisados e apresentados os resultados da aplicação do modelo *Perceptron* (redes neurais), SVM e LDA com resultado mais favorável ao uso do modelo SVM no aspecto de classificação (GIRALDI et al., 2008a).

Em Chang et al. (2003c), o problema da classificação em dois grupos (benigno e maligno) foi endereçado utilizando a característica de textura das imagens mamárias de ultrassom com tumor. Positivamente as características de textura refletem a correlação entre os pixels dentro da imagem e é mensurada pelo cálculo dos coeficientes de auto correlação e auto covariância. Neste estudo comparativo foram analisados os resultados da aplicação dos métodos de classificação SVM e rede neural do tipo *Multi-Layered Perceptron (MLP)*. As informações de textura foram utilizadas como entrada para o classificador e o resultado indicou a técnica SVM como mais robusta e eficaz na classificação de tumores mamários baseada em características de textura.

Já em Su e Wang (2010), a classificação de tumores mamários em imagens de ultrassom foi proposta utilizando uma técnica denominada agrupamento de Propagação de Afinidade (*Affinity Propagation (AP) clustering*). Também foram extraídas características morfológicas e de textura das imagens a fim de servirem de entrada para o classificador. Neste estudo o agrupamento AP é apresentado como um classificador primário de tumores benignos e malignos cujo parâmetro de entrada é uma coleção real de valores de similaridade entre os pontos fornecidos. Neste estudo também os resultados foram comparados com os obtidos na aplicação de classificadores populares como SVM e redes neurais. Nessa comparação dos métodos foi destacada uma característica muito interessante do classificador AP que atingiu um índice de acurácia de 94%. O AP não necessitou de processo de treinamento para alcançar um bom desempenho na classificação de tumores mamários.

No estudo de Shi et al. (2010) foi proposto também um novo método de classificação denominado *Fuzzy Support Vector Machine (FSVM)*. Assim como na maioria dos métodos de classificação este também recebe como entrada as características extraídas numa fase anterior à classificação. Este estudo considerou três tipos de características: de textura, fractais e baseadas em histograma. Em seguida, por meio de técnica de regressão, foi selecionado um subconjunto ótimo de características para avaliação do classificador. O FSVM é um classificador SVM com kernel polinomial baseado em lógica fuzzy. Neste modelo obrigatoriamente é necessário definir uma associação para cada uma das amostras de entrada. Como resultado final, o FSVM demonstrou robustez e acurácia na classificação de tumores mamários em imagens de ultrassom.

Em um trabalho recente utilizando redes neurais, Menechelli et al. (2010) propôs um método de classificação para separar as lesões, presentes nas imagens, em dois grupos

distintos (lesão benigna e maligna). Nesta pesquisa foram utilizadas imagens mamográficas não reais já segmentadas e também imagens mamográficas reais cuja segmentação da lesão foi feita pela aplicação da técnica *Watershed* comentada anteriormente. A rede neural, do tipo MLP, aplicada com algoritmo de *Back Propagation*, apresentou bons resultados na classificação de tumores benignos e malignos com base na característica *forma* da lesão. A segmentação e a análise da lesão foram orientadas pelo padrão de formas de nódulos mamográficos estabelecido pelo BI-RADS (*Breast Image Reporting and Data System*).

Dentre outros, os resultados promissores dos trabalhos de Thomaz et al. (2007), Giraldi et al. (2008b, 2008a) e Sato et al. (2009) motivaram outros estudos, como o que se desenvolve aqui.

3.5.1 Análise estatística univariada

A análise estatística univariada, discutida nessa seção, é baseada em um modelo linear geral (*General Linear Model - GLM*), que reúne vários modelos estatísticos diferentes. Dentre outros, pode-se citar a regressão linear, o teste t, a análise da variância (ANOVA) e a análise da covariância (ANCOVA).

Teste de hipóteses

O teste de hipóteses, aqui detalhado, será adotado para identificação das diferenças entre os grupos de imagens de tumores benignos e malignos. Usando o teste t como teste de significância, temos que esta diferença é dada basicamente pelas diferenças entre as médias de cada grupo ponderada por um desvio padrão do espalhamento das amostras (LEÃO et al., 2009), ou seja:

$$t_k = \frac{\bar{x}_{1k} - \bar{x}_{2k}}{\sigma_k \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}, \quad (3.2)$$

onde t_k é o t-valor da variável k, $\bar{x}_{1k}$ e $\bar{x}_{2k}$ são as médias respectivas da variável k do grupo 1 e do grupo 2. σ_k é o desvio padrão ponderado da variável k, N_1 e N_2 são respectivamente o total de amostras do grupo 1 e do grupo 2. O desvio padrão ponderado do conjunto de amostras é dado pela equação:

$$\sigma_k = \sqrt{\frac{(N_1 - 1)(\sigma_{1k})^2 + (N_2 - 1)(\sigma_{2k})^2}{N_1 + N_2 - 2}}, \quad (3.3)$$

onde σ_{1k} e σ_{2k} são respectivamente o desvio padrão da variável k para os grupos 1 e 2.

Trata-se de um teste que visa a identificação de diferenças entre os grupos com diferentes níveis de significância. Neste teste é verificada a significância de cada pixel em relação ao pixel correspondente em todas as imagens amostrais. O teste de hipótese estatística fornece uma afirmação acerca dos parâmetros de uma ou mais populações (testes paramétricos) ou acerca da distribuição da população (MAGALHÃES; LIMA, 2010).

Calculando o mapa de t-valores para os grupos analisados, ou seja, os t-valores de cada um dos pixels do conjunto amostral, o conceito de hipótese nula (H_0) e hipótese alternativa (H_1) pode ser utilizado para definir as variáveis que apresentam diferenças significativas. A hipótese nula sugere que as médias das variáveis amostrais do grupo 1 (tumores benignos) são iguais às médias correspondentes das amostras do grupo 2 (tumores malignos). Para que o resultado obtido indique uma diferença significativa, uma hipótese nula (H_0) deve ser rejeitada para que uma hipótese alternativa (H_1) seja validada. As hipóteses H_0 e H_1 podem ser descritas matematicamente pelas seguintes equações:

$$H_0 : \text{Média do grupo 1} = \text{Média do grupo 2}$$

$$H_1 : \text{Média do grupo 1} \neq \text{Média do grupo 2}$$

Entretanto, podem ocorrer dois tipos de erros no teste de hipóteses: erro tipo I e tipo II. No erro do tipo I rejeita-se a H_0 quando esta é verdadeira, ou seja, afirma-se que existe diferença estatisticamente significativa quando, na verdade, não existe. No erro do tipo II não rejeita-se a H_0 quando esta é falsa, ou seja, afirma-se que não existe diferença estatisticamente significativa quando na verdade existe (MAGALHÃES; LIMA, 2010). Nesse teste determina-se um valor de significância p que indique que a diferença é estatisticamente significativa com um determinado grau de certeza ou confiança. Para exemplificar considere um nível de confiança de 99% ($p=0,01$). Para esse nível, espera-se que em 99% dos casos testados as diferenças significativas sejam válidas.

Por meio da tabela *t de student* pode-se obter o t-valor correspondente a um nível de significância desejado orientado pelos graus de liberdade do conjunto de amostras (MAGALHÃES; LIMA, 2010). O grau de liberdade é obtido pela diferença entre a quantidade total de amostras e a quantidade de grupos analisados. O teste consiste em verificar se o t-valor calculado é superior ao t-valor encontrado na tabela *t de student*. Se o t-valor calculado for maior que o t-valor da tabela, então a hipótese alternativa é aceita e esta

diferença encontrada é considerada relevante do ponto de vista estatístico (LEÃO et al., 2009).

Na análise univariada, aqui utilizada, o valor de intensidade de cada pixel da imagem é comparado com o valor de intensidade do pixel correspondente das demais imagens de amostra. Essa análise é realizada de forma independente e não considera as dimensões das imagens e a correlação entre pixels. Este método também é conhecido como massivamente univariado (*mass-univariate*) (STEPHENS LARRY J. AND SPIEGEL, 2007).

3.5.2 Análise estatística multivariada

A análise multivariada avalia todas as variáveis das amostras simultaneamente. Dessa forma, cada amostra representa um ponto no espaço n -dimensional, onde n é o número de variáveis. No caso das imagens analisadas aqui, com dimensão 70x70, n é igual a 4900. Ao contrário da análise univariada, a multivariada não analisa as variáveis independentemente. Esta análise considera a influência que uma variável exerce sobre a outra.

Entretanto, problemas podem surgir na análise de amostras com uma grande quantidade de variáveis. Esse tipo de análise exige um esforço computacional elevado e proporcional ao aumento da quantidade de tais variáveis. Este problema onde o número de variáveis é muito maior que o número de amostras é conhecido como problema de poucas amostras (FUKUNAGA, 1990).

Existem vários métodos adotados na análise multivariada mas, nessa seção serão abordados os métodos: PCA, SVM, LDA e MLDA (GIRALDI et al., 2008b; SATO et al., 2009; VAPNIK, 1998; FISHER, 1936; HASTIE et al., 2009).

Análise das Componentes Principais (PCA)

A utilização desta conhecida técnica tem como finalidade básica a análise dos dados visando sua redução, eliminação de sobreposições e a seleção das formas mais representativas dos dados a partir de combinações lineares das variáveis originais. O objetivo do PCA é encontrar a direção de maior variância dos dados (FUKUNAGA, 1990).

A Figura 3.10 apresenta um esboço geométrico da aplicação do PCA destacando os autovetores. O primeiro autovetor representa a direção de maior variância do conjunto

de amostras. O segundo autovetor representa a direção da segunda maior variância do conjunto de amostras, perpendicular a primeira.

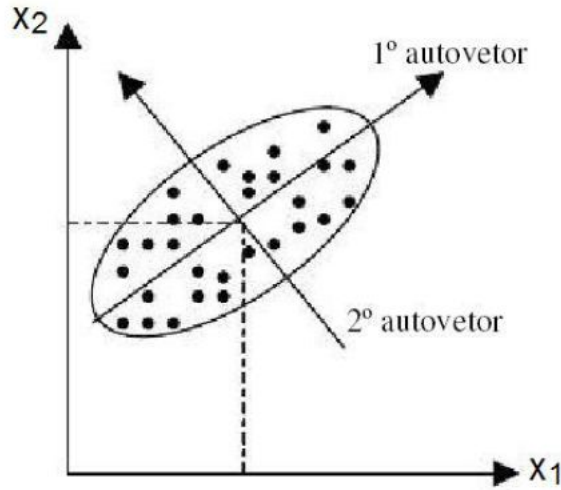


Figura 3.10: Projeção das amostras no espaço do PCA. Fonte: (LEÃO, 2009).

A seguir são enumerados os passos para o cálculo das componentes principais:

1. calcular o vetor médio dos dados;
2. subtrair o vetor médio dos dados;
3. calcular a matriz de covariância de ordem $n \times n$;
4. calcular os autovalores e autovetores da matriz de covariância;
5. projetar os dados na base de autovetores ordenados pelos respectivos maiores autovalores.

O cálculo da matriz de covariância mencionada no passo 3 é dado por:

$$S^* = X^T X, \quad (3.4)$$

onde X é a matriz de dados com média removida e de ordem $N \times n$, onde N é o número de amostras e n o número de variáveis. No entanto, o cálculo dessa matriz de covariância pode se tornar computacionalmente custoso em decorrência da quantidade muito superior de variáveis perante o número de amostras. Geralmente, na análise de imagens de ultrassom, a quantidade de variáveis é bem maior que a quantidade de amostras. Para exemplificar, considera-se as 250 imagens de ultrassom utilizadas aqui nesse trabalho. Cada imagem

possui resolução de 70x70 pixels totalizando 4900 variáveis. Esse é um típico exemplo de poucas amostras onde a quantidade de amostras, de número 250, é muito menor que a quantidade de variáveis, de número 4900. A solução desse problema pode ser alcançada, por uma manobra matemática, substituindo a Equação (3.4) pela Equação (3.5).

$$S = XX^T, \quad (3.5)$$

Esta nova matriz de covariância é de ordem $N \times N$. No estudo que se desenvolve aqui, onde N é igual a 250 e n é igual 4900 a matriz de covariância tem sua dimensão reduzida de 24.010.000 (4900 x 4900) para 62.500 (250x250). Nota-se uma redução muito relevante que exerce influência direta sobre a capacidade computacional desprendida.

Com os autovetores calculados no passo 4, obtém-se a matriz de transformação para o espaço do PCA (P_{pca}). Esta matriz é formada pela ordenação dos autovetores com seus respectivos autovalores em ordem decrescente, ou seja, $P_{pca} = [\phi_1, \phi_2, \dots, \phi_{N-1}]$. Assim, após a ordenação, os autovetores que não trazem informação relevante sobre a variabilidade das amostras podem ser descartados.

Análise de Discriminantes Lineares (LDA)

A proposta do método LDA, também conhecido com método de Fisher, é encontrar o hiperplano de maior separação entre os grupos analisados. O cálculo desse hiperplano de separação considera o conhecimento prévio da classe/grupo de cada amostra. A análise LDA é paramétrica, ou seja, considera distribuição de probabilidade. O hiperplano é um subespaço que divide o espaço em duas metades. Essas metades são formadas pelas entradas das duas classes distintas (tumores benignos e malignos).

Segundo Fisher (1936), o método baseia-se na diminuição do espalhamento das amostras com relação ao grupo ao qual elas pertencem e, também, na maximização da distância da média entre estes grupos. Em outras palavras, calcula-se as matrizes de espalhamento inter-classes e intra-classes com objetivo de separar os grupos de amostragem pela maximização da separabilidade entre classes enquanto minimiza-se a variabilidade dentro das mesmas. A Figura 3.11 apresenta a idéia geométrica da LDA.

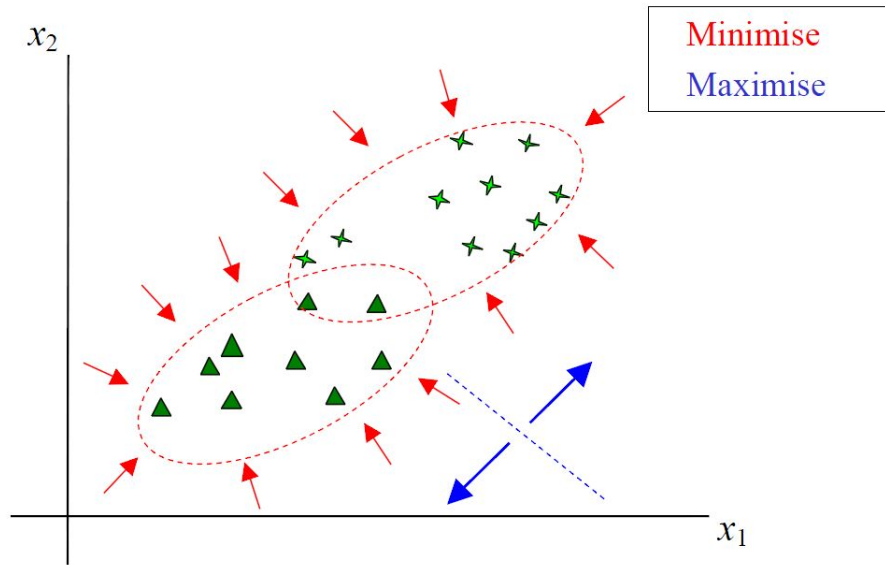


Figura 3.11: Idéia geométrica da LDA proposta por Fisher. Fonte: (THOMAZ, 2008).

Matematicamente, as matrizes de espalhamento inter-classes (S_b) e intra-classes (S_w) são dadas por:

$$S_b = \sum_{i=1}^g N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T, \quad (3.6)$$

$$S_w = \sum_{i=1}^g \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T, \quad (3.7)$$

onde g é o número de grupos analisados, N_i a quantidade de amostras do grupo i , $\bar{x}$ e $\bar{x}_i$ são a média total e a média das amostras do conjunto i respectivamente e x_{ij} é a amostra j do grupo i . E a relação proposta por Fisher que deve ser maximizada é dada pela Equação 3.8 a seguir:

$$P_{lda} = \underset{P}{\operatorname{argmax}} \frac{|P^T S_b P|}{|P^T S_w P|}. \quad (3.8)$$

Essa razão é viabilizada pelo cálculo da inversa da matriz de espalhamento intra-classes S_w . Exatamente nesse momento deve ser observado a proporção entre o número de amostras e variáveis. Em cenários onde o número de amostras é bem menor que o número de variáveis, ocorre uma instabilidade no cálculo da matriz inversa (FUKUNAGA, 1990).

Segundo Jain e Chandrasekaran (1982), a quantidade de amostras necessárias para evitar essa instabilidade no cálculo da matriz inversa de S_w deve ser igual ou superior a 5 vezes a quantidade de variáveis destas. Considerando o cenário do trabalho desenvolvido

aqui, certamente ocorreria esse problema, pois a quantidade de amostras é bem inferior a quantidade de variáveis de cada amostra. Na próxima seção é apresentada a técnica MLDA, que visa o tratamento desse problema de instabilidade na matriz inversa de S_w .

Análise de Discriminantes Lineares de Máxima Incerteza (MLDA)

Conforme mencionado anteriormente, podem ocorrer problemas na adoção do LDA padrão em cenários com amostragem limitada e com alta dimensionalidade. Nesse contexto foi então utilizado o método denominado MLDA (THOMAZ et al., 2006). Essa técnica consiste em substituir a matriz de espalhamento S_w por outra matriz regularizada S_w^* , gerando um aumento no espalhamento dos dados e mantendo as variações mais relevantes existentes nas amostras. A idéia geométrica do MLDA pode ser observada na Figura 3.12.

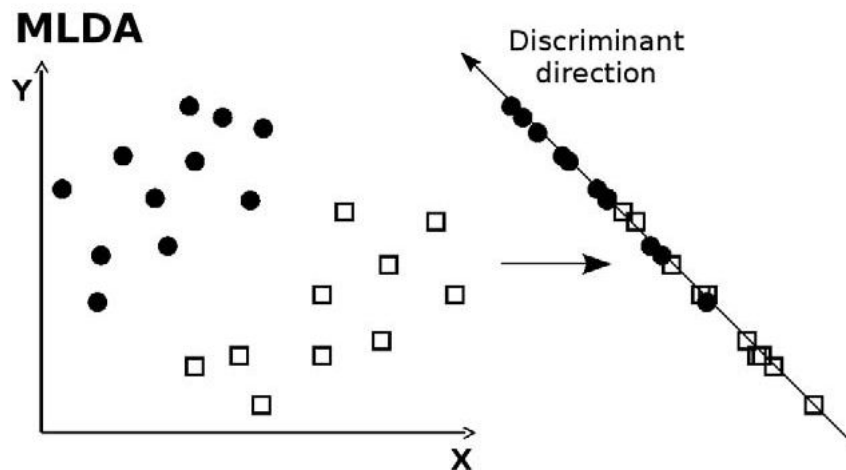


Figura 3.12: Idéia geométrica do classificador MLDA. Fonte: (SATO et al., 2009, p 106).

A nova matriz regularizada S_w^* é calculada segundo as seguintes etapas:

1. Selecionar os autovetores Φ e autovalores Λ de S_p , onde $S_p = \frac{S_w}{N-g}$;
2. Calcular a média dos autovalores $\bar{\lambda}$;
3. Gerar uma nova matriz de autovalores baseada na dispersão dos maiores $\Lambda^* = \text{diag}[\max(\lambda_1, \bar{\lambda}), \dots, \max(\lambda_n, \bar{\lambda})]$;
4. Calcular a matriz de espalhamento intra-classes regularizada $S_w^* = (\Phi \Lambda^* \Phi^T)(N-g)$.

onde S_p é uma matriz auxiliar para cálculo dos autovetores e autovalores, S_w é a matriz de espalhamento intra-classes, N é o número total de amostras, g é o número de grupos ou classes, Φ é a matriz de autovetores de S_p , Λ^* é a matriz de autovalores de S_p com os valores inferiores à média substituídos pela média e S_w^* é a nova matriz de espalhamento intra-classes.

O MLDA foi baseado na idéia de seleção de covariância de máxima entropia desenvolvida para melhorar o desempenho de classificadores Bayesianos no tratamento de amostras de número limitado (THOMAZ et al., 2004). Em outro trabalho de classificação de imagens faciais, Thomaz et al. (2006) compararam o MLDA com outras técnicas também baseadas em LDA como Chen et al.'s Method (CLDA), Yu and Yang's Method (DLDA) e Yang and Yang's Method (YLDA). Nessa pesquisa a técnica MLDA apresentou resultados superiores às demais.

Máquinas de Vetores de Suporte (SVM)

Com o mesmo propósito do LDA, o método denominado Máquina de Vetores de Suporte, ou simplesmente SVM, também visa encontrar o hiperplano de maior separação entre os grupos de amostras. De forma análoga ao LDA, o cálculo do hiperplano de separação considera o conhecimento prévio da classe de cada amostra investigada. Porém, o método SVM não é paramétrico, ou seja, não considera distribuição de probabilidade.

O SVM é uma técnica de reconhecimento de padrões com sólido embasamento teórico e que tem apresentado resultados muito satisfatórios mesmo quando comparado a métodos clássicos como redes neurais e árvores de decisão. Trata-se essencialmente de um classificador de duas classes mas que também pode ser estendido para tratamento de mais de duas classes (LORENA; CARVALHO, 2007). Este método é baseado na teoria de aprendizado estatístico, inicialmente apresentada por Vapnik e colaboradores (VAPNIK, 1998). Como vantagens, o SVM apresenta boa capacidade de generalização, robustez em grandes dimensões e convexidade da função objetivo. Para encontrar a solução ótima do classificador é usada uma função quadrática, em que não há presença de vários mínimos locais, e sim apenas um mínimo global, o que facilita a obtenção do valor ótimo.

As técnicas de aprendizado estatístico empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto de exemplos. Existem dois tipos principais de aprendizado indutivo: supervisionado e não-

supervisionado. No trabalho desenvolvido aqui será adotado o aprendizado supervisionado onde tem-se conhecimento, a priori, da classe de cada amostra. Em outras palavras, é conhecido previamente se uma imagem ultrassonográfica mamária de treinamento pertence ao grupo dos tumores benignos ou malignos.

A Figura 3.13 apresenta a idéia geométrica do SVM aplicado num cenário com amostras de treinamento de duas dimensões. O propósito é encontrar o hiperplano com margem máxima de separação entre os dados de classes distintas representadas no espaço vetorial do SVM. As margens que devem ser maximizadas, as linhas pontilhadas, podem ser vistas na mesma figura.

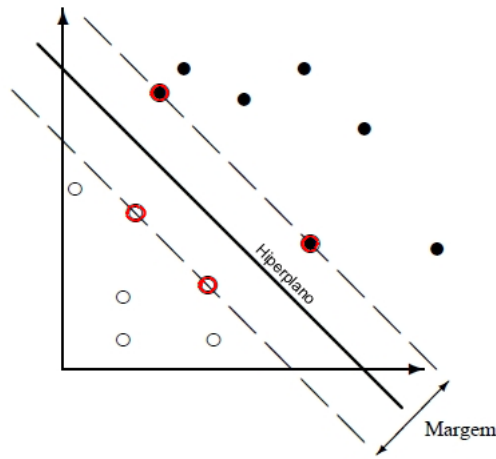


Figura 3.13: Idéia geométrica do classificador SVM. Adaptado de (BURGES, 1998)

O método SVM é descrito por:

$$W_{svm} = \sum_{i=1}^N \alpha_i y_i x_i, \quad (3.9)$$

onde α_i são os coeficientes de Lagrange não-negativos obtidos pela solução de um problema de otimização quadrático com restrições de desigualdade linear (LORENA; CARVALHO, 2007). As observações de treinamento x_i , com α_i não-zero, ficam na fronteira da margem e são chamadas vetores de suporte. Esses vetores de suporte, destacados pelos círculos vermelhos, podem ser visualizados na Figura 3.13.

O SVM pode fazer uso de vários tipos de kernel, incluindo o polinomial, o Gaussiano-RBF e o linear. Nos experimentos desenvolvidos no presente trabalho, foi adotado o kernel linear.

3.6 Considerações complementares

Os sistemas de detecção e classificação ainda possuem muitos pontos em aberto e novas pesquisas são essenciais para endereçar tais pontos. Conforme apresentado neste capítulo, não existe predominância de uma única técnica para tratamento de imagens. Assim, pode-se concluir que não existe uma metodologia única para tratar todos os tipos de imagens. Os diferentes tipos de imagens possuem particularidades que são totalmente influentes no processo analítico destas. Como exemplo, pode-se citar as imagens de ultrassom. A qualidade desse tipo de imagem é totalmente dependente do aparelho gerador e a presença comum de manchas deve ser considerada na análise. Além disso, geralmente as imagens de ultrassom não capturam as microcalcificações. Tal fato acontece em decorrência do próprio tipo da imagem que apresenta ruídos e têm baixa qualidade.

As microcalcificações são frequentemente detectadas nas mamografias de rastreamento e, em alguns casos, podem estar associadas a presença de câncer de mama (MARTINS et al., 2010). Tais microcalcificações podem ser indícios de formação do tumor e devem ser analisadas com precisão. Uma discussão recente acerca de todas as fases do processo de detecção e classificação de tumores é a utilização de imagens de ultrassom de alta resolução. Muitos problemas atuais relacionados à qualidade da imagem, incluindo a ocultação de microcalcificações, poderiam ser minimizados e/ou elucidados (SAVAGE, 2010).

Mesmo com algumas limitações, o emprego das imagens de ultrassom no diagnóstico do câncer de mama é uma opção de exame mais acessível que vêm ganhando notável espaço. Além disso, a natureza não invasiva do ultrassom favorece ainda mais sua aplicação.

O estudo desenvolvido aqui utiliza um conjunto de imagens ultrassonográficas pré-processadas e segmentadas pelo trabalho de Rodrigues et al. (2006a). São investigados métodos visando a classificação automática de tumores de mama e a identificação dos fatores que influenciaram a distinção entre os grupos de tumores benignos e malignos.

Capítulo 4

Experimentos

Neste capítulo são descritos os passos adotados na realização dos experimentos que visam a análise e classificação de imagens ultrassonográficas de tumores mamários. No processo de investigação foram utilizados métodos estatísticos univariado e multivariado com objetivo de encontrar as diferenças mais discriminantes entre os grupos de imagens de tumores malignos e benignos. A aplicação desses métodos estatísticos foi baseada na intensidade dos pixels das imagens e também em valores de pré-processamento gerados a partir destas nos trabalhos de Rodrigues et al. (2006a) e Giraldi et al. (2008b). As características utilizadas foram circularidade, heterogeneidade e sombra acústica. Nestes experimentos cada imagem representa uma amostra e cada pixel desta imagem representa uma variável da amostra.

Inicialmente as imagens foram submetidas à uma etapa de pré-processamento para, em seguida, serem analisadas por métodos estatísticos.

A seguir são descritos detalhadamente os dados utilizados nos experimentos, a etapa de pré-processamento das imagens e os passos adotados nas análises estatísticas univariada e multivariada.

4.1 Dados Experimentais

Foram utilizadas 250 imagens ultrassonográficas de tumores mamários e um conjunto de valores de pré-processamento referentes às características circularidade, heterogeneidade e sombra acústica destes tumores. As imagens de tumores benignos totalizam 100 e as de tumores malignos 150. Deve ser observada a existência de 5 imagens diferentes do mesmo tumor como mostrado na Figura 4.1. Em decorrência da existência dessas 5 imagens de cada tumor tem-se 20 imagens diferentes de tumores benignos e 30 imagens

diferentes de tumores malignos. Os valores de pré-processamento das características foram atribuídos por radiologistas e calculados nos estudos de Rodrigues et al. (2006a) e Giraldi et al. (2008b). As 250 imagens também foram obtidas desses estudos. As imagens foram adquiridas por meio de um equipamento modelo Voluson 730 (General Electric, USA) com um transdutor S-VNW5-10. As especificações técnicas deste são: frequência de varredura de 5-10 MHz, largura de varredura de 40 mm e ângulo de varredura de 20 a 30 graus.

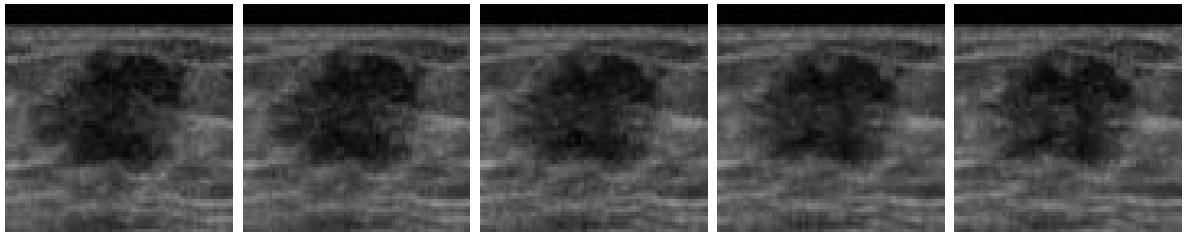


Figura 4.1: Cinco imagens do mesmo tumor.

As características das imagens tumorais investigadas são listadas na Tabela 4.1 e os valores atribuídos às mesmas podem ser visualizados posteriormente nos histogramas da Seção 5.1.

Os valores geométricos de circularidade foram obtidos a partir de um ponto central da região de interesse (tumor). Calculou-se a distância entre cada ponto da borda da lesão e o centro da imagem. Os valores também foram normalizados pela área total da imagem. Em geral, as lesões malignas apresentam valores mais altos de desvio padrão em relação a distância média quando comparadas às lesões benignas. As duas outras características, heterogeneidade e sombra acústica, são de textura. Os valores de heterogeneidade, utilizando imagens em escala de cinza, foram calculados por meio da entropia BGS (*Boltzman-Gibbs-Shannon*). Geralmente as lesões malignas são mais heterogêneas que as benignas. Já a sombra acústica tem relação direta com a região inferior da imagem. No caso das imagens de tumores mamários, a sombra acústica está extremamente relacionada ao tipo de tumor. Na maioria dos casos de tumor benigno ocorre a formação de um reforço acústico abaixo do região do tumor em decorrência da existência de muitas partículas de água. Os tumores malignos, que são geralmente mais sólidos, tendem a apresentar uma sombra acústica. Nos tumores malignos a sombra acústica é mais intensa (cor mais branca) que o reforço acústico presente nos tumores benignos. Então, para calcular os valores dessa característica foram comparados os histogramas da região da lesão

e da região logo abaixo desta. Quanto mais escuro é a região abaixo da lesão, maior é a probabilidade desta ser benigna (RODRIGUES et al., 2006b).

Esses valores foram gerados na fase de pré-processamento do trabalho de Rodrigues et al. (2006a) e são simplesmente analisados aqui nesse trabalho.

Nome	Tipo
Circularidade da área da lesão	Geométrica
Heterogeneidade da lesão	Textura
Sombra acústica da lesão	Textura

Tabela 4.1:

Características das imagens ultrassonográficas analisadas nos experimentos.

Primeiramente os valores de pré-processamento são analisados com objetivo de tentar discriminar os tipos de tumores somente por esses valores. Para observar a separação dos grupos foram construídos os histogramas a partir desses valores de pré-processamento. Na sequência são realizadas análises estatísticas considerando a intensidade de cor das imagens e também os mesmos valores de pré-processamento analisados anteriormente.

4.2 Pré-processamento de Imagens

Com o propósito de ajustar as diferenças de resolução, as imagens foram submetidas a uma etapa de pré-processamento. As imagens fornecidas, em escala de cinza, estão no formato BMP (Bitmap) e possuem resoluções diferentes que variam de 57 a 161 pixels na altura e de 75 a 199 pixels na largura. Independentemente da resolução, a maioria dos tumores está localizada no centro das imagens. Cada pixel (ponto) da imagem é representado dentro de uma escala de 0 a 255, que são as variações de tons de cinza em um sistema de representação de luminância com 8 bits de resolução. As Figuras 4.2 e 4.3 apresentam exemplos aleatórios das imagens que foram submetidas à fase de pré-processamento.



Figura 4.2: Imagens ultrassonográficas de tumores benignos.



Figura 4.3: Imagens ultrassonográficas de tumores malignos.

Na sequência foi feito o redimensionamento automático das imagens para a resolução de 70 pixels de altura. Esse ajuste automático de altura da imagem foi acompanhado do ajuste proporcional da largura da mesma. O número 70 foi escolhido por se tratar do primeiro valor, a partir de 57, que não demandaria também ajustes na largura das imagens e, conseqüentemente, alteração de proporção. Para concluir essa etapa de pré-processamento, foi feito o recorte manual da região do tumor considerando uma janela de 70 x 70 pixels. Os processos de redimensionamento e recorte foram implementados utilizando o software *Matlab*. A Figura 4.4 apresenta o resultado desses processos em três imagens aleatórias da base amostral.

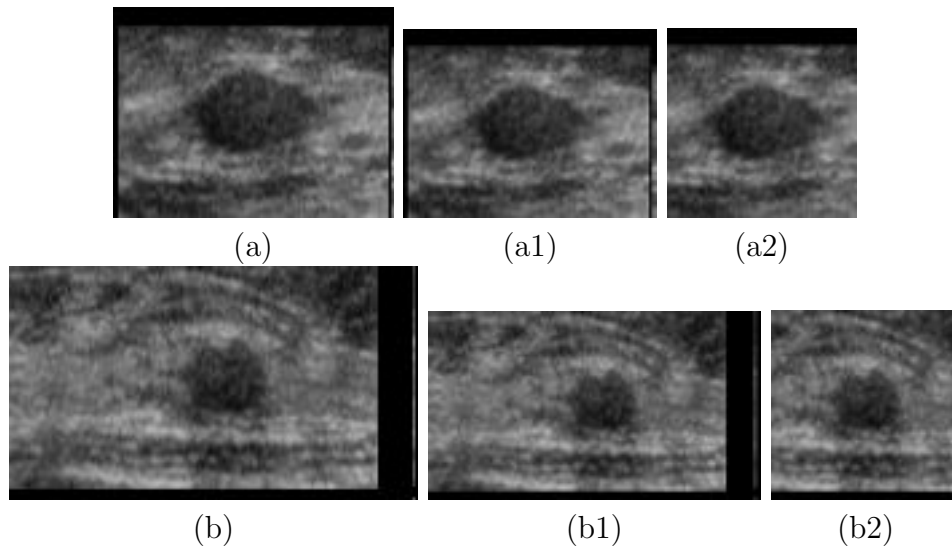


Figura 4.4: Imagens redimensionadas e recortadas para adequação ao processo de análise. As imagens (a1) e (b1) são resultantes do processo de ajuste de resolução de suas respectivas imagens originais (a) e (b). As imagens uniformes (a2) e (b2) são resultantes do processo de corte executado sobre as imagens (a1) e (b1).

4.3 Análise Estatística Univariada

Na comparação entre grupos amostrais, uma subtração simples da imagem média de um grupo com relação à imagem média do outro poderia indicar as diferenças existentes

entre estes. Com as imagens ajustadas na fase de pré-processamento calcula-se então a imagem média de cada grupo. Obrigatoriamente as imagens precisam ter a mesma resolução para implementação do cálculo da imagem média. A imagem média de cada grupo é obtida pelo cálculo da média dos valores de intensidade de cada pixel (ponto) em todas as imagens do grupo. Os valores de intensidade de cada pixel representam a cor. A Figura 4.5 demonstra a operação com duas imagens. Nos experimentos, a imagem média dos tumores malignos foi calculada a partir de 150 imagens e a imagem média dos tumores benignos foi calculada a partir de 100 imagens.

Entretanto, no caso de imagens de ultrassom, algumas das diferenças encontradas podem ser derivadas de ruídos e, portanto, podem não representar uma diferença real entre os grupos.

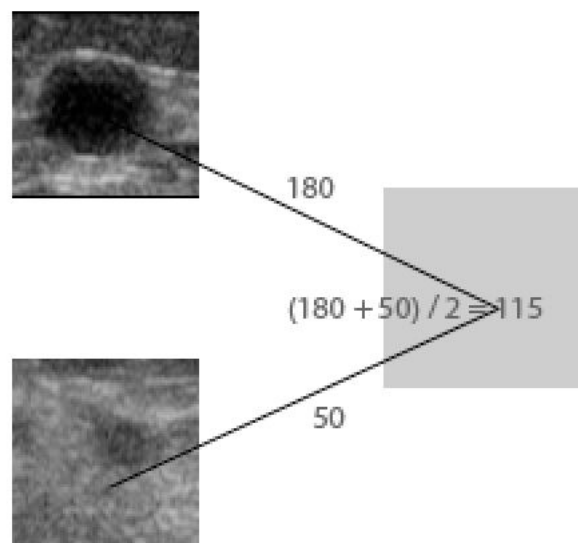


Figura 4.5: Representação do cálculo da imagem média.

Na análise estatística univariada, aqui aplicada, é feita a comparação pixel a pixel das intensidades de cor das imagens médias recém calculadas. Ao final dessa comparação é gerado um mapa de *t-valores* representando as variações entre as imagens dos grupos de tumores malignos e benignos. Conforme detalhado no capítulo anterior, este mapa é computado pela média e desvio padrão dos valores de intensidade de cor das imagens analisadas.

Com esse mapa torna-se possível identificar as regiões mais discriminantes da imagem. Tais regiões são obtidas pelo cálculo de um *t-valor* de limiar obtido em ordem de um nível de significância desejado e do grau de liberdade do conjunto de amostras. Assim, é possível selecionar os pixels cujo *t-valor* do mapa de *t-valores* for maior que o *t-valor* correspondente ao nível de significância desejado. Nesse experimento, o grau de liberdade calculado, de valor 248, foi obtido pela subtração do valor 2, representando a quantidade de grupos distintos, do valor 250 correspondente ao número de amostras. O *t-valor* de limiar foi calculado pelo software estatístico *R*.

4.4 Análise Estatística Multivariada

Ao contrário da análise estatística univariada, a análise multivariada é computacionalmente mais custosa. Tendo em vista tal fato, as imagens foram submetidas primeiramente a um processo de redução de dimensionalidade utilizando a conhecida técnica PCA. Conforme descrito no capítulo anterior, são calculados os autovalores e os autovetores da matriz de covariância de ordem $N \times N$. Os cálculos dos autovetores e autovalores foram realizados utilizando o software estatístico *R*.

Na sequência, as imagens projetadas no espaço do PCA são então projetadas no espaço dos classificadores lineares SVM e MLDA para obter a separação do conjunto amostral de tumores malignos e benignos. Neste processo de separação linear empregando os classificadores SVM e MLDA foram adotados dois tipos de parâmetros de entrada. No primeiro a classificação das imagens de tumores mamários considera a intensidade dos pixels, ou seja, a análise é feita na imagem como um todo. Já no segundo processo os classificadores são executados utilizando os valores de pré-processamento atribuídos às características dos tumores nas imagens (circularidade, heterogeneidade e sombra acústica). Os valores atribuídos a cada característica são tratados separadamente nos classificadores. A implementação do classificador SVM adotou kernel linear e parâmetro de regularização igual a 1 ($C=1$).

Após a execução dos classificadores os resultados são comparados considerando os diferentes tipos de entrada. Como parte do resultado da análise, considerando a intensidade dos pixels, são obtidos os pesos discriminantes de cada pixel da imagem. Com estes é também possível, considerando um limiar estipulado, identificar as regiões das imagens com diferenças mais significativas nas análises multivariada e univariada.

Assim como a implementação dos autovetores e autovalores do PCA, a implementação dos classificadores SVM e MLDA foi adaptada do programa *Darth Brain* (versão 1.0)¹

4.4.1 Desempenho dos classificadores SVM e MLDA

O cálculo da acurácia dos classificadores lineares SVM e MLDA foi baseado na abordagem de validação cruzada (*cross-validation*). Esta abordagem consiste em dividir o conjunto de dados de N exemplos em K subconjuntos, onde K deve ser maior ou igual a 1 (um). O valor do erro de validação é medido testando o subconjunto corrente contra o subconjunto deixado de fora. Este procedimento é repetido K vezes submetendo ao teste um subconjunto de validação por vez. O desempenho do modelo é então verificado pelo cálculo da média do erro quadrático da validação sobre todas as tentativas executadas no experimento.

Nos experimentos, foi adotada a forma extrema da abordagem de validação cruzada (*multifold-cross-validation*). Este método, conhecido como *leave-one-out* ($K=1$), foi aplicado em decorrência do pequeno número de exemplos rotulados. Exemplos rotulados são aqueles cuja classe é conhecida a priori. Nos experimentos aqui realizados tem-se conhecimento da classe de cada uma das imagens analisadas. Foi adotada a quantidade de 20.000 permutações (cruzamentos).

Ao final desse processo é possível verificar se o hiperplano de separação classificou, de forma adequada, as amostras analisadas.

4.4.2 Comparação dos resultados das análises estatísticas univariada e multivariada

A comparação dos resultados entre as análises estatísticas se dá pela análise do mapa de *t-valores* da análise univariada e pelo mapa de pesos discriminantes obtido na análise multivariada das intensidades dos pixels.

Na análise univariada, conforme será visto na Seção 5.2, é possível identificar as regiões das imagens com diferenças mais significativas em acordo com um nível de sig-

¹Desenvolvido pelo professor Dr. João Ricardo Sato no software estatístico *R* e pode ser utilizado academicamente com autorização do mesmo.

nificância desejado. Já na análise multivariada esta abordagem não é válida. Nesta a média e o desvio padrão são calculados em apenas uma dimensão no espaço do SVM e do MLDA. Neste espaço o espalhamento das amostras é diferente do espalhamento do espaço original das imagens. O cálculo do mapa de *t-valores*, na análise univariada é feito no espaço original das imagens.

Dessa forma a comparação dos resultados pode ser feita de maneira quantitativa considerando a quantidade de pixels identificados como mais discriminantes em cada uma das análises. Para exemplificar, considere que na análise univariada foram ativados 500 pixels resultantes da escolha de um grau de significância de 95%. Então para efeito de comparação, na análise multivariada será necessário selecionar os 500 pixels de maior valor (em módulo) constantes no mapa de pesos discriminantes obtido nesta.

Capítulo 5

Análise dos Resultados

Neste capítulo são apresentados e discutidos os resultados obtidos nos experimentos. São projetadas sobre uma imagem de referência as regiões das imagens identificadas correspondentemente mais discriminantes nas investigações. Para melhor visualização, as imagens resultantes dos experimentos tiveram sua resolução aumentada para 300x300 pixels. As imagens investigadas têm resolução de 70x70 pixels. Todas as análises foram feitas com as imagens nessa resolução. Somente as imagens resultantes das análises foram redimensionadas.

5.1 Representação do Conjunto de Características

Primeiramente, foram confeccionados os histogramas dos valores de pré-processamento atribuídos às características das imagens (circularidade, heterogeneidade e sombra acústica). Estes histogramas foram construídos na tentativa de separar os grupos de tumores malignos e benignos somente por estas características. As Figuras 5.1, 5.2 e 5.3 apresentam os histogramas dos valores de pré-processamento atribuídos às características circularidade, heterogeneidade e sombra acústica respectivamente. Observando os gráficos, nota-se que não foi possível separar plenamente o grupo de tumores benignos do grupo de malignos em nenhum deles.

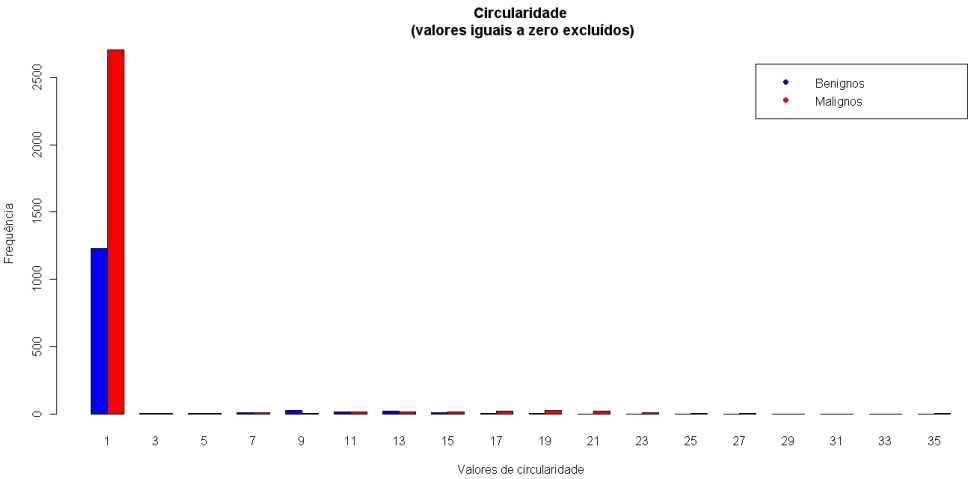


Figura 5.1: Histograma dos valores de circularidade.

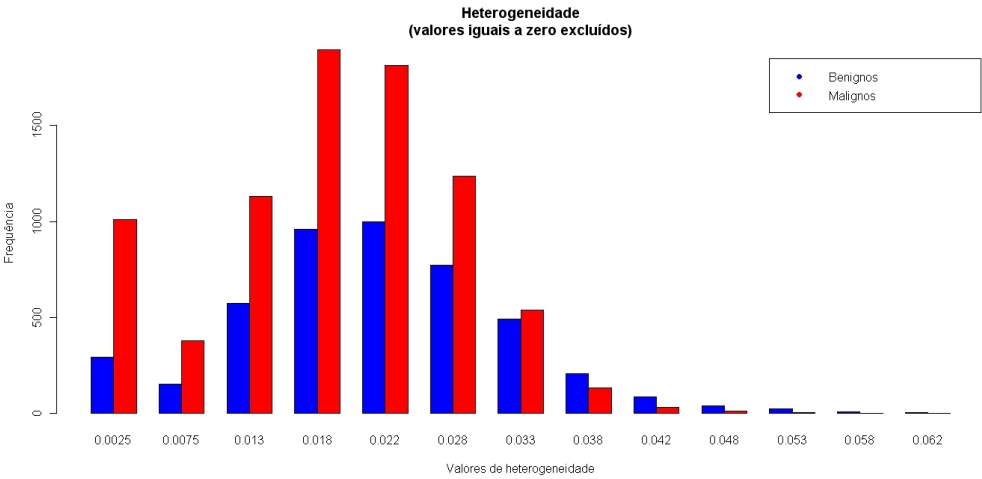


Figura 5.2: Histograma dos valores de heterogeneidade.

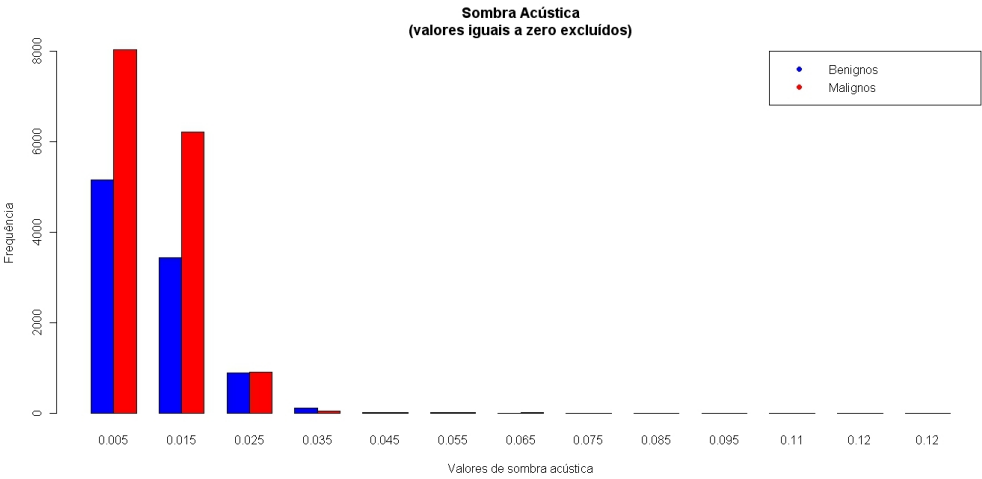


Figura 5.3: Histograma dos valores de sombra acústica.

5.2 Análise Univariada das Imagens Médicas

A Figura 5.4 apresenta as imagens médias das amostras dos grupos benigno e maligno. Claramente observa-se uma diferença no formato e tamanho dos tumores nas imagens. Na imagem média das lesões benignas (Figura 5.4a) o tumor é menor e mais concentrado. Já o tumor das imagens das lesões malignas (Figura 5.4b) parece ser maior e mais espalhado.

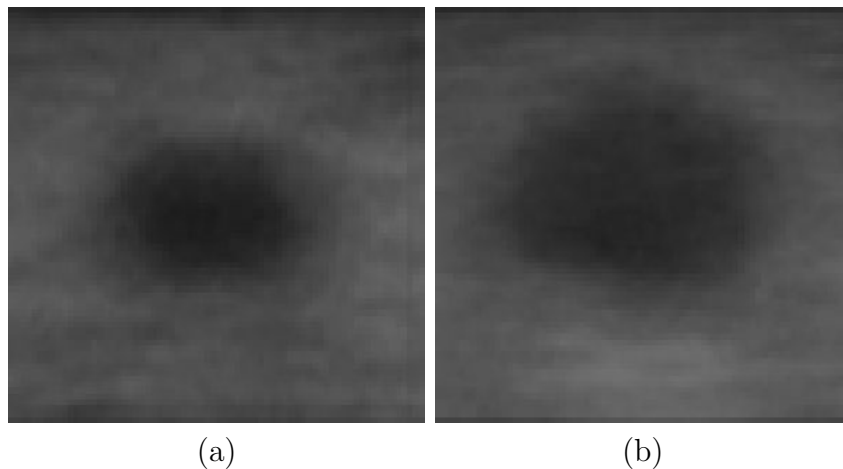


Figura 5.4: Imagem (a) resultante da média de todas imagens de tumores benignos e imagem (b) resultante da média de todas imagens de tumores malignos.

Na sequência foi feita a subtração das imagens médias dos dois grupos com o propósito de visualizar possíveis regiões discriminantes (mais diferentes). Nesse processo de subtração foi feita a normalização dos valores de intensidade dos pixels considerando o tipo da imagem onde tais valores podem variar entre 0 e 255. O resultado apresentado na Figura 5.5 indica que existe uma possível diferença entre as imagens na região inferior e também na região central direita. Provavelmente a sombra do tumor exerce influência sobre estas áreas. Esse resultado é bastante interessante pois será contrastado posteriormente com os resultados obtidos nas análises univariada e multivariada.

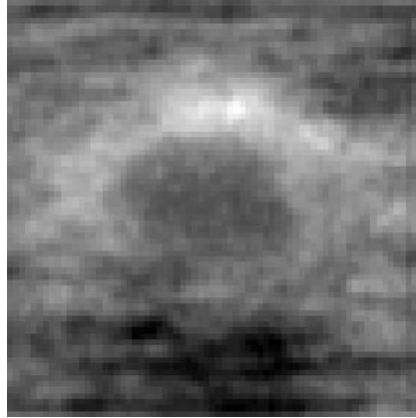


Figura 5.5: Imagem resultante da subtração da Imagem 5.4(a) pela 5.4(b).

Em complemento às diferenças observadas na subtração das imagens médias, foi implementado o teste de hipóteses com a distribuição de probabilidade *t-student*. Nesse experimento, a finalidade foi obter as variações mais discriminantes entre as imagens dos grupos de tumores malignos e benignos com níveis distintos de significância, conforme descrito anteriormente nas Equações (3.2) e (3.3).

A partir da tabela *t-student* obtém-se o valor correspondente à determinado grau de confiança e, com este, seleciona-se os valores do mapa de *t-valores* que satisfaçam determinada condição de igualdade. Nos experimentos, aqui apresentados, as regiões discriminantes foram definidas pelos maiores *t-valores* em módulo. Para obter o valor da probabilidade de cada nível de significância (tabela *t-student*) foi utilizado o software *R*¹.

Nas Figuras 5.6, 5.7 e 5.8 pode-se observar as regiões mais discriminantes selecionadas de acordo com os níveis de significância de 95%, 99% e 99.9% da tabela *t-student* respectivamente. Do ponto de vista estatístico, estas são as regiões de maior discriminância entre os dois grupos de imagens de tumor mamário. As regiões mais discriminantes foram projetadas sobre uma imagem de referência (aqui a imagem resultante da subtração das imagens médias dos dois grupos) por meio do software *Matlab*.

¹Software estatístico sob licença GNU - <http://www.r-project.org/>

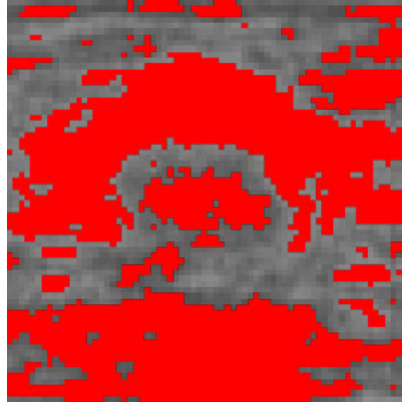


Figura 5.6: Projeção das regiões discriminantes com nível de significância de 95%.

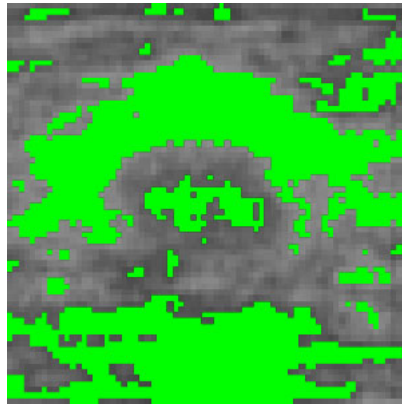


Figura 5.7: Projeção das regiões discriminantes com nível de significância de 99%.

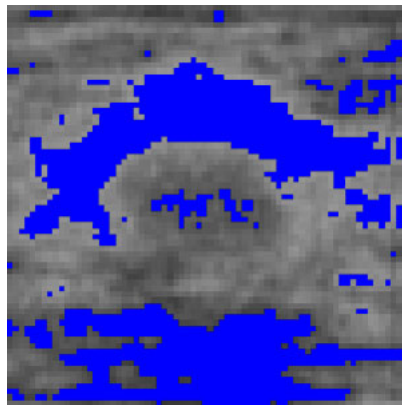


Figura 5.8: Projeção das regiões discriminantes com nível de significância de 99.9%.

Para efeito de melhor visualização, a Figura 5.9 apresenta a projeção transparente da região discriminante, com 99.9% de significância, sobre a mesma imagem de referência anterior. Nesta é possível contrastar a região identificada como mais discriminante na análise univariada com a região destacada na subtração das imagens médias dos grupos de tumores malignos e benignos (ver Figura 5.5).

A Figura 5.10 demonstra a projeção simultânea das três imagens relacionadas nas Figuras 5.6, 5.7 e 5.8 representando os níveis de significância de 95%, 99% e 99.9% da tabela *t-student*. A cor branca é resultante da intersecção das cores vermelho, verde e azul e representa a projeção simultânea dos três níveis de significância adotados (95%, 99% e 99.9%). A cor amarela é resultante da intersecção das cores vermelho e verde e representa a projeção simultânea das regiões discriminantes com percentuais de significância de 95% e 99% respectivamente.

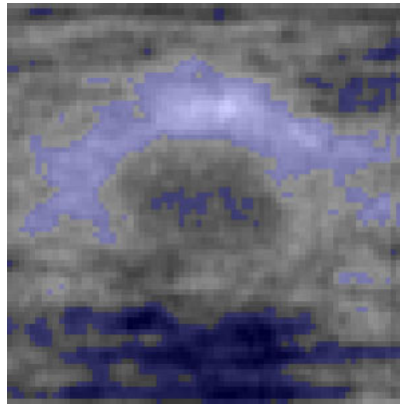


Figura 5.9: Projeção de regiões transparentes representando nível de significância de 99.9%.

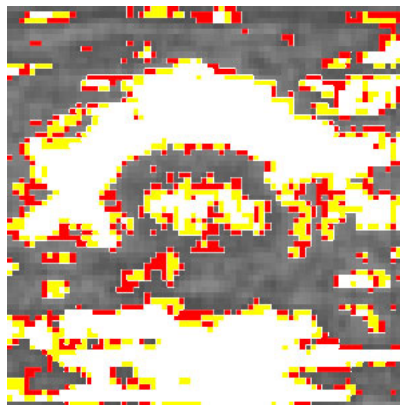


Figura 5.10: Projeção simultânea de regiões representando todos os níveis de significância experimentados (95, 99 e 99.9%).

5.3 Análise Multivariada das Imagens Médicas

Os resultados da análise multivariada são apresentados em dois passos. Primeiramente as imagens e seus valores de pré-processamento (referentes às características circularidade, heterogeneidade e sombra acústica) foram submetidos aos classificadores lineares a fim de se obter os hiperplanos de separação e também avaliar o desempenho dos mes-

mos. O segundo passo fez uso de um mapa de pesos discriminantes, gerado no processo de classificação, para identificar e mostrar as regiões mais discriminantes das imagens de tumor mamário analisadas.

5.3.1 Desempenho dos classificadores SVM e MLDA

Nessa subseção são apresentados os resultados referentes à taxa de classificação dos hiperplanos de separação linear obtidos na análise das imagens e de seus respectivos valores de pré-processamento. Esses resultados, calculados a partir de testes de validação cruzada extrema (*leave-one-out*), permitiram medir a acurácia dos classificadores na análise das imagens. O processo de validação cruzada foi implementado com 20.000 permutações. As imagens (referenciadas aqui por números) tem sua posição determinada pelo valor discriminante atribuído a esta no processo de classificação. Em outras palavras, os valores discriminantes de cada imagem são projetados no espaço discriminante dos classificadores SVM e MLDA. Os valores negativos (lado esquerdo do hiperplano de separação) foram atribuídos às imagens do grupo de tumores malignos e os valores positivos (lado direito do hiperplano de separação) foram atribuídos às imagens do grupo de tumores benignos.

As Figuras 5.11 e 5.12 apresentam os hiperplanos de separação obtidos na aplicação dos classificadores SVM e MLDA na análise de intensidade de todos os pixels das imagens simultaneamente. Claramente pode-se observar que os dois classificadores separaram muito bem os grupos. No entanto, verifica-se que a forma de separação não é idêntica para ambos classificadores. Em outras palavras, não são utilizadas as mesmas características das imagens em análises empregando modelos diferentes, como o SVM e o MLDA.

A Figura 5.13 apresenta, de maneira simultânea, os hiperplanos obtidos dos dois classificadores com o intuito de observar que a ordem de classificação das amostras, no entanto, ficou diferente entre os classificadores. Por exemplo, as imagens 95 e 96, indicadas pelas setas de cor verde na Figura 5.13 e ilustradas na Figura 5.14, ficaram bem próximas da fronteira de decisão no resultado gerado pelo classificador SVM. Uma possível explicação para tal posicionamento das imagens 95 e 96, de tumores benignos, seria a presença de características de tumores malignos como a sombra acústica e formato irregular.

Já no resultado do classificador MLDA as mesmas imagens 95 e 96, indicadas pelas setas de cor azul, ficaram posicionadas mais distantes da fronteira de decisão indicando

uma classificação menos complexa. O mesmo ocorre com as imagens 93 e 60 do grupo de tumores benignos (lado direito). Estas ficaram posicionadas no extremo do grupo indicando uma classificação mais fácil e/ou mais simples. De forma análoga, também se apresentam as imagens 206, 208, 209 e 210 do grupo de tumores malignos (Figura 5.14). Estas, indicadas pelas setas de cor vermelha, ficaram posicionadas no extremo do grupo de tumores malignos (lado esquerdo). Devido à sobreposição, as imagens 208, 209 e 210 foram também indicadas pelo círculo vermelho no lado esquerdo do hiperplano de separação.

É válido destacar ainda a posição da imagem 166 no hiperplano gerado pelo classificador MLDA (Figura 5.12). Esta imagem ficou posicionada exatamente em cima da fronteira de decisão sugerindo uma maior complexidade de classificação. Observando a imagem 166 na Figura 5.14 nota-se que esta apresenta características dos dois grupos de tumores, ou seja, esta apresenta características de tumor benigno como a ausência de sombra acústica mas, também apresenta formato irregular, geralmente característico dos tumores malignos.

A proximidade da fronteira de decisão pode ser interpretada como uma maior dificuldade de classificação. Já a maior distância desta pode indicar uma maior facilidade de classificação.

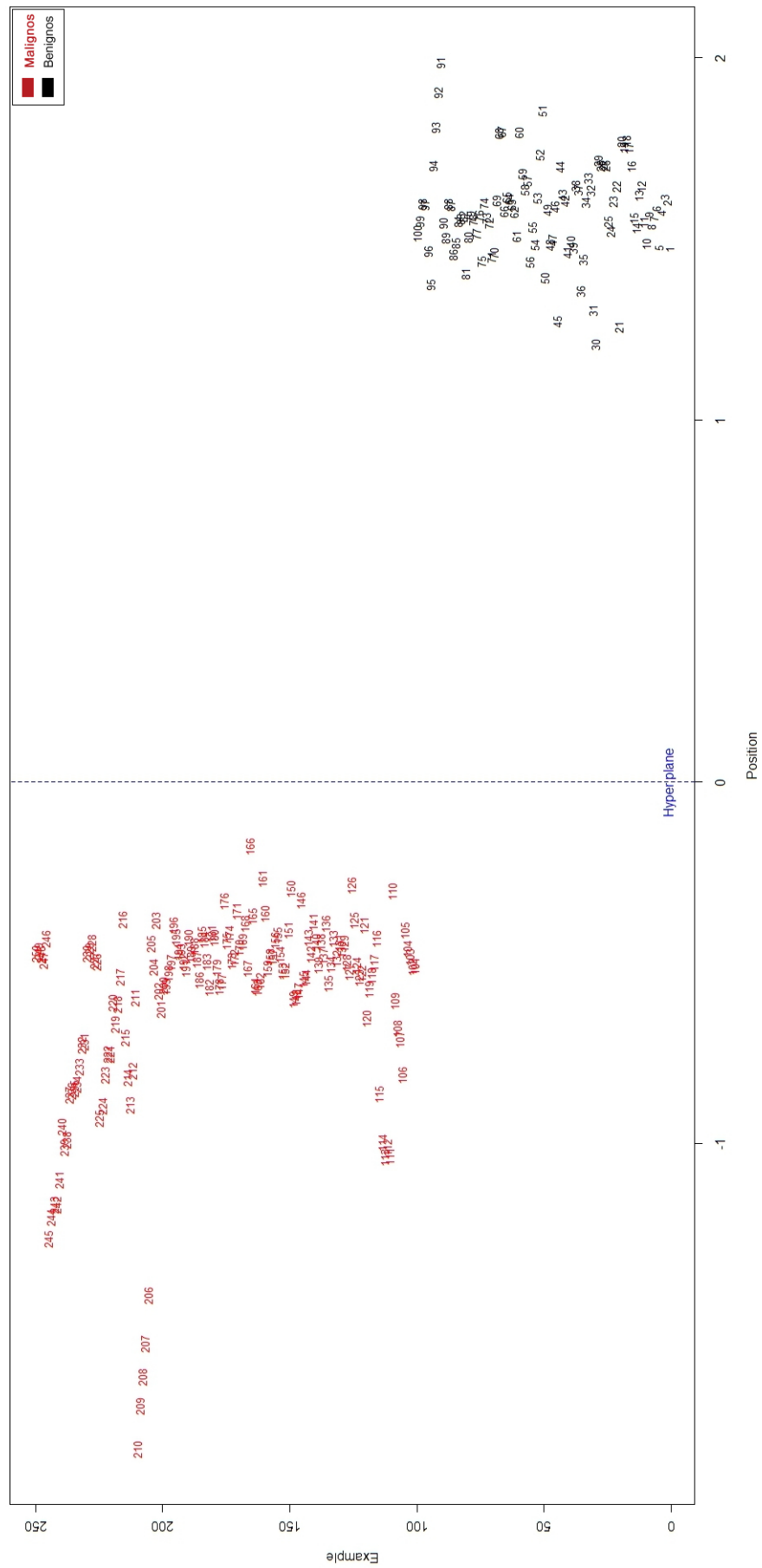


Figura 5.11: Hiperplano de separação obtido pelo classificador SVM na análise das imagens como um todo.

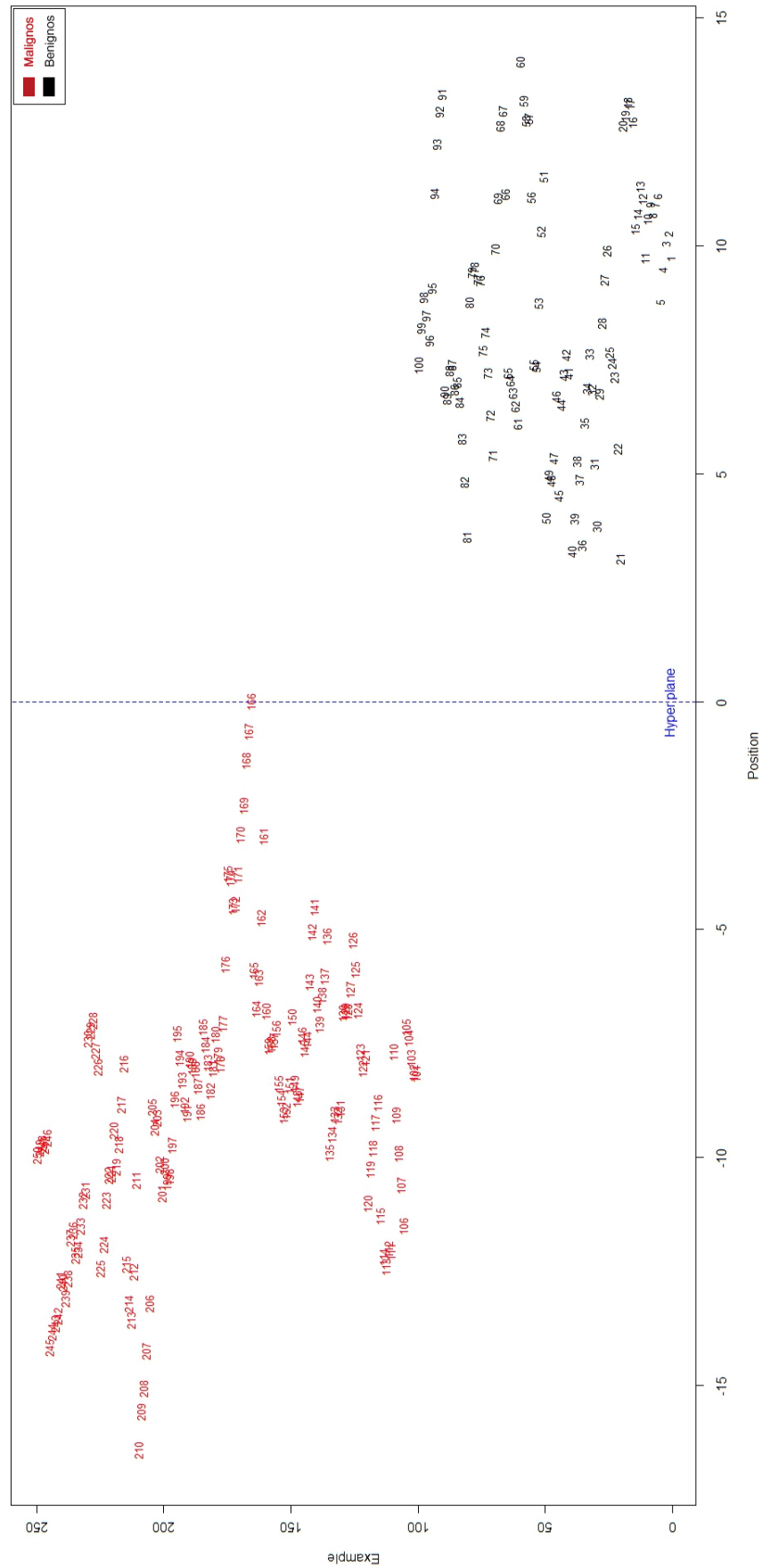


Figura 5.12: Hiperplano de separação obtido pelo classificador MLDA na análise das imagens como um todo.

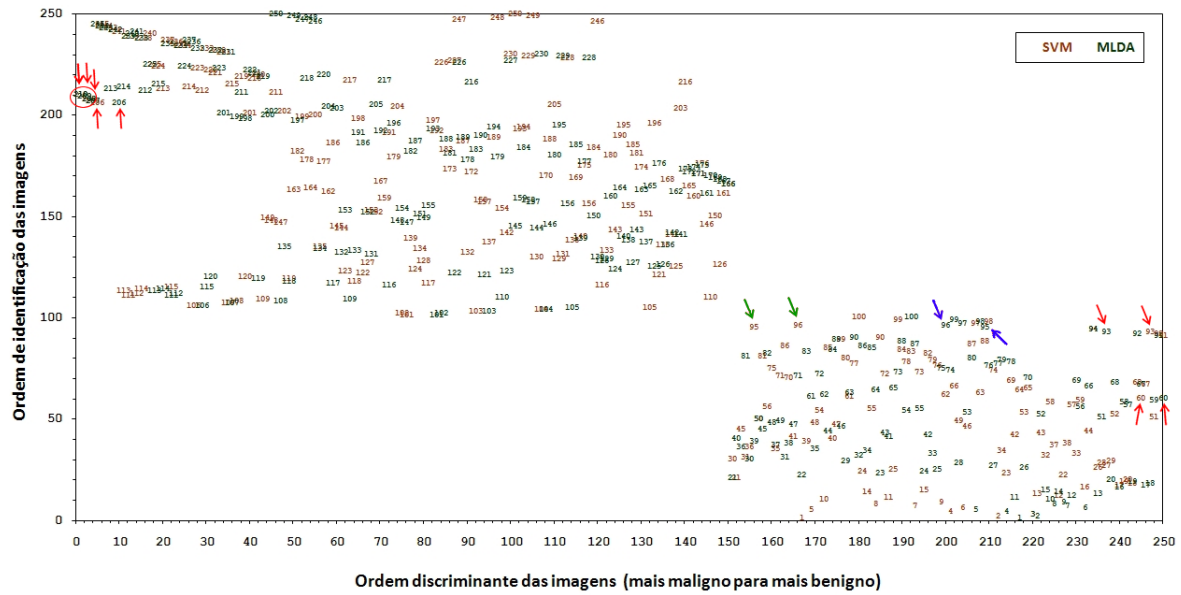


Figura 5.13: Hiperplanos de separação obtidos pelos classificadores SVM e MLDA na análise das imagens como um todo.

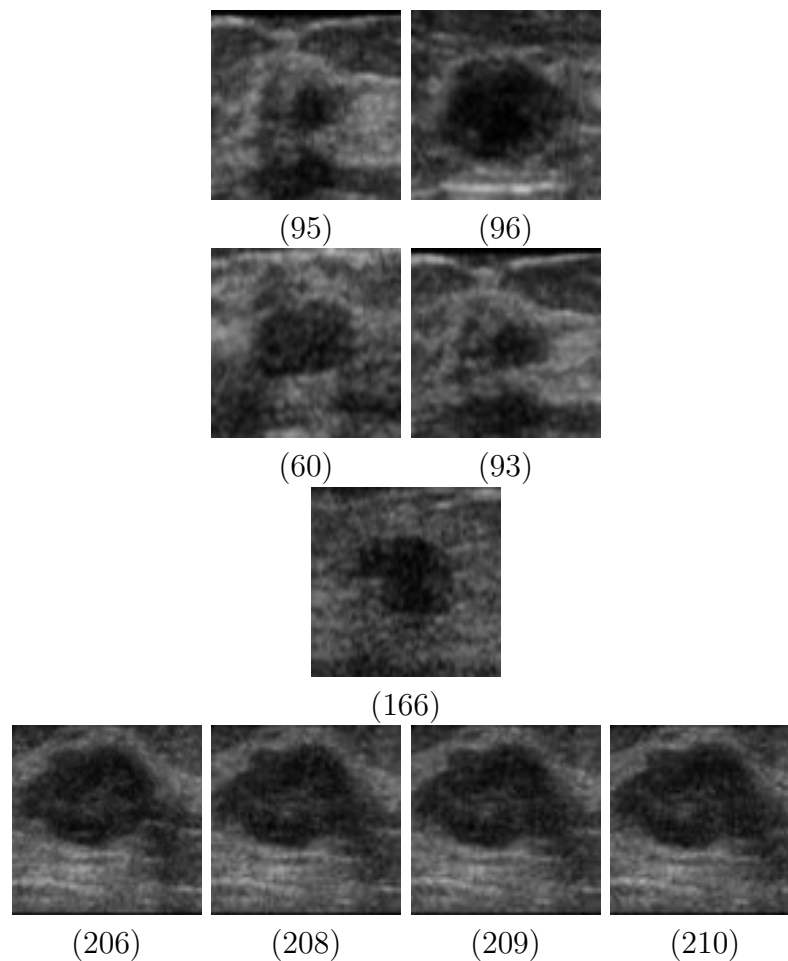


Figura 5.14: Imagens analisadas nos hiperplanos dos classificadores SVM e MLDA.

A seguir são apresentados os hiperplanos de separação obtidos na aplicação dos

classificadores lineares SVM e MLDA na análise de valores de pré-processamento atribuídos às três características das imagens mencionadas anteriormente (circularidade, heterogeneidade e sombra acústica).

As Figuras 5.15, 5.16 e 5.17 apresentam os hiperplanos de separação obtidos na aplicação do classificador SVM na análise dos valores de pré-processamento atribuídos as características circularidade, heterogeneidade e sombra acústica das imagens respectivamente. Nas Figuras 5.18, 5.19 e 5.20 são apresentados os hiperplanos de separação obtidos na aplicação do classificador MLDA na análise dos mesmos valores de pré-processamento atribuídos as características das imagens.

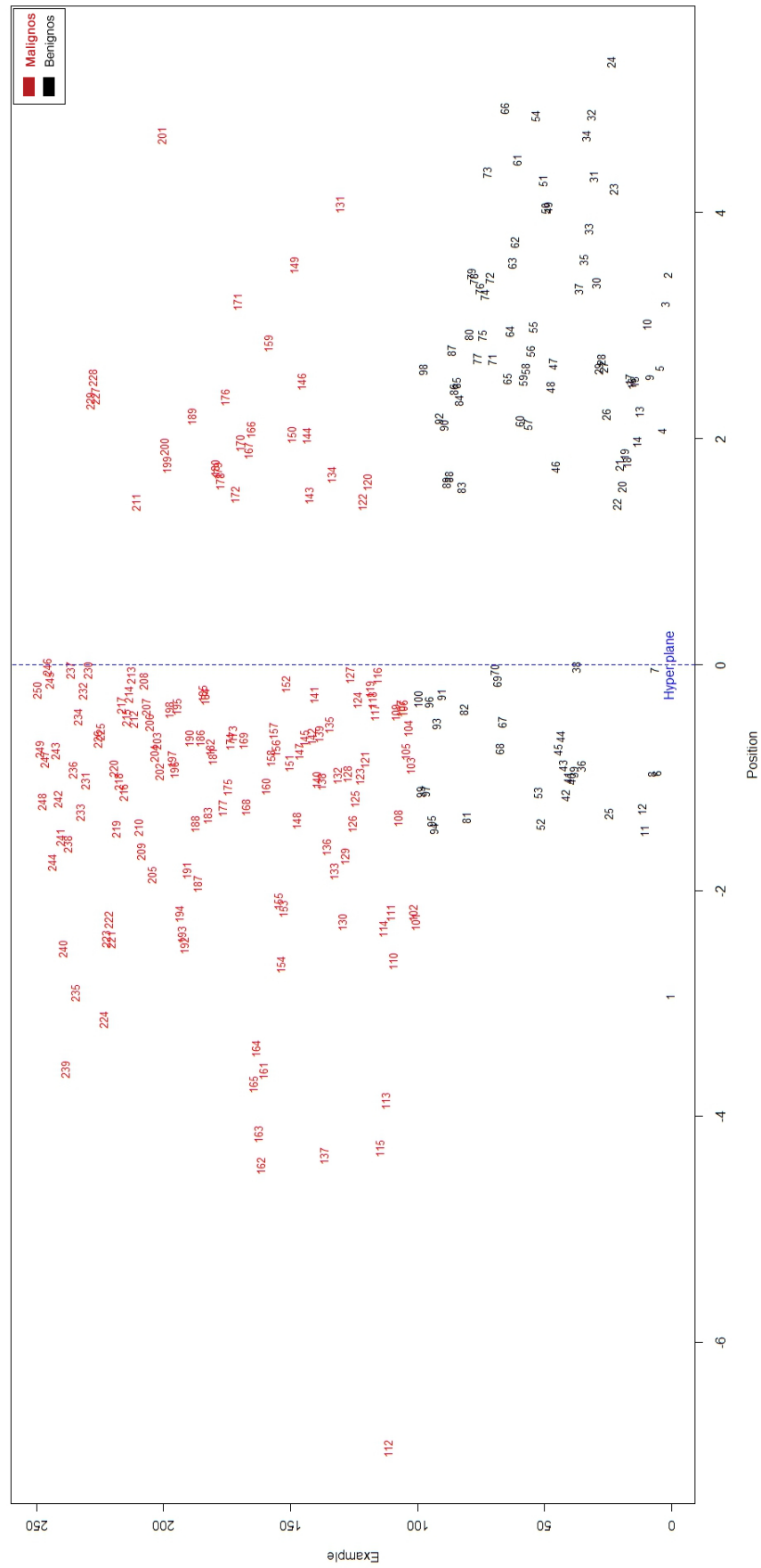


Figura 5.15: Hiperplano de separação do classificador SVM - Circularidade.

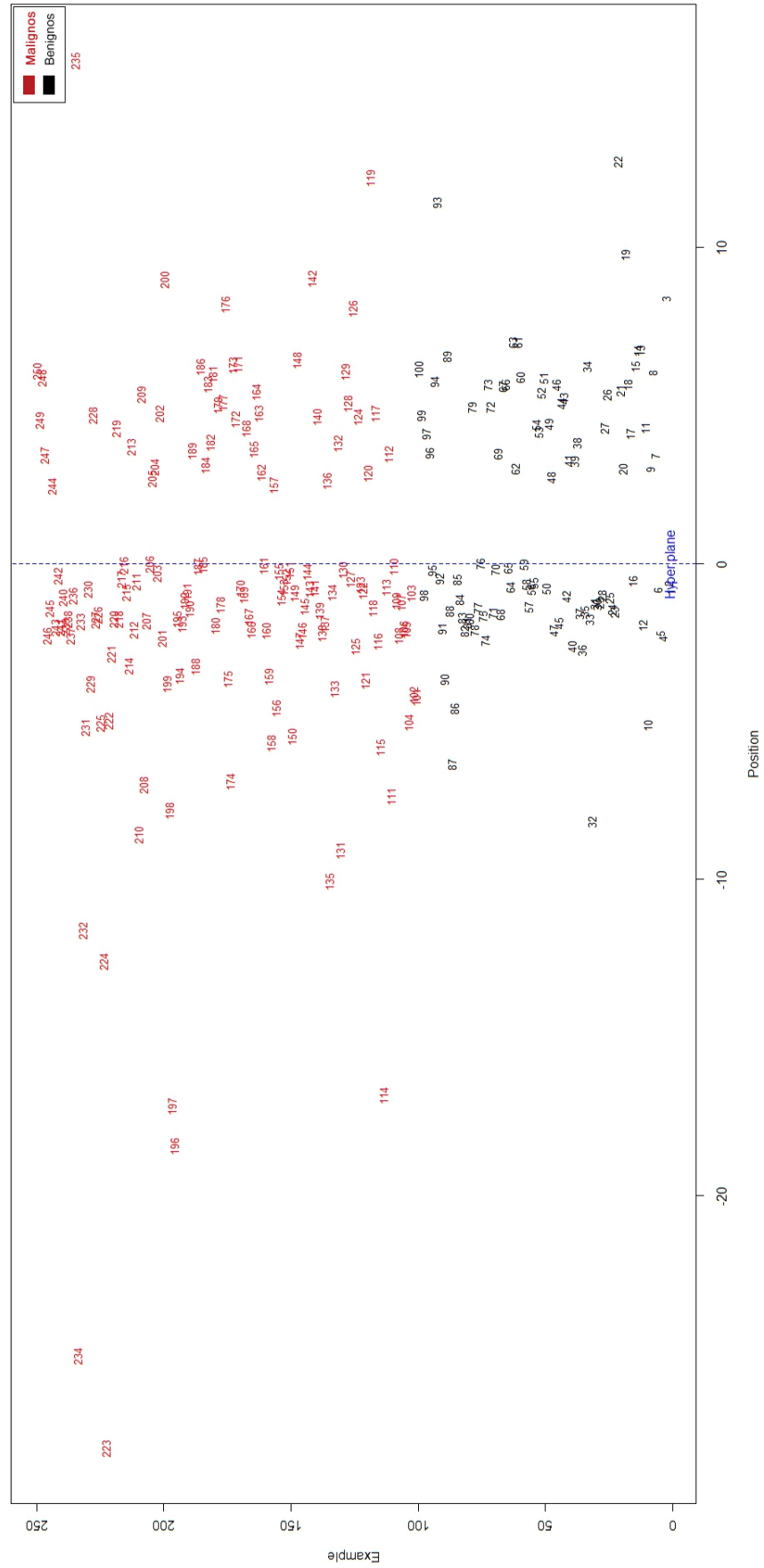


Figura 5.16: Hiperplano de separação do classificador SVM - Heterogeneidade.

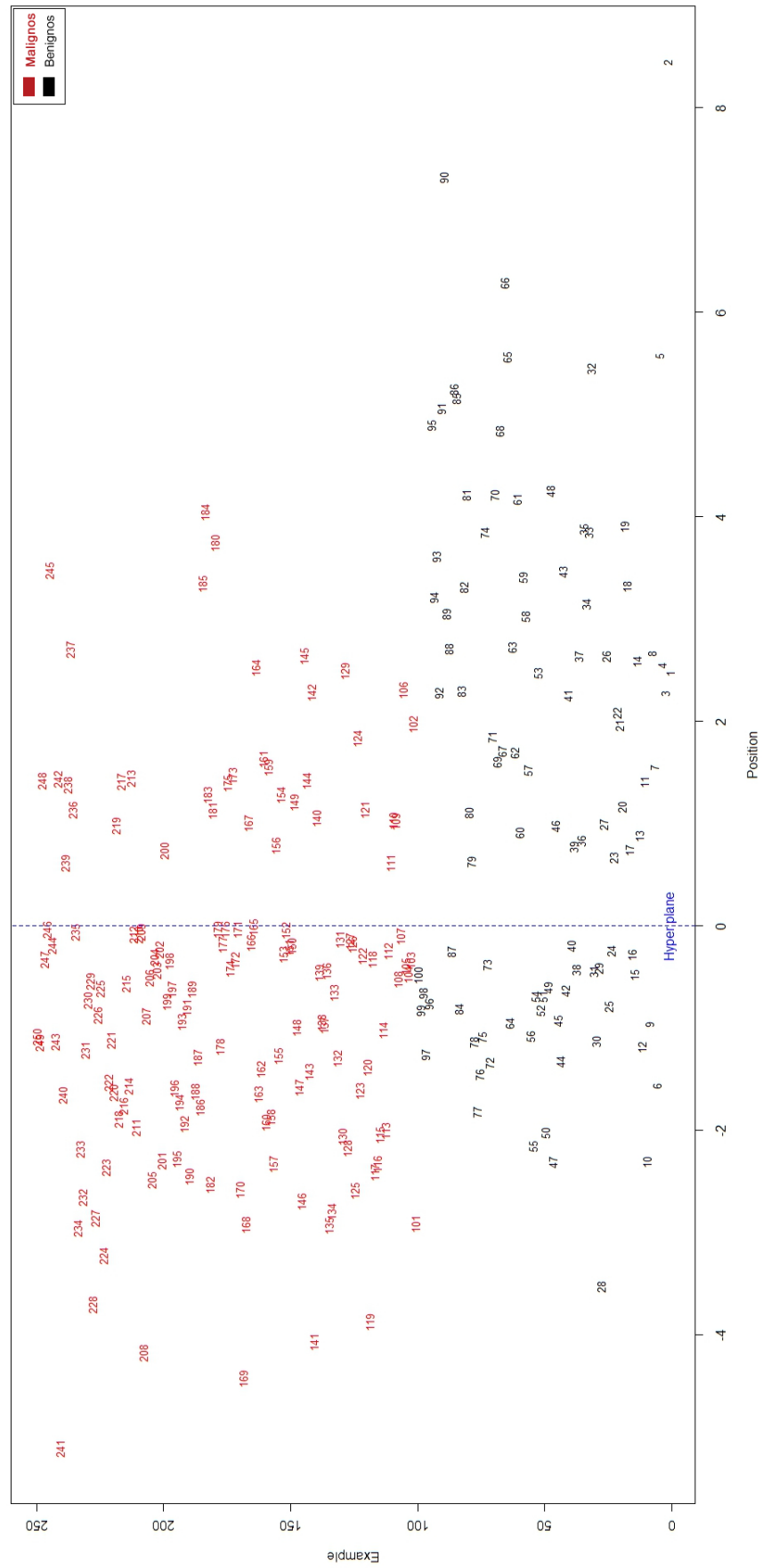


Figura 5.17: Hiperplano de separação do classificador SVM - Sombra acústica.

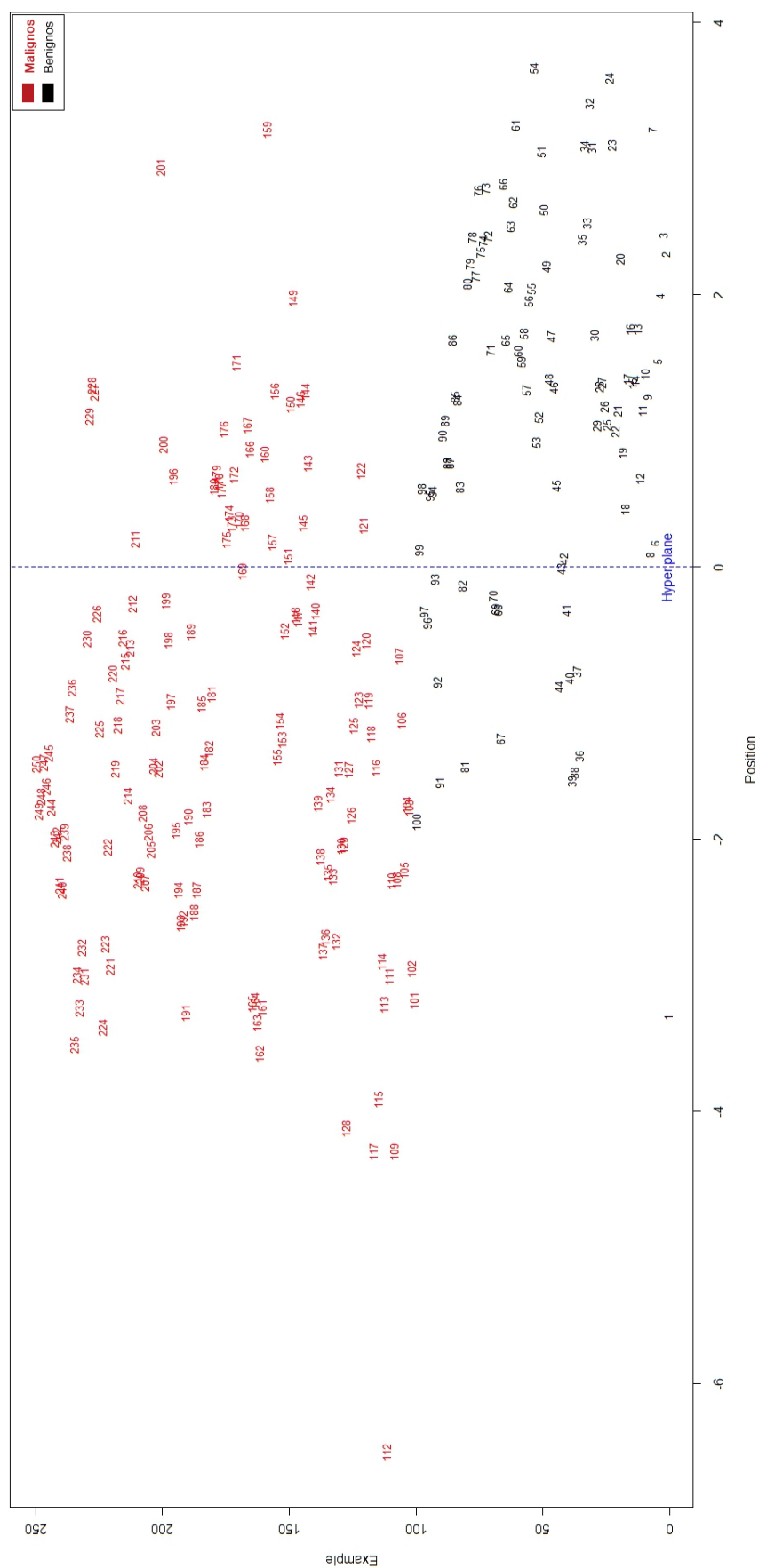


Figura 5.18: Hiperplano de separação do classificador MLDA - Circularidade.

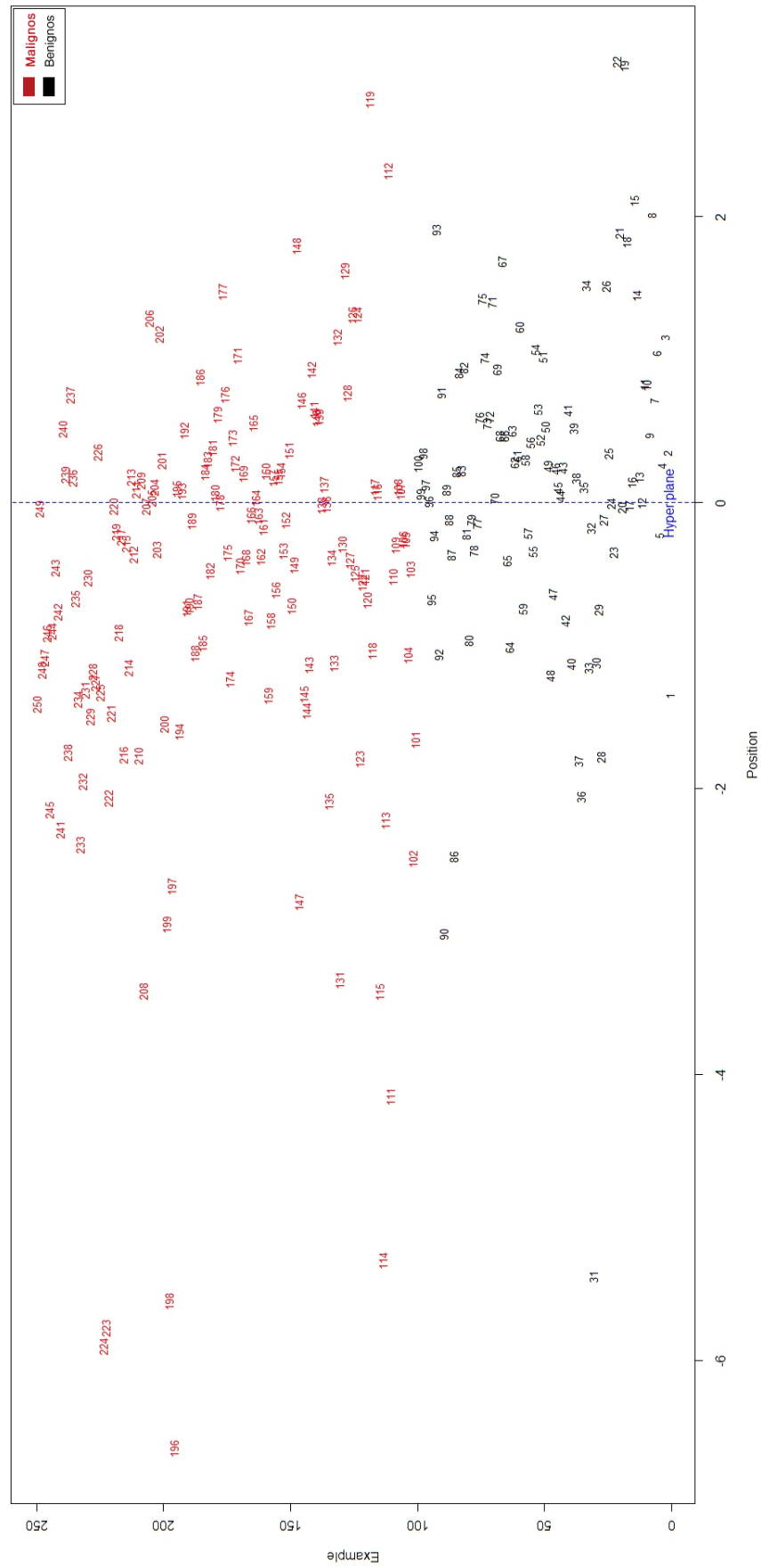
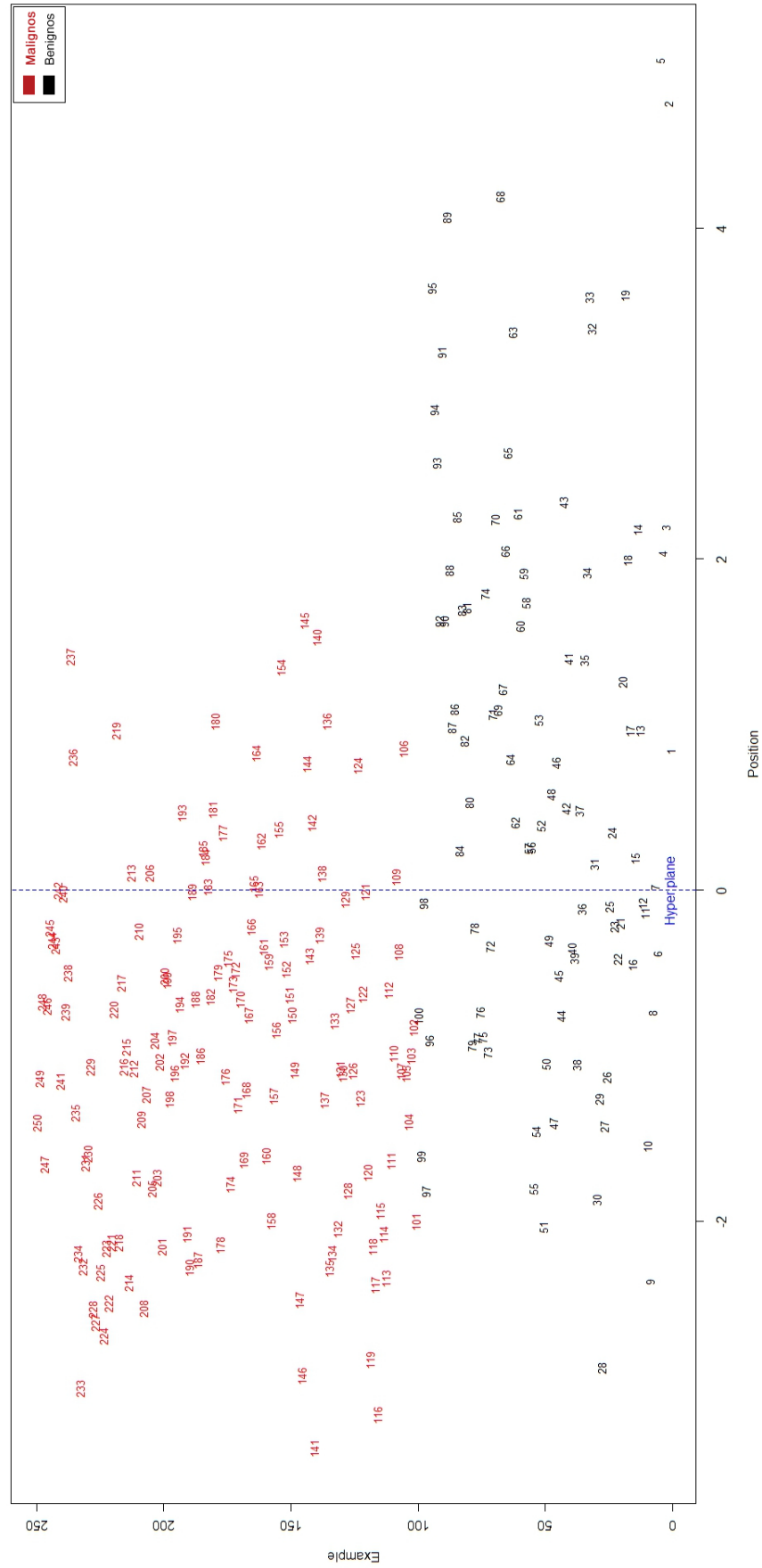


Figura 5.19: Hiperplano de separação do classificador MLDA - Heterogeneidade.



Em contraste com os resultados obtidos na análise de intensidade de todos os pixels simultaneamente, a classificação considerando os valores das características apresentou resultado não satisfatório. A maior taxa de acerto, de 78%, foi obtida na avaliação da característica circularidade pelo classificador MLDA.

A Tabela 5.1 apresenta a acurácia dos classificadores nas análises das imagens como um todo e das características exclusivamente. O desempenho dos classificadores considerando a intensidade de todos os pixels da imagem, de forma simultânea, foi notadamente superior ao desempenho desses mesmos classificadores analisando as características extraídas das imagens. Esses resultados indicam que os fatores determinantes na classificação dos grupos vão além de características como circularidade, heterogeneidade e sombra acústica, principalmente para separações lineares.

	MLDA			SVM		
	Total	Benignos	Malignos	Total	Benignos	Malignos
Imagem	100%	100%	100%	100%	100%	100%
Circularidade	78%	79%	77%	76%	68%	82%
Heterogeneidade	64%	63%	65%	60%	46%	70%
Sombra Acústica	73%	60%	82%	70%	61%	75%

Tabela 5.1:

Desempenho dos classificadores SVM e MLDA na análise de imagens ultrassonográficas da mama.

Na Tabela 5.2 são apresentados os índices de correlação entre os resultados dos classificadores nas análises de intensidade de todos os pixels e de valores referentes às características das imagens.

Como resultado da classificação foi gerado um conjunto de valores representando o peso discriminante de cada imagem nos hiperplanos de separação dos classificadores SVM e MLDA. Estes valores podem ser consultados no Anexo II. Na sequência esses valores foram ordenados dos maiores negativos (imagens com tumores mais malignos) para os maiores positivos (imagens com tumores mais benignos). Os coeficientes de correlação foram calculados sobre esse ranking e exibidos na Tabela 5.2. Foram calculados os coeficientes de correlação considerando a ordenação das imagens obtida na classificação pela análise de intensidade de todos os pixels simultaneamente e também pela análise dos valores de pré-processamento atribuídos às características circularidade, heterogeneidade e sombra acústica.

Na análise de intensidade de todos os pixels simultaneamente, a correlação entre os resultados dos classificadores foi bem expressiva alcançando o valor de 75,63%. Já na comparação dos resultados das análises de intensidade dos pixels e dos valores de pré-processamento das características, os índices foram menores mas ainda alcançaram valores de correlação da ordem de 40% na comparação dos resultados da análise da característica circularidade em ambos classificadores SVM e MLDA. Isso pode indicar que a característica circularidade exerce maior influência na classificação dos tumores estudados neste trabalho.

Relação entre classificadores	Coeficiente de correlação
SVM x MLDA	75.63
SVM x SVM circularidade	44.62
SVM x SVM sombra acústica	29.77
SVM x SVM heterogeneidade	15.62
MLDA x MLDA circularidade	41.22
MLDA x MLDA sombra acústica	27.38
MLDA x MLDA heterogeneidade	9.96

Tabela 5.2:

Correlação entre resultados dos classificadores por ordenação dos valores discriminantes de cada imagem analisada.

5.3.2 Análise visual de resultados dos classificadores SVM e MLDA

Nesta subseção são apresentadas e comparadas as projeções das regiões mais discriminantes obtidas no processo de classificação com os métodos SVM e MLDA. Ao final do processo de classificação foram obtidos os mapas dos pesos discriminantes de cada pixel da imagem, ou seja, a contribuição de cada ponto da imagem no resultado do processo de classificação.

As imagens apresentadas na Figura 5.21 mostram as regiões mais discriminantes obtidas no processo de classificação pelo método SVM. Estas regiões foram projetadas considerando a mesma quantidade de pixels selecionada no resultado da análise univariada que adotou os níveis de significância de 95%, 99% e 99.9%.

Com aparência visual similar, a Figura 5.22 apresenta as regiões mais discriminantes obtidas no processo de classificação adotando com o método MLDA. Estas regiões também foram projetadas considerando a mesma quantidade de pixels selecionada no re-

sultado da análise univariada adotando os níveis de significância de 95%, 99% e 99.9%. A imagem de referência utilizada na projeção das regiões discriminantes foi a mesma utilizada na apresentação dos resultados da análise univariada na Seção 5.2 anterior.

A Figura 5.23 apresenta a projeção transparente das regiões discriminantes obtidas na aplicação dos métodos SVM e MLDA considerando o nível de significância de 99.9%. Nas imagens, é possível contrastar as regiões identificadas como mais discriminantes na análise multivariada com as regiões destacadas na subtração das imagens médias dos grupos de tumores malignos e benignos.

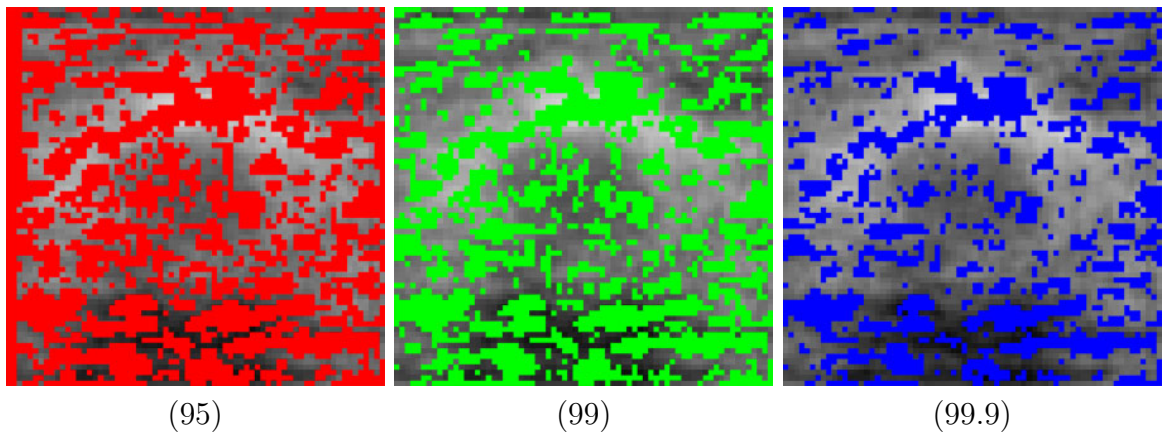


Figura 5.21: Projeção das regiões mais discriminantes segundo o classificador SVM considerando a mesma quantidade de pixels projetada na análise univariada com graus de significância de 95%, 99% e 99.9%.

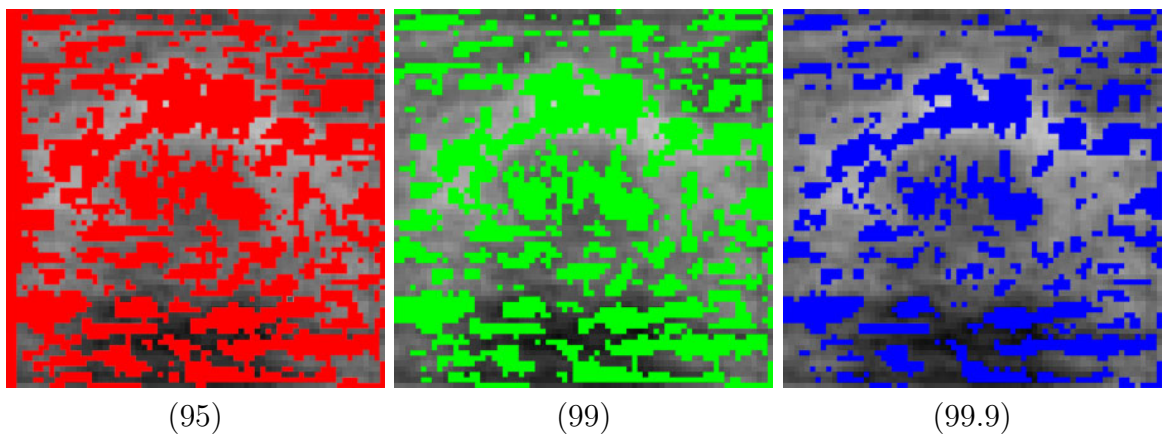


Figura 5.22: Projeção das regiões mais discriminantes segundo o classificador MLDA considerando a mesma quantidade de pixels projetada na análise univariada com graus de significância de 95%, 99% e 99.9%.

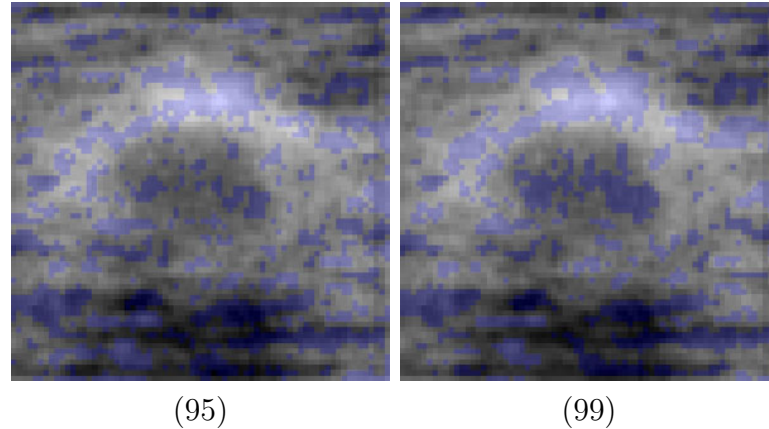


Figura 5.23: Projeção transparente das regiões mais discriminantes segundo os classificadores SVM e MLDA considerando a mesma quantidade de pixels projetada na análise univariada com grau de significância de 99.9%.

5.3.3 Comparação visual dos resultados entre as análises estatísticas univariada e multivariada

Para a comparação visual dos resultados dos métodos estatísticos univariado e multivariado, seleciona-se as regiões de maior discriminância. Conforme visto anteriormente, na análise univariada foi gerado o mapa de *t-valores* e na análise multivariada das imagens como um todo foi gerado um mapa com os pesos discriminantes de cada pixel (ponto) da imagem analisada.

Na análise univariada foram selecionados 2697 pixels correspondentes ao grau de confiança de 95%, 2145 pixels correspondentes ao grau de confiança de 99% e finalmente, 1592 pixels correspondentes ao grau de confiança de 99.9%. Estas mesmas quantidades de pixels foram selecionadas na análise multivariada considerando os maiores valores discriminantes dos pixels em módulo. Do ponto de vista estatístico, quanto menor a quantidade de pixels selecionados maior é a relevância das regiões encontradas.

O resultado comparativo das análises é apresentado na Figura 5.24. Nesta é possível visualizar as diferenças nas regiões selecionadas como mais discriminantes nos dois tipos de análise, univariada e multivariada. Enquanto que, na análise univariada, as regiões discriminantes são mais homogêneas e concentradas, na análise multivariada essas regiões são mais heterogêneas e esparsas. Nos resultados dos classificadores SVM e MLDA foram ativados pontos discriminantes em praticamente toda a imagem. Já no resultado da análise univariada, as regiões discriminantes ficaram concentradas na parte inferior e na parte central superior. Provavelmente a região discriminante inferior destacada tem forte

relação com a sombra acústica do tumor.

Essa comparação visual evidencia que as diferenças entre as imagens dos grupos não estão limitadas à observação de características como circularidade e sombra acústica dos tumores. As diferenças vão além, e outras partes da imagem devem ser observadas para alcançar um bom resultado na classificação linear dos tumores.

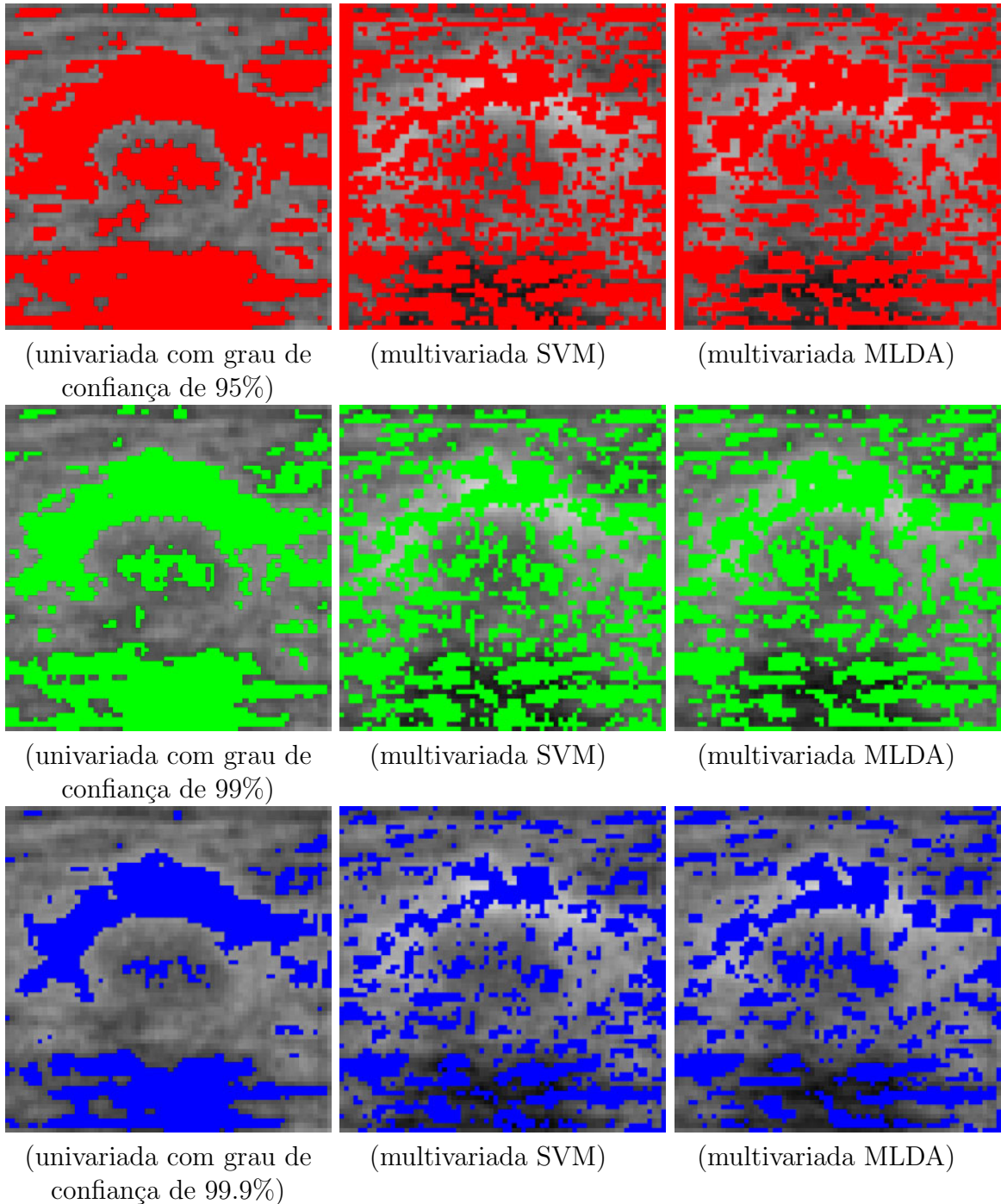


Figura 5.24: Comparação visual entre as regiões mais discriminantes obtidas na análise univariada com teste t-student e multivariada segundo os classificadores SVM e MLDA.

5.3.4 Navegação nos hiperplanos

Em caráter complementar, a Figura 5.25 apresenta uma padronização das formas de nódulos mamográficos proposta pelo BI-RADS. Nesta é possível observar a transição do nódulo benigno para o maligno. A diferença no formato dos nódulos é bastante expressiva. Deve-se lembrar ainda que esta padronização é uma proposta, e não um consenso geral. Existem casos de tumores totalmente contrários a esta padronização. Mas no âmbito geral, trata-se de um bom guia para o estudo dos tumores mamográficos em questão.

Para ilustrar essa transição, apresentada pelo BI-RADS, foi gerada a navegação entre os grupos orientada a protótipos. Tal navegação foi construída com as imagens médias calculadas a partir de intervalos estabelecidos na ordenação das imagens, obtida na aplicação dos métodos SVM e MLDA. O intervalo adotado foi de dez imagens, ou seja, a cada dez imagens ordenadas calcula-se a imagem média. Como a ordenação é composta pelas 250 imagens amostrais, obtém-se então vinte e cinco imagens para navegação.

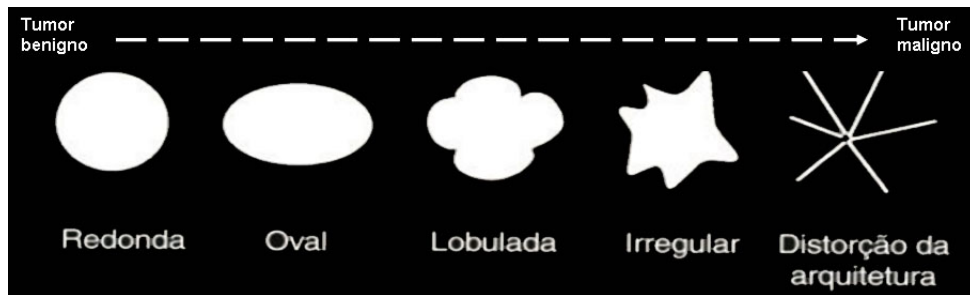


Figura 5.25: Formas dos tumores mamários propostas pelo BI-RADS.

As Figuras 5.26 e 5.27 ilustram a transição das imagens entre os grupos de tumores benignos e malignos, para os classificadores SVM e MLDA. É possível notar diferenças no formato dos tumores e também na região inferior destes. A região mais clara na parte inferior das imagens é mais intensa nas imagens que pertencem ao grupo dos tumores malignos (lado direito). O tamanho dos tumores também apresenta notável variação na transição entre os grupos. O tamanho dos tumores benignos (lado esquerdo) é menor e segue aumentando a medida que a navegação se aproxima dos tumores malignos (lado direito).

Nota-se ainda uma boa semelhança nas sequências de imagens geradas pelos classificadores SVM e MLDA. Apesar da ordem das imagens não ser igual, em ambos resultados, o tamanho do tumor aumenta na navegação em direção aos tumores malignos e a possível sombra acústica de cor branca, na parte inferior da imagem, diminui de tamanho na

navegação em direção aos tumores benignos (esquerda).

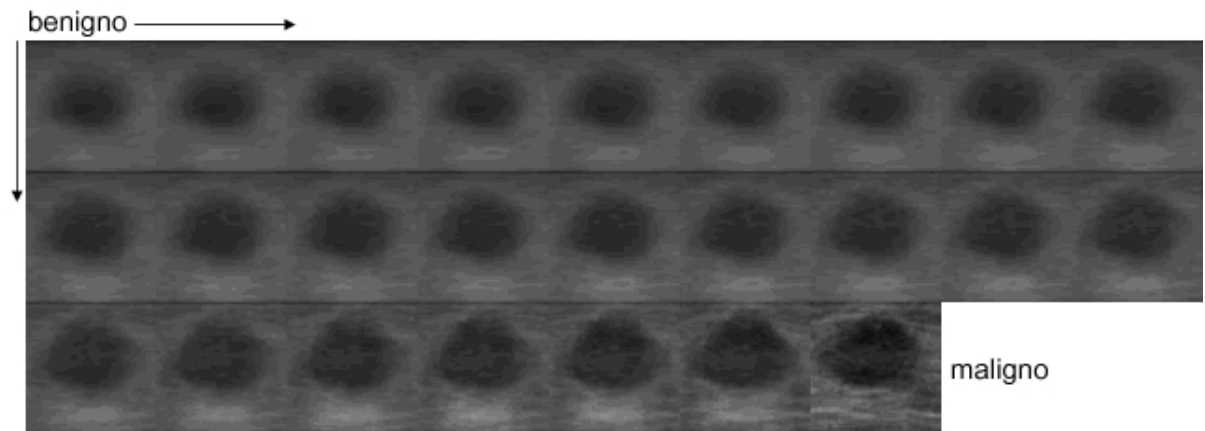


Figura 5.26: Transição de tumores benignos para malignos por meio de protótipos segundo classificador SVM.

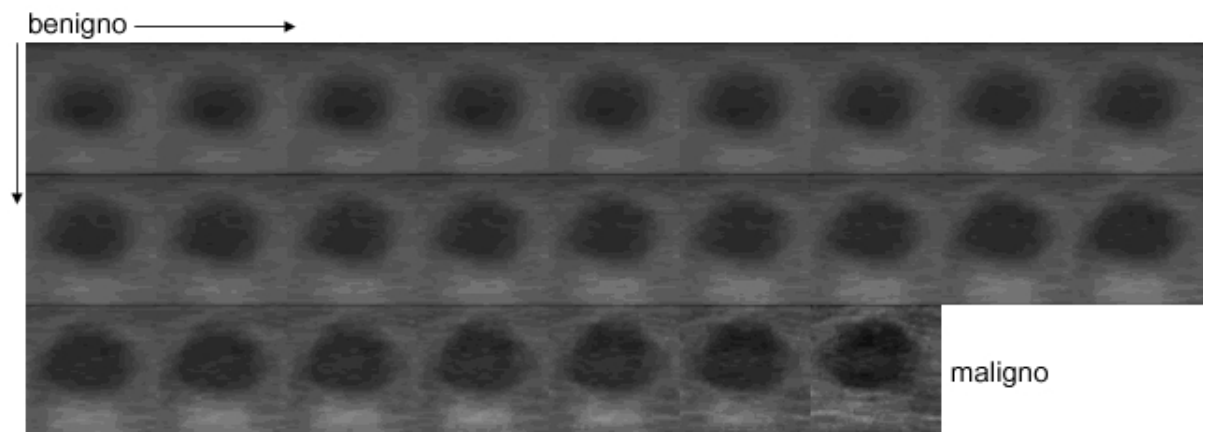


Figura 5.27: Transição de tumores benignos para malignos por meio de protótipos segundo classificador MLDA.

Capítulo 6

Conclusão

Este trabalho realizou comparações quantitativas e qualitativas entre os resultados das aplicações dos modelos estatísticos univariado e multivariado na análise de imagens ultrassonográficas de mama. O sucesso da aplicação desses modelos estatísticos é fortemente dependente das fases de pré-processamento e segmentação das imagens. Contudo, muitas vezes a aplicação de técnicas de melhoramento de imagens, na fase de pré-processamento, pode favorecer um determinado tipo de análise mas prejudicar outro. Como exemplo, pode-se citar as técnicas de suavização que sugerem a minimização das diferenças entre imagens, mas que podem também afetar o resultado de análises baseadas em intensidade dos pixels.

Os resultados obtidos na análise multivariada considerando a intensidade de todos os pixels da imagem simultaneamente, por meio dos classificadores SVM e MLDA, alcançaram desempenho superior perante a mesma análise utilizando características extraídas das imagens como circularidade e sombra acústica. Enquanto que a classificação linear utilizando a imagem como um todo alcançou a taxa de acerto absoluta em ambos classificadores, obteve-se a maior taxa de acerto de 78% analisando a característica circularidade pelo classificador MLDA.

Vale ressaltar que todas as imagens analisadas apresentavam tumores e que estes estavam posicionados na região central das mesmas. Em outras palavras, essas imagens foram normalizadas espacialmente com o intuito de evidenciar a região do tumor e eliminar regiões de fundo da imagem. A obtenção desse alto índice de acurácia foi certamente favorecida por estes ajustes feitos na etapa de pré-processamento das imagens. Na literatura o índice médio de acurácia dos classificadores é da ordem de 90%. Talvez a aplicação dos mesmos métodos em imagens de ultrassom sem pré-processamento possa apresentar

taxas de acertos menores e mais próximas desta, mas a análise estatística da imagem como um todo normalizada espacialmente apresentou resultado de classificação extremamente promissor.

Adicionalmente, a análise univariada indicou as regiões central e inferior da imagem como mais diferentes estatisticamente. Na navegação entre os grupos de tumores benignos e malignos, foi possível notar as diferenças entre os grupos justamente nessas regiões indicadas pela análise univariada. Muito provavelmente as características de circularidade e sombra acústica exerceram influência na discriminância dos grupos. Entretanto, a análise final dos resultados indicou que as diferenças entre as imagens de tumores benignos e malignos não são concentradas em determinadas regiões da imagem. Na verdade, essas diferenças estão espalhadas por toda a área da imagem. Pode-se concluir que as diferenças entre os grupos não são limitadas à observação de determinadas regiões ou características exclusivas. As diferenças são difusas e complexas, e, portanto, outras partes da imagem devem ser observadas para alcançar um bom resultado na identificação correta de cada classe ou grupo de tumores.

Por fim, acredita-se que os resultados promissores alcançados na análise das imagens como um todo, tanto para classificadores não-paramétricos e paramétricos como o SVM e o MLDA, podem ser de grande valia no treinamento e suporte para radiologistas. No caso de radiologistas iniciantes, esse tipo de análise possibilitaria a confirmação de determinada classificação do ponto de vista estatístico, principalmente em casos mais simples. Já no caso de radiologistas experientes, os resultados dos classificadores poderiam servir de apoio na tomada de decisões, principalmente nos casos mais complexos e duvidosos. Por exemplo, nos casos mais duvidosos, o radiologista poderia verificar o resultado do classificador antes de solicitar um novo tipo de exame ou mesmo determinar um tipo de tumor. A natureza não invasiva, o baixo custo e a formação da imagem em tempo real fazem da imagiologia ultrassonográfica, aliada a análise estatística de imagem, uma ferramenta essencial no diagnóstico médico.

6.1 Trabalhos Futuros

Com base na pesquisa desenvolvida nesse trabalho, algumas sugestões de estudos futuros são propostas com intuito de melhorar os resultados dos processos de classificação

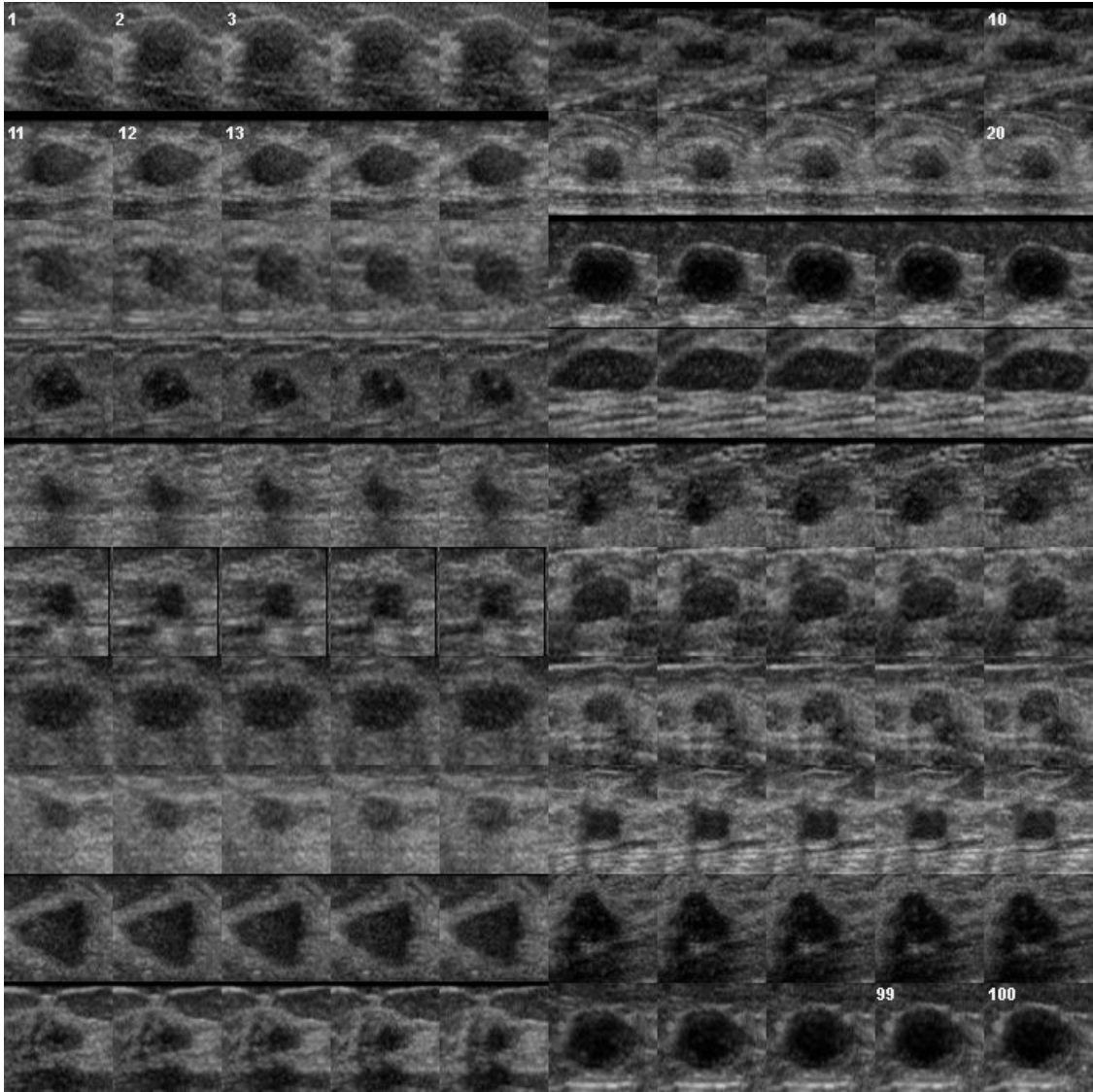
e extração de informações discriminantes em imagens de ultrassom.

Tendo em vista o resultado menos favorável da classificação linear utilizando, de forma isolada, um conjunto de características extraídas das imagens de ultrassom (circularidade, heterogeneidade ou sombra acústica), sugere-se o agrupamento dessas características visando melhorar o resultado em uma nova classificação. Ainda nesse contexto, poderia também ser investigada a utilização de classificadores não-lineares.

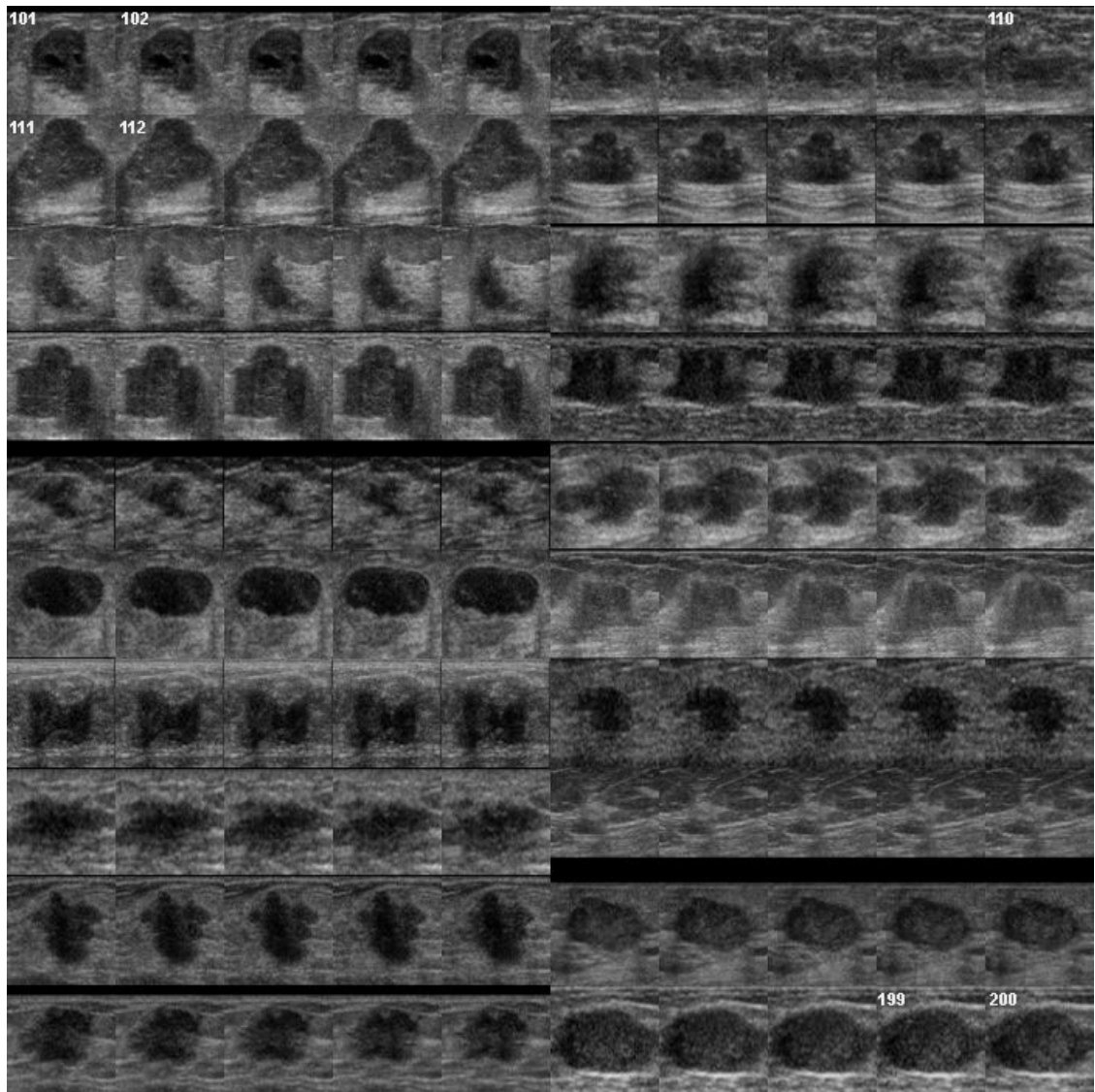
Outra possibilidade de desenvolvimento futuro se refere a questão da qualidade das imagens de ultrassom. Segundo Savage (2010), as imagens ultrassonográficas apresentam ruídos e têm baixa qualidade. Uma sugestão de estudo é a validação da classificação, linear e não-linear, com imagens de ultrassom de alta resolução. Ultimamente, a utilização desse tipo de imagem vem crescendo de forma considerável, influenciando os sistemas de detecção e classificação em decorrência da grande diferença de qualidade e dimensão da imagem.

Anexo I

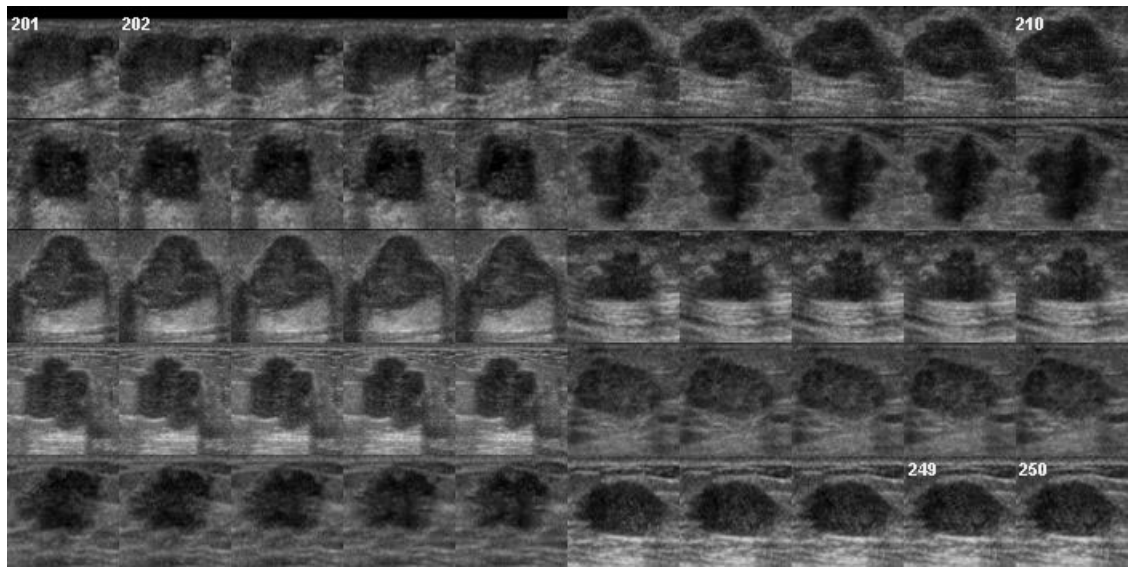
Amostras de tumores benignos (100 imagens de ultrassom).



Amostras de tumores malignos (150 imagens de ultrassom).



Amostras de tumores malignos (continuação).



Anexo II

A seguir são listados os valores discriminantes das imagens projetadas nos hiperplanos de separação dos classificadores lineares SVM e MLDA. As imagens são referenciadas por números e podem ser observadas no Anexo I.

Imagem	Valor	G	Imagem	Valor	G	Imagem	Valor	G	Imagem	Valor	G
1	1,47219087680286	1	66	1,57659802010555	1	131	-0,44481213579487	2	196	-0,396244261480607	2
2	1,60012193194138	1	67	1,79707919828980	1	132	-0,486639438335124	2	197	-0,498469782473469	2
3	1,61624800290647	1	68	1,79232858254924	1	133	-0,432544447256411	2	198	-0,528442742262001	2
4	1,57374260551676	1	69	1,60585884613984	1	134	-0,502540703818212	2	199	-0,565897888455227	2
5	1,47689467220784	1	70	1,45956838561057	1	135	-0,556000041018369	2	200	-0,562128875940659	2
6	1,58509136507206	1	71	1,44947298522426	1	136	-0,391727012276412	2	201	-0,627363020489628	2
7	1,55622478146372	1	72	1,54434122244702	1	137	-0,481619321345244	2	202	-0,576142105754132	2
8	1,53499204966221	1	73	1,55672484623596	1	138	-0,441546963929707	2	203	-0,381563730789566	2
9	1,56933298057366	1	74	1,59963253852312	1	139	-0,505737320378323	2	204	-0,511631425831798	2
10	1,48499149424396	1	75	1,43853785866623	1	140	-0,440802371086582	2	205	-0,446724602248222	2
11	1,54510136077058	1	76	1,56446891157452	1	141	-0,387070698606632	2	206	-1,41859939278848	2
12	1,644193229064	1	77	1,51304829964152	1	142	-0,475134311654746	2	207	-0,55061935731246	2
13	1,61988850186443	1	78	1,55252797734478	1	143	-0,430449031647076	2	208	-1,64365200255614	2
14	1,52850837076169	1	79	1,5621260602049	1	144	-0,540574002549721	2	209	-1,72280961151548	2
15	1,55826658194483	1	80	1,50475032680435	1	145	-0,545560382148072	2	210	-1,84329952952153	2
16	1,69876102860527	1	81	1,40288735815404	1	146	-0,32596554998624	2	211	-0,595132467230681	2
17	1,75076718247642	1	82	1,5607814169808	1	147	-0,577441245987505	2	212	-0,79724773658769	2
18	1,77011267376264	1	83	1,55395157004853	1	148	-0,595933652726198	2	213	-0,892511191165148	2
19	1,75193167003416	1	84	1,54839022404857	1	149	-0,601169354560229	2	214	-0,816668141854957	2
20	1,76783171260901	1	85	1,48891061539521	1	150	-0,295115444207261	2	215	-0,706010416428243	2
21	1,25600470590775	1	86	1,45674967143099	1	151	-0,409647671581329	2	216	-0,380123931214309	2
22	1,64531447480524	1	87	1,59202549163600	1	152	-0,522686575642354	2	217	-0,537073082031469	2
23	1,60338987496069	1	88	1,59518594311004	1	153	-0,523087147363568	2	218	-0,612836844137953	2
24	1,52156080125137	1	89	1,50115976981577	1	154	-0,476371102793276	2	219	-0,66974734561833	2
25	1,54692179423490	1	90	1,54430968130891	1	155	-0,422399838563018	2	220	-0,610138810681587	2
26	1,70271470211413	1	91	1,98790745112993	1	156	-0,43616442262741	2	221	-0,751712057293747	2
27	1,70370252083243	1	92	1,90478884654665	1	157	-0,482606961645772	2	222	-0,752202183738744	2
28	1,7028209719512	1	93	1,80727939538878	1	158	-0,482694858144226	2	223	-0,809922029285688	2
29	1,71460986253486	1	94	1,70209838060872	1	159	-0,514722191179639	2	224	-0,893532132193873	2
30	1,20745288428109	1	95	1,37412135954632	1	160	-0,363573869453471	2	225	-0,92851516111069	2
31	1,30316284307129	1	96	1,46470037030802	1	161	-0,267171828903803	2	226	-0,496085900832893	2
32	1,63263408361376	1	97	1,59282113325912	1	162	-0,54923037271554	2	227	-0,491426890889363	2
33	1,6678380741966	1	98	1,59670949410297	1	163	-0,568219120557439	2	228	-0,44331347569006	2
34	1,60153854339795	1	99	1,54704183516785	1	164	-0,563596287923586	2	229	-0,466096226534598	2
35	1,44196488228662	1	100	1,51519927320884	1	165	-0,370710536850507	2	230	-0,473410936218039	2
36	1,35558148218455	1	101	-0,507379776555501	2	166	-0,177055618494340	2	231	-0,716581997233315	2
37	1,63533150052769	1	102	-0,510679010903197	2	167	-0,514939342073882	2	232	-0,725920986688531	2
38	1,64650527371592	1	103	-0,48344196609535	2	168	-0,390132547289874	2	233	-0,785549508692907	2
39	1,47387524257047	1	104	-0,461088054341999	2	169	-0,441237067092126	2	234	-0,834125896421907	2
40	1,49272444889804	1	105	-0,408688129153396	2	170	-0,457851984491186	2	235	-0,845815093837347	2
41	1,46189414715303	1	106	-0,810109689365187	2	171	-0,356100053068244	2	236	-0,85361681270628	2
42	1,60621149234951	1	107	-0,711473288871328	2	172	-0,4844165852396	2	237	-0,864652492426436	2
43	1,62062597815265	1	108	-0,679737402090899	2	173	-0,495092674358816	2	238	-0,989186221229314	2
44	1,70073066903373	1	109	-0,604339716142874	2	174	-0,41768804427204	2	239	-1,00926386160572	2
45	1,27339200663231	1	110	-0,302288590117992	2	175	-0,439213028391676	2	240	-0,950969691003752	2
46	1,58938064430199	1	111	-1,03516094642983	2	176	-0,330143126116806	2	241	-1,09796186128476	2
47	1,49783524636022	1	112	-1,00901499081801	2	177	-0,55241021386345	2	242	-1,16658173597896	2
48	1,48302179911178	1	113	-1,03675610796357	2	178	-0,56474527228927	2	243	-1,16740346522653	2
49	1,58031231157821	1	114	-0,994547579645037	2	179	-0,513668665696715	2	244	-1,20182663043696	2
50	1,39174956802066	1	115	-0,861300753383734	2	180	-0,430903314035073	2	245	-1,26078137881542	2
51	1,85445654914875	1	116	-0,433249433153259	2	181	-0,419694022726824	2	246	-0,433490998180713	2
52	1,73195133077598	1	117	-0,498891038222351	2	182	-0,568154865714093	2	247	-0,491194595693705	2
53	1,61147667447306	1	118	-0,536073255878778	2	183	-0,495851853220906	2	248	-0,47660542288386	2
54	1,48432020055129	1	119	-0,572275610390252	2	184	-0,434308586530748	2	249	-0,466047366944988	2
55	1,53291401498563	1	120	-0,653522011383448	2	185	-0,421399438099507	2	250	-0,47295730210996	2
56	1,43465931978307	1	121	-0,395283931538610	2	186	-0,547432735141578	2			
57	1,66039877659389	1	122	-0,52824458696681	2	187	-0,490919976410341	2			
58	1,63483605122857	1	123	-0,540356269984201	2	188	-0,453646959676176	2			
59	1,67956512959766	1	124	-0,505074069863827	2	189	-0,477249663568448	2			
60	1,79378829525455	1	125	-0,38247854977819	2	190	-0,430437489277409	2			
61	1,50636268552150	1	126	-0,286233113667469	2	191	-0,514423682557664	2			
62	1,57276694236829	1	127	-0,523780503133194	2	192	-0,49747805211097	2			
63	1,59370171592256	1	128	-0,499824308585261	2	193	-0,468971707754529	2			
64	1,60920403031039	1	129	-0,446294193787982	2	194	-0,468116253760465	2			
65	1,61430309446109	1	130	-0,463340807103546	2	195	-0,427119078251695	2			

Tabela 6.1: Pesos discriminantes das imagens projetadas no hiperplano de separação do classificador SVM

Imagem	Valor	G	Imagem	Valor	G	Imagem	Valor	G	Imagem	Valor	G
1	9,7284059157539	1	66	11,1300345777713	1	131	-8,92514235259343	2	196	-8,72562063416285	2
2	10,2738603205805	1	67	12,9541470074769	1	132	-9,07859878699207	2	197	-9,72516898512367	2
3	10,0451968619981	1	68	12,6364839847386	1	133	-9,05231742848272	2	198	-10,4083318120288	2
4	9,47944885046577	1	69	11,0450087374306	1	134	-9,48249968295938	2	199	-10,5274386689239	2
5	8,76856484162634	1	70	9,92556083192392	1	135	-9,87804688946	2	200	-10,1813688347838	2
6	11,1019152202826	1	71	5,39524562793689	1	136	-5,14533472417663	2	201	-10,7856905487262	2
7	10,9103418345018	1	72	6,27241941004223	1	137	-6,02152831727718	2	202	-10,1392044639402	2
8	10,6679288244675	1	73	7,20942954044203	1	138	-6,42710534321675	2	203	-9,11904613628255	2
9	10,9033213615429	1	74	8,09750373975729	1	139	-7,07365107655136	2	204	-9,31546447307892	2
10	10,5802760251961	1	75	7,69106943846836	1	140	-6,63821230719411	2	205	-8,87544560966677	2
11	9,71854083895925	1	76	9,23602288492684	1	141	-4,49128063862812	2	206	-13,1952798030271	2
12	11,0301045079639	1	77	9,26031936413033	1	142	-5,04395066245039	2	207	-14,2273730232349	2
13	11,2975164180249	1	78	9,51703537818521	1	143	-6,13994777339101	2	208	-15,0574544829906	2
14	10,6910658805302	1	79	9,41475425968892	1	144	-7,39812427442245	2	209	-15,5621430173509	2
15	10,3562465294799	1	80	8,76404658167273	1	145	-7,60881165940151	2	210	-16,4056365088714	2
16	12,69837079819	1	81	3,60342969432709	1	146	-7,30806576925997	2	211	-10,4966414049129	2
17	13,100625752626	1	82	4,81710395855809	1	147	-8,60581641964855	2	212	-12,4879853663393	2
18	13,1361863144849	1	83	5,77052429027563	1	148	-8,70241720876415	2	213	-13,5626772345212	2
19	12,8419295258395	1	84	6,57365926913049	1	149	-8,37741539085679	2	214	-13,1821126940798	2
20	12,6229097999747	1	85	7,00289833509559	1	150	-6,91781067497068	2	215	-12,3350055862076	2
21	3,13328501809774	1	86	6,83913574059499	1	151	-8,40468203557241	2	216	-7,94077309889358	2
22	5,5479432667912	1	87	7,39036410376613	1	152	-8,97439189788699	2	217	-8,82277048998992	2
23	7,10884158137132	1	88	7,2539568597654	1	153	-9,0668656072049	2	218	-9,69823968242789	2
24	7,41764364674439	1	89	6,64560617608926	1	154	-8,6739405516434	2	219	-10,199704133567	2
25	7,66163969061256	1	90	6,80264982224199	1	155	-8,36107020443096	2	220	-9,39156956052534	2
26	9,89846710837772	1	91	13,328888886252	1	156	-7,15845045037463	2	221	-10,3449583895974	2
27	9,2597444359477	1	92	12,9317215787251	1	157	-7,4323695906063	2	222	-10,3788945402934	2
28	8,3075747246271	1	93	12,2277868425103	1	158	-7,50571806884679	2	223	-10,9176930804564	2
29	6,74213339828995	1	94	11,14128101777074	1	159	-7,54733871529177	2	224	-11,9026073793839	2
30	3,84292366100954	1	95	9,06647262589562	1	160	-6,78379587227057	2	225	-12,4301273300138	2
31	5,21988219409022	1	96	7,91972059588166	1	161	-2,945987169741	2	226	-7,9983447416001	2
32	6,83791660110146	1	97	8,47237099529476	1	162	-4,74191266967476	2	227	-7,64771570381898	2
33	7,6319223101288	1	98	8,87615228347893	1	163	-6,04331865482886	2	228	-6,98820399397357	2
34	6,87031940737374	1	99	8,19074264976642	1	164	-6,71802701953833	2	229	-7,20129506000355	2
35	6,10879765319032	1	100	7,38849781866228	1	165	-5,89230164430338	2	230	-7,38561761346815	2
36	3,42505766406412	1	101	-8,15004461252086	2	166	0,0051722572281796	1	231	-10,7110152480183	2
37	4,86140038481753	1	102	-8,12766049358068	2	167	-0,648336118505445	2	232	-10,9323966937092	2
38	5,27035311580731	1	103	-7,80464342414912	2	168	-1,29092659370192	2	233	-11,4945011458527	2
39	4,01199528915809	1	104	-7,36850847007937	2	169	-2,26781913603724	2	234	-11,9997797227327	2
40	3,29993272814196	1	105	-7,13187885772434	2	170	-2,91004769912819	2	235	-12,1339917674151	2
41	7,20053178496392	1	106	-11,4896807353275	2	171	-3,77950612967979	2	236	-11,5791154756519	2
42	7,60367443527107	1	107	-10,597817368609	2	172	-4,43611180214823	2	237	-11,7448271759749	2
43	7,16063920449655	1	108	-9,90863061755314	2	173	-4,47188515223159	2	238	-12,6474764040651	2
44	6,50933625658284	1	109	-9,05584594106869	2	174	-3,85814663801299	2	239	-13,0830218658189	2
45	4,52906493915793	1	110	-7,67229698431311	2	175	-3,77304885071392	2	240	-12,7412511743174	2
46	6,70261915688326	1	111	-12,0607735411434	2	176	-5,75387293424992	2	241	-12,6743325857412	2
47	5,34339171772943	1	112	-12,0068417099303	2	177	-7,05418930418166	2	242	-13,4711982248685	2
48	4,85543582110581	1	113	-12,3920723397627	2	178	-7,95340592816362	2	243	-13,6369866655219	2
49	4,9539448684533	1	114	-12,1593808805702	2	179	-7,74854895929691	2	244	-13,8092323324611	2
50	4,03494027750322	1	115	-11,2767208755484	2	180	-7,29101342494093	2	245	-14,1537577337441	2
51	11,5103995834753	1	116	-8,79311471033138	2	181	-8,0422968144981	2	246	-9,55076677176304	2
52	10,3016838857503	1	117	-9,22410809393523	2	182	-8,5270215710819	2	247	-9,72300716168615	2
53	8,7310843017716	1	118	-9,78561328692639	2	183	-7,91818348221702	2	248	-9,68681447845304	2
54	7,3456118464431	1	119	-10,2591766074981	2	184	-7,51310320740288	2	249	-9,78453069480657	2
55	7,39594681687133	1	120	-10,9927222857439	2	185	-7,12691818313005	2	250	-9,94011604908137	2
56	11,0599615923646	1	121	-7,81196663134759	2	186	-8,99093171021984	2			
57	12,7984270587201	1	122	-8,03866990307588	2	187	-8,42246068240552	2			
58	12,736704738184	1	123	-7,66919004060679	2	188	-8,05754749928311	2			
59	13,1785273780826	1	124	-6,7543679727482	2	189	-7,9936141335644	2			
60	14,0434348757221	1	125	-5,8729215150213	2	190	-7,83785866988543	2			
61	6,09454462214747	1	126	-5,23189551160133	2	191	-9,03635871977973	2			
62	6,4814824421071	1	127	-6,30580700370143	2	192	-8,84126515002236	2			
63	6,7597750769653	1	128	-6,80685061106726	2	193	-8,29497466970747	2			
64	7,04304601927009	1	129	-6,78935115136961	2	194	-7,79939625707754	2			
65	7,20862101118445	1	130	-6,82591196450319	2	195	-7,26937476312	2			

Tabela 6.2: Pesos discriminantes das imagens projetadas no hiperplano de separação do classificador MLDA

Glossário

ACM	Active Contour Model, 29
AEM	Auto-Exame Clínico da Mama, 13
Anisotropia	Fenômeno no qual as propriedades físicas da imagem variam conforme a direção da mesma, 22
AP	Affinity Propagation, 48
ARGF	Aggressive Region Growing Filtering, 22
Assintomáticas	Sem sintoma da doença em questão, 13
AWMF	Adaptive Weighted Median Filter, 22
BI-RADS	Breast Image Reporting and Data System. Sistema padronizado, originário nos Estados Unidos, utilizado para uniformizar os relatos de radiologia referentes às imagens de mamografia, 48
CAD	Computer-Aided Diagnosis, 17
CDIS	Carcinoma Ductal In Situ, 10
CLIS	Carcinoma Ductal Lobular In Situ, 10
DWT	Discrete Wavelet Transform, 23
ECM	Exame Clínico da Mama, 13
EDP	Equação Diferencial Parcial, 22
FSVM	Fuzzy Support Vector Machine, 48
GF	Geometric Filter, 22
GVF	Gradient Vector Flow, 29
IDWT	Inverse Discrete Wavelet Transform, 23
INCA	Instituto Nacional de Câncer, 10
Isotropia	Fenômeno no qual as propriedades físicas da imagem não sofrem influência da direção da mesma, 22
LDA	Linear Discriminant Analysis, 48
MAP	Maximum a Posteriori, 26

MAP	Maximum a posteriori, 22
MLDA	Maximum uncertainty Linear Discriminant Analysis, 48
MLP	Multi-Layered Perceptron, 48
MRF	Markov Random Field, 26
MSD	Minimum Side Difference, 34
NCD	Non linear Coherence Enhancement Diffusion, 22
Ncut	Normalized Cut Technique, 31
Neoplasia maligna	Sinônimo de tumor maligno, 5
PCA	Principal Component Analysis, 34
RM	Ressonância Magnética, 14
SAR	Synthetic Aperture Radar, 23
SOM	Self-Organizing Map, 29
SRAD	Speckle Reducing Anisotropic Diffusion, 22
SVM	Support Vector Machine, 29
TC	Tomografia Computadorizada, 14

Referências Bibliográficas

ABD-ELMONIEM, K.Z.; YOUSSEF, A.-B.M.; KADAH, Y.M. Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion. **Biomedical Engineering, IEEE Transactions on**, v. 49, n. 9, p. 997 –1014, sept. 2002.

ACHIM, A.; BEZERIANOS, A.; TSAKALIDES, P. Novel bayesian multiscale method for speckle removal in medical ultrasound images. **Medical Imaging, IEEE Transactions on**, v. 20, n. 8, p. 772 –783, aug. 2001.

AMERICAN CANCER SOCIETY. **Breast Cancer: Cancer Facts and Figures**. 2010.

BELLMAN, Richard E. **Adaptive control processes**. [S.l.]: Princeton University Press, 1961.

BOUKERROUI, Djamal; BASKURT, Atilla; NOBLE, J. Alison; BASSET, Olivier. Segmentation of ultrasound images–multiresolution 2d and 3d algorithm based on global and local statistics. **Pattern Recognition Letters**, v. 24, n. 4-5, p. 779 – 790, 2003.

BOUKERROUI, Djamal; BASSET, Olivier; GUÉRIN, Nicole; BASKURT, Attila. Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation. **European Journal of Ultrasound**, v. 8, n. 2, p. 135 – 144, 1998.

BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. 1998.

CALÓOPE, P.B.; MEDEIROS, F.N.S.; MARQUES, R.C.P.; COSTA, R.C.S. A comparison of filters for ultrasound images. In: SOUZA, José Neuman de; DINI, Petre; LORENZ, Pascal (Ed.). **Telecommunications and Networking - ICT 2004**. [S.l.]: Springer Berlin / Heidelberg, 2004, (Lecture Notes in Computer Science, v. 3124). p. 1035–1040.

CAVINATO, Marcos Vinicius. **Realce de imagens utilizando transformações de intensidade**. [S.l.], 2009.

CHANG, Ruey-Feng; WU, Wen-Jie; MOON, Woo; CHEN, Dar-Ren. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. **Breast Cancer Research and Treatment**, Springer Netherlands, v. 89, p. 179–185, 2005.

CHANG, Ruey-Feng; WU, Wen-Jie; MOON, Woo Kyung; CHEN, Dar-Ren. Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. **Ultrasound in Medicine & Biology**, v. 29, n. 5, p. 679 – 686, 2003.

CHANG, Ruey-Feng et al. Segmentation of breast tumor in three-dimensional ultrasound images using three-dimensional discrete active contour model. **Ultrasound in Medicine & Biology**, v. 29, n. 11, p. 1571 – 1581, 2003.

CHANG, Ruey-Feng et al. Support vector machines for diagnosis of breast tumors on us images. **Academic Radiology**, v. 10, n. 2, p. 189 – 197, 2003.

CHEN, Yan; YIN, Ruming; FLYNN, Patrick; BROCHAT, Shira. Aggressive region growing for speckle reduction in ultrasound images. **Pattern Recogn. Lett.**, Elsevier Science Inc., New York, NY, USA, v. 24, n. 4-5, p. 677–691, 2003.

CHENG, H.D. et al. Automated breast cancer detection and classification using ultrasound images: A survey. **Pattern Recognition**, v. 43, n. 1, p. 299 – 317, 2010.

COHN, Wendy F.; ROPKA, Mary E.; JONES, Susan M.; MIESFELDT, Susan. Information needs about hereditary breast cancer among women with early-onset breast cancer. **Cancer Detection and Prevention**, v. 27, n. 5, p. 345 – 352, 2003.

CRIMMINS, Thomas R. Geometric filter for speckle reduction. **Appl. Opt.**, OSA, v. 24, n. 10, p. 1438–1443, 1985.

DOMG, Y.; MILNE, A.K.; FORSTER, B.C. Toward edge sharpening: a sar speckle filtering algorithm. **Geoscience and Remote Sensing, IEEE Transactions on**, v. 39, n. 4, p. 851 –863, apr. 2001.

DRUKKER, Karen; GIGER, Maryellen L.; METZ, Charles E. Robustness of computerized lesion detection and classification scheme across different breast us platforms. **Radiology**, v. 237, n. 3, p. 834–840, 2005.

DUPONT, William D.; PAGE, David L. Risk factors for breast cancer in women with proliferative breast disease. **New England Journal of Medicine**, 1985.

DUTT, V.; GREENLEAF, J.F. Adaptive speckle reduction filter for log-compressed b-scan images. **Medical Imaging, IEEE Transactions on**, v. 15, n. 6, p. 802 –813, dec. 1996.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, n. 7, p. 179–188, 1936.

FORSBERG, F. Ultrasonic biomedical technology; marketing versus clinical reality. **Ultrasonics**, v. 42, n. 1-9, p. 17 – 27, 2004. ISSN 0041-624X. Proceedings of Ultrasonics International 2003.

FROST, Victor S.; STILES, Josephine Abbott; SHANMUGAN, K. S.; HOLTZMAN, Julian C. A model for radar images and its application to adaptive digital filtering of multiplicative noise. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, PAMI-4, n. 2, p. 157 –166, mar. 1982.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**. 2. ed. [S.l.]: Academic Press, 1990.

GERIG, G.; KUBLER, O.; KIKINIS, R.; JOLESZ, F.A. Nonlinear anisotropic filtering of mri data. **Medical Imaging, IEEE Transactions on**, v. 11, n. 2, p. 221 –232, jun. 1992.

GIRALDI, Gilson et al. Dimensionality reduction, classification and reconstruction problems in statistical learning approaches. **Revista de Informática Teórica e Aplicada**, v. 15, n. 1, 2008.

GIRALDI, Gilson A. et al. Statistical learning approaches for discriminant features selection. **Journal of the Brazilian Computer Society**, v. 14, n. 2, p. 7–22, 2008.

GOMEZ, W.; LEIJA, L.; PEREIRA, W.C.A.; INFANTOSI, A.F.C. Morphological operators on the segmentation of breast ultrasound images. In: . [S.l.: s.n.], 2009. p. 67–71.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital Image Processing**. Second edition. [S.l.]: Addison-Wesley, 2002.

HAO, Xiaohui; GAO, Shang kai; GAO, Xiaorong. A novel multiscale nonlinear thresholding method for ultrasonic speckle suppressing. **Medical Imaging, IEEE Transactions on**, v. 18, n. 9, p. 787–794, sep. 1999.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Second edition. [S.l.]: Springer, 2009.

HORSCH, Karla; GIGER, Maryellen L.; VENTA, Luz A.; VYBORNÝ, Carl J. Automatic segmentation of breast lesions on ultrasound. **Medical Physics**, AAPM, v. 28, n. 8, p. 1652–1659, 2001.

HORSCH, Karla; GIGER, Maryellen L.; VENTA, Luz A.; VYBORNÝ, Carl J. Computerized diagnosis of breast lesions on ultrasound. **Medical Physics**, AAPM, v. 29, n. 2, p. 157–164, 2002.

HUANG, Yu-Len; CHEN, Dar-Ren. Watershed segmentation for breast tumor in 2-d sonography. **Ultrasound in Medicine & Biology**, v. 30, n. 5, p. 625–632, 2004. ISSN 0301-5629.

HUANG, Yu-Len; CHEN, Dar-Ren. Automatic contouring for breast tumors in 2-d sonography. In: . [S.l.: s.n.], 2005. p. 3225–3228.

HUANG, Yu-Len; WANG, Kao-Lun; CHEN, Dar-Ren. Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. **Neural Computing & Applications**, Springer London, v. 15, p. 164–169, 2006. 10.1007/s00521-005-0019-5.

INSTITUTO NACIONAL DE CÂNCER. **Estimativa 2010: incidência de câncer no Brasil**. 2009.

JAIN, A.K.; DUIN, R.P.W.; MAO, Jianchang. Statistical pattern recognition: a review. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 22, n. 1, p. 4–37, jan. 2000.

JAIN, A. K.; CHANDRASEKARAN, B. Dimensionality and sample size considerations in pattern recognition practice. In: KRISHNAIAH, P. R.; KANAL, L. N. (Ed.). **Handbook of Statistics**. The Netherlands, North-Holland: Eds. Amsterdam, 1982. v. 2, p. 835–855.

JOHNSON, Richard A.; WICHERN, Dean W. **Applied Multivariate Statistical Analysis**. Sixth edition. [S.l.]: Pearson Prentice Hall, 2007.

JUMAAT, A.K.; RAHMAN, W.E.Z.W.A.; IBRAHIM, A.; MAHMUD, R. Comparison of balloon snake and gvf snake in segmenting masses from breast ultrasound images. In: . [S.l.: s.n.], 2010. p. 505–509.

KASS, Michael; WITKIN, Andrew; TERZOPOULOS, Demetri. Snakes: Active contour models. **International Journal of Computer Vision**, v. 1, n. 4, p. 321–331, 1987.

KIM, Yong Sun; RA, Jong Beom. Improvement of ultrasound image based on wavelet transform: speckle reduction and edge enhancement. In: **in Proc. SPIE Medical Imaging 2005**. [S.l.: s.n.], 2005. p. 1085–1092.

KOPANS, Daniel B. **Imagem da Mama**. 2^a. ed. [S.l.]: Medsi, 2000.

KOTROPOULOS, Constantine; PITAS, Ioannis. Segmentation of ultrasonic images using support vector machines. **Pattern Recognition Letters**, v. 24, n. 4-5, p. 715 – 727, 2003.

KUAN, Darwin T.; SAWCHUK, Alexander A.; STRAND, Timothy C.; CHAVEL, Pierre. Adaptive noise smoothing filter for images with signal-dependent noise. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, PAMI-7, n. 2, p. 165 –177, mar. 1985.

KUO, Wen-Jia et al. Retrieval technique for the diagnosis of solid breast tumors on sonogram. **Ultrasound in Medicine & Biology**, v. 28, n. 7, p. 903 – 909, 2002.

LEÃO, R. D. **Extração multilinear de informações discriminantes em imagens de ressonância magnética do cérebro humano**. Dissertação (Mestrado) — Centro Universitário da FEI, 2009.

LEÃO, Rafael D.; SATO, João R.; THOMAZ, Carlos E. **Comparação entre as análises estatísticas univariada e multivariada para extração de informação discriminante em imagens de ressonância magnética do cérebro humano**. [S.l.], 2009.

LEE, Carol H. et al. Breast cancer screening with imaging: Recommendations from the society of breast imaging and the acr on the use of mammography, breast mri, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. **Journal of the American College of Radiology**, v. 7, n. 1, p. 18 – 27, 2010.

LEE, Jong-Sen. Digital image enhancement and noise filtering by use of local statistics. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, PAMI-2, n. 2, p. 165 –168, mar. 1980.

LIHUA, Li; JIANGLI, Lin; DEYU, Li; TIANFU, Wang. Segmentation of medical ultrasound image based on markov random field. In: . [S.l.: s.n.], 2007. p. 968 –971.

LIU, Bo et al. Probability density difference-based active contour for ultrasound image segmentation. **Pattern Recognition**, v. 43, n. 6, p. 2028 – 2042, 2010.

LIU, Haihua; XIE, Changsheng; CHEN, Zhouhui; LEI, Yi. Segmentation of ultrasound image based on morphological operation and fuzzy clustering. In: . [S.l.: s.n.], 2006. p. 4 pp.

LIU, Xu; HUO, Zhimin; ZHANG, Jiwu. Automated segmentation of breast lesions in ultrasound images. In: . [S.l.: s.n.], 2005. p. 7433 –7435.

LORENA, A. C.; CARVALHO, A. C. P. L. F. **Uma Introdução às Support Vector Machines**. [S.l.], 2007.

LOUPAS, T.; MCDICKEN, W.N.; ALLAN, P.L. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. **Circuits and Systems, IEEE Transactions on**, v. 36, n. 1, p. 129 –135, jan. 1989.

MAGALHÃES, Marcos N.; LIMA, Antonio Carlos P. **Noções de Probabilidade e Estatística**. 6^a. ed. [S.l.]: EDUSP, 2010.

MARTINS, Letícia Athayde Linhares; BARRA, Alexandre de Almeida; LUCENA, Clécio Ênio Murta de. Microcalcificações mamárias suspeitas de malignidade. **Revista Brasileira de Cancerologia** 2010, v. 56(2), p. 251–258, 2010.

MENECELLI, Renan Caldeira; RIBEIRO, Patricia Bellin; SCHIABEL, Homero. Desenvolvimento de um software classificador da forma de nódulos mamográficos segmentados utilizando a rede neural artificial multi-layer perceptron (mlp). **VI Workshop de Visão Computacional**, 2010.

NOBLE, J.A.; BOUKERROUI, D. Ultrasound image segmentation: a survey. **Medical Imaging, IEEE Transactions on**, v. 25, n. 8, p. 987 –1010, aug. 2006.

PANJWANI, D.K.; HEALEY, G. Markov random field models for unsupervised segmentation of textured color images. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 17, n. 10, p. 939 –954, oct. 1995.

PERONA, P.; MALIK, J. Scale-space and edge detection using anisotropic diffusion. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 12, n. 7, p. 629 –639, jul. 1990.

PRAGER, Richard W.; GEE, Andrew H.; TREECE, Graham M.; BERMAN, Laurence H. Analysis of speckle in ultrasound images using fractional order statistics and the homodyned k-distribution. **Ultrasonics**, v. 40, n. 1-8, p. 133 – 137, 2002.

QUISTGAARD, J.U. Signal acquisition and processing in medical diagnostic ultrasound. **Signal Processing Magazine, IEEE**, v. 14, n. 1, p. 67 –74, jan. 1997.

RANGSANSERI, Y.; PRASONGSOOK, W. Speckle reduction using wiener filtering in wavelet domain. In: . [S.l.: s.n.], 2002. v. 2, p. 792 – 795 vol.2.

ROBSON, Mark E.; OFFIT, Kenneth. Considerations in genetic counseling for inherited breast cancer predisposition. **Seminars in Radiation Oncology**, v. 12, n. 4, p. 362 – 370, 2002. Management of Breast Cancer in the Molecular Era.

RODRIGUES, Paulo S.; CHANG, Ruey-Feng; SURI, Jasjit S. Non-extensive entropy for cad systems of breast cancer images. **Computer Graphics and Image Processing, Brazilian Symposium on**, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, n. 3, p. 121–128, 2006.

RODRIGUES, Paulo S.; GIRALDI, Gilson A.; CHANG, Ruey-Feng; SURI, Jasjit S. Automatic classification of breast lesions in 3-d ultrasound images. In: _____. [S.l.]: Biomedical Imaging and Bioinformatics, 2006.

SATO, João Ricardo et al. Evaluating svm and mllda in the extraction of discriminant regions for mental state prediction. **NeuroImage**, v. 46, n. 1, p. 105 – 114, 2009.

SAVAGE, Neil. **Medical Imagers Lower the Dose**. mar. 2010. Disponível em: <<http://spectrum.ieee.org/biomedical/imaging/medical-imagers-lower-the-dose>>.

SHAH, Biren A.; FUNDARO, Gina M.; MANDAVA, Sabala. **Breast Imaging Review: A Quick Guide to Essential Diagnoses**. [S.l.: s.n.], 2010.

SHI, Jianbo; MALIK, J. Normalized cuts and image segmentation. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 22, n. 8, p. 888 –905, aug. 2000.

SHI, Xiangjun et al. Detection and classification of masses in breast ultrasound images. **Digital Signal Processing**, v. 20, n. 3, p. 824 – 836, 2010.

STEPHENS LARRY J.AND SPIEGEL, Murray R. **Statistics**. [S.l.]: McGraw-Hill, 2007.

SU, Yanni; WANG, Yuanyuan. Computer-aided classification of breast tumors using the affinity propagation clustering. In: . [S.l.: s.n.], 2010. p. 1 –4.

THOMAZ, Carlos et al. Multivariate statistical differences of mri samples of the human brain. **Journal of Mathematical Imaging and Vision**, Springer Netherlands, v. 29, p. 95–106, 2007.

THOMAZ, C.E.; GILLIES, D.F.; FEITOSA, R.Q. A new covariance estimate for bayesian classifiers in biometric recognition. **Circuits and Systems for Video Technology, IEEE Transactions on**, v. 14, n. 2, p. 214 – 223, feb. 2004.

THOMAZ, Carlos E. **Topicos Especiais em Aprendizagem - Notas de aula**. 2008.

THOMAZ, C. E.; KITANI, E. C.; GILLIES, D. F. A maximum uncertainty lda-based approach for limited sample size problems - with application to face recognition. **Journal of the Brazilian Computer Society**, v. 12, p. 7–18, September 2006.

VAPNIK, Vladimir N. **Statistical Learning Theory**. [S.l.]: Wiley-Interscience, 1998.

WOODS, John W.; BIEMOND, Jan. Comments on "a model for radar images and its application to adaptive digital filtering of multiplicative noise". **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, PAMI-6, n. 5, p. 658 –659, sep. 1984.

YU, Yongjian; ACTON, S.T. Speckle reducing anisotropic diffusion. **Image Processing, IEEE Transactions on**, v. 11, n. 11, p. 1260 – 1270, nov. 2002.

ZHI, Xiaona; WANG, Tianfu. An anisotropic diffusion filter for ultrasonic speckle reduction. In: . [S.l.: s.n.], 2008. p. 327 –330.