

CENTRO UNIVERSITÁRIO DA FEI
VANESSA SANTOS LESSA

**CLASSIFICAÇÃO DE IMAGENS DE ULTRASSOM DE CÂNCER DE MAMA BASEADA
EM INFORMAÇÕES HÍBRIDAS UTILIZANDO TEORIA DA INFORMAÇÃO**

São Bernardo do Campo, SP

2010

VANESSA SANTOS LESSA

**CLASSIFICAÇÃO DE IMAGENS DE ULTRASSOM DE CÂNCER DE MAMA BASEADA
EM INFORMAÇÕES HÍBRIDAS UTILIZANDO TEORIA DA INFORMAÇÃO**

Dissertação de Mestrado apresentada ao Centro Universitário da FEI para obtenção do título de Mestre em Engenharia Elétrica, orientada pelo Prof. Dr. Paulo Sérgio Rodrigues.

São Bernardo do Campo, SP
2010

Lessa, Vanessa Santos

Classificação de imagens de ultrassom de câncer de mama baseada em informações híbridas utilizando teoria da informação / Vanessa Santos Lessa. São Bernardo do Campo, 2010.

90 f. : il.

Dissertação - Centro Universitário da FEI.

Orientador: Prof. Dr. Paulo Sérgio Rodrigues

1. Câncer de Mama. 2. Diagnóstico Assistido por Computador. 3. Teoria da Informação. I. Rodrigues, Paulo Sérgio, 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: Vanessa Santos Lessa

Matrícula: 1071125

Título do Trabalho: CLASSIFICAÇÃO DE IMAGENS DE ULTRASSOM DE CÂNCER DE MAMA BASEADA EM INFORMAÇÕES HÍBRIDAS UTILIZANDO TEORIA DA INFORMAÇÃO.

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues

Data da realização da defesa: 27 / outubro / 2010

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

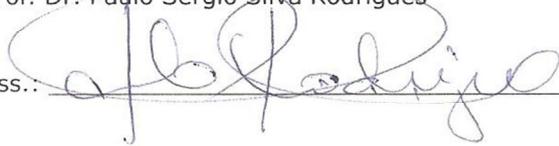
APROVADO

REPROVADO

São Bernardo do Campo, 27/10/2010

MEMBROS DA BANCA JULGADORA

Prof. Dr. Paulo Sérgio Silva Rodrigues

Ass.: 

Profª. Drª. Maria Cláudia Ferrari de Castro

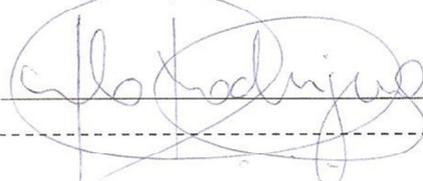
Ass.: 

Prof. Dr. Flavius Portella Ribas Martins

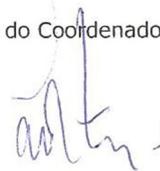
Ass.: 

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

Aos meus pais, Luiz e Maria, e a meu irmão Wagner. A todos aqueles que me incentivaram a seguir os meus sonhos.

Agradecimentos

Primeiramente gostaria de agradecer a Deus que está sempre comigo e iluminando meu caminho, aos meus queridos pais, Luiz e Maria e a meu irmão Wagner por todo o amor e dedicação, por sempre me incentivarem a seguir em frente.

Gostaria de agradecer a minha avó Zulmira pelo apoio nos momentos difíceis, aos meus padrinhos, Jorge e Iracema pelas orientações, ao meu namorado Marcelo pela compreensão nos momentos em que não podemos estar juntos e a todos os meus familiares que contribuíram para a realização do meu sonho.

Gostaria de agradecer também à todos os colegas do mestrado, Valquiria, Horst, Caruso, Danilo Nunes, Danilo Eduardo, Rodolpho Freire, Rodolfo Fussek, Thiago, Giuliano, Matheus e Rafael por terem contribuído com sugestões, discussões enriquecedoras sobre todo o conteúdo estudado e pela força sempre compartilhada.

Agradeço ao meu orientador, Paulo Sérgio Rodrigues pela paciência, motivação, dedicação e o fornecimento de seu conhecimento para realização deste trabalho. Ao professor Carlos Eduardo Thomaz por acreditar no meu trabalho.

"A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original."

Albert Einstein

Resumo

Esta proposta de dissertação de mestrado apresenta uma metodologia para a classificação de câncer de mama em imagem de ultrassom do ponto de vista da teoria de informação, incluindo informação mútua, entropia e entropia relativa, estendendo os estudos para a entropia Tsalliana. Os dados gerados a partir de uma base de dados normalizada compõem um espaço vetorial de informações híbridas, envolvendo textura e geometria espacial das lesões. Um dos objetivos principais é avaliar os resultados de Rodrigues et al.(2006c), bem como Thomaz et al.(2008) e Giraldi et al.(2008) com relação às características discriminantes mínimas.

Palavras-chave: Câncer de mama, Diagnóstico Assistido por Computador, Teoria da Informação.

Abstract

This proposed dissertation presents a methodology to the classification breast tumor ultrasound image from the view point of information theory, including mutual information, entropy and relative entropy, extending the study to the entropy Tsalliana. The data generated from a standardized database comprising a vector space hybrid information involving spatial geometry and texture of the goals of lesion. The key is to evaluate the results of Rodrigues et al.(2006c) and Thomas et al.(2008) and Giraldi et al.(2008) regarding the minimum discriminant features.

Keywords: Breast Tumor, Computer-Aided Diagnosis, Information Theory.

Lista de Figuras

2.1	Entropia de Shannon $H(x)$ pra diferentes valores de $p(x_1)$. A entropia é máxima para $p(x_1) = p(x_2) = 0,5$ e mínima para os casos extremos $[p(x_1) = 0; p(x_2) = 1$ e $p(x_1) = 1; p(x_2) = 0]$ (Adaptado de (Cabella, 2008))	23
2.2	Entropia de Tsallis $H_q(x)$ em função de $p(x_1)$ para diferentes valores de q . Para valores negativos de q , a entropia diverge para casos extremos $[p(x_1) = 0; p(x_2) = 1$ e $p(x_1) = 1; p(x_2) = 0]$ e é mínima para $[p(x_1) = p(x_2) = 0,5]$. (Adaptado de (Cabella, 2008))	24
2.3	Entropia de Tsallis $H_q(x)$ em função de $p(x_1)$ para diferentes valores de q . Para q positivo, a entropia é máxima para $[p(x_1) = p(x_2) = 0,5]$ e mínima pra $[p(x_1) = 0; p(x_2) = 1$ e $p(x_1) = 1; p(x_2) = 0]$. Quando $q = 0$ a entropia é constante independente do valor de $p(x_1)$, e para $q = 1$ obtemos a entropia de Shannon. (Adaptado de (Cabella, 2008))	25
2.4	Duas abordagens para segmentação. (A) Imagem original em níveis de cinza. (B) Imagem segmentada através de uma binarização. (C) Imagem segmentada por detecção de borda. Adaptado de (Esquef, 2002)	27
2.5	Cálculo automático do valor de q	32
2.6	Exemplo gráfico medido a partir de uma distância "d" e de um ângulo que pode ser de 0, 45, 90, 135 graus.	33
2.7	Imagem representada em três níveis de cinza como uma matriz de <i>pixels</i> 5x5 e a respectiva matriz de coocorrência para a direção 0° e a distância de um.	33
2.8	(a) Maior diâmetro de um tumor benigno e (b) Maior diâmetro de um tumor maligno.	35
2.9	(a) Área de um tumor benigno e (b) Área de um tumor maligno.	35
2.10	(a) Protuberância de um tumor benigno e (b) Protuberância de um tumor maligno.	36
2.11	Análise do volume de um ultrassom em 3D.(Adaptado de (Dar et al., 2006))	37
2.12	Exemplo de Árvore de Decisão. Adaptado de: (Sobral, 2003)	38
2.13	Hiperplano dividindo a base de treinamento.	40
2.14	Solução para um cenário de 2 dimensões linearmente separável. Os pontos pelos quais a linha pontilhada passa são chamados de vetores suporte. Adaptado de: (Oguri, 2006)	41
2.15	Representação de um neurônio artificial. Adaptado de:(Silva, 2005)	43
2.16	Exemplo de uma rede <i>perceptron</i> . Adaptado de: (Milare, 2003)	44
2.17	Interpretação geométrica para a rede <i>perceptron</i> . Adaptado de: (Milare, 2003)	45
2.18	Rede Neural composta por várias camadas de neurônios. Adaptado de: (Silva, 2005)	45
2.19	Curva de Características de Operação do Receptor (Curva ROC) (Adaptado de (Margotto, 2010))	48
2.20	Curva ROC de comparação (Adaptado de (Margotto, 2010))	49

3.1	Imagem digital de 640 x 480 <i>pixels</i> capturada de um scanner ultrassom (US). A região de interesse é um retângulo de 1.65 cm x 0.96 cm e 155 x 90 <i>pixels</i> . [Adaptada de [(Dar et al., 2002)]]	51
3.2	Exemplo de uma imagem digitalizada 3-D US. [Adaptada de [(Wei et al., 2005)]]	52
3.3	A) Demonstra os planos 2-D em uma imagem com dimensão 3-D. B) Reconstrução do Volume em 3-D. C) Os planos cartesianos extraídos. [Adaptado de [(Dar et al., 2006)]]	53
3.4	a) Imagem original de um tumor. b) Resultado após o detector de speckle. c) A escala de cinza dos <i>pixels</i> para classificação os speckles. [Adaptado de [(Ruey et al., 2002)]]	54
3.5	Gráfico de comparação dos resultados obtidos das pesquisa estudadas.	58
3.6	Média dos resultados obtidos nas pesquisas estudadas baseado no protocolo de validação cruzada.	58
3.7	Segmentação não-extensiva (Adaptado de (Rodrigues et al., 2006c))	62
3.8	(a) Imagem original de um tumor benigno. (b) Resultado com duas recursividades e $q = 0.5$.(Adaptado de (Rodrigues et al., 2006c))	62
3.9	Cadeia Morfológica.(Adaptado de (Rodrigues et al., 2006c))	62
3.10	Arquitetura Level-Sets.(Adaptado de (Rodrigues et al., 2006c))	63
3.11	MSV com um Kernel B-Spline. A linha branca é um polinômio de separação dos tumores malignos (círculos cinza) dos benignos (círculos pretos). Os círculos com as fronteiras brancas são aqueles dentro da região de confiança.	64
3.12	Curva ROC (Adaptado de (Rodrigues et al., 2006c))	65
4.1	(a) Localização do Centróide (b) Extração dos raios	67
4.2	Histograma de Raios (Azul - Tumores Benignos e Vermelho - Tumores Malignos)	68
4.3	(a) Segmentada com localização da fronteira. (b) Imagem Original. (c) Extração da Região de Interesse	68
4.4	Histograma da Região de Interesse (Azul - Tumores Benignos e Vermelho - Tumores Malignos)	69
4.5	(a) Localização dos pontos extremos de x e da fronteira. (b) Imagem Original. (c) Sombra acústica	69
4.6	Histograma da Sombra Acústica (Azul - Tumores Benignos e Vermelho - Tumores Malignos)	70

Lista de Tabelas

2.1	Lista de <i>kernel</i> mais utilizado (Adaptado de (Bisognin, 2007))	42
3.1	Desempenho para diferentes parâmetros de textura de (Dar et al., 2002).	52
3.2	Resultado do trabalho de (Ruey et al., 2002)	54
3.3	Lista dos trabalhos relacionados com informações do tipo de imagem e classificador utilizado.	59
4.1	Resultados da classificação utilizando entropia relativa	72
4.2	Classificação utilizando entropia relativa	73
4.3	Resultados da classificação utilizando entropia relativa usando características híbridadas	74
4.4	Classificação utilizando entropia relativa usando características híbridadas	74
4.5	Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para o histograma da área do tumor	75
4.6	Classificação utilizando distância de Kullback-Leiber não-extensiva para o histograma da área do tumor	76
4.7	Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para o histograma da sombra acústica	77
4.8	Classificação utilizando distância de Kullback-Leiber não-extensiva para o histograma da sombra acústica	78
4.9	Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para o histograma de raios.	79
4.10	Classificação utilizando distância de Kullback-Leiber não-extensiva para o histograma de raios	80
4.11	Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para características híbridadas.	81
4.12	Classificação utilizando distância de Kullback-Leiber não-extensiva para características híbridadas	82

Lista de Símbolos e Abreviaturas

$(A * B)$	— Entropia da distribuição composta, não é igual a soma das entropias
Λ	— Área sobre as funções $p(x)$
ΔQ	— Variação de calor
ΔS	— Variação de entropia
$2D$	— bidimensional, com duas dimensões.
$3D$	— tridimensional, com três dimensões.
A	— Área do nódulo
A_c	— Acuracidade
AD	— Árvore de Decisão
C	— Circularidade
D	— Distância entre duas distribuições de probabilidade, também conhecida como Divergência de Kullback-Leibler
DAC	— Diagnóstico assistido por computador.
D_q	— Generalização da Divergência de Kullback-Leibler (D) associada ao parâmetro q , também conhecida como Divergência de Kullback-Leibler não-extensiva
E_s	— Especificidade
f	— Vetor de características
FN	— Falso Negativo
FP	— Falso Positivo
h	— Histograma
H	— Entropia (Teoria da Informação)
H_{max}	— Entropia Máxima
H_q	— Entropia não-extensiva
k	— Constante de Boltzmann
K	— Kernel
MSV	— Máquina de Suporte Vetorial
N	— Total de elementos de uma distribuição
$NESRA$	— Non-Extensive Segmentation Recursive Algorithm.
$p(x)$	— Probabilidades associadas aos resultados possíveis de um acontecimento (x)
PDI	— Processamento digital da imagem.
$pixel$	— Picture element.
q	— Coeficiente de não extensividade.
RN	— Rede Neural
S	— Entropia (Termodinâmica)
Se	— Sensibilidade

t	—	Valor fixo de comparação
T	—	Temperatura constante
t_{opt}	—	Ponto de comparação ótimo
TFP	—	Taxa Falso Positivo
TVP	—	Taxa Verdadeiro Positivo
VN	—	Verdadeiro Negativo
VP	—	Verdadeiro Positivo
VPN	—	Valor Predito Negativo
VPP	—	Valor Predito Positivo
w	—	Número de microestados do sistema

Sumário

LISTA DE FIGURAS	9
LISTA DE TABELAS	11
LISTA DE SÍMBOLOS E ABREVIATURAS	12
1 Introdução	16
1.1 Imagem e Entropia	17
1.2 Apresentação dos Capítulos	18
2 Fundamentos Teóricos	19
2.1 Teoria da Informação	19
2.1.1 Entropia de Shannon	22
2.1.2 Entropia de Tsallis	24
2.1.3 Divergência de Kullback-Leiber	25
2.1.4 Divergência de Kullback-Leiber Não-Extensiva	26
2.2 Segmentação	26
2.3 Extração de Características	32
2.3.1 Informações de Textura	33
2.3.2 Informações Geométricas	34
2.4 Classificação	36
2.4.1 Árvore de Decisão	37
2.4.2 Máquina de Suporte Vetorial	39
2.4.3 Redes Neurais Artificiais	42
2.5 Testes Diagnósticos	46
2.6 Curva ROC	48
3 Trabalhos Relacionados	50
3.1 Extração de Características	50
3.1.1 Extração da Região de Interesse	51

	15
3.1.2	Descritores de Textura 52
3.1.3	Descritores de Forma 54
3.2	Classificação do Tumor 55
3.3	Discussões e Resultados 57
3.4	Trabalho base 60
4	Metodologia e experimentos 66
4.1	Banco de Imagens 66
4.2	Proposta para a Extração das Características 66
4.2.1	Histograma de Raios 67
4.2.2	Histograma da Região de Interesse 67
4.2.3	Histograma da Sombra Acústica 68
4.3	Classificação das Imagens 69
4.4	Metodologia de comparação de vetores 70
4.4.1	Classificação por distância de Kullback-Leiber 72
4.4.2	Classificação por distância de Kullback-Leiber usando características híbrid das 73
4.4.3	Classificação por distância de Kullback-Leiber não-extensiva 74
4.4.4	Classificação por distância de Kullback-Leiber não-extensiva usando ca racterísticas híbridadas 76
5	Conclusões 83
	GLOSSÁRIO 90

Capítulo 1

Introdução

De acordo com a Sociedade Americana de Câncer, o câncer de mama é a segunda causa de morte entre as mulheres de todo mundo. Mesmo com o crescente número de casos, há três décadas, a taxa de mortalidade vem caindo consideravelmente. Essa queda dá-se principalmente devido ao diagnóstico precoce e à melhoria no tratamento.

Entre os fatores que mais têm contribuído para a melhoria do diagnóstico está o maior acesso a exames e à informação, e o avanço da tecnologia disponível, das quais fazem parte os equipamentos de ultrassom. Estes podem ser encontrados em versões 2D ou 3D, sendo relativamente baratos e fáceis de manipular. Os scanners 2D são mais baratos que os 3D; sendo assim, mais utilizados (Rodrigues et al., 2006c).

São diversos os fatores que são levados em conta antes de um diagnóstico final, que muitas vezes é dado por uma junta médica. Alguns desses fatores, como histórico familiar, hábitos alimentares, tabagismo e fatores ambientais, não dependem de informações de exames que usam imagens digitais como elemento básico; outros, ao contrário, são tomados a partir de uma cautelosa avaliação de aspectos geométricos da lesão, tonalidades de cores de textura, variação em tamanho e forma durante um período de tempo. A necessidade de realizar exames em imagens tem diversas vantagens. Entre elas, o fato de não ser um exame invasivo é uma das maiores, uma vez que pode evitar biopsias onde a sensibilidade da camada de tecido que envolve a lesão é alta com risco de agravamento da doença em caso de perfuração.

Considerando esses fatores clínicos, mais o vasto número de exames com scanners 2D, além da necessidade de treinamento de estudantes e técnicos em exames de ultrassonografia para diagnóstico de lesões mamárias, o desenvolvimento de sistemas computacionais para auxílio ao diagnóstico automático é uma ajuda cada vez mais necessária. Sendo assim, a produção de pesquisas na área médica, que buscam o diagnóstico automático, ganham cada vez mais importância. Convencionou-se chamar essa área de Diagnóstico Assistido por computador (DAC).

1.1 Imagem e Entropia

Em DAC são várias as metodologias utilizadas para o diagnóstico. Normalmente usam informações de textura, cor ou geométricas. Uma vez extraídas tais informações, um espaço vetorial é montado através de processos de quantização, o que leva ao uso de modelos de classificação bem estabelecidos.

A classificação automática de bases de dados de ultrassom distinguindo entre tumores malignos e benignos é feita geralmente sobre base de imagens normalizadas em resolução e processo de captura. Apesar disso, por diversos motivos, os resultados ainda não são satisfatórios.

Entre as informações de textura e cor, as mais utilizadas são: contraste, covariância, dissimilaridade, heterogeneidade, homogeneidade e sombra acústica; e entre as informações geométricas, as mais usadas são: área, circularidade e protuberância.

O estudo da correlação entre essas características ainda é um desafio e nada foi provado ainda sobre o poder discriminante de cada uma delas, nem de maneira individual, nem de maneira conjunta. A medida que novas características vão sendo apresentadas, melhor estudadas ou quantificadas de maneira mais precisa, melhoram os métodos aplicados aos sistemas de DAC.

Apesar da grande necessidade do estudo conjunto de informações híbridas, poucos trabalhos abordam o tema diretamente. A maioria como Dar et al. (1999), Dar et al. (2002), Ruey et al.(2002) e etc, focam nas características de textura, por serem mais fáceis de manipular, e alguns

trabalhos como Wei et al.(2005), Rodrigues et al.(2006), Giraldi et al.(2008) e etc, usam também informações geométricas.

A suspeita de correlação entre essas características para a classificação, tanto geométricas quanto de textura, sugere o uso de teoria de informação mútua para o estudo dessas relações.

Teoria da Informação, apesar de ser uma área já bem estabelecida matematicamente, ainda não foi usada fortemente para avaliar os sistemas DAC, e poucos trabalhos têm sido encontrados na literatura, a maioria preocupando-se com o registro de imagem.

Por outro lado, desde que a entropia Tsalliana ganhou importância no mundo científico, ainda não existem trabalhos que utilizaram a teoria da informação e Entropia Tsalliana, mesmo esta última sendo naturalmente prevista como continuação da primeira. A maioria dos trabalhos que usam entropia Tsalliana para processamento de imagens são aplicados como método de segmentação e não como medida de classificação de tumores.

A proposta deste trabalho é aplicar em um modelo híbrido de características de câncer de mama e os conceitos de entropia Tsalliana e distância de Kullback-Leiber para classificação de imagens de ultrassom. O ponto de partida para o desenvolvimento desse trabalho propõe uma primeira abordagem estatística da imagem e em seguida utilizará entropia Tsalliana.

1.2 Apresentação dos Capítulos

O segundo capítulo da tese introduz os fundamentos teóricos que serão necessários para o entendimento do trabalho. O capítulo terceiro apresenta um resumo dos trabalhos relacionados e uma comparação entre eles. O quarto capítulo apresenta a metodologia desta tese e os resultados. Por fim no quinto capítulo temos as conclusões sobre a contribuição deste trabalho.

Capítulo 2

Fundamentos Teóricos

Este capítulo visa introduzir os conceitos e termos de processamento digital de imagens que serão necessários para o entendimento do trabalho e à evolução dos estudos da entropia, ferramenta base desta tese.

2.1 Teoria da Informação

A Primeira Lei da Termodinâmica ou Lei da Conservação da Energia foi proposta no século XIX, uma vez que esse princípio pôs uma limitação importante aos processos que se podem produzir na natureza, sendo apenas possíveis aqueles em que a energia permanece inalterável.

Sadi Carnot e Lord Kelvin desenvolveram uma construção teórica sobre engenho a vapor, que utilizava uma máquina a vapor para estudar a Termodinâmica onde Carnot concluiu que, quando se fornece calor a uma máquina a vapor, somente uma parte dessa energia se transforma no movimento de rotação necessário ao motor, o resto é transformado em calor; isto é, somente uma parte da energia da máquina pode ser transformada em trabalho mecânico. Assim, Carnot concluiu que somente com a variação da temperatura poderíamos gerar trabalho.

Em 1865, Rudolph Clausius dedicou uma atenção ao estudo sobre o engenho a vapor para a distinção entre a fração de calor que se pode transformar em energia mecânica e a que se deve desprezar como calor perdido. Assim, descreveu a primeira por "energia livre" e aquela perda de

energia (que não pode ser transformada em trabalho útil) por meio de um novo termo "entropia", que tem origem numa palavra grega *entrepion* que significa degradação, que é definido como o quociente entre a quantidade de calor transferido e a temperatura absoluta a que o processo é realizado.

Clausius, com esse trabalho, a descoberta de Carnot e a Primeira Lei da Termodinâmica escreveu a Segunda Lei da Termodinâmica que afirma que, na natureza, só podem ocorrer aqueles processos em que a entropia aumenta ou, quando muito, permanece constante. A entropia não era medida em valor absoluto. Apenas se media a variação de entropia de um sistema quando este passava de um estado para outro. Considerando uma transformação reversível, temos a seguinte relação:

$$\Delta S = \frac{\Delta Q}{T} \quad (2.1)$$

onde, ΔS representa a variação de entropia, ΔQ representa a variação de calor e temperatura constante (T).

Ainda no século XIX, prevalecia a idéia de que o calor era uma substância que fluía de um objeto para outro, o "calórico". Uma minoria de físicos sustentava a visão de que o calor era a energia resultante do movimento aleatório dos átomos. Quanto maior a temperatura de uma matéria, mais vigoroso é o movimento das moléculas que a formam.

Posteriormente, James Maxwell tratou os gases como uma coleção de pequenas partículas movimentando-se a diferentes velocidades, colidindo umas com as outras. Assim, ele determina as propriedades de um gás. Apesar de não ter sido o autor da Teoria Cinética dos Gases, foi o primeiro a aplicar métodos estatísticos e probabilísticos para descrever as propriedades das moléculas dos gases, sendo considerado um dos fundadores da Teoria Cinética dos Gases.

Em 1872, Ludwig Boltzmann, utilizando os métodos estatísticos de Maxwell, usou a probabilidade para descrever as propriedades da matéria a partir das propriedades dos átomos. Boltzmann transpôs a entropia para o mundo reversível da Física Newtoniana, onde cada uma das partículas que formam a matéria pode retroceder no seu estado e voltar ao lugar de onde começou. Dessa

forma, a totalidade do sistema também deve voltar ao estado inicial, o que contradiz a irreversibilidade de um sistema termodinâmico com máxima entropia permanecer nesse estado. Para resolver esse problema, Boltzmann interpretou a entropia como uma propriedade estatística. Inicialmente, não há nada que impeça todas as partículas de reverterem os seus percursos e irem de um estado de desordem para um estado de ordem, porém, o enorme número de partículas tornaria o evento raro, sendo apenas provável. Boltzmann ao introduzir a Teoria das Probabilidades na Segunda Lei da Termodinâmica, apresentou pela primeira vez uma lei de natureza probabilística.

Em 1877, no artigo "*On the relation between the second law of mechanical theory of heat and probability calculus with respect to the theorems on thermal equilibrium*" Boltzmann formula uma expressão de entropia baseado num sistema fechado. A entropia (S) seria proporcional ao volume do espaço de fase ocupado pelo macroestado do sistema (Boltzmann, 1877), ou seja,

$$S = k \ln \omega \quad (2.2)$$

onde k é a chamada constante de Boltzmann e ω é o número de microestados do sistema. Na Equação (2.2) quanto maior for o número de microestados possíveis, maior é a desordem e maior é a entropia. Em 1884, Boltzmann afirmou que "a entropia está relacionada com a diminuição de conhecimento"(Entropy, 2008), cruzando a ponte entre informação e entropia.

Willard Gibbs, observando o trabalho de Maxwell e de Boltzmann, desenvolveu uma expressão mais completa para descrever a entropia dada pela Equação (2.3)

$$S = -k \sum_{i=1}^{\omega} p_i * \ln(p_i) \quad (2.3)$$

onde p é a probabilidade do sistema estar no microestado i e ω o número total de microestados. Assim, temos que a máxima entropia corresponde à máxima desordem possível.

A entropia de Boltzmann-Gibbs, também conhecida por entropia clássica, tem como propriedade a regra da adição mostrada na Equação 2.4 , de forma que, para duas distribuições $P_A = [p_1, \dots, p_n]$ e $P_B = [q_1, \dots, q_m]$ a entropia da distribuição composta $(P_A * P_B) = \{p_i q_j\}_{i,j}$ onde $1 \leq i \leq n$ e $1 \leq j \leq m$ é igual a soma das entropias de P_A e P_B (Tavares, 2003).

$$S(P_A * P_B) = S(P_A) + S(P_B) \quad (2.4)$$

2.1.1 Entropia de Shannon

Claude Shannon(1916-2001), ao procurar um sistema telefônico que transferisse o máximo de quantidade de informação e uma forma de retificar distorções nas mensagens, associando esse desafio à entropia, desenvolveu uma teoria matemática para a comunicação.

No seu trabalho, uma mensagem que parte de uma fonte, é codificada e emitida por um transmissor, passa por um canal de comunicação, sofre perturbações, designadas de ruídos, e chega ao receptor passando por um sistema de decodificação. Para medir a quantidade de informação, Shannon criou o conceito de entropia partindo do conceito de que, se a possibilidade de uma mensagem ocorrer for pequena então ela contém muita informação, mas, caso contrário, se a mensagem for previsível, então conterá pouca informação (Shannon, 1948).

Assim, Shannon apresentou em 1948 no seu trabalho "*A Mathematical Theory of Communication*" uma expressão para a entropia que prova ser única,

$$H(x) = -k \sum p(x) * \log(p(x)) \quad (2.5)$$

sendo $p(x)$ as probabilidades associadas aos resultados possíveis de um acontecimento e k uma constante positiva. A escolha da base do logaritmo corresponde à escolha da unidade de medida da informação (Shannon, 1948).

A base da função logaritmo determina a unidade de medida da informação. A rigor, ela pode ser qualquer número maior que 1, sendo usual a utilização da base 2 para sistemas digitais de informação. Ao utilizar essa base, a unidade resultante é a unidade binária(*bit*) (Sturzbecher, 2006).

Para $k = 1$, esta quantidade mede a informação em *bits* contida em uma mensagem. O seu valor é a quantidade (comprimento) mínimo de *bits*, necessária para transmitir informações. Quanto

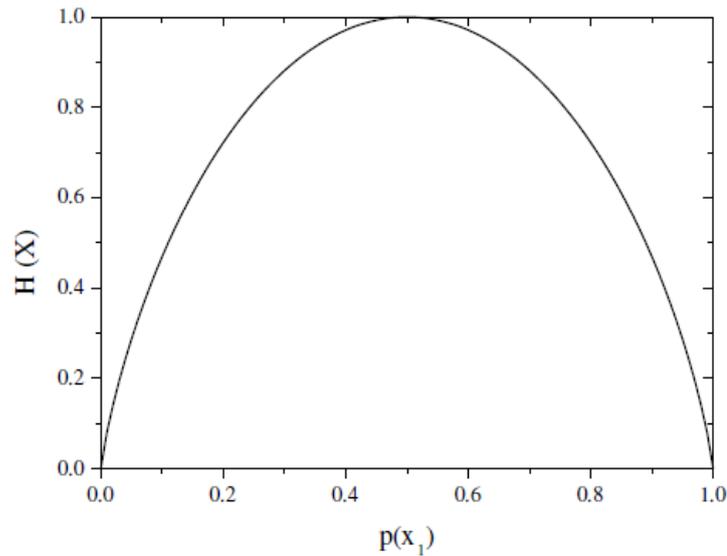


Figura 2.1: Entropia de Shannon $H(x)$ pra diferentes valores de $p(x_1)$. A entropia é máxima para $p(x_1) = p(x_2) = 0,5$ e mínima para os casos extremos [$p(x_1) = 0; p(x_2) = 1$ e $p(x_1) = 1; p(x_2) = 0$] (Adaptado de (Cabella, 2008))

mais previsível é a mensagem, menor o comprimento necessário para transmiti-la e vice-versa. (Cabella, 2008)

Considere como exemplo o lançamento de uma moeda e a observação do resultado (cara ou coroa). A entropia da jogada é maximizada se cara e coroa têm iguais probabilidades de ocorrer (50%). Esta é a situação de máxima incerteza, uma vez que é mais difícil prever o resultado da jogada. No entanto, se sabemos que a moeda não têm iguais probabilidades, então há menos incerteza. Todo o tempo, um dos lados tem mais probabilidade de ocorrer. A incerteza e, consequentemente a entropia são reduzidas. Nesse caso extremo teríamos uma moeda totalmente viciada onde todas as jogadas revelassem apenas uma face. Não havendo incerteza. A entropia é zero: cada jogada da moeda não fornece nenhuma informação.

Portanto, para estados equiprováveis a entropia é máxima, enquanto que para situações extremas (probabilidade máxima para um estado e mínima para os outros), a entropia é mínima. Considere a variável aleatória (X) que pode assumir dois estados, x_1 e x_2 , com probabilidades $p(x_1)$ e $p(x_2) = [1 - p(x_1)]$ o valor da entropia $H(X)$ em função de $p(x_1)$, está ilustrado no gráfico da figura 2.1 (Cabella, 2008).

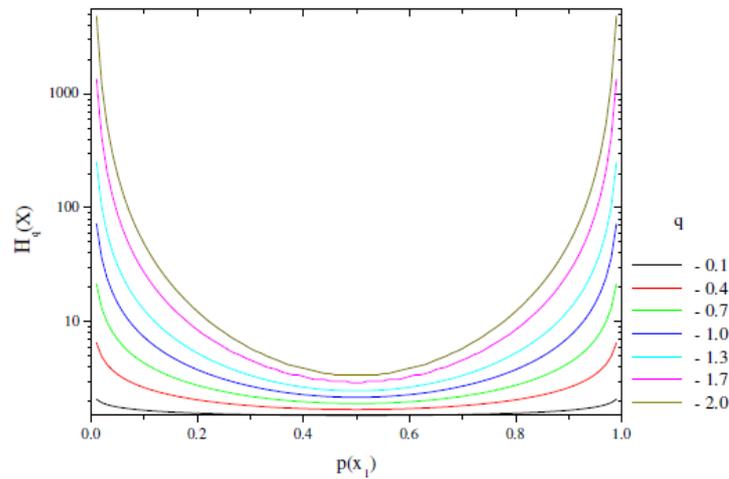


Figura 2.2: Entropia de Tsallis $H_q(x)$ em função de $p(x_1)$ para diferentes valores de q . Para valores negativos de q , a entropia diverge para casos extremos [$p(x_1) = 0; p(x_2) = 1$ e $p(x_1) = 1; p(x_2) = 0$] e é mínima para [$p(x_1) = p(x_2) = 0,5$]. (Adaptado de (Cabella, 2008))

2.1.2 Entropia de Tsallis

Sabe-se, no entanto, há mais de 100 anos que a entropia tradicional de Boltzman-Gibbs-Shannon não descreve corretamente muitos sistemas físicos. Por esse motivo, Constantino Tsallis propôs uma nova fórmula para a entropia. Conhecida como entropia de Tsallis não-extensiva, ou q -entropia, devido à inclusão do parâmetro q , chamado de coeficiente de não extensividade.

A entropia clássica de Shannon considera, implicitamente, as partículas de um sistema estatisticamente independentes, já a entropia de Tsallis considera essas partículas interdependentes entre si. Estes conceitos podem ser aplicados em imagens digitais considerando a correlação entre os diferentes pixels e o padrão fractal, encontrado em várias imagens (Diniz, 2008).

A fórmula para a entropia de Tsallis é dada por:

$$H_q(x) = \frac{1 - \sum p^q(x)}{1 - q} \quad (2.6)$$

para uma variável aleatória x com distribuição de probabilidade $p(x)$.

O comportamento da entropia de Tsallis, como função do parâmetro q , é representado nos gráficos das figuras 2.2 e 2.3.

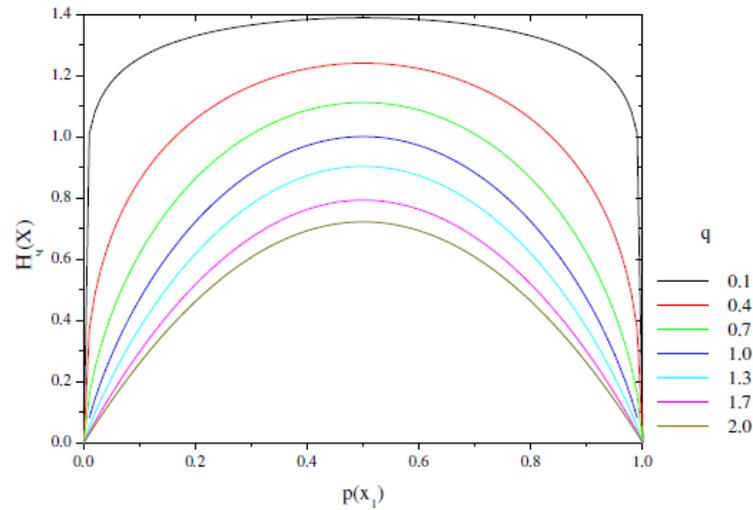


Figura 2.3: Entropia de Tsallis $H_q(x)$ em função de $p(x_1)$ para diferentes valores de q . Para q positivo, a entropia é máxima para $[p(x_1) = p(x_2) = 0,5]$ e mínima pra $[p(x_1) = 0; p(x_2) = 1$ e $p(x_1) = 1; p(x_2) = 0]$. Quando $q = 0$ a entropia é constante independente do valor de $p(x_1)$, e para $q = 1$ obtemos a entropia de Shannon. (Adaptado de (Cabella, 2008))

A Equação 2.7 mostra que a entropia de Tsallis não é aditiva, uma vez que, para duas distribuições aleatórias independentes $A = [p_1, \dots, p_n]$ e $B = [q_1, \dots, q_m]$, a entropia da distribuição composta ($A * B$) não é igual a soma das entropias de A e B , diferente da entropia clássica. Essa propriedade é denominada pseudo-aditividade (Shannon, 1948), (Tsallis, 2001), (Albuquerque e Esquef, 2004) e (Rodrigues e Giraldi, 2009a).

$$H_q(A * B) = H_q(A) + H_q(B) + (1 - q) * H_q(A) * H_q(B) \quad (2.7)$$

2.1.3 Divergência de Kullback-Leiber

Em teoria da informação, entropia relativa é a medida de "distância" entre distribuições de probabilidade p_1 e p_2 de duas variáveis discretas X_1 e X_2 , respectivamente. Esta medida de informação é conhecida como divergência de Kullback-Leibler (D) (Cover, 1991) e é definida por:

$$D(p_1||p_2) = \sum_x p_1(x) * \log \frac{p_1(x)}{p_2(x)} \quad (2.8)$$

A entropia relativa não é simétrica, isto é, $D(p_1||p_2) \neq D(p_2||p_1)$ e seu valor é sempre positivo. É zero se e somente se $p_1 = p_2$.

2.1.4 Divergência de Kullback-Leiber Não-Extensiva

À semelhança da generalização da entropia de Shannon com a Divergência de Kullback-Leibler também pode ser generalizada e associada a um parâmetro q . Em 1988, Borland e colaboradores definiram essa generalização, designada por q -informação de Kullback-Leibler, através do ganho de informação associada à ocorrência de um evento não extensiva de cada evento (Borland et al., 1998).

As técnicas não-extensivas permitem um ajuste da entropia através do parâmetro q . Esta modificação nos permite avaliar o processo de classificação em função de q .

$$D_q(p_1||p_2) = \sum_x p_1(x) * \log_q \frac{p_1(x)}{p_2(x)} \quad (2.9)$$

2.2 Segmentação

Segmentar uma imagem significa, de modo geral, separar a imagem como um todo nas partes que a constituem e que se diferenciam entre si (Esquef, 2002).

Ainda não foi descoberta uma forma definitiva de segmentação capaz de separar com exatidão todos os objetos em uma imagem (Jain, 2009), pois é um processo empírico e adaptativo; isto é, depende da imagem e da informação que desejamos extrair. Existem várias técnicas de segmentação e diversos estudos para desenvolvimento e melhoria destas técnicas.

De um modo geral, as técnicas de segmentação utilizam duas abordagens principais: a similaridade entre os *pixels* e a descontinuidade entre eles (Esquef, 2002). A técnica denominada binarização, baseada em similaridade, é simples e eficiente do ponto de vista computacional, sendo muito utilizada em sistemas de visão computacional. Esta técnica baseia-se na amplitude do nível de cinza para caracterizar os objetos constituintes estabelecendo um limiar de separação entre os valores que serão considerados como sendo brancos e os valores que serão considerados como

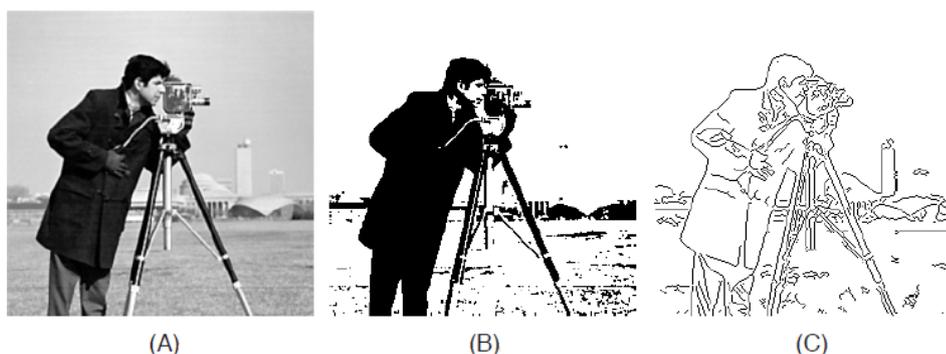


Figura 2.4: Duas abordagens para segmentação. (A) Imagem original em níveis de cinza. (B) Imagem segmentada através de uma binarização. (C) Imagem segmentada por detecção de borda. Adaptado de (Esquef, 2002)

sendo pretos. O resultado é uma imagem binária isto é, em preto e branco.

As técnicas baseadas em descontinuidade entre os *pixels* procuram determinar variações abruptas do nível de luminância entre os *pixels* vizinhos. Estas variações, em geral, permitem detectar o grupo de pixels que delimitam os contornos ou bordas dos objetos na imagem. O histograma é utilizado com frequência como sendo uma distribuição estatística dos *pixels* ("luminância") da imagem (Esquef, 2002).

A técnica denominada detecção de bordas utiliza as variações entre os *pixels* vizinhos para detectar os contornos ou bordas dos objetos. A Figura 2.4 apresenta dois exemplos, sendo (B) utilizando binarização e (C) detecção de bordas.

Existem diversos algoritmos de segmentação, cada um com sua peculiaridade. Dentre os mais conhecidos estão o *k*-means (Jain, 2009), a segmentação por entropia (Albuquerque e Esquef, 2004) e o limiar adaptativo (Sezgin e Sankur, 2004).

Os algoritmos de segmentação que usam o limiar são os mais utilizados devido à sua simplicidade, e também por ser o algoritmo mais veloz, com complexidade $O(n)$, pois cada pixel é acessado somente uma vez. Este algoritmo classifica os *pixels* com maior intensidade com um valor máximo e com menor intensidade com um valor mínimo usando uma função degrau com um valor fixo de comparação t . A definição de t pode ser feita manualmente, de forma adaptativa, ou ainda utilizando entropia não-extensiva (Albuquerque e Esquef, 2004).

Para definir t utilizando entropia não-extensiva Albuquerque et al.(2004) separou o histograma (h) em duas partes h_A e h_B variando o ponto de separação do início ao fim do histograma (h). São calculadas as entropias das duas partes do histograma H_{h_A} e H_{h_B} para cada ponto de separação. Podemos então descobrir o ponto de comparação t_{opt} ótimo, no qual a entropia conjunta é máxima utilizando a Equação 2.7 e observando o Algoritmo 1.

Algoritmo 1 Segmentação com Entropia Não-extensiva

```

 $h$  = histograma da imagem
 $L$  = posições do histograma
 $H_{Max} = -\infty$ 
 $t_{opt}=1$ 
for ( $t=1$  até  $L$ ) do
   $A = h[1 \dots t]$ 
   $B = h[t + 1 \dots L]$ 
  normaliza( $A$ )
  normaliza( $B$ )
   $H_A$  = calcular a q-Entropia de  $H_A$  segundo Equação (2.6)
   $H_B$  = calcular a q-Entropia de  $H_B$  segundo Equação (2.6)
   $H_T$  = calcular a entropia conjunta de ( $A * B$ ) segundo Equação (2.7)
  if  $H_T > H_{Max}$  then
     $H_{Max} = H_T$ 
     $t_{opt} = t$ 
  end if
end for

```

Rodrigues et al. (2009) propuseram uma extensão para a equação de Tsallis adicionando mais um parâmetro entrópico na Equação (2.6). Pode-se dizer que para uma distribuição contínua aleatória X e sua função de distribuição de probabilidades $p(X)$, a entropia de Tsallis é definida por (Rodrigues e Giraldi, 2009b):

$$H_q = \frac{1 - \int_{-\infty}^{+\infty} [p(x)]^q dx}{q - 1} \quad (2.10)$$

Podemos definir então as entropias do *foreground* ($H_q(A)$) e do *background* ($H_q(B)$) da seguinte forma:

$$H_q(A) = \frac{1 - \int_{-\infty}^t [p_A(x)]^q dx}{q - 1} \quad (2.11)$$

$$H_q(B) = \frac{1 - \int_{-\infty}^t [p_B(x)]^q dx}{q - 1} \quad (2.12)$$

Considerando as seguintes condições podemos validar as Equações 2.11 e 2.12:

$$\int_{-\infty}^t p_A(x) dx = 1 \quad (2.13)$$

$$\int_t^{+\infty} p_B(x) dx = 1 \quad (2.14)$$

$$p_B(x) \geq 0, \quad p_A(x) \geq 0, \quad \forall x \in \mathfrak{R} \quad (2.15)$$

Substituindo as Equações (2.11) e (2.12) na Equação (2.7), obtém-se a medida de informação entre o *foreground* e o *background*. Assim, é possível desenvolver uma formulação contínua para o problema de otimização definido na Equação (2.7). Uma condição necessária para maximizar $H_q(A * B)$ é $dH_q(A * B)/dt = 0$, que pode ser decomposto utilizando álgebra simples (Erdmann, 2009):

$$(-[p_A(t)]^q + [p_B(t)]^q) + (1 - q)(-[p_A(t)]^q H_q(B) + [p_B(t)]^q H_q(A)) = 0 \quad (2.16)$$

Podemos reecrevar a Equação (2.16) da seguinte maneira:

$$\left(-\left[\frac{p(t)}{\Lambda_A}\right]^q + \left[\frac{p(t)}{\Lambda_B}\right]^q\right) + (1 - q)\left(-\left[\frac{p(t)}{\Lambda_A}\right]^q H_q(B) + \left[\frac{p(t)}{\Lambda_B}\right]^q H_q(A)\right) = 0 \quad (2.17)$$

onde Λ_A e Λ_B são as áreas sobre as funções $p(x)$, antes e depois da solução $t = t_{opt}$, respectivamente.

Uma solução independente de q para esta expressão é $H_q(A) = H_q(B) = H$ e $\Lambda_A = \Lambda_B$. A partir das expressões (2.11) e (2.12), nota-se que:

$$\begin{aligned} H_q(A) = H_q(B) &\Leftrightarrow \int_{-\infty}^t \left[\frac{p(t)}{\Lambda_A}\right]^q dx = \int_t^{+\infty} \left[\frac{p(t)}{\Lambda_B}\right]^q dx \Leftrightarrow \\ &\int_{-\infty}^t \frac{[p(x)]^q}{\Lambda_A^q} dx = \int_t^{+\infty} \frac{[p(x)]^q}{\Lambda_B^q} dx, \\ \frac{\Lambda_A^q}{\Lambda_B^q} &= \frac{\int_{-\infty}^t [p(x)]^q dx}{\int_t^{+\infty} [p(x)]^q dx}. \end{aligned} \quad (2.18)$$

As Equações (2.17) e (2.18) podem ser utilizadas em uma solução independente de q para maximização de $H_q(A * B)$. Esta forma tem um custo computacional alto, além do fato de que o valor t pode não existir e geralmente a solução é um valor de t_{opt} que pode não ser o mais eficiente para segmentação. Sendo assim, Rodrigues e Giraldi propuseram um método para evitar tais problemas (Rodrigues e Giraldi, 2009b), que redefine as métricas $H_q(A * B)$ como:

$$H_{q_A, q_B}(A * B) = H_{q_A}(A) + H_{q_B}(B) + \left(1 - \left(\frac{q_A + q_B}{2}\right)\right) H_{q_A}(A)H_{q_B}(B) \quad (2.19)$$

uma generalização da Equação (2.16) pode ser dada pois, q_A e q_B são dois valores independentes do parâmetro entrópico:

$$\left(\frac{-[p_A(t)]^{q_A}}{q_A - 1} + \frac{[p_B(t)]^{q_B}}{q_B - 1}\right) + \left(1 - \left(\frac{q_A + q_B}{2}\right)\right) \quad (2.20)$$

$$\left(\frac{-[p_A(t)]^{q_A}}{q_A - 1} H_{q_B}(B) + \frac{[p_B(t)]^{q_B}}{q_B - 1} H_{q_A}(A)\right) = 0.$$

Observa-se que $H_{q_A}(A) = H_{q_B}(B) = H$ e $p_A(t) = p_B(t)$ não é uma solução direta se $q_A \neq q_B$. A solução para o problema de otimização não está limitada à $H_{q_A}(A) = H_{q_B}(B) = H$, o que pode introduzir mais um grau de liberdade no algoritmo de segmentação com q -entropia, se $H_q(A * B)$ for substituído por $H_{q_A, q_B}(A * B)$, deve-se colocar dois valores para o parâmetro entrópico q . Para atender a este problema, retorna-se à Equação (2.18) e se propõe a generalização abaixo (Erdmann, 2009):

$$\frac{\Lambda_A^{q_A}}{\Lambda_B^{q_B}} = \frac{\int_{-\infty}^t [p(x)]^{q_A} dx}{\int_t^{+\infty} [p(x)]^{q_B} dx}. \quad (2.21)$$

Para simplificar esta equação, é possível considerar algumas heurísticas com o objetivo de obter uma expressão envolvendo a razão q_A/q_B . Supondo que:

$$\frac{\int_{-\infty}^t [p(x)]^{q_A} dx}{\int_t^{+\infty} [p(x)]^{q_B} dx} = \alpha^{q_B}, \quad (2.22)$$

onde α é uma constante. Substituindo a Equação (2.22) na Expressão (2.21) e tomando-se o log

dos dois lados, tem-se:

$$\frac{q_B}{q_A} = \frac{\ln \Lambda_A}{\ln (\alpha \Lambda_B)} = \frac{\ln \int_{-\infty}^t [p(x)] dx}{\ln \left(\alpha \int_t^{+\infty} [p(x)] dx \right)} \quad (2.23)$$

A alteração proposta por Rodrigues et al. (2009) na entropia de Tsallis é dada por (Rodrigues e Giraldi, 2009b):

$$\begin{aligned} H_{q_A, q_B}(A * B)(t) &= \frac{1 - \sum_{i=1}^t (p_A)^{q_A}}{q_A - 1} + \frac{1 - \sum_{i=t+1}^k (p_B)^{q_B}}{q_B - 1} \\ &+ \left(1 - \frac{q_A + q_B}{2}\right) \cdot \frac{1 - \sum_{i=1}^t (p_A)^{q_A}}{q_A - 1} \cdot \frac{1 - \sum_{i=t+1}^k (p_B)^{q_B}}{q_B - 1} \end{aligned} \quad (2.24)$$

Para $q_A = q_B = q$, a Equação (2.24) volta a forma de Tsallis dada pela Equação (2.6). Iniciando o algoritmo com $q = q_A$, pode-se calcular o valor correspondente de q_B por meio da Equação (2.23), e assim, resolver o problema de maximização definido na Equação (2.7) com $H_q(A * B)$ substituído por $H_{q_A, q_B}(A * B)$. O algoritmo para segmentação proposto por Rodrigues et al. (2009) foi denominado *Non-Extensive Segmentation Recursive Algorithm* (NESRA) ou Segmentação com Entropia Não-Extensiva Estendida e é dado pelo Algoritmo 2 :

Algoritmo 2 Segmentação com Entropia Não-Extensiva Estendida

- Calcular $topt$ para q_A utilizando o Algoritmo (1);
 - Calcular q_B utilizando $topt$ e q_A por meio da Equação (2.23);
 - Segmentar novamente com o Algoritmo (1), utilizando $H_{q_A, q_B}(A * B)$;
-

Rodrigues et al.(2009), propuseram um método para cálculo automático do valor de q (Rodrigues e Giraldi, 2009b). Dado que a entropia máxima de uma distribuição X ocorre quando todos os estados de X , (x_1, x_2, \dots, x_n) possuem a mesma probabilidade, dessa forma a entropia máxima de uma distribuição X , H_{MAX} é dada pela Equação (2.25).

$$H_{MAX} = \frac{1}{q-1} [1 - n(p^q(x))] \quad (2.25)$$

onde: q é o parâmetro entrópico e n é o total de elementos da distribuição X .

Do ponto de vista da Teoria da Informação, quanto menor for a relação entre a entropia H_q produzida por um valor q e a entropia máxima H_{MAX} de um sistema, maior é a informação con-

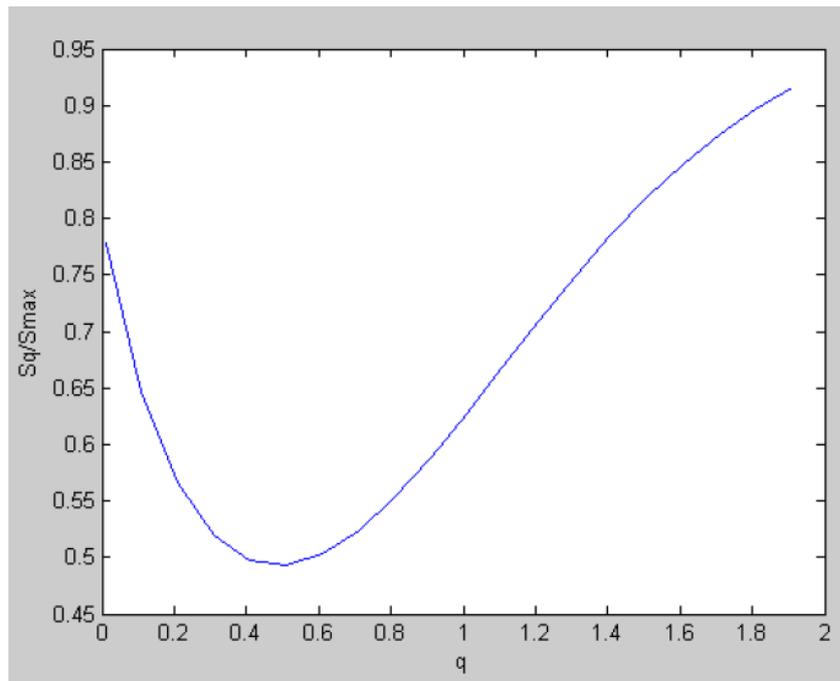


Figura 2.5: Adaptada de (Rodrigues e Giraldi, 2009b). Cálculo automático do valor de q .

tida no sistema. Este é um princípio conhecido da Teoria da Informação e podemos verificar que o valor ótimo de q pode ser calculado minimizando a função H_q/H_{MAX} (Rodrigues e Giraldi, 2009b). Dessa maneira, para cada imagem, é calculada a relação entre a entropia H_q e a entropia máxima H_{MAX} variando-se o valor de q na faixa de $[0.01, 0.02, \dots, 2.0]$, e verificando para qual valor esta relação é minimizada. Na Figura 2.5, pode observar-se como a relação entre H_q e H_{MAX} se comporta com a variação do valor de q . (Rodrigues e Giraldi, 2009b)

2.3 Extração de Características

Nesta etapa, extraímos as informações úteis da imagem analisada. Existem basicamente duas classes de medidas: i) atributos da imagem como um todo, como exemplo, número de objetos, área total de objetos, etc. e ii) atributos de região, que se referem aos objetos independentemente, como exemplo, área, perímetro, forma, etc (Esquef, 2002).

Nas próximas seções serão apresentadas as características mais utilizadas nos trabalhos relacionados.

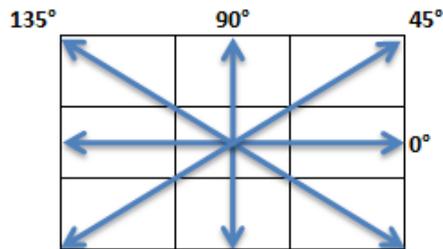


Figura 2.6: Exemplo gráfico medido a partir de uma distância "d" e de um ângulo que pode ser de 0, 45, 90, 135 graus.

Imagem					Matriz de coocorrência				
1	0	2	2	0					
1	1	2	0	1		j			
1	0	0	1	1		0	1	2	
1	2	0	1	0	i	0	2	6	4
2	1	2	2	1		1	6	4	5
						2	4	5	4

Figura 2.7: Imagem representada em três níveis de cinza como uma matriz de *pixels* 5x5 e a respectiva matriz de coocorrência para a direção 0° e a distância de um.

2.3.1 Informações de Textura

As três principais abordagens utilizadas em processamento de imagens por textura são: abordagem estrutural, abordagem espectral e abordagem estatística (Gonzalez e Woods, 1992).

A abordagem estrutural visualiza as texturas como sendo uma repetição de padrões, enquanto que a abordagem espectral baseia-se nas propriedades do espectro de Fourier avaliando os picos de alta frequência no espectro. A abordagem estatística analisa a imagem através de regras estatísticas que distribuem os níveis de cinza e é a mais utilizada nos trabalhos relacionados.

Os atributos relacionados à textura podem ser baseados na análise de um conjunto de matrizes de coocorrências de níveis de cinza para calcular a probabilidade de ocorrência combinando a direção e a distância entre pares de *pixels* com valores de intensidade semelhantes. A Figura 2.6 indica as direções possíveis para análise de um *pixel* considerando uma distância "d" do *pixel* analisado, onde "d" é igual a um. Na Figura 2.7 temos uma imagem representada em três níveis de

cinza: 0, 1 e 2 como uma matriz de *pixels* 5x5 e a respectiva matriz de coocorrência para a direção 0 graus e a distância de um *pixel*.

A principal desvantagem de se usar informações de textura em exames de ultrassom para diagnóstico de câncer de mama é a grande dependência dessas informações dos aparelhos transdutores. Esses equipamentos são baseados em reflexão sonora, são extremamente suscetíveis a ruídos e interferência de outras imagens próximas às lesões, sendo então a qualidade da imagem muito dependente do fabricante e tecnologia do transdutor usado. Informações de textura feitas com scanners de fabricantes diferentes podem ser estatisticamente muito diferentes.

2.3.2 Informações Geométricas

Informações geométricas são retiradas geralmente das bordas (polígono) do tumor e podem ser associadas ao tamanho, forma e conteúdo da lesão.

1. Diâmetro

A medição do diâmetro em uma lesão é um dos métodos utilizados pelos médicos para verificar a evolução do nódulo. Os sistemas computacionais de análise de imagens identificam as fatias da imagem que possuem maior diâmetro e o medem de uma extremidade a outra. A Figura 2.8 apresenta o maior diâmetro de duas imagens onde (a) é um tumor benigno e (b) um tumor maligno.

2. Área

A medida da área de um tumor é extraída analisando fatias das imagens extraídas no exame do ultrassom. Um sistema computacional de análise de imagens pode somar todos os *pixels* da imagem de ultrassom. Essa informação também é importante na avaliação da evolução do tumor. A Figura 2.9 mostra a área de um tumor benigno (a) e um tumor maligno (b).

3. Circularidade

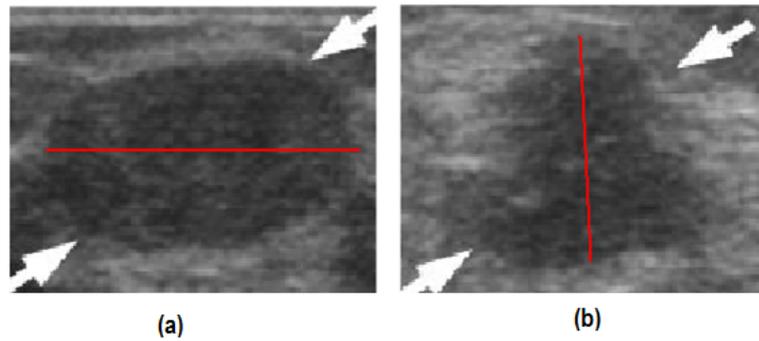


Figura 2.8: (a) Maior diâmetro de um tumor benigno e (b) Maior diâmetro de um tumor maligno.

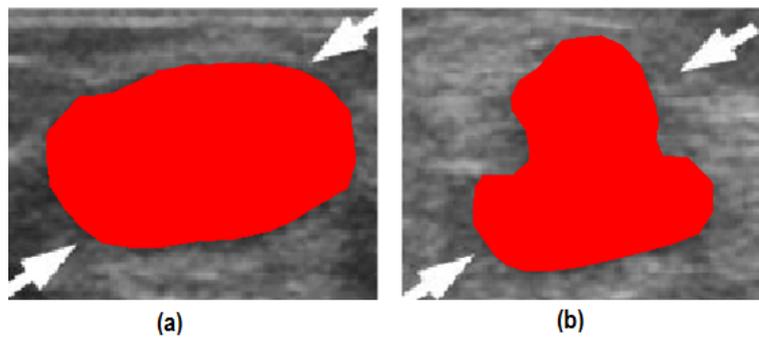


Figura 2.9: (a) Área de um tumor benigno e (b) Área de um tumor maligno.

A forma arredondada de uma lesão fornece indicações sobre sua malignidade ou benignidade. Desta forma a medida da característica denominada circularidade fornece a informação que pode sugerir um diagnóstico ao médico. A circularidade pode ser dada por:

$$C = \frac{P^2}{4\pi A} \quad (2.26)$$

onde P é o perímetro da lesão e A é a área do nódulo. Quanto mais próximo de 1 for o valor da circularidade, maior será a probabilidade de benignidade e, quanto mais distante, maior será a probabilidade de malignidade (Silva e Carvalho, 2002).

4. Protuberância

Os tumores podem não ser totalmente circulares, existem protuberâncias em sua superfície. Nas imagens de ultrassom 3D ou até mesmo nas sub-imagens em 2D podemos notar a irregularidade do tumor. Na Figura 2.10 apresentamos a protuberância de um tumor benigno e um tumor maligno. Podemos notar que o tumor benigno possui menos protuberâncias que o

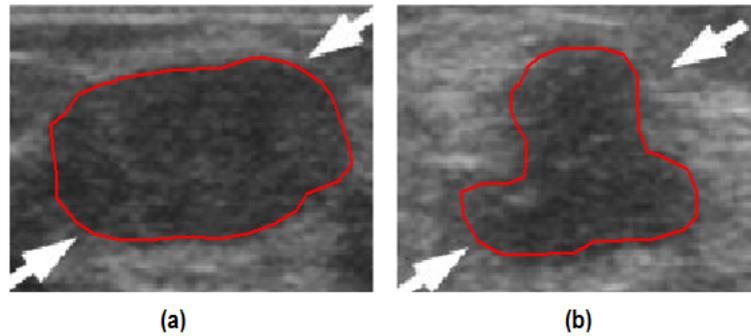


Figura 2.10: (a) Protuberância de um tumor benigno e (b) Protuberância de um tumor maligno.

tumor maligno, conseqüentemente para as lesões do exemplo, o tumor benigno é mais circular que o maligno.

5. Volume

O volume também é bastante utilizado para medir a evolução de um tumor. Entretanto não fornece uma visão geral e nem uma avaliação tão precisa, pois a lesão pode se desenvolver em qualquer direção. Por esta razão e principalmente com o desenvolvimento de recursos computacionais, a comunidade médica tem procurado utilizar a medida de volume somente como forma de avaliação da evolução da lesão (Reeves e Kostis, 2000) (Yankeleitz, 2000). A Figura 2.11 mostra a análise tridimensional do volume de interesse de um tumor. São necessários 3 *frames* (um no início da lesão, um no meio e outro no final) em forma retangular para extrair as subimagens do tumor.

2.4 Classificação

O objetivo da classificação é gerar uma forma automática de identificação dos objetos segmentados e das informações da imagem. Podemos dividir em duas etapas o processo de classificação: o aprendizado e o reconhecimento.

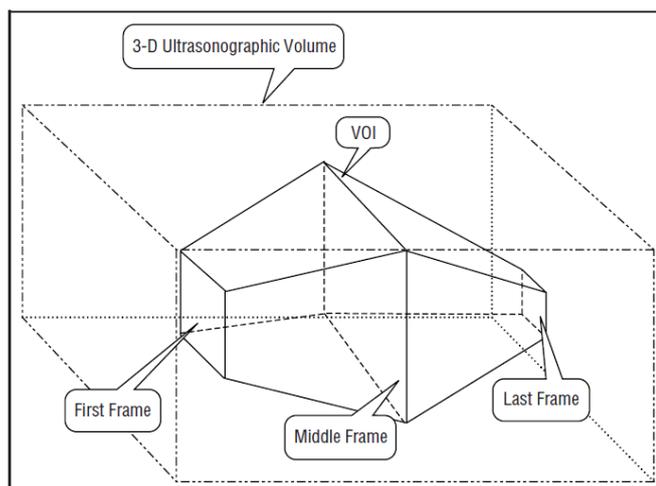


Figura 2.11: Análise do volume de um ultrassom em 3D.(Adaptado de (Dar et al., 2006))

2.4.1 Árvore de Decisão

O modelo de Árvore de Decisão (AD) é usado para examinar um dado em uma árvore com regras para se fazer uma predição. A AD é uma maneira de se representar um conjunto de regras. Ela contém um nó, filhos e folhas. Dependendo do algoritmo, um nó pode conter dois ou mais filhos. Cada filho conduzirá a outro nó de decisão ou ao fundo da árvore chamado de nó folha. As árvores de decisão são usadas geralmente para classificar a qual grupo pertence o dado.

Na Figura 2.12 temos uma representação gráfica de árvore de decisão que pode ser utilizada para classificação de imagens, onde o vetor que representa a imagem é composto das características (f) da mesma e representado por $\theta = [f_1, f_2, f_3, f_4, f_5]$. O valor (x) de cada característica é submetido às regras da árvore e classificado. Os círculos representam os nós internos (intermediários ou decisórios); os quadrados representam os nós folhas ou terminais, as linhas representam os ramos que interligam dois nós; e x_1 e x_2 representam as variáveis decisórias. Chama-se de variável decisória a variável de entrada que levará a uma nova divisão da árvore de decisão, em relação a um possível valor. (Sobral, 2003)

Para interpretar a árvore de decisão da Figura 2.12, analisamos a condição de satisfação (por exemplo: $x_1 \leq 0.7$). O ramo segue para o lado esquerdo se a condição for verdadeira e, caso contrário, segue para o lado direito. Os nós são numerados (por exemplo: f_1 nó 1, f_2 nó 2, f_3 nó

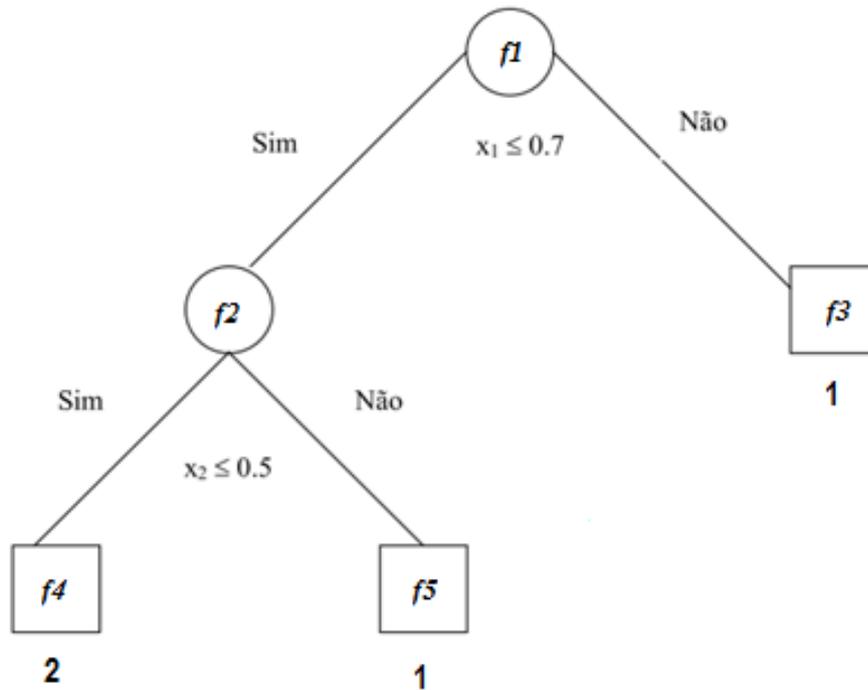


Figura 2.12: Exemplo de Árvore de Decisão. Adaptado de: (Sobral, 2003)

3, f_4 nó 4, f_5 nó 5). Essa árvore de decisão possui 3 níveis, onde temos no nível 0 o nó 1, no nível 1 os nós 2 e 3 e no nível 2 os nós 4 e 5.

O aprendizado de uma árvore de decisão é supervisionado, ou seja, o método aproxima funções alvo de valor discreto, na qual a função aprendida é representada por uma árvore de decisão. (Sobral, 2003)

Após o aprendizado com a árvore treinada são obtidas as regras para classificarmos em duas classes, 1 e 2. Para o nosso exemplo da Figura 2.12 temos a regra 1 para a classe 1, se $(x_1 \geq 0.7)$ ou se $(x_1 \leq 0.7$ e $x_2 \geq 0.5)$. Para a classe 2 temos a regra 2, se $(x_1 \leq 0.7$ e $x_2 \leq 0.5)$.

Nas árvores de decisão utilizadas para classificação, também conhecidas como árvores de classificação, cada nó terminal ou folha leva a uma classe que indica um determinado conjunto de dados. Nesse tipo de árvore podem existir dois ou mais nós terminais com a mesma classe.

Um algoritmo para indução de árvores de decisão trata-se de um exemplo de algoritmo de estrutura. Este algoritmo utiliza a estratégia "dividir-para-conquistar", ou seja, um problema complexo é decomposto em subproblemas mais simples (Monard e A., 2003). A idéia básica do Algoritmo 3 que induz uma árvore de decisão é:

1. Escolher um atributo;
2. Estender a árvore adicionando um ramo para cada valor do atributo;
3. passar os exemplos para as folhas;
4. para cada nó folha - se todos os exemplos são da mesma classe, associar esta classe ao nó folha, senão, repetir os itens (1), (2) e (3).

Porém, a árvore assim construída pode estar ajustada demais aos dados de treinamento. Uma árvore de decisão a está ajustada demais aos dados se existir uma árvore a' tal que a tem menor erro que a' no conjunto de treinamento, porém a' tem menor erro no conjunto de teste (Carvalho, 2005).

Para corrigir o fato de uma árvore estar ajustada demais, deve-se executar um procedimento de poda da árvore, a fim de reduzir o excesso de ajuste aos dados de treinamento.

A realização da poda pode ser realizada de duas maneiras, pré-poda, onde paramos o crescimento da árvore mais cedo e pós-poda, a árvore cresce completamente e depois podamos. A pós-poda é mais lenta, porém mais confiável que a pré-poda (Quinlan, 1987).

2.4.2 Máquina de Suporte Vetorial

Máquina de Suporte Vetorial (MSV) é uma ferramenta eficaz no reconhecimento de padrões e aprendizado de máquina. O objetivo é definir um hiperplano como mostra a Figura 2.13 que divida uma base de treinamento, tal que todos os pontos com a mesma característica estão de um mesmo lado, formando dois grupos divididos pelo hiperplano.

Dados N exemplos de treinamento $\{x_i, y_i\}$, $i = 1, 2, \dots, N$, onde $x_i \in R$ é uma representação vetorial de um documento e $y_i \in \{-1, 1\}$ sua classe correspondente, suponha que seja possível

Algoritmo 3 Procedimento para a construção da árvore de decisão (Mitchell, 1997)

```

ConjExemplos = o conjunto de treinamento
AtributoMeta = atributo a ser predito pela árvore
ListaAtributos = lista dos outros atributos a serem testados
START(ConjExemplos,AtributoMeta,ListaAtributos)
/* Selecionar o melhor atributo para o nó raiz da árvore, de acordo com a função avaliação. */
if (todos os exemplos em ConjExemplos são de uma única classe) then
    Retorna um único nó com valor classe
else
    if (ListaAtributos = NULL) then
        Retorna um único nó com o valor de AtributoMeta mais frequente em ConjExemplos
    else
        A = O atributo de ListaAtributos que melhor classifica ConjExemplos
        for (cada valor ( $v_i$ ) possível de A) do
            A =  $v_i$  /* Adiciona uma nova ramificação */
            Criar o subconjunto  $ConjExemplos_{v_i}$  contendo os exemplos de ConjExemplos que satisfazem o teste A =  $v_i$ 
            if ( $ConjExemplos_{v_i}$  = NULL) then
                Criar uma ramificação subordinada ao novo nó com o valor de AtributoMeta mais frequente.
            else
                START( $ConjExemplos_{v_i}$ ,AtributoMeta,ListaAtributos - A)
            end if
        end for
    end if
end if
Retornar Raiz
  
```

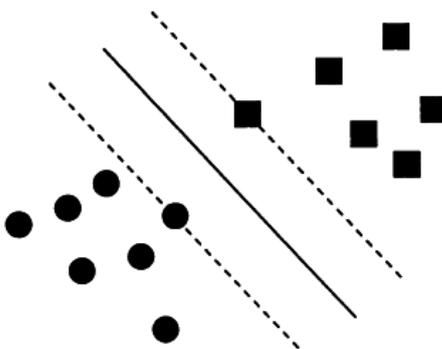


Figura 2.13: Hiperplano dividindo a base de treinamento.

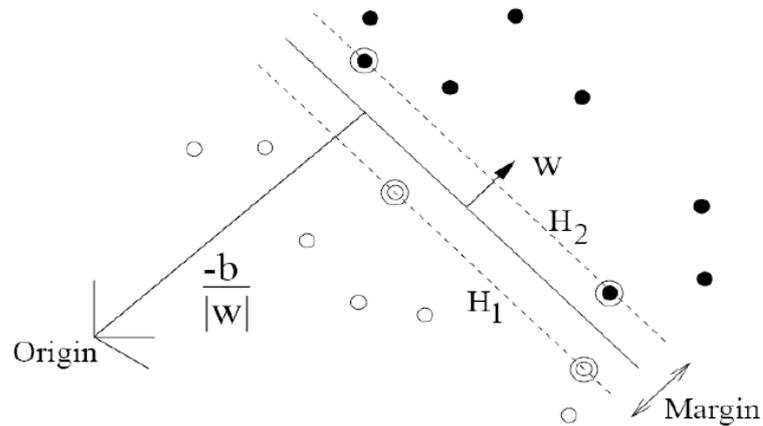


Figura 2.14: Solução para um cenário de 2 dimensões linearmente separável. Os pontos pelos quais a linha pontilhada passa são chamados de vetores suporte. Adaptado de: (Oguri, 2006)

traçar um hiperplano que separe os exemplos positivos dos negativos. Este hiperplano pode ser definido por $x \cdot w + b = 0$ onde x são pontos sobre o hiperplano, w é a normal do hiperplano, e $|b|/|w|$ é a distância perpendicular do hiperplano à origem. (Oguri, 2006)

Analisando a Figura 2.14, onde o conjunto de treinamento é linearmente separável, o algoritmo busca o hiperplano separador tal que a margem seja maximizada. Para formular este algoritmo, suponha, que após o treinamento, o conjunto esteja linearmente separado, satisfazendo as seguintes equações:

$$x_i \cdot w + b \geq +1 \quad \text{para } y_i = +1 \quad (2.27)$$

$$x_i \cdot w + b \leq -1 \quad \text{para } y_i = -1 \quad (2.28)$$

Podemos combinar estas desigualdades e obter:

$$y_i(x_i \cdot w + b) - 1 \geq 0, \quad i = 1, 2, \dots, D \quad (2.29)$$

Considere agora os exemplos para os quais a igualdade na Equação (2.27) é válida. Tais exem-

plos são pontos sobre o hiperplano $x \cdot w + b = 1$ com normal w e distância perpendicular à origem de $\frac{|1-b|}{\|w\|}$. Simetricamente, os exemplos para os quais a igualdade na Equação (2.28) é válida são pontos sobre o hiperplano $x \cdot w + b = -1$ com normal w e distância perpendicular à origem $\frac{|-1-b|}{\|w\|}$. O objetivo do algoritmo pode se formulado como um problema de otimização quadrático sujeito a restrições definidas na Equação (2.29). (Oguri, 2006)

As Máquinas de Suporte Vetorial podem aprender a separar dados do tipo linearmente separáveis quanto inseparáveis através de funções *Kernel*. Um *Kernel* (K) é uma função que recebe dois pontos x e y do espaço de entradas e computa o produto escalar desses dados no espaço de características (Herbrich, 2001).

Alguns dos *Kernels* mais utilizados na prática são Polinômio de Grau d , Gaussiano, Perceptron Multi Camadas e Splines listados na Tabela 2.1.

Tabela 2.1: Lista de *kernel* mais utilizado (Adaptado de (Bisognin, 2007))

Tipo de Classificador	Função <i>Kernel</i>
Polinômio de Grau d	$K(x, y) = ((x * y) + 1)^d$
Gaussiano	$K(x, y) = \exp(-\frac{\ (x-y)\ ^2}{2\sigma^2})$
Perceptron Multi Camadas	$\tanh(x * y) - \theta$
Splines	$1 + (x * y) + \frac{1}{2}(x * y)\min(x * y) - \frac{1}{6}(x * y)^3$

O *Kernel* possui parâmetros que influenciam no poder de generalização da superfície, como por exemplo o grau d do *Kernel* (Polinômio de grau d).

2.4.3 Redes Neurais Artificiais

Redes Neurais Artificiais (RN) são modelos matemáticos que se assemelham às estruturas neuronais biológicas e que têm capacidade computacional adquirida por meio de aprendizado e generalização (Braga, 2003). Essa estrutura tenta imitar um cérebro humano com ligações entre os neurônios e entrada e saída de sinais, como demonstra a Figura 2.15. O neurônio artificial é composto por um conjunto de conexões de entrada, sendo cada entrada ponderada por um peso, possui uma junção de soma, responsável pela combinação aditiva das entradas ponderada pelos pesos e uma função de ativação, geralmente não linear, que é responsável pela ativação da saída ou resposta do elemento processador.

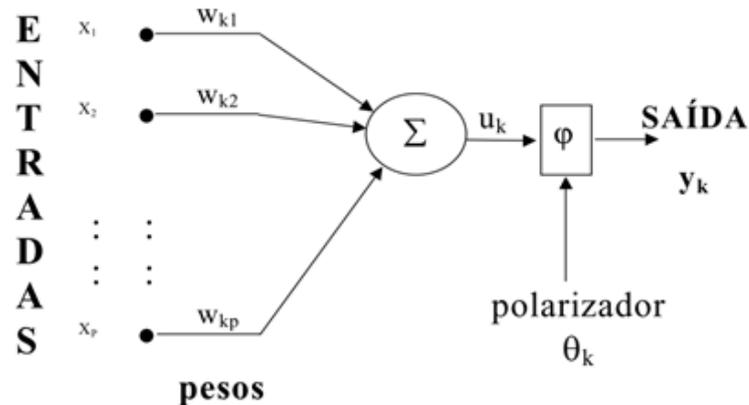


Figura 2.15: Representação de um neurônio artificial. Adaptado de:(Silva, 2005)

A função básica do neurônio artificial, ou elemento processador, é realizar o somatório, ponderado por fatores denominados pesos sinápticos, dos elementos do vetor de entrada e aplicar este resultado como entrada de uma função não linear denominada função de ativação. (Silva, 2005)

RN multicamadas são arquiteturas em que neurônios são organizados em duas ou mais camadas. As entradas são conectadas aos elementos processadores básicos, que são por sua vez interconectados com elementos de outras camadas e/ou a saída da rede (Silva, 2005).

A camada de entrada não realiza processamento, apenas distribui os valores de entrada para os neurônios de primeira camada de processamento.

Geralmente, a solução de problemas utilizando uma Rede Neural Artificial inicia-se pela fase de aprendizado ou treinamento, no qual exemplos representativos das classes do problema são apresentados à rede (Milare, 2003). Por meio de um algoritmo de treinamento são extraídas automaticamente as características para representar implicitamente os exemplos. Após a extração das características, a rede utiliza durante a sua fase de uso o grau de similaridade com as representações armazenadas na rede para a classificação de outros exemplos.

Os pesos associados às conexões são calculados e modificados até a rede encontrar uma con-

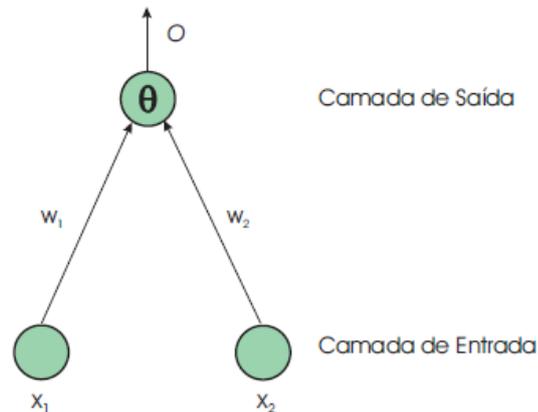


Figura 2.16: Exemplo de uma rede *perceptron*. Adaptado de: (Milare, 2003)

figuração onde as saídas estejam com um grau mínimo de erro para os exemplos apresentados à rede durante o treinamento.

Para descrever o treinamento de uma Rede Neural Artificial será utilizado um modelo bem simples de uma rede denominada *perceptron* (Rosenblatt, 1962). A Figura 2.16 possui uma camada de entrada e uma camada de saída, x_1 e x_2 representam a entrada da rede, w_1 e w_2 são os pesos, θ é uma constante conhecida como limiar.

Os pesos são inicializados com números randômicos entre $[-1,1]$. Em seguida, os exemplos são apresentados à rede até que um critério de parada seja satisfeito. A saída (O) é calculada utilizando o seguinte critério: Se $\sum_i w_i x_i - \theta > 0$ então $O = 1$ senão $O = 0$.

A saída da rede (O) é comparada à saída desejada (y). Se essas saídas forem iguais, os pesos da rede não são modificados e o próximo exemplo é apresentado à rede; caso contrário, os pesos devem ser modificados utilizando a Equação 2.30 (Milare, 2003).

$$w_i = w_i + \Delta w_i \quad (2.30)$$

onde

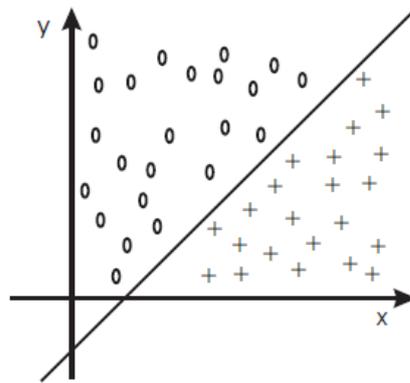


Figura 2.17: Interpretação geométrica para a rede *perceptron*. Adaptado de: (Milare, 2003)

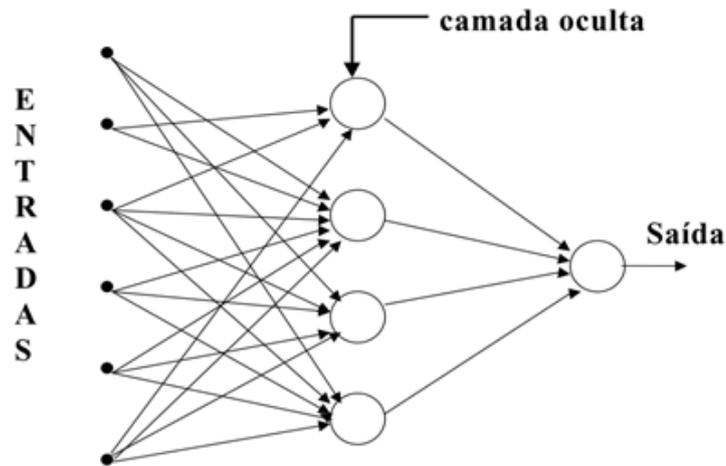


Figura 2.18: Rede Neural composta por várias camadas de neurônios. Adaptado de: (Silva, 2005)

$$\Delta w_i = (y - O)x_i$$

O objetivo do treinamento de uma rede *perceptron* é encontrar uma combinação linear de pesos e limiar que consiga separar os exemplos de uma classe C_1 dos exemplos de uma classe C_2 (Weiss e Kulikowski, 1991). O treinamento de uma Rede Neural se resume a encontrar uma reta que separe o exemplo em duas classes como mostrado na Figura 2.17, onde podemos observar nesse caso que o limiar indica em que ponto a reta intercepta o eixo y .

A principal limitação das redes *perceptron* é que elas somente solucionam problemas linearmente separáveis. Problemas mais complexos necessitam de redes multicamada como ilustrada na Figura 2.18. Esse tipo de rede pode ser treinada com uma regra de aprendizado conhecida como algoritmo *backpropagation* (Rumelhart et al., 1986).

2.5 Testes Diagnósticos

Testes diagnósticos são descritos como métodos capazes de indicar a presença ou ausência de uma doença, com uma determinada chance de erro (Linnet, 1987). Podemos errar de duas formas, classificando o indivíduo doente como não doente e classificando um indivíduo não doente como doente. Nos testes diagnósticos observamos quatro situações relacionadas aos tipos de acerto e de erro.

1. Verdadeiro positivo (*VP*) - Quantidade de indivíduos doentes que são classificados como doentes.
2. Verdadeiro negativo (*VN*) - Quantidade de indivíduos não doentes que são classificados como não doentes.
3. Falso positivo (*FP*) - Quantidade de indivíduos não doentes que são classificados como doentes.
4. Falso negativo (*FN*) - Quantidade de indivíduos doentes que são classificados como não doentes.

Os valores de *VP* e *VN*, são interpretados como "acerto" do teste sob investigação, e são importantes medidas de desempenho. Dentre as situações interpretadas como "erro" do teste, temos os valores de *FN* e *FP*.

No estudo de testes diagnósticos, um procedimento importante é a obtenção das medidas de desempenho do teste sob investigação. Valores preditos, sensibilidade, especificidade, bem como outros elementos estatísticos da teoria da decisão, são medidas de desempenho usadas para avaliar testes diagnósticos (Fleiss, 1981).

Valor predito positivo (*VPP*) de um teste diagnóstico é definido como a proporção de indivíduos portadores da doença, dado que o resultado do teste é positivo. Em outras palavras, *VPP* expressa a probabilidade de um paciente com resultado positivo no teste ter a doença (Vaz, 2009). Segundo o trabalho de (Dunn e Everitt, 1995), podemos estimar diretamente o valor utilizando a equação 2.31.

$$VPP = \frac{VP}{VP + FP} \quad (2.31)$$

Valor predito negativo (*VPN*) de um teste diagnóstico é definido como a proporção de indivíduos não portadores da doença, dado que o resultado do teste é negativo. Em outras palavras, *VPN* expressa a probabilidade de um paciente com resultado negativo no teste não ter a doença (Vaz, 2009). Segundo o trabalho de (Dunn e Everitt, 1995), podemos estimar diretamente o valor utilizando a Equação 2.32.

$$VPN = \frac{VN}{FN + VN} \quad (2.32)$$

Acuracidade (*Ac*) é definida como a probabilidade do teste que está sendo investigado fornecer um resultado correto, dado todos os indivíduos, isto é, portador ou não portador da doença. A quantidade *Ac* avalia o quão confiável é o teste para detectar a doença e informar corretamente se ela está ou não está presente. O valor pode ser estimado utilizando a Equação 2.33.

$$Ac = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.33)$$

Sensibilidade (*Se*) é definida como a probabilidade do teste que está sendo investigado fornecer um resultado positivo, dado que o indivíduo é portador da doença. A quantidade *Se* avalia a capacidade do teste detectar a doença quando ela está de fato presente. O valor de *Se* pode ser estimado utilizando a Equação 2.34.

$$Se = \frac{VP}{VP + FN} \quad (2.34)$$

Especificidade (*Es*) é definida como a probabilidade do teste que está sendo investigado fornecer um resultado negativo, dado que o indivíduo não é portador da doença. A quantidade *Es* avalia a capacidade do teste "afastar" a doença quando ela está de fato ausente. O valor de *Es* pode ser estimado utilizando a equação 2.35.

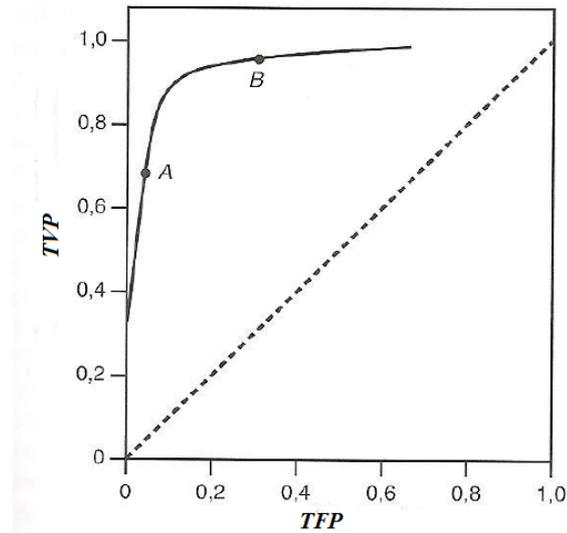


Figura 2.19: Curva de Características de Operação do Receptor (Curva ROC) (Adaptado de (Margotto, 2010))

$$Es = \frac{VN}{VN + FP} \quad (2.35)$$

2.6 Curva ROC

As Curvas ROC foram desenvolvidas no campo das comunicações como uma forma de demonstrar as relações entre sinal-ruído. Interpretando o sinal como os verdadeiros positivos (sensibilidade) e o ruído, os falsos positivos (1-especificidade) (Margotto, 2010).

A Curva ROC é um gráfico representado na Figura 2.19 da taxa de verdadeiros positivos(TVP) que é dada pela Equação 2.36 versus a taxa de falsos positivos(TFP) que é dada pela Equação 2.37.

$$TVP = \frac{VP}{VP + FN} \quad (2.36)$$

$$TFP = \frac{FP}{FP + VN} \quad (2.37)$$

A Curva ROC evidencia os valores para os quais existe maior sensibilidade em função da

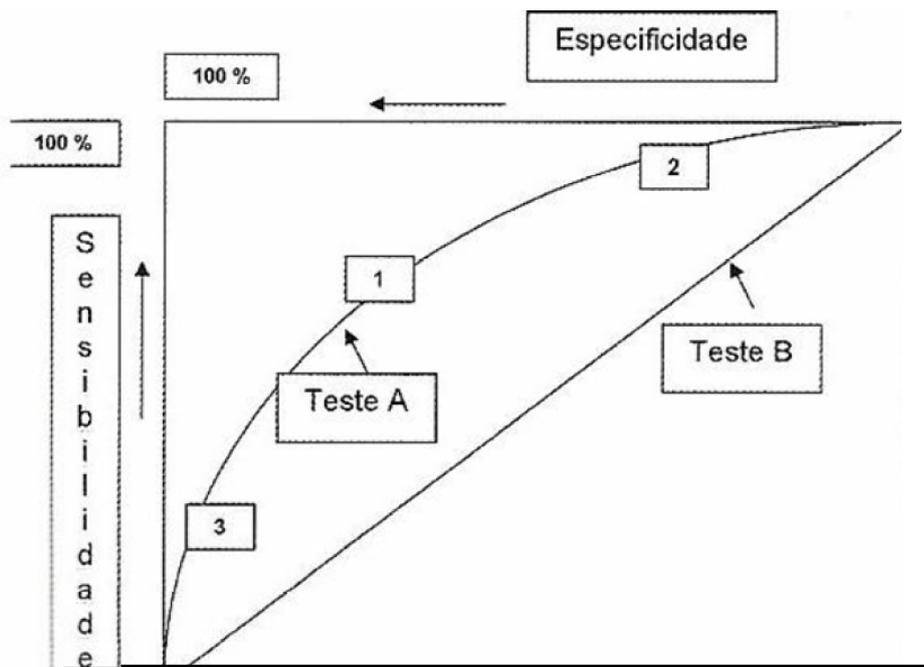


Figura 2.20: Curva ROC de comparação (Adaptado de (Margotto, 2010))

especificidade que corresponde ao ponto em que se encontra mais próxima do canto superior esquerdo do diagrama, uma vez que TVP é um e o de TFP é zero.

A Curva ROC permite comparar dois ou mais exames diagnósticos que constitui uma das maiores virtudes das curvas ROC. No gráfico da Figura 2.20, o Teste A tem melhor acurácia que o teste B. O ponto 1 possui maior valor de sensibilidade e especificidade; o ponto 2 possui maior sensibilidade, porém menor especificidade e já o ponto 3, maior especificidade, porém, menor sensibilidade.

Capítulo 3

Trabalhos Relacionados

A maioria dos trabalhos na literatura científica para DAC tem focado em informações de textura, por serem mais simples de implementar com métodos estatísticos já bem fundamentados. Por outro lado, as informações geométricas, que independem de hardware têm sido pouco exploradas.

Este capítulo apresenta uma análise de diversos trabalhos da área estudada sob três pontos de vista: aqueles que utilizam descritores de textura; descritores geométricos e os diferentes métodos de classificação utilizados para detectar os tipos (maligno e benigno) de lesões mamárias.

Todos os trabalhos citados neste capítulo possuem o mesmo objetivo de melhor classificar imagens de ultrassom mamário diferenciando tumores malignos e benignos, mas com propostas de metodologias diferentes.

3.1 Extração de Características

Essa seção tem como objetivo explicar a extração de características de uma imagem de ultrassom e a maneira como são classificadas as informações extraídas.

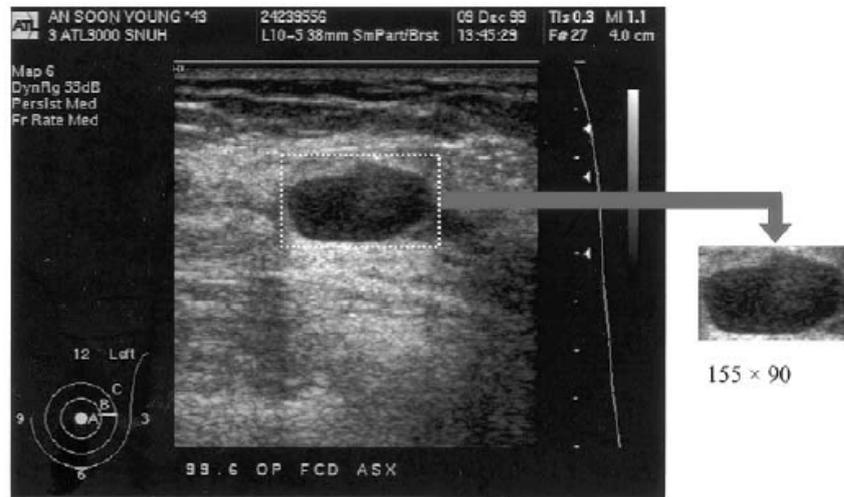


Figura 3.1: Imagem digital de 640 x 480 *pixels* capturada de um scanner ultrassom (US). A região de interesse é um retângulo de 1.65 cm x 0.96 cm e 155 x 90 *pixels*. [Adaptada de [(Dar et al., 2002)]]

3.1.1 Extração da Região de Interesse

Na Figura 3.1 observamos a imagem de um tumor capturado por um scanner que gera imagens em duas dimensões (2D). O médico traça um retângulo em torno do tumor na imagem original gerando uma sub-imagem denominada de região de interesse. Essa sub-imagem é analisada pelo computador considerando informações de variação da intensidade e informações de textura (Dar et al., 2002) e (Ruey et al., 2002).

A Figura 3.2 apresenta uma lesão utilizando um scanner que gera imagens em três dimensões. Esse tipo de imagem nos fornece muito mais informações do que as imagens extraídas em um scanner 2D, pois as informações podem ser obtidas no planos transversal, longitudinal e superior simultaneamente.

A Figura 3.3 demonstra graficamente a divisão de uma imagem 3D em diversas imagens 2D e a reconstrução do volume através dos planos cartesianos extraídos.

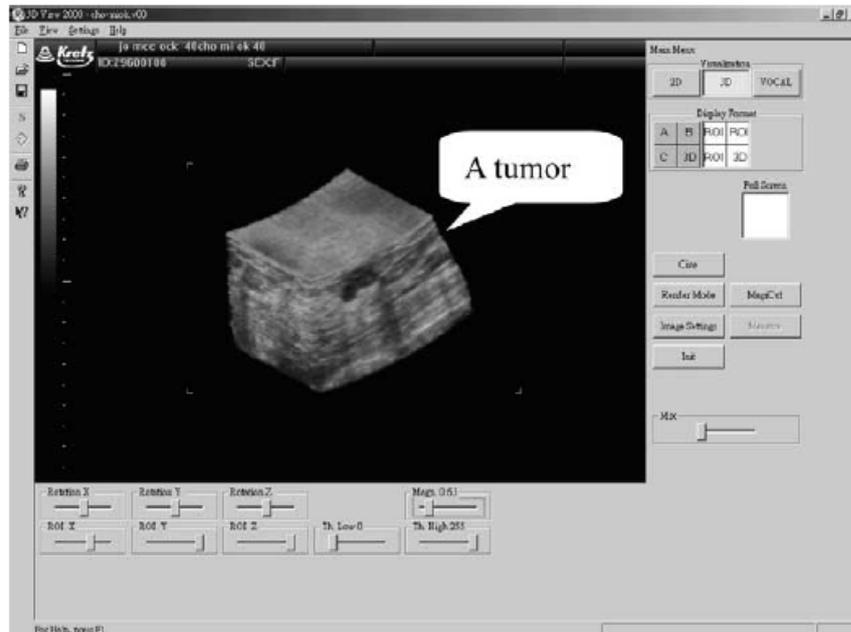


Figura 3.2: Exemplo de uma imagem digitalizada 3-D US. [Adaptada de [(Wei et al., 2005)]]

3.1.2 Descritores de Textura

Os trabalhos relacionados que utilizaram imagens 2D fizeram uso de métodos estatísticos para analisar as informações das imagens, como Dar et al.(2002) que usou contraste, covariância e dissimilaridade que utilizam a distância entre *pixels*, por isso é vantajoso termos imagens grandes e não com um elevado número na escala de cinza. A Tabela 3.1 mostra uma avaliação do desempenho para três diferentes parâmetros de textura utilizados no trabalho.

Tabela 3.1: Desempenho para diferentes parâmetros de textura de (Dar et al., 2002).

Item	Contraste	Covariância	Dissimilaridade
Acuracidade	86.94	88.89	54.86

Ruey et al.(2002) utilizaram o coeficiente *speckle* de autocovariância. Para explicar o método, primeiramente temos que definir *speckle*. Quando um feixe de luz é transmitido ou espalhado por uma superfície com rugosidades da ordem do comprimento de onda incidente, ocorre a formação de uma estrutura granular no espaço livre à qual é dado o nome de *speckle* (Silva, 2007). É um fenômeno natural de interferência, tipicamente ondulatório, visto não apenas no visível, mas também no espectro eletromagnético e na acústica.

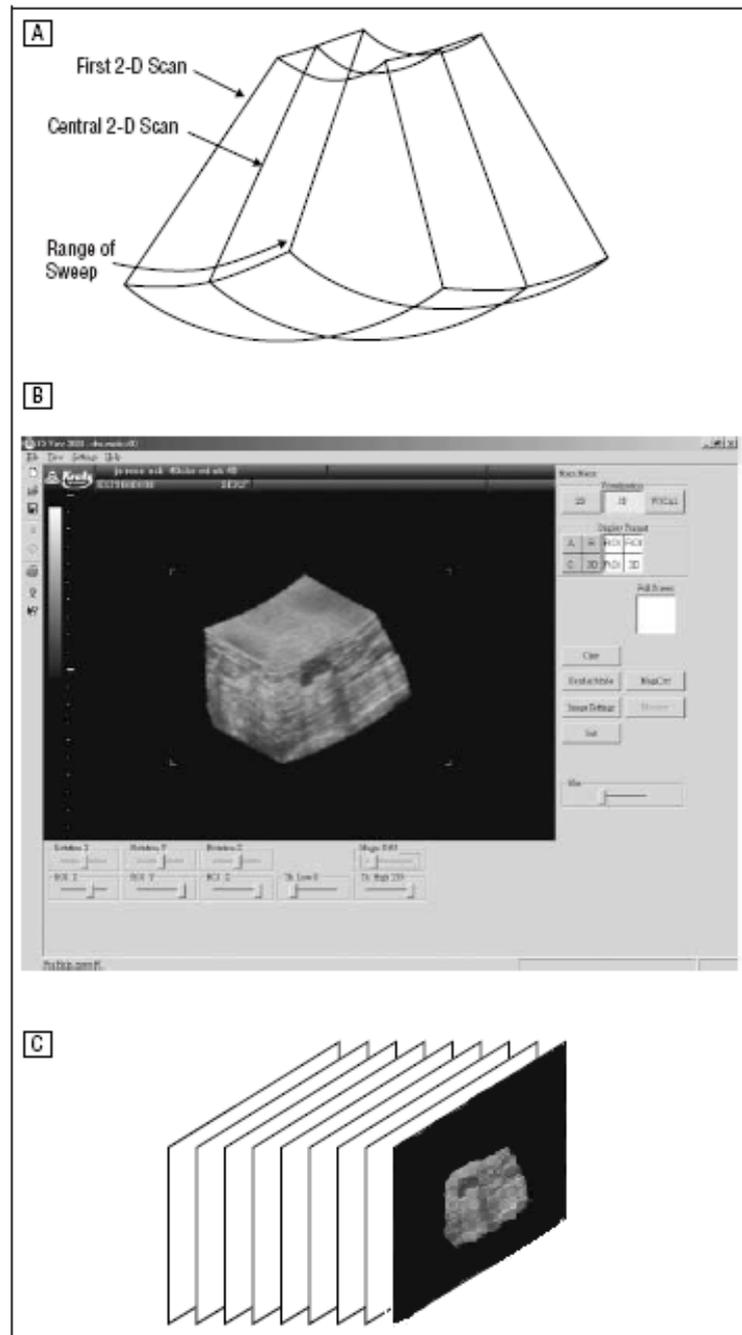


Figura 3.3: A) Demonstra os planos 2-D em uma imagem com dimensão 3-D. B) Reconstrução do Volume em 3-D. C) Os planos cartesianos extraídos. [Adaptado de [(Dar et al., 2006)]].

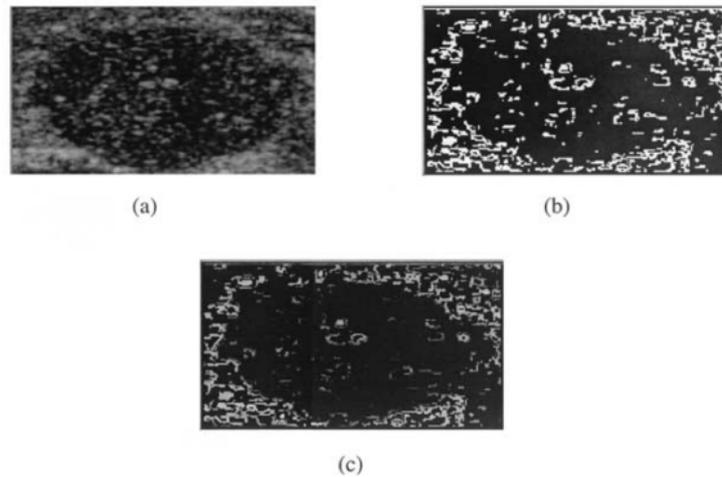


Figura 3.4: a) Imagem original de um tumor. b) Resultado após o detector de speckle. c) A escala de cinza dos *pixels* para classificação os speckles. [Adaptado de [(Ruey et al., 2002)]]].

O trabalho de Ruey et al.(2002) utilizou um detector de *speckle* antes da extração de informações da textura. A Figura 3.4 demonstra a aplicação do detector de *speckle*. Após a extração dos *speckle* foi utilizado coeficiente de autocorrelação que reflete a correlação entre os *pixels* da imagem. A Tabela 3.2 apresenta os resultados do trabalho. Utilizando o detector de *speckle* o resultado foi em média melhor do que o método convencional com todos os *pixels*.

Tabela 3.2: Resultado do trabalho de (Ruey et al., 2002)

Método	<i>Acuracidade</i>	<i>Sensibilidade</i>	<i>Especificidade</i>	<i>VPP</i>	<i>VPN</i>
Com detector de <i>Speckle</i>	93.20	95.45	91.43	89.74	96.24
Convencional (todos <i>pixels</i>)	85.60	89.08	82.86	80.33	90.63
Sem detector <i>speckle</i>	82.80	86.36	80.00	77.24	88.19

Segundo (Dar et al., 1999) tumores são classificados pela discriminação de quanto a massa interna é regular, discriminando também as bordas e a heterogeneidade interna. Baseado nisso (Dar et al., 1999) utilizou coeficiente de autocorrelação entre os *pixels*.

3.1.3 Descritores de Forma

Outros trabalhos como (Wei et al., 2005), (Dar et al., 2006), (Wei et al., 2003), (Rodrigues et al., 2006a), (Rodrigues et al., 2006d) e (Rodrigues et al., 2006b) utilizaram imagens 3D, onde são extraídas diversas imagens em 2D passando cada uma por uma avaliação de características

trazendo informações do plano transversal, longitudinal e superior.

Técnicas para extrair essas informações geralmente envolvem a normalização e o pré-processamento das imagens, aplicação de operadores morfológicos e algoritmos de geometria computacional. Em geral, entre as maiores dificuldades está a necessidade da separação inicial da lesão do fundo da imagem. Como se trata geralmente de imagens com alto grau de ruído, fundo heterogêneo, com baixo contraste e formas geométricas muito flexíveis, há a necessidade de algoritmos de extração da lesão mais eficientes, tendo como contrapartida um maior consumo computacional. Essa necessidade de extração adequada da lesão do fundo é uma das razões pelas quais muitos pesquisadores têm optado por métodos que consideram somente informações de texturas. No entanto, informações geométricas tendem a ser mais igualitárias para scanners de fabricantes diversos do que informações de textura. Todas essas considerações levam a várias maneiras de modelar sistemas de Diagnóstico Assistido por Computador (DAC).

3.2 Classificação do Tumor

Cada informação extraída de uma imagem é transformada em um vetor montando uma base de informações. Alguns trabalhos como o de Wen et al (2002) utilizam uma técnica conhecida como *Data Mining* para treinamento do sistema após a extração das informações. O *Data Mining* é uma das técnicas utilizadas para a realização da descoberta de conhecimento em bases de dados. Aspectos específicos incluem investigação e criação de conhecimento, processos, algoritmos e mecanismos de recuperação de conhecimento potencial de estoques de dados (Norton, 1999).

A idéia geral de sistemas DAC para lesões mamárias é a extração de informações de textura e/ou geométricas seguida de algum método de classificação que considere a lesão como maligna ou benigna. Entre os métodos de classificação mais usados estão às técnicas baseadas em árvore de decisão, máquina de suporte vetorial e as baseadas em redes neurais.

(Dar et al., 2002) estudou a eficiência de um classificador de imagens de ultrassom entre tumores malignos e benignos com uma pequena base para treinamento, utilizando matriz de coocor-

rência, a qual depende da escala de cinza de cada imagem e os parâmetros estatísticos dessa matriz como contraste, covariância e dissimilaridade. A técnica de bootstrap foi utilizada para compensar a base de teste. Para a classificação das imagens utilizou-se uma árvore de decisão que atingiu um percentual de acuracidade de 86.94% baseada no contraste, 88.89% considerando covariância e 54.86% baseada na dissimilaridade.

Utilizando quatro matrizes de coocorrência (0° , 45° , 90° e 135° com distância 1) e os mesmos descritores de Dar et al.(2002) e Wen et al (2002) apresentaram acuracidade de 86.9% considerando o contraste, 88.9% considerando covariância e 54.7% baseado na dissimilaridade. Para a classificação dos tumores, esse trabalho utilizou árvore de decisão (Seção 2.4.1).

Aproveitando um fenômeno comum nas imagens de ultrassom, conhecido como *speckle*, Ruey et al (2002) desenvolveram seu trabalho baseando-se em um coeficiente de autocovariância que é um método estatístico. Para utilizar as informações de *speckle*, primeiramente foi usado um detector de *Speckle* para determinar a localização de cada um; em seguida foi calculado o coeficiente de autocovariância dos *pixels* com os *speckle*. A ferramenta utilizada para classificar os tumores foi um modelo de Máquina de Suporte Vetorial. Esse trabalho resultou em uma acuracidade de 93.20% usando o *speckle*, 85.60% usando o método convencional de autocorrelação entre os *pixels* e 82.80% sem *speckle*.

Em Rodrigues et al (2006) foi apresentado um método baseado em entropia não-extensiva de 5 passos para classificação de tumores malignos e benignos utilizando imagens de ultrassom. Os cinco passos são os seguintes: Segmentação com entropia não-extensiva; Operação Morfológica para separação da lesão do background; Ajuste da região com *level-sets*; Extração de cinco características da lesão: área, circularidade, protuberância, homogeneidade e sombra acústica, e classificação usando Máquina de Suporte Vetorial (Seção 2.4.2). O método proposto atingiu 95% de acuracidade, 97% de sensibilidade, 94% de especificidade, 91.51% de VPP e 97.92% de VPN para uma base de 250 imagens, sendo 100 benignos e 150 malignos. Posteriormente Thomaz et al (2008) e Giraldi et al (2008) mostraram que o número de características discriminantes nessa base é bem menor do que se supunha, e propuseram um método semiautomático para determinar essas características.

Correlacionando os *pixels* com seus vizinhos Dar et al (1999) normalizaram imagens 2D e calcularam o coeficiente de autocorrelação entre *pixels*. A ferramenta escolhida para classificação foi Redes Neurais e a acuracidade para classificar tumores malignos foi de 95.0% (133 de 140 tumores).

Chung et al (2003) usaram uma base de imagens colecionadas durante um período de 4 anos e, após análise, foi utilizado Rede Neural para classificação dos Tumores. As características morfológicas como, forma, contorno e tamanho foram extraídas de cada lesão e separadas para análise.

Utilizando imagens 3D, Dar et al.(2006) propôs um método de trabalho composto de 2 estágios. O primeiro estágio de segmentação é onde geramos sub-imagens a partir do volume de interesse para determinar o coeficiente de autocorrelação utilizando 3 *frames* para definir a área do tumor analisada. Com as sub-imagens é gerada uma matriz de autocorrelação de dimensão 3 e uma matriz de normalização. O segundo estágio é a classificação dessas matrizes onde foram utilizadas Redes Neurais (Seção 2.4.3). O trabalho comparou a utilização de coeficiente de autocorrelação gerando matrizes para imagens em 3D e 2D, onde obteve uma acuracidade de 91.9% vs 77.0%.

3.3 Discussões e Resultados

Todos os trabalhos relacionados possuem base de imagens diferentes e por isso devemos avaliar seus resultados individualmente.

O gráfico da Figura 3.5 demonstra os resultados obtidos em cada trabalho considerando a Acuracidade, Sensibilidade, Especificidade, Valor Predito Positivo e Valor Predito Negativo agrupados pelas técnicas de classificação de lesões mamárias.

Analisando o gráfico da Figura 3.5 podemos verificar que o VPN é o avaliador no protocolo de validação cruzada que apresentou o maior percentual na avaliação das pesquisas. Nos trabalhos de Dar et al. (1999) e Chung et al.(2003) a linha do Valor Predito Negativo chegou a 99% quando

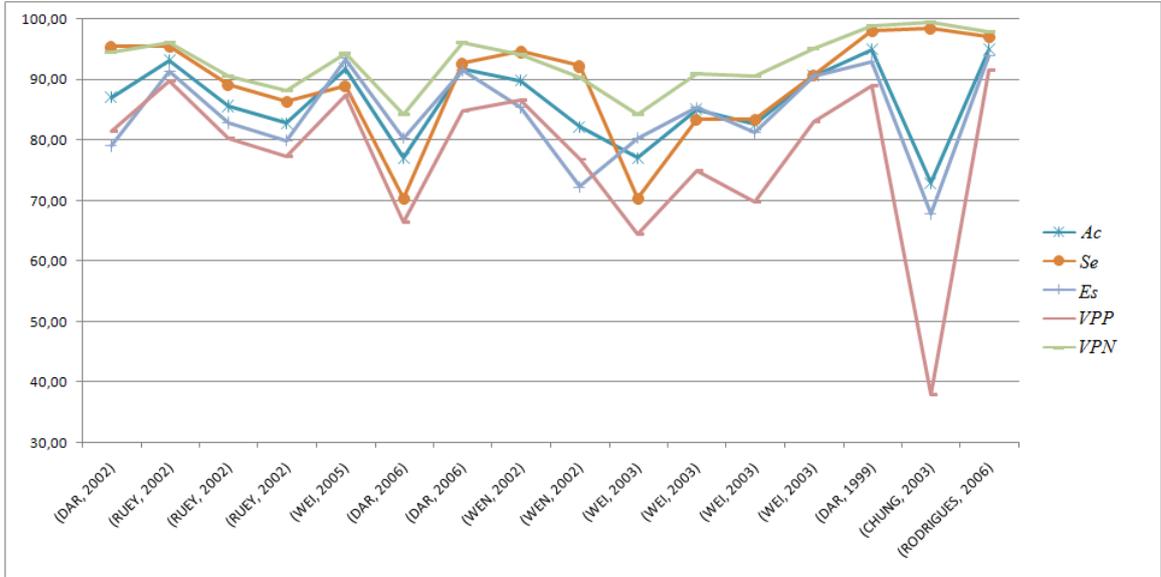


Figura 3.5: Gráfico de comparação dos resultados obtidos das pesquisa estudadas.

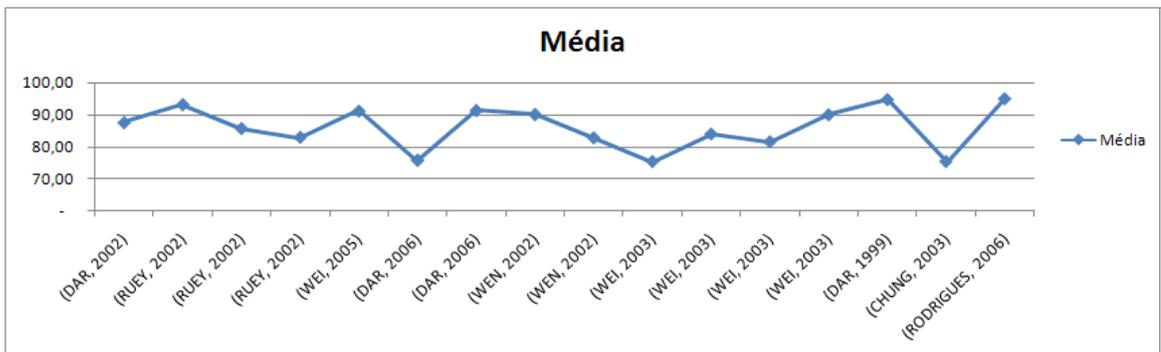


Figura 3.6: Média dos resultados obtidos nas pesquisas estudadas baseado no protocolo de validação cruzada.

Tabela 3.3: Lista dos trabalhos relacionados com informações do tipo de imagem e classificador utilizado.

Artigo	Imagem	Classificador
Art.1 (Dar et al., 2002)	2-D	Arvore de decisão
Art.2-A (Ruey et al., 2002)	2-D	MSV
Art.2-B (Ruey et al., 2002)	2-D	MSV
Art.2-C (Ruey et al., 2002)	2-D	MSV
Art.3 (Wei et al., 2005)	3-D	Rede Neural
Art.4-A (Dar et al., 2006)	3-D	Rede Neural
Art.4-B (Dar et al., 2006)	3-D	Rede Neural
Art.5-A (Wen et al., 2002)	2-D	Arvore de decisão
Art.5-B (Wen et al., 2002)	2-D	Arvore de decisão
Art.6-A (Wei et al., 2003)	3-D	Rede Neural
Art.6-B (Wei et al., 2003)	3-D	Rede Neural
Art.6-D (Wei et al., 2003)	3-D	Rede Neural
Art.6-E (Wei et al., 2003)	3-D	Rede Neural
Art.7 (Dar et al., 1999)	2-D	Rede Neural
Art.8 (Chung et al., 2003)	2-D	Rede Neural
Art.9 (Rodrigues et al., 2006c)	3-D	MSV

utilizadas imagens 2D e Redes Neurais para classificação, sendo se destacado também quando avaliada a curva da Sensibilidade. Na curva da Acuracidade e da Especificidade o trabalho de Dar et al (1999) também é o de melhor resultado. Para avaliação da curva Valor Predito Positivo o melhor resultado foi o trabalho de Ruey et al (2002) utilizando imagens 2-D com Máquina de Suporte Vetorial para classificação utilizando o método de speckle-emphasis.

O trabalho de Chung et al.(2003) apresentou o pior resultado avaliado pelas curvas de Valor Predito Positivo, Especificidade e Acuracidade. Nos itens Sensibilidade e Valor Predito Negativo temos os resultados mais baixos nos trabalhos de Dar et al.(2006) e Wei et al.(2003) onde foram implementados utilizando imagens 3D e Rede Neural para classificação dos tumores. Podemos observar no gráfico da Figura 3.6 que o trabalho de Rodrigues et al.(2006) foi o que obteve melhor resultado na média de avaliação, atingindo um valor médio de satisfação de 95.09%.

A Tabela 3.3 discrimina todos os trabalhos relacionados do ponto de vista dos tipos de imagens utilizadas e classificadores. Podemos observar analisando a terceira coluna que a utilização de Rede Neural é o preferido entre os trabalhos relacionados para a classificação, seguido de Máquina de Suporte Vetorial e Árvore de Decisão.

3.4 Trabalho base

Os trabalhos na área de DAC podem ser diferenciados com relação a quantidade e o tipo de características usadas nos vetores de características e do método de classificação.

Com relação aos tipos de características, essas podem ser baseadas em textura ou informações geométricas. Exemplos de informações de textura utilizados em Dar et al. (2002), Ruey et al. (2002), Wen et al. (2002), Dar et al. (1999) e Chung et al. (2003) podem ser: contraste, covariância, dissimilaridade, heterogeneidade, homogeneidade, sombra acústica, entre outros. Por outro lado, exemplos de informações geométricas utilizado em Wei et al. (2005), Dar et al. (2006), Wei et al. (2003) e Rodrigues et al. (2006), podem ser área, circularidade e protuberância.

Existem alguns sistemas híbridos como os de Dar et al. (2006) e Rodrigues et al. (2006) que utilizaram características de textura e características geométricas. Como não há consenso entre os pesquisadores de quais características mais influenciam no diagnóstico, os sistemas híbridos ganham em importância.

Um sistema híbrido foi proposto nos trabalhos de Rodrigues et al. (2006a), Rodrigues et al. (2006b) e Rodrigues et al. (2006c), onde as características geométricas usadas foram área, circularidade e protuberância; e as características de textura foram homogeneidade e sombra acústica. Recentemente, nos trabalhos de Giraldi et al. (2008) e Thomaz et al. (2008) foi mostrado que nos trabalhos de Rodrigues et al. (2006a), Rodrigues et al. (2006b) e Rodrigues et al. (2006c) as características mais discriminantes eram área, sombra acústica e homogeneidade; uma geométrica e duas de textura. Como resultado, a metodologia proposta pelos autores apresentou 95% de acuracidade, 97% de sensibilidade, 94% de especificidade, 91.51% de *VPP* e 97.92% de *VPN*. Em nosso trabalho, como será visto mais adiante, essas três características mais discriminantes são usadas com o objetivo de reduzir a dimensão do espaço vetorial para viabilizar uma menor complexidade computacional.

A base desta dissertação é o trabalho de Rodrigues et al. (2006) que sugeriu informações híbridas e apresentou uma metodologia de 5 passos para classificações de tumores malignos e benignos.

Os 5 passos são os seguintes:

1. Segmentação não-extensiva

No caso das imagens de ultrassom, a primeira segmentação é um processo complexo, uma vez que as imagens têm um baixo SNR (*Signal-to-Noise Ratio* ou razão Sinal Ruído - a relação entre o sinal obtido de determinado objeto dividido pelo ruído capturado durante a exposição.(Jacques, 2003)), baixa resolução, baixo contraste e várias regiões pequenas e espúrias. Particularmente, a primeira tarefa fundamental é a segmentação uma vez que o êxito das etapas restantes depende dela (Rodrigues et al., 2006c).

Rodrigues et al.(2006) propôs uma extensão natural para o algoritmo de segmentação proposto por Albuquerque Esquef (2004) que, recursivamente, realiza mais de uma segmentação por região, em imagens de ultrassom, além de ter um baixo tempo computacional. O algoritmo proposto foi denominado *Non-Extensive Segmentation Recursive Algorithm* (NESRA) (Rodrigues et al., 2006c).

Na Figura 3.7 mostramos o resultado de segmentação para 3 valores de q . Podemos observar nas imagem que, quanto maior o valor de q , melhor o resultado da segmentação.

2. Cadeia morfológica

No segundo passo foi usada uma cadeia morfológica para extração da lesão do background. Exemplificando, temos a Figura 3.8 que mostra a imagem original em (a) e a imagem final em (b) após ser identificado a área da região de interesse.

Na Figura 3.9 temos na fileira (a) três imagens originais passando pelo algoritmo NESRA na fileira (b) e a utilização da cadeia morfológica em (c).

3. Arquitetura Level-sets

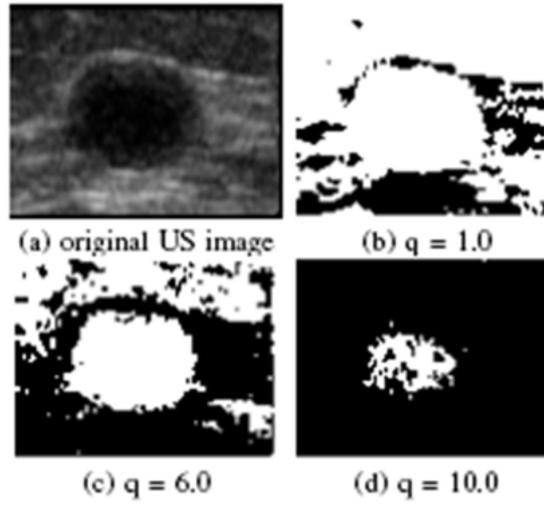


Figura 3.7: Segmentação não-extensiva (Adaptado de (Rodrigues et al., 2006c))

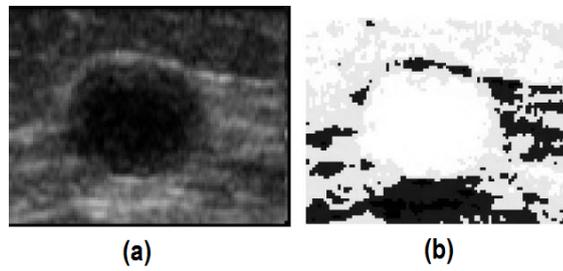


Figura 3.8: (a) Imagem original de um tumor benigno. (b) Resultado com duas recursividades e $q = 0.5$.(Adaptado de (Rodrigues et al., 2006c))

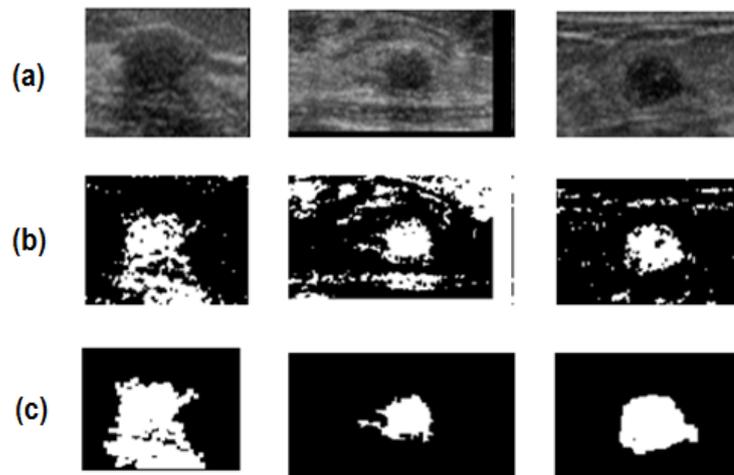


Figura 3.9: Cadeia Morfológica.(Adaptado de (Rodrigues et al., 2006c))



Figura 3.10: Arquitetura Level-Sets.(Adaptado de (Rodrigues et al., 2006c))

Após a extração da lesão do background, no terceiro passo foi usada uma arquitetura de level-sets para suavizar a curva em relação a sua borda real, uma vez que a cadeia morfológica tende a degradar a borda original ao extrair a lesão do background. Na Figura 3.10 mostramos em (a) a imagem após a cadeia morfológica e em (b) a aplicação da arquitetura de level-sets. Podemos notar a suavização da borda da lesão.

4. Extração de características da lesão mamária

Na etapa quatro, foram utilizados valores absolutos das seguintes características:

- Área

Em geral os tumores malignos possuem maior área que os tumores benignos tornando uma característica com um alto poder discriminante.

- Circularidade

Nessa característica as lesões benignas geralmente têm mais áreas circulares comparado com as lesões malignas. Também pode ser um bom discriminante.

- Homogeneidade

Lesões malignas tendem a ser menos homogêneas do que as lesões benignas.

- Protuberância

Lesões malignas possuem mais protuberâncias em relação à média das áreas benignas.

- Sombra Acústica

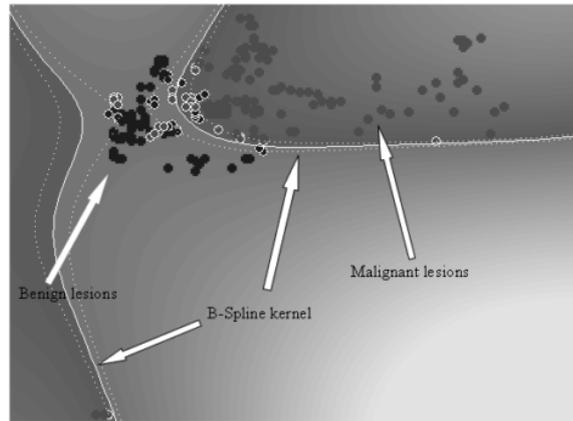


Figura 3.11: MSV com um Kernel B-Spline. A linha branca é um polinômio de separação dos tumores malignos (círculos cinza) dos benignos (círculos pretos). Os círculos com as fronteiras brancas são aqueles dentro da região de confiança.

Lesões benignas são mais aquosas tendendo assim a formação de um reforço acústico abaixo da lesão. Por outro lado, quando a lesão é mais sólida (uma característica maligna), há uma tendência em formar sombra acústica.

5. Classificação por Máquina de Suporte Vetorial

Estas cinco características são compostas em um vetor de características. O espaço de todos os vetores de características é, finalmente, submetido a um classificador MSV para classificar a lesão como maligna ou benigna. O MSV é uma ferramenta discriminante, principalmente quando não temos uma separação linear entre os dados. Uma vez que o espaço de característica não é linearmente separável, Rodrigues et al. (2006) utilizaram uma curva B-Spline como Kernel. A Figura 3.11 mostra o resultado da aplicação do MSV com Kernel B-Spline para classificação da base.

Para avaliar os resultados obtidos Rodrigues et al.(2006) utilizaram uma curva ROC - (*Receiver Operating Characteristic*), apresentada na Figura 3.12 para comparar qual combinação de característica tem melhor resultado. Como resultado, a curva ROC obtida apresentou melhor desempenho para Área, Heterogenidade e Sombra Acústica.

As curvas ROC facilitam o entendimento entre a relação da sensibilidade e a especificidade de um teste de diagnóstico quantitativo ao longo de valores contínuos de ponto de corte (Rodrigues

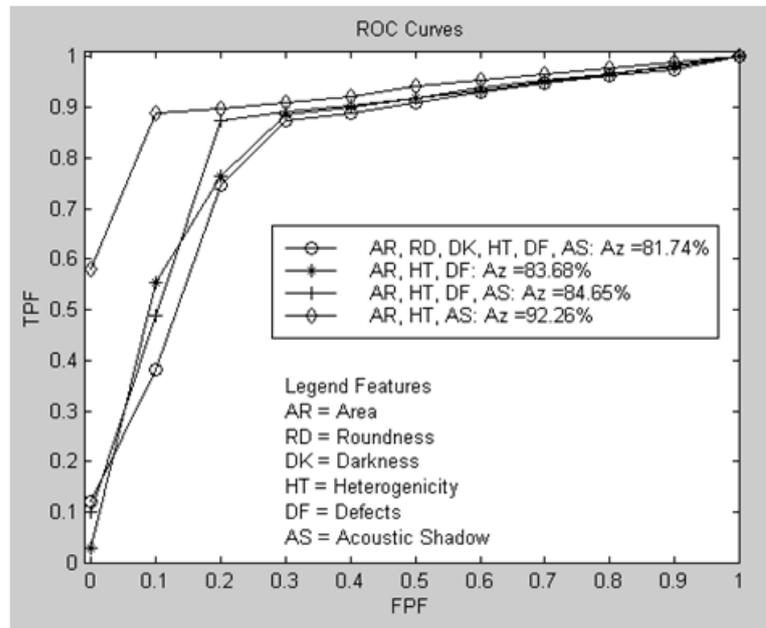


Figura 3.12: Curva ROC (Adaptado de (Rodrigues et al., 2006c))

et al., 2006c).

A curva ROC da Figura 3.12 também pode ser plotada representando por equivalência a fração da taxa de verdadeiro positivo e taxa falsos positivos, pois trata-se de uma comparação de duas características operacionais (Rodrigues et al., 2006c).

Capítulo 4

Metodologia e experimentos

Com base nos conceitos previamente discutidos, a proposta deste trabalho é realizar a classificação de imagens de câncer de mama geradas por ultrassom, com uma abordagem estatística para extração das características e utilização dos conceitos da Teoria da Informação para classificar.

4.1 Banco de Imagens

A base de dados utilizada foi a mesma utilizada no trabalho de Rodrigues et al.(2006) que consta de 250 imagens de tumores mamários gerados por ultrassom, sendo 150 imagens de tumores malignos e 100 imagens de tumores benignos. Também foram fornecidas para este trabalho todas as imagens após o processamento; isto é, depois de serem segmentadas, extraída a lesão do background e a suavização com a arquitetura de *level-sets*.

4.2 Proposta para a Extração das Características

Do trabalho de Rodrigues et al (2006), analisando o gráfico da Figura 3.12, foram obtidos os melhores resultados para a Área, Heterogenidade e Sombra Acústica. Assim, baseando-se neste resultado, escolhemos essas três características para serem extraídas com uma abordagem estatística.



Figura 4.1: (a) Localização do Centróide (b) Extração dos raios

4.2.1 Histograma de Raios

A primeira característica a ser extraída das imagens foi o histograma de raios. Para a extração desta característica utilizamos as imagens segmentadas. Localizamos primeiramente o centróide "C" do tumor que é o ponto cujas coordenadas são as médias das coordenadas dos pontos extremos dos eixos x e y, como demonstrado na Figura 4.1(a).

Após a localização do ponto "C", calculamos a distância entre esse ponto e todos os pontos da fronteira desta imagem utilizando distância euclidiana. A Figura 4.1(b) exemplifica a localização de 2 pontos na borda da região do tumor.

Utilizando todas as imagens segmentadas separadas em dois grupos, criamos dois vetores base, apresentados na Figura 4.2 que demonstra o comportamento desta característica para as imagens dos grupos maligno e benigno.

O gráfico da Figura 4.2 apresenta o comportamento estatístico dos raios para cada tipo de tumor. No eixo x temos o tamanho do raio e o eixo y representa a frequência com que aparece nas imagens.

4.2.2 Histograma da Região de Interesse

A segunda característica a ser extraída das imagens foi o histograma de escala de cinza da região de interesse. Para a extração desta característica utilizamos a imagem segmentada do tu-

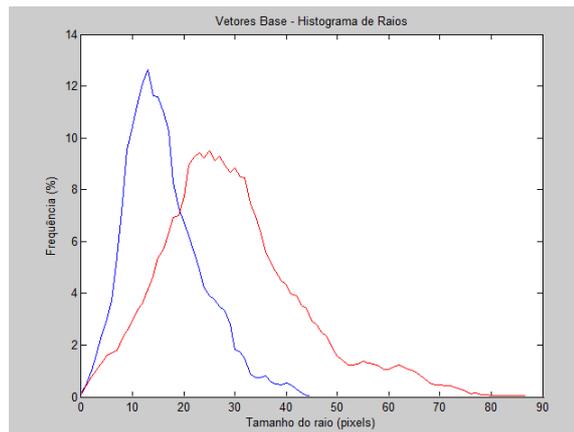


Figura 4.2: Histograma de Raios (Azul - Tumores Benignos e Vermelho - Tumores Malignos)



Figura 4.3: (a) Segmentada com localização da fronteira. (b) Imagem Original. (c) Extração da Região de Interesse

mor para localizar todos os pontos da fronteira, como mostra a Figura 4.3(a). Após a localização da fronteira, utilizamos como um delimitador para localizar o tumor na imagem original (Figura 4.3(b) e Figura 4.3(c)).

Após a extração de todas as regiões de interesse das 250 imagens, geramos dois vetores base representados na Figura 4.4, que demonstra o comportamento do histograma de escala de cinza para a região de interesse das 150 imagens malignas e 100 imagens benignas. No eixo x temos os valores de cinza (0,255) e no eixo y a frequência com que cada valor aparece no grupo de imagens.

4.2.3 Histograma da Sombra Acústica

A terceira e última característica a ser extraída das imagens foi o histograma de escala de cinza da sombra acústica. Para a extração desta característica utilizamos a imagem segmentada do tumor

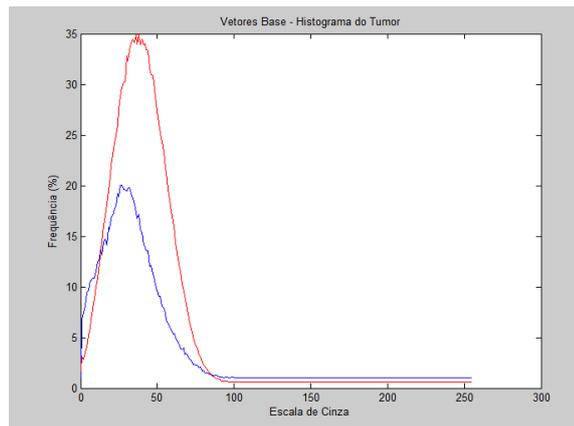


Figura 4.4: Histograma da Região de Interesse (Azul - Tumores Benignos e Vermelho - Tumores Malignos)



Figura 4.5: (a) Localização dos pontos extremos de x e da fronteira. (b) Imagem Original. (c) Sombra acústica

para localizar os pontos extremos do eixo x e a fronteira inferior da imagem como mostra a Figura 4.5(a). Após a localização destes pontos utilizamos como um delimitador para identificar a sombra acústica (Figura 4.5(b) e a Figura 4.5(c)).

Depois de identificar todas as sombras acústicas, separamos os 2 grupos e geramos os vetores base para o histograma de escala de cinza para a sombra acústica, como mostra o gráfico da Figura 4.6, onde no eixo x temos os valores de cinza (0,255) e no eixo y a frequência com que cada valor aparece no grupo de imagens.

4.3 Classificação das Imagens

Para atender a proposta do trabalho de realizarmos a classificação de imagens de câncer de mama, após abordarmos estatisticamente a extração das características, analisamos dois métodos diferentes para realizar a classificação das imagens:

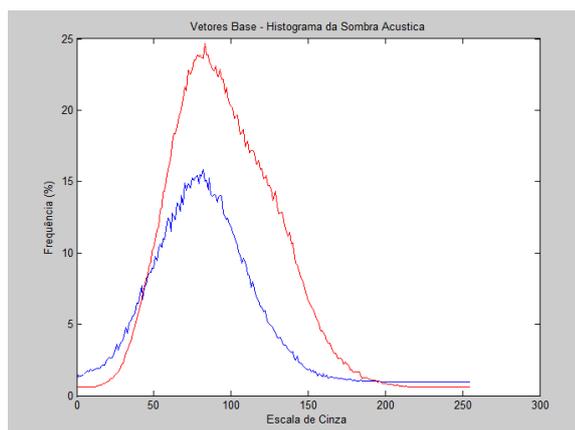


Figura 4.6: Histograma da Sombra Acústica (Azul - Tumores Benignos e Vermelho - Tumores Malignos)

1. Distância de Kullback-Leiber;
2. Distância de Kullback-Leiber não-extensiva.

Depois de extrairmos os seis vetores base em forma de distribuição probabilística para as três características analisadas, utilizamos as variações da distância de Kullback-Leiber para classificar cada imagem em relação ao vetor base.

O desempenho do classificador depende principalmente das características extraídas, pois quanto maior a variação entre os grupos; isto é, quanto maior a diferença entre as imagens de tumores malignos e benignos, melhor será o desempenho do classificador.

Antes dos resultados de cada classificador, a seguir apresentamos a metodologia para comparar os dois vetores (vetor base e vetor extraído da imagem a ser classificada).

4.4 Metodologia de comparação de vetores

A idéia geral da dissertação é evitar usar informações de intensidade absoluta das variáveis em troca do uso da frequência com que essas variáveis ocorrem. Uma maneira de conseguir isso é através da comparação das distribuições das probabilidades das variáveis ao invés da comparação de suas intensidades, por exemplo: sejam os vetores de características $\partial = [f_1, f_2, f_3, \dots, f_n]$ e $\partial' = [f'_1, f'_2, f'_3, \dots, f'_n]$ que desejamos comparar. Cada variável f_i e f'_i em nosso trabalho é agora uma função de distribuição de probabilidade. Sendo assim, é possível comparar f_i e f'_i usando a

teoria da informação, através do cálculo da divergência, ou entropia relativa:

$$D(p_\alpha(f_i)||p_\beta(f'_i)) = \sum_{i=1}^n p_\alpha(f_i) * \log \frac{p_\alpha(f_i)}{p_\beta(f'_i)} \quad (4.1)$$

onde $p_\alpha(f_i)$ e $p_\beta(f'_i)$ são as respectivas distribuições de probabilidades das características f_i e f'_i .

Sendo o sistema proposto multi-variado, é possível usar a Equação (4.1) em sua forma conjunta:

$$D(p_\alpha(f_1, f_2, \dots, f_n)||p_\beta(f'_1, f'_2, \dots, f'_n)) = \sum_{i=1}^n \dots \sum_{i=1}^n p_\alpha(f_1, f_2, \dots, f_n) * \log \frac{p_\alpha(f_1, f_2, \dots, f_n)}{p_\beta(f'_1, f'_2, \dots, f'_n)} \quad (4.2)$$

Quando comparamos duas distribuições probabilísticas utilizamos a Equação (4.1). Quando o sistema é multi-variado com distribuição probabilística interdependente entre elas, utilizamos a forma conjunta dada pela Equação (4.2). Para desenvolver o cálculo entre dois vetores, utilizamos a regra da cadeia aplicada na Equação (4.2) e que representa um algoritmo de complexidade exponencial $O(m^n)$, onde n é a dimensão do espaço vetorial e m é o número de discretizações das funções p_α e p_β .

Para o método de classificação que estamos propondo, esse problema é contornado de duas maneiras:

- Considerando que os trabalhos de Thomaz et al. (2008) e Giraldi et al. (2008) mostraram que o espaço vetorial de características para a nossa base é no máximo de ordem 3 a complexidade do algoritmo indexado pela Equação (4.2) é no máximo $O(m^3)$.
- Podemos determinar uma outra fórmula, diferente da regra da cadeia para calcular a divergência entre as 3 características. Propomos utilizar a média das diferenças entre as características extraídas nos vetores base (θ) e os vetores extraídos das imagens a serem classificadas (θ').

$$D(\theta|\theta') = \frac{\sum_{i=1}^n D_A((f_i)||f'_i) + \sum_{i=1}^n D_S((f_i)||f'_i) + \sum_{i=1}^n D_R((f_i)||f'_i)}{3} \quad (4.3)$$

O uso da teoria da informação, através da entropia relativa para classificação de nossos vetores, ainda não usados na literatura, trás uma nova perspectiva e abre caminho para a aplicação de ferramenta matemático bem estabelecido, dado pela entropia não-extensiva Tsalliana. Assim, uma outra contribuição da dissertação é o estudo da classificação do câncer de mama do ponto de vista não-extensivo.

4.4.1 Classificação por distância de Kullback-Leiber

O primeiro método de classificação proposto foi executado utilizando distância de Kullback-Leiber clássica, também conhecida como entropia relativa. Assim, comparamos a distância entre os vetores base e a informação extraída estatisticamente de cada imagem a ser classificada.

Cada vetor de característica a ser analisado foi comparado com o vetor base maligno e o vetor base benigno extraídos inicialmente. Classificamos a imagem analisada no grupo em que a distância apresentasse o menor valor.

Para analisar o desempenho do classificador, apresentamos primeiramente os resultados na Tabela 4.1.

Tabela 4.1: Resultados da classificação utilizando entropia relativa

	Hist. da área	Hist Sombra Acústica	Hist. Raios
VP	56	78	62
VN	93	84	74
FP	57	66	76
FN	44	22	38
TVP	0,56	0,78	0,62
TFP	0,38	0,44	0,51
TVP/TFP	1,47	1,77	1,22

Como o método proposto nessa tese para classificação das imagens não possui um limiar para traçarmos a curva ROC, utilizamos apenas para avaliar os resultados calculados a razão entre TVP

e TFP, que demonstra o melhor desempenho para o histograma da sombra acústica com 1,77.

Para uma melhor comparação dos resultados, utilizamos o protocolo de validação cruzada onde podemos analisar a acuracidade, sensibilidade, especificidade, valor predito positivo e valor predito negativo do sistema. Na Tabela 4.2 podemos verificar que as informações extraídas da área do tumor tiveram o maior valor para especificidade com 62%, a sombra acústica apresentou maior resultado para acuracidade, sensibilidade, VPP e VPN com 65%, 78% 54% e 79% respectivamente.

Tabela 4.2: Classificação utilizando entropia relativa

	Área	Sombra Acústica	Raios
Acuracidade	0,60	0,65	0,54
Sensibilidade	0,56	0,78	0,62
Especificidade	0,62	0,56	0,49
Valor Predito Positivo	0,50	0,54	0,45
Valor Predito Negativo	0,68	0,79	0,66

Analisando as Tabelas 4.1 e 4.2 podemos concluir que a utilização da distância de Kullback-Leiber para classificação utilizando a característica do histograma da sombra acústica apresentou os melhores resultados perdendo apenas na especificidade para o histograma da área. Observando a razão TVP/TFP o histograma do raio apresentou o pior desempenho na classificação.

4.4.2 Classificação por distância de Kullback-Leiber usando características híbridas

Utilizando a Equação (4.3) para classificar as imagens considerando as 3 características ao mesmo tempo, obtivemos os resultados apresentados na Tabela 4.3. Podemos concluir que, como apresentado no trabalho de Rodrigues et al.(2006) a utilização de um sistema híbrido resulta em um desempenho melhor do classificador pois, a razão de TVP e TFP é maior (2,85) que a apresentada individualmente.

Na avaliação dos itens do protocolo de validação cruzada apresentados na Tabela 4.4 também podemos notar a melhora, pois a acuracidade, a especificidade e o VPP também tiveram maiores resultados comparado aos resultados individuais.

Tabela 4.3: Resultados da classificação utilizando entropia relativa usando características híbridas

	Híbrido
VP	59
VN	119
FP	31
FN	41
TVP	0,59
TFP	0,21
TVP/TFP	2,85

Tabela 4.4: Classificação utilizando entropia relativa usando características híbridas

	Híbrido
Acuracidade	0,71
Sensibilidade	0,59
Especificidade	0,79
Valor Predito Positivo	0,66
Valor Predito Negativo	0,74

4.4.3 Classificação por distância de Kullback-Leiber não-extensiva

O segundo método de classificação proposto foi executado utilizando distância de Kullback-Leiber não-extensiva. O resultado apresentado com esta técnica depende do valor de q utilizado. Assim, comparamos a distância entre os vetores base e a informação extraída estatisticamente de cada imagem a ser classificada variando o valor de q no intervalo de 0,1 a 3.

A Tabela 4.5 apresenta todos os resultados das variações de q e podemos observar que para q igual a 1,01 os resultados são próximos aos classificados com distância de Kullback-Leiber ou entropia relativa. Quando q é igual a 1 a entropia de Tsallis se iguala a entropia relativa.

Os melhores resultados para a classificação utilizando o histograma da área estão acima de $q = 1$; isso significa que as frequências mais baixas estão influenciando positivamente nos resultados.

Para avaliarmos o desempenho do classificador utilizando o protocolo de validação cruzada, geramos a Tabela 4.6 que avalia a acuracidade, sensibilidade, especificidade, VPP e VPN variando

o valor de q . Quando q está entre 0,1 e 0,2 o valor de sensibilidade e VPN são os maiores.

Tabela 4.5: Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para o histograma da área do tumor

q	FP	FN	VP	VN	TVP	TFP	TVP/TFP
0,10	142	1	99	8	0,99	0,95	1,05
0,20	142	1	99	8	0,99	0,95	1,05
0,30	140	2	98	10	0,98	0,93	1,05
0,40	137	12	88	13	0,88	0,91	0,96
0,50	132	23	77	18	0,77	0,88	0,88
0,60	124	29	71	26	0,71	0,83	0,86
0,70	114	39	61	36	0,61	0,76	0,80
0,80	106	48	52	44	0,52	0,71	0,74
0,90	100	44	56	50	0,56	0,67	0,84
0,99	93	56	44	57	0,44	0,62	0,71
1,01	57	44	56	93	0,56	0,38	1,47
1,10	56	48	52	94	0,52	0,37	1,39
1,20	61	46	54	89	0,54	0,41	1,33
1,30	60	47	53	90	0,53	0,40	1,33
1,40	58	51	49	92	0,49	0,39	1,27
1,50	58	54	46	92	0,46	0,39	1,19
1,60	56	56	44	94	0,44	0,37	1,18
1,70	56	58	42	94	0,42	0,37	1,13
1,80	54	61	39	96	0,39	0,36	1,08
1,90	50	63	37	100	0,37	0,33	1,11
2,00	44	64	36	106	0,36	0,29	1,23
3,00	42	70	30	108	0,30	0,28	1,07

Os resultados para o histograma da sombra acústica estão na Tabela 4.7, de maneira análoga à validação do histograma da área, os valores para q igual 1,01 também são idênticos aos obtidos com o classificador por entropia relativa. Observamos que a razão entre TVP e TFP aumenta para os maiores valores de q .

Para verificarmos o desempenho deste classificador com o histograma da sombra acústica, a Tabela 4.8 foi criada e apresenta o melhor resultado de acuracidade para q igual a 2 e 2,5 com 66%, sensibilidade com q igual a 0,1 e 0,2 resultou em 88%, especificidade com q igual a 2,5 teve 61%, VPP e VPN com 56% e 78% respectivamente com q igual a 2.

A Tabela 4.9 apresenta os resultados para o histograma de raios. Podemos verificar que, para

Tabela 4.6: Classificação utilizando distância de Kullback-Leiber não-extensiva para o histograma da área do tumor

q	Acuracidade	Sensibilidade	Especificidade	VPP	VPN
0,10	0,43	0,99	0,05	0,41	0,89
0,20	0,43	0,99	0,05	0,41	0,89
0,30	0,43	0,98	0,07	0,41	0,83
0,40	0,40	0,88	0,09	0,39	0,52
0,50	0,38	0,77	0,12	0,37	0,44
0,60	0,39	0,71	0,17	0,36	0,47
0,70	0,39	0,61	0,24	0,35	0,48
0,80	0,38	0,52	0,29	0,33	0,48
0,90	0,42	0,56	0,33	0,36	0,53
0,99	0,40	0,44	0,38	0,32	0,50
1,01	0,60	0,56	0,62	0,50	0,68
1,10	0,58	0,52	0,63	0,48	0,66
1,20	0,57	0,54	0,59	0,47	0,66
1,30	0,57	0,53	0,60	0,47	0,66
1,40	0,56	0,49	0,61	0,46	0,64
1,50	0,55	0,46	0,61	0,44	0,63
1,60	0,55	0,44	0,63	0,44	0,63
1,70	0,54	0,42	0,63	0,43	0,62
1,80	0,54	0,39	0,64	0,42	0,61
1,90	0,55	0,37	0,67	0,43	0,61
2,00	0,57	0,36	0,71	0,45	0,62
3,00	0,55	0,30	0,72	0,42	0,61

valores de q maiores que 1,01 aumenta a razão TVP/TFP consideravelmente comparado às outras características.

O desempenho do classificador com o histograma de raios está na Tabela 4.10. Podemos observar nos resultados que para q igual a 1,5 os melhores valores de acuracidade, sensibilidade e VPN são 91%, 94% e 96% respectivamente. Quando q está entre 2 e 3 os valores de especificidade e VPP são 91% e 86%.

4.4.4 Classificação por distância de Kullback-Leiber não-extensiva usando características híbridas

Utilizando a Equação (4.3) para classificar as imagens considerando as 3 características ao mesmo tempo, obtivemos os resultados apresentados na Tabela 4.11. O desempenho do classifi-

Tabela 4.7: Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para o histograma da sombra acústica

q	FP	FN	VP	VN	TVP	TFP	TVP/TFP
0,10	143	12	88	7	0,88	0,95	0,92
0,20	140	12	88	10	0,88	0,93	0,94
0,30	137	19	81	13	0,81	0,91	0,89
0,40	133	34	66	17	0,66	0,89	0,74
0,50	122	49	51	28	0,51	0,81	0,63
0,60	108	65	35	42	0,35	0,72	0,49
0,70	92	73	27	58	0,27	0,61	0,44
0,80	90	77	23	60	0,23	0,60	0,38
0,90	88	77	23	62	0,23	0,59	0,39
0,99	84	78	22	66	0,22	0,56	0,39
1,01	66	22	78	84	0,78	0,44	1,77
1,10	66	22	78	84	0,78	0,44	1,77
1,20	66	22	78	84	0,78	0,44	1,77
1,30	68	22	78	82	0,78	0,45	1,72
1,40	68	22	78	82	0,78	0,45	1,72
1,50	70	23	77	80	0,77	0,47	1,65
1,60	65	23	77	85	0,77	0,43	1,78
1,70	64	23	77	86	0,77	0,43	1,80
1,80	64	25	75	86	0,75	0,43	1,76
1,90	61	25	75	89	0,75	0,41	1,84
2,00	60	25	75	90	0,75	0,40	1,88
3,00	58	28	72	92	0,72	0,39	1,86

cadador estabiliza em seu melhor resultado após q chegar ao valor de 1,01 em 2,85.

A Tabela 4.12 apresenta o protocolo de validação cruzada para as informações híbridas e, conforme os resultados obtidos na Tabela 4.11, os valores também são estabilizados após q chegar em 1,01.

Os desempenhos desse classificador para q igual a 1,5 foi de 56%, 45% e 64% para acuracidade, VPP e VPN respectivamente. Para sensibilidade com q igual a 0,1 temos 88% e para especificidade utilizando q igual a 0,9 ou a 3,0 temos 73%.

Tabela 4.8: Classificação utilizando distância de Kullback-Leiber não-extensiva para o histograma da sombra acústica

q	Acuracidade	Sensibilidade	Especificidade	VPP	VPN
0,10	0,38	0,88	0,05	0,38	0,37
0,20	0,39	0,88	0,07	0,39	0,45
0,30	0,38	0,81	0,07	0,37	0,41
0,40	0,33	0,66	0,09	0,33	0,33
0,50	0,32	0,51	0,12	0,29	0,36
0,60	0,31	0,35	0,17	0,24	0,39
0,70	0,34	0,27	0,24	0,23	0,44
0,80	0,33	0,23	0,29	0,20	0,44
0,90	0,34	0,23	0,33	0,21	0,45
0,99	0,35	0,22	0,44	0,21	0,46
1,01	0,65	0,78	0,56	0,54	0,79
1,10	0,65	0,78	0,56	0,54	0,79
1,20	0,65	0,78	0,56	0,54	0,79
1,30	0,64	0,78	0,55	0,53	0,79
1,40	0,64	0,78	0,55	0,53	0,79
1,50	0,63	0,77	0,53	0,52	0,78
1,60	0,65	0,77	0,57	0,54	0,79
1,70	0,65	0,77	0,57	0,55	0,79
1,80	0,64	0,75	0,57	0,54	0,77
1,90	0,66	0,75	0,59	0,55	0,78
2,00	0,66	0,75	0,60	0,56	0,78
3,00	0,66	0,72	0,61	0,55	0,77

Tabela 4.9: Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para o histograma de raios.

q	FP	FN	VP	VN	TVP	TFP	TVP/TFP
0,10	130	86	14	20	0,14	0,87	0,16
0,20	130	88	12	20	0,12	0,87	0,14
0,30	126	93	7	24	0,07	0,84	0,08
0,40	126	95	5	24	0,05	0,84	0,06
0,50	126	95	5	24	0,05	0,84	0,06
0,60	125	95	5	25	0,05	0,83	0,06
0,70	131	94	6	19	0,06	0,87	0,07
0,80	131	94	6	19	0,06	0,87	0,07
0,90	74	62	38	76	0,38	0,49	0,77
0,99	131	94	6	19	0,06	0,87	0,07
1,01	76	38	62	74	0,62	0,51	1,22
1,10	20	6	94	130	0,94	0,13	7,05
1,20	19	6	94	131	0,94	0,13	7,42
1,30	18	6	94	132	0,94	0,12	7,83
1,40	17	6	94	133	0,94	0,11	8,29
1,50	16	6	94	134	0,94	0,11	8,81
1,60	16	7	93	134	0,93	0,11	8,72
1,70	15	8	92	135	0,92	0,10	9,20
1,80	14	11	89	136	0,89	0,09	9,54
1,90	14	12	88	136	0,88	0,09	9,43
2,00	14	13	87	136	0,87	0,09	9,32
3,00	14	14	86	136	0,86	0,09	9,21

Tabela 4.10: Classificação utilizando distância de Kullback-Leiber não-extensiva para o histograma de raios

q	Acuracidade	Sensibilidade	Especificidade	VPP	VPN
0,10	0,14	0,14	0,13	0,10	0,19
0,20	0,13	0,12	0,13	0,08	0,19
0,30	0,12	0,07	0,07	0,05	0,21
0,40	0,12	0,05	0,09	0,04	0,20
0,50	0,12	0,05	0,12	0,04	0,20
0,60	0,12	0,05	0,17	0,04	0,21
0,70	0,10	0,06	0,24	0,04	0,17
0,80	0,10	0,06	0,29	0,04	0,17
0,90	0,46	0,38	0,33	0,34	0,55
0,99	0,10	0,06	0,13	0,04	0,17
1,01	0,54	0,62	0,49	0,45	0,66
1,10	0,90	0,94	0,87	0,82	0,96
1,20	0,90	0,94	0,87	0,83	0,96
1,30	0,90	0,94	0,88	0,84	0,96
1,40	0,91	0,94	0,89	0,85	0,96
1,50	0,91	0,94	0,89	0,85	0,96
1,60	0,91	0,93	0,89	0,85	0,95
1,70	0,91	0,92	0,90	0,86	0,94
1,80	0,90	0,89	0,91	0,86	0,93
1,90	0,90	0,88	0,91	0,86	0,92
2,00	0,89	0,87	0,91	0,86	0,91
3,00	0,89	0,86	0,91	0,86	0,91

Tabela 4.11: Resultados do classificador utilizando distância de Kullback-Leiber não-extensiva para características híbridas.

q	FP	FN	VP	VN	TVP	TFP	TVP/TFP
0,10	112	59	41	38	0,41	0,75	0,55
0,20	112	59	41	38	0,41	0,75	0,55
0,30	112	59	41	38	0,41	0,75	0,55
0,40	112	59	41	38	0,41	0,75	0,55
0,50	112	59	41	38	0,41	0,75	0,55
0,60	112	59	41	38	0,41	0,75	0,55
0,70	112	59	41	38	0,41	0,75	0,55
0,80	112	59	41	38	0,41	0,75	0,55
0,90	112	59	41	38	0,41	0,75	0,55
0,99	112	59	41	38	0,41	0,75	0,55
1,01	31	41	59	119	0,59	0,21	2,85
1,10	31	41	59	119	0,59	0,21	2,85
1,20	31	41	59	119	0,59	0,21	2,85
1,30	31	41	59	119	0,59	0,21	2,85
1,40	31	41	59	119	0,59	0,21	2,85
1,50	31	41	59	119	0,59	0,21	2,85
1,60	31	41	59	119	0,59	0,21	2,85
1,70	31	41	59	119	0,59	0,21	2,85
1,80	31	41	59	119	0,59	0,21	2,85
1,90	31	41	59	119	0,59	0,21	2,85
2,00	31	41	59	119	0,59	0,21	2,85
3,00	31	41	59	119	0,59	0,21	2,85

Tabela 4.12: Classificação utilizando distância de Kullback-Leiber não-extensiva para características híbridas

q	Acuracidade	Sensibilidade	Especificidade	VPP	VPN
0,10	0,32	0,41	0,25	0,27	0,39
0,20	0,32	0,41	0,25	0,27	0,39
0,30	0,32	0,41	0,25	0,27	0,39
0,40	0,32	0,41	0,25	0,27	0,39
0,50	0,32	0,41	0,25	0,27	0,39
0,60	0,32	0,41	0,25	0,27	0,39
0,70	0,32	0,41	0,25	0,27	0,39
0,80	0,32	0,41	0,25	0,27	0,39
0,90	0,32	0,41	0,25	0,27	0,39
0,99	0,32	0,41	0,25	0,27	0,39
1,01	0,71	0,59	0,79	0,66	0,74
1,10	0,71	0,59	0,79	0,66	0,74
1,20	0,71	0,59	0,79	0,66	0,74
1,30	0,71	0,59	0,79	0,66	0,74
1,40	0,71	0,59	0,79	0,66	0,74
1,50	0,71	0,59	0,79	0,66	0,74
1,60	0,71	0,59	0,79	0,66	0,74
1,70	0,71	0,59	0,79	0,66	0,74
1,80	0,71	0,59	0,79	0,66	0,74
1,90	0,71	0,59	0,79	0,66	0,74
2,00	0,71	0,59	0,79	0,66	0,74
3,00	0,71	0,59	0,79	0,66	0,74

Capítulo 5

Conclusões

Este trabalho teve como objetivo o estudo de classificadores de câncer de mama em imagens de ultrassom do ponto de vista da teoria de informação, incluindo informação mútua, entropia e entropia relativa, estendendo os estudos para a entropia Tsalliana, utilizando um sistema de características híbridas para posterior comparação com os resultados obtidos por (Rodrigues et al., 2006c), bem como (Thomaz et al., 2008) e (Giraldi et al., 2008) com relação às características discriminantes mínimas.

O método proposto para a classificação dos tumores utiliza o conceito de entropia para classificar as imagens que foram utilizadas na pesquisa de (Rodrigues et al., 2006c). Para a classificação, a entropia relativa, ou distância de Kullback-Leiber e a Kullback-Leiber não-extensiva, são utilizadas como métrica de similaridade entre os vetores de características extraídas.

Os experimentos executados neste trabalho mostraram que, seguindo a metodologia de classificação utilizada na pesquisa de (Rodrigues et al., 2006c) obtem-se um desempenho superior, quando comparado ao classificador utilizando os conceitos de Teoria da Informação. Não foi possível traçar a curva ROC pois nossa metodologia não possui um limiar. Assim, ao longo do trabalho, comparamos os resultados da razão entre TVP e TFP.

Os resultados dos sistemas híbridos para o classificador utilizando entropia relativa foi melhor que o individual. Para o classificador utilizando entropia de Tsallis o sistema híbrido não superou os melhores resultados do histograma de raios. Como os maiores desempenhos foram para os

q 's acima de 1,01 podemos concluir que as informações com pequenas frequências possuem uma influência positiva no resultado do classificador.

Referências Bibliográficas

- Albuquerque, M.; Esquef, I. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, 25:1056–1065, 2004.
- Bisognin, G. Utilização de máquinas de suporte vetorial para predicação de estruturas terciárias de proteínas. *Universidade do Vale do Rio dos Sinos*, 2007.
- Boltzmann, W. On the relation between the second law of mechanical theory of heat and probability calculus with respect to the theorems on thermal equilibrium. *University Vienna*, 1877.
- Borland, L.; A., Plastino; C., Tsallis. Information gain within nonextensive thermostatics. *Journal of Mathematical Physics*, vol 39, n 12, 1998.
- Braga, A. e Carvalho, A. e Ludermir T. Redes neurais artificiais. 1 ed. Vol 1 Cap 6 p 141-168, 2003.
- Cabella, B. Inferência estatística em métodos de análise de ressonância magnética funcional. *Universidade de São Paulo*, 2008.
- Carvalho, D. Árvore de decisão / algoritmo genético para tratar o problema de pequenos disjuntos em classificação de dados. *Universidade Federal do Rio de Janeiro*, 2005.
- Chung, M.; Yi, H.; Ko, C.; Guo, S.; Chui, M.; Hong, J.; See, Y. Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology*, 2003.
- Cover, T. e Thomas J. Elements of information theory. *Wiley-Interscience*, 1991.
- Dar, R.; Ruey, F.; YU, L. Computer-aided diagnosis applied to us of solid breast nodules by using neural networks. *Computer Applications. Radiology*, 1999.

- Dar, R.; Wen, J.; Ruey, F.; Moon, W.; Lee, C. Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound. *World Federation for Ultrasound in Medicine and Biology. USA*, 2002.
- Dar, R.; Ruey, F.; Wei, M.; Woo, K. Computer-aided diagnosis for 3-dimensional breast ultrasonography. *American Medical Association. USA.*, 2006.
- Diniz, P. Segmentação de tecidos cerebrais usando entropia q em imagens de ressonância magnética de pacientes com esclerose múltipla. *Faculdade de Medicina de Ribeirão Preto*, 2008.
- Dunn, G.; Everitt, B. Clinical biostatistics an introduction to evidencebased medicine. *London Edward Arnold*, 1995.
- Entropy. A brief history of entropy. <http://www.geocities.com/chemforum/entropyhistory.htm>, 2008.
- Erdmann, H. Detecção e classificação de imagens sintéticas utilizando entropia não-extensiva. *Centro Universitário da FEI*, 2009.
- Esquef, I. Técnicas de entropia em processamento de imagens. *Centro Brasileiro de Pesquisas Físicas*, 2002.
- Fleiss, J. Statistical methods for rates and proportions. *2ed New York John Wiley*, 1981.
- Giraldi, G.; Rodrigues, P.; Thomaz, C.; Kitani, E. Statistical learning approaches for discriminate features selection. *Journal of the Brazilian Computer Society*, 20:1–16, 2008.
- Gonzalez, R.; Woods, E. *Digital Image Processing Using Matlab*. 1992.
- Herbrich, R. Learning kernel classifiers: Theory and algorithms. *MIT Press*, 2001.
- Jacques, C. Observando asteróides/cometas com a técnica track and stack. *REA/Brasil/MG*, 2003.
- Jain, A.K. Data clustering: 50 years beyond k-means. *Technical Report MSU-CSE-09-11, Department of Computer Science, Michigan State University, East Lansing, Michigan*, 2009.
- Linnet, K. Comparison of quantitative diagnostic tests: Type i error, power, and sample size. *Statistics in Medicine*, 1987.

- Margotto, P. Curva roc como fazer e interpretar no spss. *Escola Superior de Ciências da Saúde (ESCS)/SES/DF*, 2010.
- Milare, C. Extração de conhecimento de redes neurais artificiais utilizando sistemas de aprendizado simbólico e algoritmos genéticos. *Instituto de Ciências Matemáticas e de Computação - ICMC-USP*, 2003.
- Mitchell, T. Machine learning. *New York: McGraw-Hill*, 1997.
- Monard, M.; A., Baranauskas. Indução de regras e Árvores de decisão. *Sistemas Inteligentes, Editora Manole Ltda.*, 2003.
- Norton, M. Knowledge discovery in databases. *Library Trends*, v. 48, n. 1, p. 9-21, 1999.
- Oguri, P. Aprendizado de máquina para o problema de sentiment classification. *Tese de mestrado - Pontifícia Universidade Católica-Rio*, 2006.
- Quinlan, J. Simplifying decision trees. *International Journal of Man-Machine Studies*, 1987.
- Reeves, A.; Kostis, W. Computer-aided diagnosis for lung cancer. *Radiologic Clinics of North America*, 32:964–974, 2000.
- Rodrigues, P; Giraldi, G. Computing the q-index for tsallis non-extensive image segmentation. *In SBC ed XXIIInd Brazilian Symposium on Computer Graphics and Image Processing (Sibgrapi 2009) vol To Appear*, 2009a.
- Rodrigues, P; Giraldi, G. Improving the non-extensive tsallis non-extensive medical image segmentation based on tsallis entropy. *Pattern Analysis and Application, Submitted*, 2009b.
- Rodrigues, P.; Giraldi, G.; Chang, R.; Suri, J. Automatic classification of breast lesions in 3d ultrasound images. *J. Suri. (Org.). Biomedical imaging and Bioinformatics*, 2006a.
- Rodrigues, P.; Giraldi, G.; Chang, R.; Suri, J. Automatic classification of breast lesions in 3d ultrasound images. *J. Suri. (Org.). Biomedical imaging and Bioinformatics*, 2006b.
- Rodrigues, P.; Giraldi, G.; Provenzano, M.; Faria, M.; Chang, R.; Suri, J. A new methodology for breast lesion classification in 3-d ultrasound: A non-extensive entropy-based algorithm. 2006c.

- Rodrigues, P.; Giraldi, G.; Suri, J.; Chang, R. Non-extensive entropy for cad systems for breast cancer images. *International Symposium on Computer Graphics and Image processing, IEEE Computer Society*, 2006d.
- Rosenblatt, F. Principles of neurodynamics. *Spartan Books - p 31*, 1962.
- Ruey, F.; WEN, J.; MOON, W.; Dar, R. Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. *World Federation for Ultrasound in Medicine and Biology. USA*, 2002.
- Rumelhart, D.; Hilton, G.; Williams, R. Learning internal representations by error propagation. in parallel distributed processing: Explorations in the microstructure of cognition. *Vol 1, Cambridge, MA: MIT Press*, 1986.
- Sezgin, M.; Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 2004.
- Shannon, C. A mathematical theory of communication. *The Bell System Technical Journal*, 1948.
- Silva, A.; Carvalho, P. Sistema de análise de nódulo pulmonar. *Pontifícia Universidade Católica do Rio de Janeiro - PUC - RJ*, 2002.
- Silva, E. e Muramatsu M. O fenômeno do speckle como introdução à metrologia óptica no laboratório didático. *Revista Brasileira de Ensino de Física*, v. 29, n. 2, p. 283-286, 2007.
- Silva, R. Redes neurais artificiais aplicadas à detecção de intrusão em redes tcp/ip. *Tese de mestrado - Pontifícia Universidade Católica-Rio*, 2005.
- Sobral, A. Previsão de carga horária - uma nova abordagem por árvore de decisão. *Tese de doutorado - Pontifícia Universidade Católica-Rio*, 2003.
- Sturzbecher, M. Detecção e caracterização da resposta hemodinâmica pelo desenvolvimento de novos métodos de processamento de imagens funcionais por ressonância magnética. *Universidade de São Paulo*, 2006.
- Tavares, A. Aspectos matemáticos da entropia. *Universidade de Aveiro - Depto de Matemática*, 2003.

- Thomaz, C.; Giraldi, G.; Rodrigues, P. Dimensionality reduction, classification and reconstruction problems in statistical learning approaches. *Revista de Informática Teórica e Aplicada*, 15: 141–173, 2008.
- Tsallis, C. Non-extensive statistical mechanics and its applications. *Series Lecture Notes in Physics, Berlin*, 2001.
- Vaz, J. Regiões de incerteza para a curva roc em testes diagnósticos. *Tese de mestrado - Universidade Federal de São Carlos*, 2009.
- Wei, M.; Ruey, F.; Woo, K.; Dar, R. Breast cancer diagnosis using three-dimensional ultrasound and pixel relation analysis. *World Federation for Ultrasound in Medicine and Biology. USA.*, 2003.
- Wei, M.; Ruey, F.; Shou, J.; Cheng, S.; Moon, W.; Shou, T.; Dar, R. 3-d ultrasound texture classification using run difference matrix. *World Federation for Ultrasound in Medicine and Biology. USA*, 2005.
- Weiss, S.; Kulikowski, C. Computer systems that learn. classification and prediction methods from statistics, neural nets, machine learning, and expert systems. *San Mateo, CA: Morgan Kaufmann*, 1991.
- Wen, J.; Ruey, F.; Woo, K.; Lee, C.; Dar, R. Computer-aided diagnosis of breast tumors with different us systems. *Association of University Radiologist. USA.*, 2002.
- Yankeleitz, D. Small pulmonary nodules: Volumetrically determined growth rates based on ct evaluation. *Radiology*, 217:251–256, 2000.

Glossário

Acuracidade	— A probabilidade do teste que está sendo investigado fornecer um resultado correto
Background	— Segundo plano da imagem
Backpropagation	— Método de ensino de Redes Neurais
Binarização	— Ato de separar uma imagem em duas classes.
Bit	— Simplificação para dígito binário, "BInary digiT", é a menor unidade de informação que pode ser armazenada ou transmitida.
Entropia	— Grandeza associada ao "grau de desordem"
Especificidade	— A probabilidade do teste que está sendo investigado fornecer um resultado negativo, dado que o indivíduo não é portador da doença
Foreground	— Primeiro plano da imagem
frames	— Resultado da divisão de uma imagem em subimagens
Histograma	— É um gráfico que apresenta o número de pontos por cada nível de cinza da imagem
Imagem Digital	— É a representação de uma imagem bidimensional usando números binários
Kernel	— Função que recebe dois pontos e realiza o produto escalar
k-means	— Método de análise de cluster
Limiar	— Criar um limite para dividir um histograma em pixels pretos e brancos
Luminância	— Distribuição estatística dos pixels de uma imagem
Pixel	— Unidade básica de amostra de uma imagem.
Polinômio	— São séries de monômios (ou termos), que por sua vez são expressões matemáticas.
Pseudo-aditividade	— Propriedade - mostra que a entropia de Tsallis não é aditiva
Segmentação	— Ato de separar por característica.
Sensibilidade	— A probabilidade do teste que está sendo investigado fornecer um resultado positivo, dado que o indivíduo é portador da doença
Termodinâmica	— É o ramo da Física que estuda os efeitos da mudança em temperatura, pressão e volume em sistemas físicos na escala macroscópica