



**Hugo Filipe de Carvalho da Palma Medeiros**

Licenciatura em Ciências de Engenharia Biomédica

**Classificação de Tumores de Cancro na  
Mama Através de Radar de Banda Ultra-  
Larga de Microondas**

Dissertação para obtenção do Grau de Mestre em  
Engenharia Biomédica

Orientador: Raquel Conceição, Doutora, Faculdade  
de Ciências da Universidade de Lisboa

Co-orientador: Pedro Almeida, Professor Doutor,  
Faculdade de Ciências da Universidade de Lisboa

Júri:

Presidente: Prof. Doutor Mário António Basto Forjaz Secca

Arguente: Prof. Doutor Hugo Filipe Silveira Gamboa

Vogais: Prof. Doutora Raquel Cruz da Conceição

Prof. Doutor Pedro Miguel Dinis de Almeida



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Outubro 2013**





**Hugo Filipe de Carvalho da Palma Medeiros**

Licenciatura em Ciências de Engenharia Biomédica

**Departamento de Física**

## **Classificação de Tumores de Cancro na Mama Através de Radar de Banda Ultra- Larga de Microondas**

Dissertação para obtenção do Grau de Mestre em  
Engenharia Biomédica

**Dissertação apresentada na Faculdade de Ciências e  
Tecnologia da Universidade Nova de Lisboa para  
obtenção de Grau de Mestre em Engenharia Biomédica.**

**A presente dissertação foi desenvolvida com a  
colaboração do Instituto de Biofísica e Engenharia  
Biomédica da Faculdade de Ciências da Universidade de  
Lisboa.**

**Orientador: Doutora Raquel Conceição**

**Co-orientador: Professor Doutor Pedro Almeida**



# Copyright

## **Classificação de Tumores de Cancro na Mama Através de Radar de Banda Ultra-Larga de Microondas**

Copyright© 2013- Todos os direitos reservados. Hugo Filipe de Carvalho da Palma Medeiros. Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa tem o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



# Agradecimentos

Os meus agradecimentos são para as pessoas que me acompanharam não só durante os últimos meses mas sim desde o início do curso. Perdoem a falta de palavras, escrever não é o meu forte como sabem.

Em primeiro lugar quero agradecer à Raquel Conceição por toda a ajuda dada ao longo destes meses e todos os conselhos e correcções detalhadas feitas a esta dissertação. Sem dúvida não poderia pedir alguém melhor como orientadora, obrigado!

Agradeço também ao Professor Pedro Almeida pela oportunidade dada de trabalhar no Instituto de Biofísica e Engenharia Biomédica.

Aos meus companheiros do IBEB com quem passei bons momentos e fizeram com que o trabalho fosse mais fácil de fazer: Carolina, Sara, Joana, Ana Ruxa e Guilherme (ainda deves um bolo ao pessoal).

Quero agradecer ao Professor Mário Secca por todo o esforço e empenho dado ao curso de Engenharia Biomédica.

A todas as pessoas e colegas que conheci nestes anos de faculdade. Um especial agradecimento àqueles com quem passei noites de estudo na faculdade recheadas com momentos de loucura geniais, me diverti nas festas académicas, fiz churrascadas, joguei horas e horas de PES, fiz viagens a Almeirim só para comer uma sopa da pedra, passei horas na esplanada da Teresa, me ri e passei momentos inesquecíveis: Rita, Zé, Marcelo, Rui, Pica, Sérgio, Açoriano, Semi, Bárbara, Mónica, Carla, Teresa, David, Quim, Leo, Heavy, Nuno, Hugo, Alentejano, Franjas. Obrigado pela vossa amizade! E claro ao Bi com quem tive o prazer de morar durante 3 belos anos.

Ao Marcelo por tudo o que passámos juntos durante um ano incrível em Itália. Como é que depois de tanto tempo, tantas aventuras e tantas histórias nunca nos fartámos um do outro?

A todos os amigos que conheci em Erasmus e que partilharam comigo aquele que, foi e será para sempre, certamente um dos melhores anos da minha vida. Não vou mencionar nomes porque me esqueceria certamente de alguém mas um especial obrigado aos portugueses, brasileiros, espanhóis e todos os da América Latina.

Aos elementos d'A Banda: André, França, Catarino, Cabecinha, Beirão e Beto por todas as actuações feitas por esse mundo fora e todo o apoio dado.

À Tata e Joana por contribuírem com a vossa boa disposição para o bom ambiente sempre vivido na Fleming.

Ao Raposo, Fitas, Jere, Marujo e Sori por serem aqueles amigos da há muitos anos com quem sei que posso sempre contar. Não só foram importantes nestes últimos 5 anos, como em todos os anteriores e seguramente nos que virão.

À Ana Rita Pereira por ser a pessoa maravilhosa que é e que esteve sempre do meu lado ao longo dos últimos 5 anos. Sem ti não era a mesma coisa.

Por fim, um especial agradecimento aos meus pais, irmão, avós, tios e primos por todo o apoio dado ao longo da minha vida e por me ajudarem a ser a pessoa que sou hoje. Sem dúvida que sem vocês, eu não estaria neste momento a escrever esta dissertação. Obrigado!



# Resumo

A imagem por microondas é uma das técnicas mais promissoras de imagem médica para detecção e classificação do cancro da mama, explorando as diferenças das propriedades dieléctricas entre tecidos da mama e massas cancerígenas quando sujeitas a frequências microondas. O radar de banda ultra-larga de microondas baseia-se na iluminação da mama com um pulso de banda ultra-larga gravando o sinal resultante por retrodispersão. Esta técnica tem um grande potencial por ser relativamente barata, não invasiva, confortável para o paciente e utilizar radiação não-ionizante. A técnica de imagem por microondas tem sido estudada via simulação, sendo que os primeiros protótipos estão agora a ser construídos.

Nesta dissertação é estudado o potencial do radar de banda ultra-larga de microondas para classificação de tumores baseada no tamanho e/ou na forma do tumor. Numa primeira parte foram estudados tumores numéricos, construídos com base em *Gaussian Random Sphere*, em simulações, e numa segunda parte foram estudados fantômas físicos de tumores com o uso de um protótipo. Três classificadores foram utilizados: Análise Discriminante Linear, Análise Discriminante Quadrática e *Support Vector Machines*. Várias arquitecturas foram estudadas, combinando a classificação por tamanho e depois por forma e vice-versa. Foram estudados cenários de fantômas da mama homogéneos e heterogéneos.

**Palavras-chave:** tumor da mama, classificação, radar banda ultra-larga de microondas, machine learning



# Abstract

Microwave Imaging is one of the most promising medical imaging techniques for breast tumour detection and classification, exploiting the dielectric contrast between normal and tumour tissues at microwave frequencies. In UWB Radar Imaging a microwave pulse is used to irradiate the breast, and resulting backscattered waves are recorded. This technique has great potential because it is relatively cheap, non-invasive, painless for the patient and uses non-ionising radiation. Microwave Imaging has been studied *via* simulation, and the first prototypes are being built.

In this dissertation the potential of UWB Radar for breast tumour classification based on the size and/or the shape of the tumour is studied. Firstly, numerical tumours, based on Gaussian Random Spheres, were studied in simulation scenarios. In the second half of the study, physical phantoms were studied with the use of a prototype. Three classification methods were used: Linear Discriminant Analysis, Quadratic Discriminant Analysis and Support Vector Machines. Several different classification architectures were considered combining size classification and shape classification. Homogeneous and heterogeneous breast phantoms scenarios were considered.

**Key Words:** breast tumour, classification, UWB Radar, machine learning



# Índice

Resumo.....	vii
Abstract.....	ix
Índice .....	xi
Índice de Figuras.....	xiii
Índice de Tabelas.....	xv
Lista de Acrónimos.....	xvii
1 Introdução.....	1
2 Conceitos Teóricos .....	9
2.1 Contexto teórico .....	9
2.1.1 Anatomia da mama .....	9
2.1.2 Cancro da mama e formação de tumores.....	10
2.1.3 Propriedades dieléctricas .....	11
2.2 Radar de Banda Ultra Larga (UWB).....	13
2.2.1 Simulação de Sistema de Radar UWB .....	13
2.2.1.1 Simulação dos tecidos e aquisição dos sinais .....	14
2.2.1.2 FDTD.....	14
2.2.1.3 TF/SF .....	15
2.2.2 Protótipos de UWB .....	17
2.2.2.1 Metodologia.....	17
2.3 Pós-processamento do sinal .....	19
2.3.1 Métodos para Extração de Características Relevantes.....	19
2.3.1.1 Análise de Componentes Principais .....	19
2.3.1.2 Análise de Componentes Independentes .....	20
2.3.1.3 Transformada Discreta de Wavelet .....	21
2.3.2 Classificadores .....	22

---

2.3.2.1	Análise Discriminante Linear .....	22
2.3.2.2	Análise Discriminante Quadrática.....	23
2.3.2.3	Support Vector Machines .....	24
3	Classificação de tumores da mama via simulação.....	27
3.1	Fantômas numéricos .....	27
3.1.1	Conceitos básicos das GRS.....	27
3.1.2	Base de dados numérica.....	29
3.2	Arquitecturas.....	32
3.3	Resultados.....	35
3.3.1	Características do PCA, ICA e DWT .....	35
3.3.2	Comparação entre os classificadores LDA e QDA .....	36
3.3.3	Avaliação global dos classificadores, dos métodos para extracção de características relevantes e das arquitecturas .....	39
4	Classificação de tumores da mama com uso de um Protótipo de UWB.....	41
4.1	Fantômas Físicos.....	41
4.1.1	Fabrico dos fantômas .....	42
4.2	Resultados.....	42
4.2.1	Número de components principais do PCA .....	42
4.2.2	Número de antenas.....	43
4.2.3	Comparação entre LDA,QDA e SVM.....	45
5	Considerações finais .....	49
5.1	Conclusões .....	49
5.2	Limitações e perspectivas futuras .....	51
	Referências Bibliográficas.....	53

# Índice de Figuras

Figura 1.1 Número de casos de cancro diagnosticados (em cima) e número de mortes causadas por cancro (em baixo) em 2008 [2].	1
Figura 1.2 Estimativa da taxa de incidência do cancro da mama (por 100 mil habitantes) em 2008.	2
Figura 1.3 Estimativa da taxa de mortalidade do cancro da mama (por 100 mil habitantes) em 2008.	2
Figura 2.1 Anatomia da mama. Corte sagital.	9
Figura 2.2 Esquema genérico de um sistema multiestático de imagem por UWB adaptado de [27].	14
Figura 2.3 Secção transversal da matriz espacial 3D do FDTD. O tumor respresentado, um tumor espiculado localizado no centro do TF, é iluminado por uma onda plana (as setas indicam a direcção da onda e a linha preta a sua origem) propagando-se na direcção de $z+$ e os sinais por retrodispersão são gravados na localização do primeiro observador: (0,0,-74) mm (representado pelo círculo preenchido). Os quatro pontos de observação estão representados por círculos.	16
Figura 2.4 Protótipo do sistema de UWB com o fantôma do tumor, verde, colocado ao centro.	18
Figura 2.5 Demonstração geométrica a duas dimensões do conceito de hiperplano ideal, margem e support vectors.	24
Figura 3.1 Amostras de GRS representando tumores benignos com raio médio de 5 mm. De cima para baixo: (a) liso e (b) macrobulado.	30
Figura 3.2 Amostras de GRS representando tumores malignos com raio médio de 5 mm. De cima para baixo: (a) microlobulado, (b) espiculado com 3 espículos, (c) espiculado com 5 espículos e (d) espiculado com 10 espículos.	32
Figura 3.3 Arquitecturas baseadas só na forma dos modelos. À esquerda GF e à direita DF.	33

Figura 3.4 Arquitecturas que classificam detalhadamente a forma de cada tumor. À esquerda DDF e à direita DDTDDF.....	34
Figura 3.5 Arquitecturas que classificam os tumores em benignos e malignos com uma separação por tamanhos inicial. À Esquerda DTGF e à direita GTGF.....	34
Figura 3.6 Relação entre número de componentes principais usados e a performance dos classificadores LDA (linha azul) e QDA (linha preta) para o GF (à esquerda) e o GT (à direita).....	35
Figura 3.7 Esquema das decomposições do DWT com duas iterações adaptado de [27].....	36
Figura 4.1 Modelos físicos de tumores. Em cima, modelos de tumores malignos. Em baixo: modelos de tumores benignos.....	42
Figura 4.2 Relação entre número de componentes principais usados e a performance dos classificadores LDA (linha azul) e QDA (linha preta) para os modelos homogéneos (à esquerda) e heterogéneos (à direita). ....	43
Figura 4.3 Relação entre número de antenas usadas para a classificação e a performance dos classificadores LDA (linha azul) e QDA (linha preta) para os modelos homogéneos (à esquerda) e heterogéneos da mama (à direita).....	44



# Índice de Tabelas

Tabela 3.1 Performance das classificações parciais de tamanho e forma e classificação geral de tamanho-forma combinando os métodos de extracção de características relevantes PCA, ICA e DWT com os classificadores LDA e QDA para as seis diferentes arquitecturas. ....	38
Tabela 3.2 Média dos resultados para cada combinação entre métodos de extracção de características relevantes e os classificadores. ....	39
Tabela 3.3 Performances médias global resultantes de todas as combinações para cada arquitectura. ....	40
Tabela 4.1 Classificação utilizando a antena mais próxima e a antena mais distante do tumor para os modelos homogéneos e heterogéneos da mama. ....	44
Tabela 4.2 Resultados das classificações (com sinais agrupados e não agrupados) dos modelos físicos de tumores combinando o PCA com os classificadores LDA, QDA e SVM para os modelos homogéneos e heterogéneos da mama. A arquitectura utilizada foi a Grosseira-Forma. ....	46



# Lista de Acrónimos

ADN	Ácido Desoxirribonucleico
ARV	Analizador de Rede Vectorial
DDF	Directa-Detalhada-Forma
DDTDDF	Directa-Detalhada-Tamanho-Directa-Detalhada-Forma
DF	Detalhada-Forma
DTGF	Detalhada-Tamanho-Grosseira-Forma
DWT	Transformada Discreta de Wavelet (do inglês, <i>Discrete Wavelet Transform</i> )
FDTD	Diferenças Finitas do Domínio do Tempo (do inglês, <i>Finite-Difference Time Domain</i> )
GF	Grosseira-Forma
GRSs	Do inglês, <i>Gaussian Random Spheres</i>
GTGF	Grosseira-Tamanho-Grosseira-Forma
ICA	Análise de Componentes Independentes (do inglês, <i>Independent Component Analysis</i> )
LDA	Análise Discriminante Linear (do inglês, <i>Linear Discriminant Analysis</i> )
LDB	Do inglês, <i>Linear Discriminant Basis</i>
MI	Imagem por Microondas (do inglês, <i>Microwave Imaging</i> )
MRI	Imagem por Ressonância Magnética (do inglês, <i>Magnetic Resonance Imaging</i> )
PCA	Análise de Componentes Principais (do inglês, <i>Principal Component Analysis</i> )
PET	Tomografia por Emissão de Positrões (do inglês, <i>Positron Emission Tomography</i> )
QDA	Análise Discriminante Quadrática (do inglês, <i>Quadratic Discriminant Analysis</i> )
RBF	Função de Base Radial (do inglês, <i>Radial Basis Function</i> )
RTS	Do inglês, <i>Radar Target Signatures</i>
SF	Do inglês, <i>Scatterde-Field</i>
SVM	Do inglês, <i>Support Vector Machines</i>
TF	Do inglês, <i>Total-Field</i>
TF/SF	Do inglês, <i>Total-Field/Scattered-Field</i>
UPML	Do inglês, <i>Uniaxial Perfectly Matched Layer</i>

UWB Banda Ultra-Larga (do inglês, *Ultra Wideband*)

# 1 Introdução

O cancro é uma das principais causas de morte no mundo, tendo sido responsável por 7,6 milhões de mortes (aproximadamente 13% de todas as mortes) em 2008. Os tipos mais frequentes de cancros diferem entre homens e mulheres. No caso das mulheres o cancro mais comum é o cancro da mama (afecta uma em oito mulheres) e é também o responsável pelo maior número de mortes de pessoas do sexo feminino com cancro [1, 2]. Na Figura 1.1 estão representadas as taxas de incidência e mortalidade dos principais tipos de cancro nas mulheres.

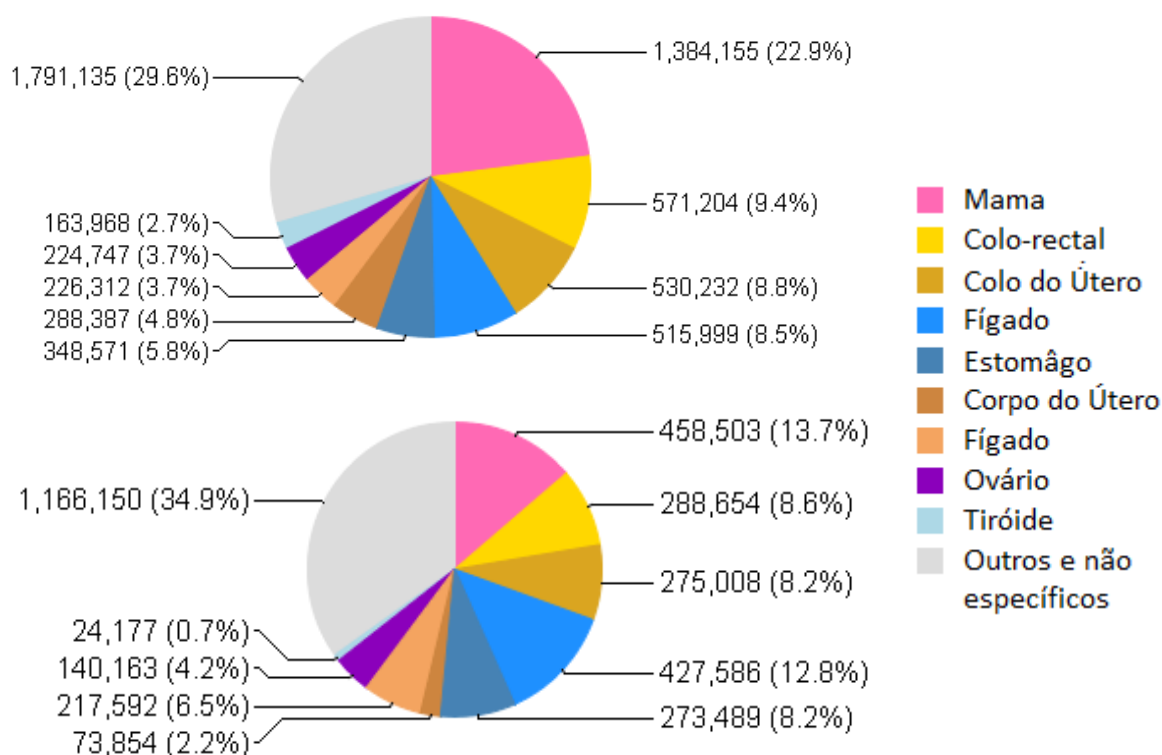


Figura 1.1 Número de casos de cancro diagnosticados (em cima) e número de mortes causadas por cancro (em baixo) em 2008 [2].

A percentagem de incidências e de mortalidade do cancro da mama não tem uma distribuição uniforme globalmente. Nas Figura 1.2 e Figura 1.3 é possível observar as taxas de incidência e de mortalidade do cancro da mama a nível global. A ocorrência de cancro não se deve só a características genéticas e idade mas também a um conjunto de factores de risco associados aos hábitos quotidianos: um estilo de vida não saudável (tabaco, álcool, dietas inadequadas, inactividade física), exposição a elementos cancerígenos, à poluição do ar, à radiação e algumas infecções (tais como a

hepatite B) [3]. No geral, em países subdesenvolvidos assiste-se a uma menor incidência do cancro, embora seja mais fatal, dado o fraco acesso a cuidados de saúde [4].

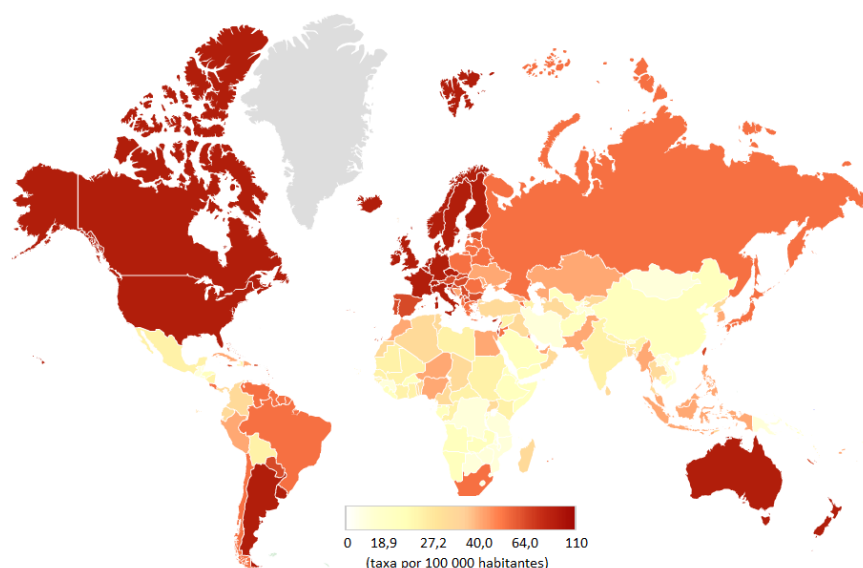


Figura 1.2 Estimativa da taxa de incidência do cancro da mama (por 100 mil habitantes) em 2008

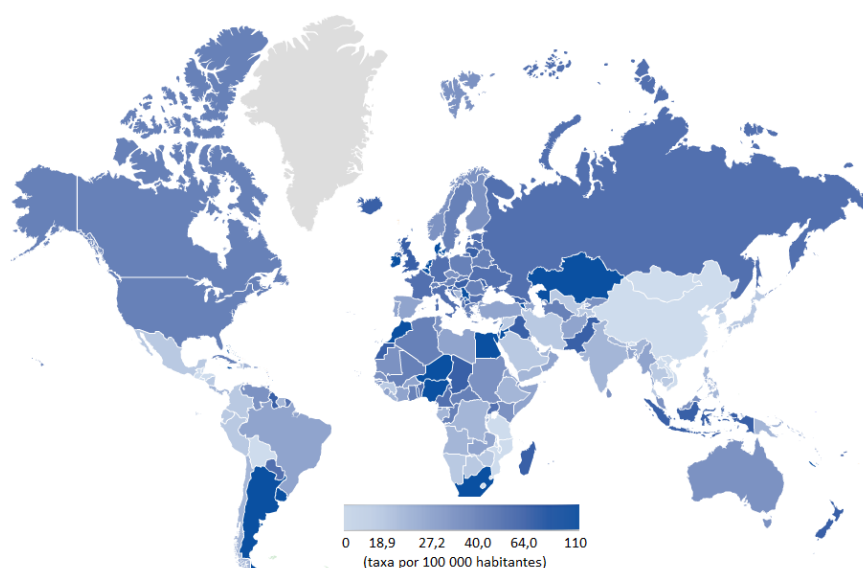


Figura 1.3 Estimativa da taxa de mortalidade do cancro da mama (por 100 mil habitantes) em 2008

No ano de 2008, em Portugal, foram registados 5333 casos de cancro da mama e 1537 mortes devido a esta neoplasia. No caso das mulheres, o cancro da mama representa 27,7% dos casos de cancro registados e 15,9% de mortalidade [5].

O cancro da mama não é um dos tipos de cancro com maior taxa de mortalidade, tendo em conta as respectivas taxas de incidência, mas como é aquele com mais casos registados mundialmente é também o responsável por maiores mortalidade em termos absolutos. Para além dos casos fatais, dentro dos casos de

superação desta neoplasia, o processo desde que é diagnosticada até estar totalmente curada pode ser bastante doloroso e dispendioso. Evron *et al.* [6] afirma que se detectado numa fase inicial, o cancro da mama é curável. Os factores chave para combater a taxa de mortalidade devido ao cancro são: diagnóstico credível numa detecção precoce, intervenção precoce e monitorização fidedigna. Este trabalho focar-se-á no estudo de um tipo de diagnóstico que é potencialmente possível numa fase precoce. Algumas técnicas para diagnóstico do cancro da mama existentes são referidas nos próximos parágrafos.

Uma das ferramentas mais comuns para rastreio e diagnóstico do cancro da mama é a imagiologia médica. Actualmente as mais comuns são a Mamografia e Ultra-Som como técnicas de rastreio, e as Imagens por Ressonância Magnética (do inglês *Magnetic Resonance Imaging, MRI*) são também bastante comuns em diagnósticos mais complexos. A Cintimamografia e a Tomografia por Emissão de Positrões (do inglês *Positron Emission Tomography, PET*) são outras técnicas a serem investigadas para imagem da mama.

A técnica não palpável mais utilizada para rastreio e detecção do cancro da mama é a Mamografia por Raio-X. Esta técnica baseia-se na variação das densidades do tecido da mama, utilizando raios-X de baixa energia que penetram a mama comprimida para obter uma imagem. Apesar da técnica demonstrar bons resultados e ter um baixo custo, apresenta também algumas limitações bem conhecidas em termos de sensibilidade e especificidade (proporção de pacientes correctamente identificados como tendo ou não cancro, respectivamente). Para além disso, devido à radiação ionizante, há sempre um risco inerente à saúde do paciente. Esta técnica causa algum desconforto ao paciente devido à compressão da mama [7-9]. Outro problema associado à Mamografia são as altas taxas de diagnósticos falsos positivos reportados [10, 11], e ainda mais preocupante, as de falsos negativos que podem ser até 34% [8]. Os falsos positivos são os diagnósticos positivos em pessoas que não têm cancro e os falsos negativos dizem respeito aos diagnósticos negativos em que o paciente tem cancro [7-10]. Quanto à classificação entre tumores benignos e malignos, vários estudos apresentaram resultados à volta dos 94% de eficácia [12-14].

Os Ultra-Sons também são bastante utilizados na detecção e classificação de cancros e tumores da mama. O seu funcionamento consiste na emissão de uma onda sonora (de frequências não audíveis ao ouvido humano) contra a mama, que são reflectidas quando encontram um objecto, criando um eco. Medindo as ondas do eco é possível determinar a distância, tamanho e forma do objecto. Esta técnica não é commumente utilizada para rastreio [15] mas é frequentemente aplicada quando massas suspeitas são observadas pela mamografia em mulheres com mamas densas

[11]. É utilizada quase exclusivamente para rastreio de mulheres mais jovens [16]. É conveniente para o paciente, em tempo real e um método de diagnóstico seguro e demonstra também ser uma boa ferramenta de diagnóstico para classificação de tumores da mama. O seu potencial de diagnóstico já foi demonstrado mas, como depende da análise da textura, é dependente do sistema e por norma só apresenta bons resultados em sistemas de ultra-som muito específicos. Outras limitações inerentes à técnica são o baixo contraste e a interferência com ruído e *speckle*. O *speckle* é causado pela interferência de ecos derivados de detectores aleatórios que são demasiado pequenos para serem integrados na formação da imagem, resultando numa estrutura granular, sobreposta à imagem. Devido a este problema, é difícil detectar tumores de dimensões pequenas com ultra-sons [17, 18]. Chang *et al.* [17] obteve uma performance de 90,95% na classificação de tumores entre benignos e malignos, com uma sensibilidade de 88,89% e uma especificidade de 92,5%.

Outra técnica comumente utilizada é recorrendo à Imagem por Ressonância Magnética (MRI). A MRI recorre ao uso de radiação não ionizante, baseada em campos magnéticos e ao efeito que estes têm nos átomos de hidrogénio no corpo, oferecendo um grande contraste e resolução espacial. O hidrogénio é abundante na água e por isso estruturas como a gordura ou os vasos sanguíneos são bem visíveis com esta técnica [19]. Porém, têm custos associados bastante elevados, fazendo com que não seja utilizada regularmente como técnica de rastreio [15], servindo principalmente como primeira técnica de diagnóstico adicional ou complementar de suspeitas lesões identificadas pela mamografia e/ou ultra-sons. Tumores com pouca vascularização induzida em tecidos vizinhos não são muitas vezes detectados. A MRI apresenta uma sensibilidade de aproximadamente 90% e uma especificidade entre os 37% e os 86% [20]. Quanto à classificação de tumores, Jacobs *et al.* [20] apresentou 100% de performance na classificação de tumores benignos e malignos em 36 pacientes.

As duas outras técnicas de imagem, Cintimamografia e a Tomografia por Emissão de Positrões (PET) apresentam custos bastante elevados, assim como é necessária a utilização de agentes radioactivos. Muito resumidamente, na Cintimamografia um radiofármaco é injectado no paciente, esse radiofármaco acumula-se nas zonas lesadas. O radiofármaco irá emitir radiação que será detectada por uma camara de raios gama que constrói a imagem [9]. Apesar da Cintimamografia detectar com sucesso tumores vascularizados, o número de falsos positivos varia entre os 66% e os 73% e o número de falsos negativos entre 2% e 8% [9]. Segundo Nass *et al.* [9] esta técnica apresente níveis de sensibilidade entre 75% e 94% e de especificidade entre 80% e 89%. Num exame de PET, um agente radioactivo, marcado com um isótopo emissor de positrões, é injectado no paciente um radiofármaco. O radioisótopo sofre



decaimentos, libertando positrões, que ao interagir com um electrão emite dois fótons de 511 keV com um ângulo de 180° entre eles. Quando os fótons não interagem com a matéria e são detectados, numa janela temporal estabelecida, em locais opostos é considerada uma coincidência verdadeira. Só estas coincidências contribuem para a formação da imagem. O PET não permite uma diferenciação entre tumores benignos e malignos [21] e os tumores pequenos (<1cm) são de difícil detecção. É uma técnica conveniente para pacientes com resultados inconclusivos nos métodos convencionais [22] e para detectar reincidências em pacientes que já efectuaram tratamento [23]. Apresenta uma sensibilidade de entre 80% e 100% e especificidade entre 75% e 85% [9]. Contudo, o seu grande problema é o alto custo e o facto do agente radioactivo ter um tempo limitado de acção pelo facto do seu fabrico ter que ser relativamente próximo do local do exame [22]. De realçar que quanto menor for o tempo de vida do agente radioactivo, menos será a radiação a que o paciente estará exposto [24].

Na tentativa de colmatar as falhas das técnicas referidas acima surge a imagem por microondas (*Microwave Imaging, MI*) [25, 26]. As principais características que fazem desta técnica uma das mais promissoras para ser utilizada como técnica de rastreio para detecção e classificação do cancro da mama é ser de baixo custo, não invasiva, utilizar radiação não-ionizante, ser confortável para a paciente e permitir alta resolução [27, 28].

A imagem por microondas baseia-se nas diferenças das propriedades dieléctricas a nível molecular em vez de anatómicas (densidade) dos tecidos da mama e das massas cancerígenas[10]. Estas propriedades têm sido estudadas até aos dias de hoje. Até 2007 diversos estudos relatavam que a vascularização nos tecidos malignos induz a presença de uma maior concentração de água nestes tecidos. Essa diferença na concentração de água entre tecidos malignos e normais, será expressa pela magnitude dos sinais recebidos conseguindo-se identificar a presença de tecido malignos [26, 29, 30]. Apesar dos estudos anteriormente referidos reportarem que os tecidos normais e os tecidos malignos têm propriedades dieléctricas diferentes, em 2007 Lazebnik *et al.* [31, 32] conduziu um estudo em larga escala sobre o contraste destas propriedades entre vários tipos de tecido saudável (adiposo e fibroglandular) e tecidos de tumores benignos e malignos. Lazebnik descobriu que as propriedades entre tecido fibroglandular e tumoral são muito semelhantes e que existem diferenças nas propriedades dieléctricas entre tecido adiposo e tecido tumoral. Entre tumores malignos e benignos foi concluído não existirem diferenças significativas. Não sendo as propriedades dieléctricas determinantes para a classificação de tumores é então necessário analisar outras características. Neste trabalho é proposta uma classificação baseada na assinatura do tumor (*Radar Target Signature, RTS*)[33].

Três diferentes abordagens têm sido propostas para obter imagem médica por microondas: Radar de Banda Ultra Larga (UWB Radar Imaging), utilizado neste trabalho, Tomografia por Microondas (*Microwave Tomography*) e métodos de *Time-Reversal Finite-Difference Time Domain* (FDTD) [27]. O *UWB Radar Imaging* envolve a iluminação da mama com um pulso de ondas de banda ultra-larga (na frequência de microondas) de uma, ou mais, antenas em localizações diferentes, gravando o sinal resultante por retrodispersão (*backscattered signal*). Os sinais são depois processados de modo a calcular a localização do tumor [25, 28, 34, 35].

Chen *et al.* [36], em 2007, começou o seu estudo sobre o impacto da morfologia dos tumores nas suas assinaturas em sistemas de *UWB Imaging* analisando a resposta tardia do sinal de retrodispersão no domínio temporal. Os fantômas 2D, homogêneos, utilizados foram baseados nas descobertas feitas por Rangayyan *et al.* [37] sobre as diferenças das formas e texturas entre tumores malignos e benignos. Rangayyan *et al.* caracterizou os tumores benignos com contornos ovais e os tumores malignos com contornos irregular, espiculados. Chen, no seu trabalho, utilizou um sistema monoestático em que a antena emissora é também receptora do sinal. Em 2008, Chen *et al.* [38], adicionou alguma heterogeneidade aos fantômas. Foram criados fantômas dos três grupos definidos por Lazebnik *et al.* [32]. Chen obteve resultados satisfatórios e promissores para tumores com grande quantidade de tecido adiposo presente mas o mesmo não se verificou para tumores com grande presença de tecido fibroglandular.

Chen *et al.* [39, 40] deu seguimento aos seus estudos anteriores melhorando o sistema de aquisição de dados para um modelo multiestático mais evoluído. Este modelo possui mais do que uma antena, emitindo cada uma delas, à vez, um pulso de UWB sendo todas as outras antenas usadas como receptoras. Os modelos dos tumores, projectados da mesma forma que nos estudos anteriores, foram separados, consoante a sua geometria em duas categorias: benignos (ovais e macrolobulados) e malignos (microlobulados e espiculados). Foi atingida, para uma classificação baseada na forma do tumor, uma performance de 86,7%.

Teo *et al.* [41] focou-se na primeira parte do sinal de retrodispersão no domínio temporal, visto conter a maior intensidade do sinal. Foram utilizadas cinco antenas, simultaneamente receptoras e emissoras, e a localização do tumor era sensivelmente conhecida. Os 60 tumores foram separados em dois grupos, 30 lisos e 30 espiculados, que foram correctamente separados.

No que a classificação 3D diz respeito, Davis *et al.* [42] classificou um conjunto de tumores numéricos, analisando as suas assinaturas, baseado na forma e tamanho do tumor. Os modelos numéricos foram construídos com base em *Gaussian Random Spheres* (GRSs) permitindo diferentes irregularidades na forma. Foram consideradas

três formas: lisos, microlobulados e espiculados; e quatro tamanhos diferentes, variando desde 2,5 a 10 mm de raio. Foram utilizados dois métodos para extracção de características, *Local Discriminant Bases* (LDB) e Análise de Componentes Principais (do inglês, *Principal Component Analysis*, PCA), combinadas com o classificador Análise Discriminante Linear (do inglês, *Linear Discriminant Analysis* LDA). Davis *et al.* considerou uma classificação só por tamanho, uma só pela forma e uma separando primeiro por tamanho e depois por forma, obtendo performances de 97%, 70% e 86% respectivamente.

Desde 2010, Conceição *et al.* [27, 33, 43-45], tem vindo a estudar as potencialidades do Radar de Banda Ultra Larga na classificação de tumores. Conceição *et al.* criou uma base de dados de 480 modelos numéricos de tumores baseados em GRSs. Quatro tamanhos e quatro formas foram considerados. Três métodos para extracção de características relevantes foram aplicados: Análise de Componentes Principais, Análise de Componentes Independentes (do inglês *Independent Component Analysis*, ICA) e Transformada Discreta de Wavelet (do inglês *Discrete Wavelet Transform*, DWT). Cada um destes métodos foi combinado com três classificadores: Análise Discriminante Linear, Análise Discriminante Quadrática (do inglês *Quadratic Discriminant Analysis*, QDA) e *Support Vector Machines* (SVM). Várias arquitecturas foram estudadas, combinando em algumas delas a separação por tamanho e depois por forma e vice-versa [27]. Foram estudados cenários homogéneos e heterogéneos [33].

As principais técnicas para a classificação de tumores detectados são os métodos para extracção de características em conjunto com os classificadores, como utilizado em [27, 42-45]. No presente trabalho foram utilizados os mesmos métodos e classificadores que Conceição [27]. Três métodos para extracção de características relevantes: PCA, ICA e DWT. Estes métodos foram comparados usando três métodos de classificação diferentes: LDA, QDA e SVM. O funcionamento destes métodos e classificadores é explicado em 2.3.

Nestes parágrafos será descrito um breve resumo do que será mostrado em cada capítulo. Esta dissertação está dividida em 4 capítulos, sendo o primeiro, a presente introdução. Na introdução foi apresentado o panorama do cancro no Mundo e em particular do cancro da mama. Foi também apresentado o estado da arte sobre as técnicas de detecção e diagnóstico de tumores mamários.

No capítulo 2 serão abordados os aspectos teóricos fundamentais para a compreensão do resto do trabalho. É feita uma breve apresentação sobre a anatomia da mama, explicado o processo da formação de tumores e são apresentadas as propriedades dieléctricas dos tecidos mamários.

Ainda no capítulo 2, é explicado o funcionamento do Radar de Banda Ultra Larga (UWB), para a parte da simulação baseada em tumores numéricos (*Gaussian Random Spheres*, GRSs) assim como para o protótipo com modelos reais de tumores. A teoria dos métodos para extracção de características relevantes e dos classificadores é explicada no final do capítulo.

O capítulo 3 é referente aos resultados do estudo realizado com os modelos numéricos de tumores. É explicado como estes modelos são gerados, a base de dados utilizada e as arquitecturas aplicadas aos classificadores. No final são apresentados e discutidos os resultados e feitas as respectivas conclusões.

No capítulo 4, referente ao estudo realizado com fantômas físicos de tumores, é explicado o seu fabrico assim como o dos modelos de tecido fibroglandular. Os resultados são mostrados e discutidos e por fim são feitas as conclusões. Algumas limitações e perspectivas futuras são apresentadas.

## 2 Conceitos Teóricos

### 2.1 Contexto teórico

#### 2.1.1 Anatomia da mama

A mama é constituída por três tipos de tecido: tecido adiposo, tecido conectivo e glândulas mamárias [46]. As proporções entre estes tecidos, a quantidade de água, gordura e tecido fibroglandular variam consoante a pessoa e o seu estado hormonal (grávida, fase do ciclo da menstruação ou na menopausa) [10, 47].

Imediatamente sob a pele encontra-se o tecido adiposo, que é composto por células vesiculares cheias de gordura e organizadas em lóbulos separados pelo ligamento de *Cooper*. O tecido mais interno da mama consiste nas glândulas mamárias e no tecido conjuntivo (colagénio e elastina)[46, 48]. A mama tem um complexo conjunto de vasos sanguíneos e uma complexa rede linfática [49]. Cada mama tem cerca de 15 a 20 secções chamadas lobos com muitas secções menores chamadas lóbulos que por sua vez contêm grupos de pequenas glândulas que podem produzir leite. A maioria das mamas contém tecido adiposo abundante, ao qual está associado o desenvolvimento de cancro na mama [50]. Internamente a mama é separada do músculo grande peitoral por uma camada fina de tecido conjuntivo laxo com tecido adiposo, conhecida por espaço retromamário [48]. Na Figura 2.1 está representado um corte sagital de uma mama feminina.

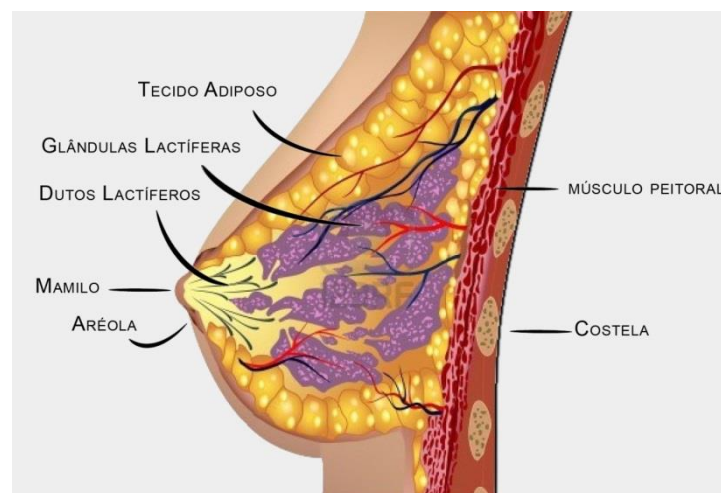


Figura 2.1 Anatomia da mama. Corte sagital.

### 2.1.2 Cancro da mama e formação de tumores

O cancro é o crescimento e metastização de células descontrolada que pode afectar qualquer parte do corpo [51]. Para tentar perceber melhor o cancro é essencial perceber como as células normais se tornam células cancerígenas. Cada célula contém no seu núcleo o material genético (Ácido Desoxirribonucleico, ADN) que controla a célula, decidindo o que faz, quando se reproduz e até quando morre. Cada tipo de célula tem a sua função, no conjunto elas são responsáveis pelas funções vitais do organismo, como garantir o transporte do oxigénio para todo o corpo, controlar a nutrição, combater doenças, armazenamento de energia, entre outras. Quando as células envelhecem ou se danificam normalmente morrem e são substituídas por novas células. Contudo, este processo pode não ocorrer como previsto, danificando ou alterando o ADN. Essas alterações são chamadas de mutações e causam um crescimento e divisão anormal das células. Estas células, que sofrerão ciclos de replicação diferentes das células normais, são denominadas de células tumorais. Conforme as mutações que sofreram, uma certa quantidade de informação genética foi perdida ou modificada, o que define o grau de malignidade presente [52, 53]. Os tumores podem ser benignos ou malignos. De notar que só os tumores malignos são cancerígenos [53]. À partida, os tumores benignos não são nocivos para a pessoa. No entanto, se afectarem glândulas hormonais próximas ou se comprimirem outros órgãos com o seu crescimento, podem representar uma ameaça.

As características principais dos tumores benignos são:

- Normalmente crescem a um ritmo lento
- Frequentemente podem ser retirados e têm taxas menores de reincidência
- Não migram para outras partes do corpo

E as dos tumores malignos:

- Crescem/replicam-se a um ritmo elevado
- Podem invadir tecidos vizinhos, destruindo-os
- Migram para outras partes do corpo [52, 53]

Com dito anteriormente, os tumores crescem localmente ou podem-se migrar para outras partes do corpo. As principais razões de ser de difícil cura é a incapacidade de isolamento das células tumorais e saudáveis e a metastização, a migração das

células tumorais para outras partes do corpo. As principais formas de os tumores se espalharem é através do sistema sanguíneo e linfático [53].

As células dos tumores sofrem diversas alterações, algumas delas a nível anatómico. O citoesqueleto fica desorganizado causando a perda da forma original da célula, que se torna mais arredondada. Devido a alterações da superfície da célula, a permeabilidade da mesma é afectada, fazendo com que os tecidos do tumor absorvam mais água do que as células normais. O aumento de água nos tecidos aumenta a condutividade e a permissividade [26, 29]. Devido ao grande crescimento das células do tumor há também um aumento da vascularização (angiogénese) de modo a fornecer nutrição às células [26].

Todas estas alterações são importantes para detecção e avaliação de tumores. Outras características que provaram ser úteis em termos de classificação são características baseadas no tamanho, forma, contornos, textura da superfície, profundidade, localização e densidade dos tumores [26, 37, 42, 44, 48]. Quanto à forma e textura, os tumores benignos têm uma textura lisa e suave, com contornos ovais ou macrolobulados, enquanto os tumores malignos apresentam uma forma mais irregular, espiculada ou microlobulada e uma textura áspera [12, 13, 35, 37].

### **2.1.3 Propriedades dieléctricas**

Como referido anteriormente, as células cancerígenas sofrem diversas alterações químicas e fisiológicas. Do ponto de vista da imagem por microondas, estes factores levam a uma mudança das propriedades dieléctricas das células cancerígenas: condutividade e permissividade [27].

Muitos estudos [48, 54-59] foram publicados com o objectivo de conhecer e perceber as diferenças destas propriedades nos tecidos normais e de tumores da mama. No entanto, até 2007, todos os estudos efectuados apresentavam limitações, originando que nem todos estivessem de acordo. Um resumo dos estudos referidos anteriormente foi feito por Sha et al. [60] e por Lazebnik *et al.* [31]. A maioria desses estudos envolviam uma pequena população de pacientes, não incluía todos os tipos de tecido normal, benigno e maligno, e, com raras excepções, não se estendiam acima dos 3,2 GHz [31]. Lazebnik *et al.* [31] atribui as discrepâncias observadas à variabilidade dos procedimentos e protocolos usados nos diferentes estudos combinado com o desafio inerente da caracterização dos tecidos da mama, que são, de facto, bastante heterogéneos. Lazebnik, fundamentando-se na existência de tal heterogeneidade, inicia um estudo a larga escala dividido em duas partes. No primeiro estudo, Lazebnik estudou as propriedades dieléctricas do tecido normal da mama obtido de cirurgias de

redução da mama, e portanto tecidos saudáveis [31]. No segundo estudo estudou as propriedades dieléctricas do tecido da mama normal, benigno e maligno, obtidos através de cirurgias para remoção de cancro [32].

Em ambos os estudos as medições das propriedades dieléctricas foram feitas desde os 0,5 aos 20 GHz usando uma sonda coaxial de volume variável. As amostras foram categorizadas em 3 grupos consoante a percentagem de tecido adiposo presente na amostra. O grupo 1 contém as amostras com 0-30% de tecido adiposo (grupo com maior presença de água), o grupo 2 contém as amostras com 31-84% de tecido adiposo e o grupo 3 contém as amostras com 85-100% de tecido adiposo (grupo com menor presença de água).

Lazebnik, no primeiro estudo [31], mostrou que as propriedades dieléctricas do tecido mamário são principalmente determinadas pela quantidade de tecido adiposo presente em cada amostra. Quanto menos tecido adiposo presente, mais é a quantidade de água e por isso maiores são os valores de condutividade e permitividade. Foi concluído também que factores secundários como a idade do paciente, a temperatura do tecido e o tempo entre a excisão e a medição dos valores de propriedades dieléctricas têm efeitos mínimos nas propriedades dieléctricas observadas. Por último, foi descoberto que não existem diferenças estatisticamente significativas entre a variabilidade da constante dieléctrica e na condutividade efectiva intra-paciente e entre-pacientes.

Para Lazebnik, as principais diferenças entre o seu estudo e estudos anteriores em pequena escala são: a maior importância dada à heterogeneidade na mama, uma maior gama de frequências utilizadas e as amostras não serem obtidas somente de cirurgias para remoção do cancro. As propriedades dieléctricas apresentaram valores mais baixos para os grupos com elevadas quantidades de tecido adiposo do que previamente reportado, e valores mais altos para os grupos com menos quantidade de tecido adiposo. Este estudo foi importante porque caracteriza os tecidos de uma mama saudável, sendo útil para comparação e validação de dados obtidos de outras amostra noutros estudos.

No segundo estudo [32], Lazebnik estuda as propriedades dieléctricas de amostras de tecido normal (adiposo, glandular e conjuntivo), benigno e maligno da mama obtidos de cirurgias para remoção do cancro da mama. Os resultados obtidos estão, na maioria, de acordo com alguns estudo publicados anteriormente [54, 55, 58].

Em [32] foi descoberto que as propriedades dieléctricas de tecidos removidos através de cirurgia ao cancro apresentam valores menores que os dos tecidos obtidos pelas cirurgias de redução mamária (mastectomia). Isto deve-se ao facto dos tumores,



por norma, se desenvolverem em tecidos glandulares. Como os tecidos obtidos nas cirurgias de remoção do cancro foram retirados de zonas afastadas do tumor, com grande concentração de tecido adiposo, apresentam valores das propriedades dieléctricas mais baixos. Quanto às diferenças entre tecido normal e maligno não foram registadas diferenças estatisticamente significativas [32].

Estes estudos [31, 32] proporcionam uma base de dados de propriedades dieléctricas para tecidos da mama e tumorais bastante detalhada, o que foi crucial para o desenvolvimento dos fantômas numéricos utilizados neste trabalho e desenvolvidos em [27].

Mais recentemente, Halter *et al.* [61], apesar de algumas limitações do seu estudo, nomeadamente em relação ao pequeno número de pacientes, observou que poderão haver algumas limitações relacionadas com medições *ex vivo* no estudo de Lazebnik *et al.* [31, 32].

## **2.2 Radar de Banda Ultra Larga (UWB)**

Técnicas de imagem e da classificação recorrendo a ondas microondas têm sido estudadas há mais de uma década. Muitos estudos foram feitos à base de simulações e, nos últimos anos, alguns protótipos foram construídos, com resultados publicados [62-66]. Outros estudos explorando as capacidades do radar UWB para classificação de tumores com base no tamanho e/ou na forma foram publicados [33, 42-45, 67]. Uma explicação do funcionamento e aquisição de dados tanto da simulação num computador como num protótipo é de seguida apresentada.

### **2.2.1 Simulação de Sistema de Radar UWB**

Tipicamente, as antenas estão colocadas à volta da mama e cada uma delas funciona como geradoras e receptoras. Um sistema só com uma antena, que envia e recebe o seu próprio sinal é chamado de monoestático e um sistema em que todas as antenas são receptoras do sinal gerado por uma delas é um sistema multiestático. Neste trabalho, foi utilizado um sistema multiestático para as simulações com modelos numéricos e um sistema monoestático na utilização do protótipo para os modelos físicos. Um esquema de um típico sistema multiestático de imagem de UWB é mostrado na Figura 2.2.

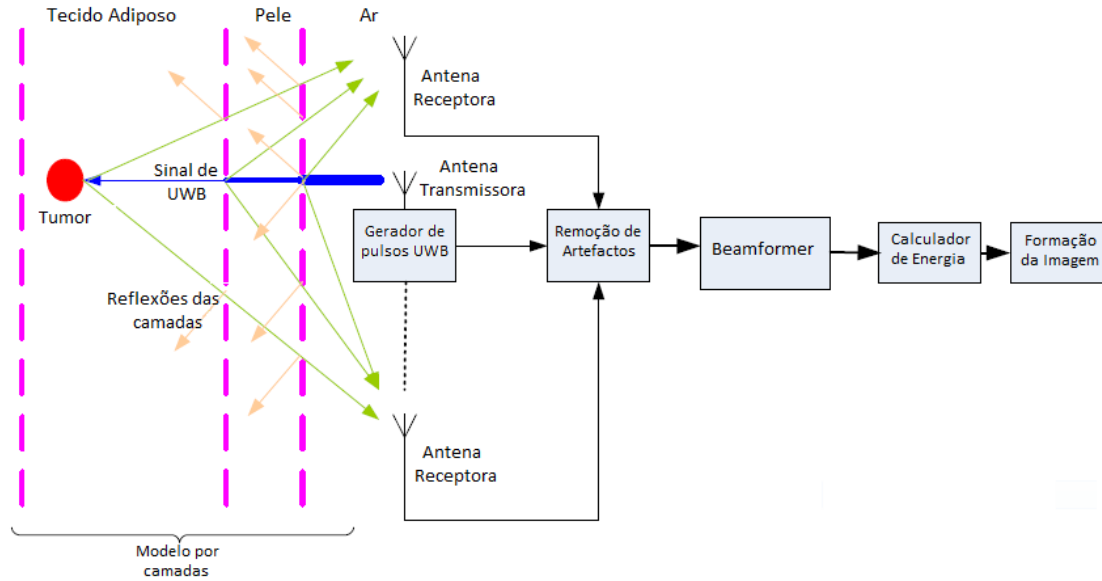


Figura 2.2 Esquema genérico de um sistema multiestático de imagem por UWB adaptado de [27].

De notar que os sinais usados neste trabalho para classificação não necessitam de ser processados por um eliminador de artefactos nem por um *Beamformer* porque são trabalhados em bruto.

### 2.2.1.1 Simulação dos tecidos e aquisição dos sinais

As simulações são efectuadas com o uso do método de modulação FDTD (Diferenças Finitas no Domínio do Tempo, do inglês *Finite-Difference Time-Domain*) para gerar modelos da mama anatomicamente correctos e com propriedades dieléctricas adequadas. O FDTD é o método numérico mais utilizado para modelação de propagação de sinais electromagnéticos em tecidos biológicos [68]. De modo a analisar a assinatura do tumor (do inglês *Radio Target Signature*, RTS), os dados resultantes por retrodispersão foram gerados por uma simulação de Campo-Total/Campo-Disperso (Total-Field/Scattered-Field, TF/SF), explicada nos próximos sub-capítulos.

### 2.2.1.2 FDTD

A técnica das Diferenças Finitas no Domínio do Tempo permite a análise de interacções de ondas electromagnéticas com o material do corpo de formas arbitrárias. O método das FDTD é caracterizado pelas equações de Maxwell. As FDTD foram aplicadas pela primeira vez às equações de Maxwell por Yee *et al.* [69]. Desde então, está técnica tem vindo a ser desenvolvida e refinada, sendo aplicada a muitos problemas de propagação, radiação e dispersão de ondas electromagnéticas. A técnica

é explicada por Taflove e Hagness *et al.* [70]. Taflove e Hagness em [43] apresentam várias razões para que esta técnica continue a ser explorada e aprimorada, como por exemplo: não usar álgebra linear, apresentar um bom grau de exactidão e tratar o comportamento não-linear de forma natural.

O método FDTD consiste em resolver as equações de Maxwell no domínio do tempo e num determinado espaço pré-definido, calculando as derivadas em ordem ao tempo e em ordem ao espaço, através de diferenças finitas.

$$\frac{dE}{dt} = \frac{1}{\epsilon_0} \nabla \times H \quad \text{Equação 2.1}$$

$$\frac{dH}{dt} = \frac{1}{\mu_0} \nabla \times E \quad \text{Equação 2.2}$$

em que  $\epsilon_0$  e  $\mu_0$  são a permissividade eléctrica e a permeabilidade magnética, respectivamente.  $dt$  e  $dx$  estão relacionadas através de

$$dt = \frac{dx}{c_0} \quad \text{Equação 2.3}$$

em que  $c_0$  representa a velocidade da luz no vácuo.

Para o presente trabalho, baseado em [27], a dependência da frequência natural do tecido biológico é incorporada no modelo numérico FDTD através do modelo de Debye, definido por Luebbers *et al.* [71]:

$$\epsilon_r^*(\omega) = \epsilon_r + \frac{\sigma}{j\omega\epsilon_0} + \frac{\chi_1}{1 + j\omega\tau} \quad \text{Equação 2.4}$$

em que  $\omega = 2\pi f$  é a frequência angular e os parâmetros específicos para cada tipo de tecido são:  $\epsilon_r$  a permissividade relativa,  $\epsilon_0$  a permissividade no vácuo,  $\chi_1$  a susceptibilidade eléctrica,  $\sigma$  a condutividade e  $\tau$  o tempo de relaxação.

Tanto os tecidos da mama como os do tumor são implementados numa simulação de FDTD utilizando os parâmetros de Debye como estabelecido em [31, 32]. Para os tecidos malignos:  $\epsilon_\infty = 6,749$ ,  $\Delta\epsilon = 50,09$ ,  $\sigma_s = 0,794 \text{ sm}^{-1}$  e  $\tau = 10,50 \text{ ps}$  e para os tecidos adiposos homogéneos:  $\epsilon_\infty = 3,140$ ,  $\Delta\epsilon = 1,708$ ,  $\sigma_s = 0,036 \text{ sm}^{-1}$  e  $\tau = 14,65 \text{ ps}$ . A resolução da matriz cúbica utilizada para o FDTD é de 0,5 mm.

### 2.2.1.3 TF/SF

Assim como Davis *et al.* [42], os dados resultantes por retrodispersão de um alvo dieléctrico são calculados através de uma aproximação a um Campo-Total/Campo-Disperso (do inglês *Total-Field/Scattered-Field*, TF/SF).

As dimensões utilizadas para este trabalho foram as mesmas utilizadas por Conceição *et al.* [44]. A matriz espacial 3D do FDTD é particionada em duas regiões, campo total (do inglês *Total Field*, TF) e campo disperso (do inglês *Scattered Field*, SF). O tumor está completamente incorporado no campo total, um cubo localizado no centro do SF com 40 mm cada lado. O campo disperso (SF) é um prisma quadrado com base de 153,5 mm de lado e altura 137,5 mm. A origem de ambos os campos é no ponto (0,0,0) mm. A região do TF/SF é constituída por uma UPML (do inglês, Uniaxial Perfectly Matched Layer) com 6 mm com o objectivo de suprimir qualquer reflexão das fronteiras [42, 44, 71].

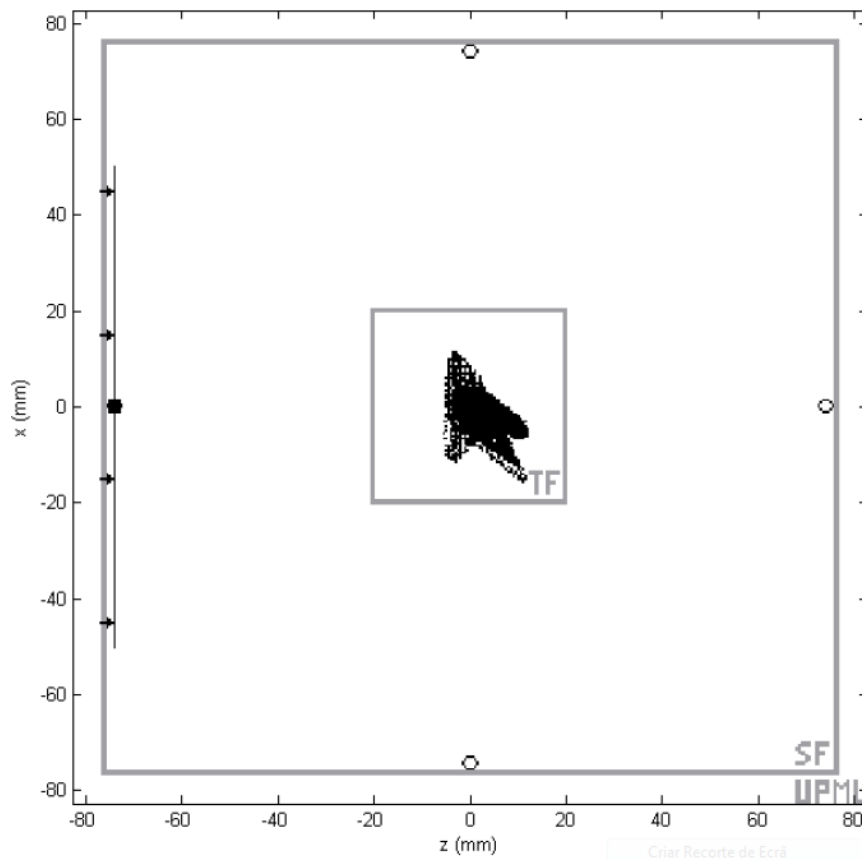


Figura 2.3 Secção transversal da matriz espacial 3D do FDTD. O tumor representado, um tumor espiculado localizado no centro do TF, é iluminado por uma onda plana (as setas indicam a direcção da onda e a linha preta a sua origem) propagando-se na direcção de  $z^+$  e os sinais por retrodispersão são gravados na localização do primeiro observador: (0,0,-74) mm (representado pelo círculo preenchido). Os quatro pontos de observação estão representados por círculos.

A frequência da simulação de FDTD é 1200 GHz. Uma onda plana é transmitida em direcção ao alvo (tumor) em quatro diferentes ângulos, equidistantes do alvo. Os quatro pontos diferentes de observação estão representados na Figura 2.3, o primeiro por um círculo preenchido e os outros por círculos vazios. Estão todos localizados no SF e as suas coordenadas são: (0,0,-74), (-74,0,0), (0,0,74) e (74,0,0) mm, nos eixos ( $x,y,z$ ), por ordem do primeiro ao quarto. Os resultantes sinais co-polarizados são

gravados e analisados pelos mesmos quatro pontos de observação. De notar que só os sinais co-polarizados são gravados uma vez que foi descoberto por ser suficiente para uma alta performance sem extra custos computacionais [42, 44].

O pulso incidente é um pulso Gaussiano modulado com centro de frequência em 6 GHz e largura temporal de 160 ps [42]. Para o primeiro e o terceiro pontos de observação, o pulso é linearmente polarizado na direcção x e y e transmitido na direcção z enquanto que para o segundo e quarto pontos de observação, o pulso é polarizado na direcção y e z e transmitido na direcção x. Cada um dos quatro sinais por retrodispersão são gravados sem a presença do tumor e subtraídos, respectivamente, a cada um dos quatro sinais adquiridos para cada tumor. O sinal sofre depois um processo de redução da amostragem, denominado de *downsample*, e é reduzido de 1200 para 75GHz.

### 2.2.2 Protótipos de UWB

Depois dos testes com simulações, a etapa seguinte inclui os estudos feitos com fantômas físicos com a utilização de protótipos. Nos últimos anos alguns protótipos e tipos de fantômas físicos foram propostos para detecção e reconstrução de imagem [62-66]. Embora esta tecnologia já tenha sido utilizada em pequenos ensaios clínicos, ainda não há nenhum sistema comercial disponível [64, 72]. No que diz respeito à classificação pela forma, este estudo é o primeiro a ser feito com um protótipo e fantômas físicos, com resultados publicados [73]. No próximo sub-capítulo é apresentada a metodologia utilizada, englobando os materiais e a aquisição de dados, sendo que o fabrico dos fantômas é explicado no capítulo 4.1.1.

#### 2.2.2.1 Metodologia

O sistema é formado por um tanque de acrílico de dimensões  $56 \times 56 \times 40$  cm, cheio com óleo de canola de modo a haver uma correspondência da impedância entre a antena e o fantôma da mama. As características dieléctricas do óleo de canola são descritas num estudo de Fear *et al.* [74].

Os sinais são gravados usando um modelo monoestático, em que a mesma antena serve como transmissora e receptora. A resposta do fantôma é gravada com o uso de um Analisador de Rede Vectorial (ARV) que funciona como gerador de ondas e gravador de sinais. Uma antena Vivaldi, como permitividade de 2,65 e tangente de perdas igual a 0,0022, é conectada ao ARV por um cabo de 50  $\Omega$  e é fixada na parede interior do tanque numa estrutura de acrílico à prova de água. Esta metodologia já

provou ser mais robusta comparada com modelos multiestáticos , onde são utilizados vários pares de antenas [75]. Em contra partida grava muito menos sinais do que os modelos multiestáticos. A antena emite um sinal com uma largura de banda de 5 GHz com um centro de frequência de 3,5 GHz. As suas dimensões são de  $12 \times 10,5$  cm. O protótipo, com o fantôma do tumor colocado no centro é mostrado na Figura 2.4.

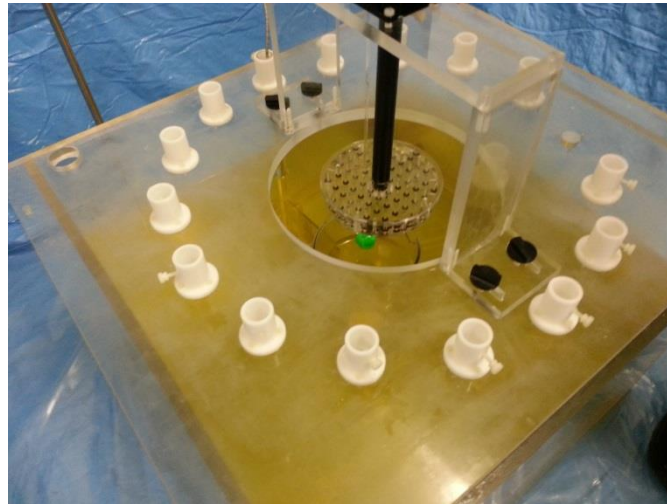


Figura 2.4 Protótipo do sistema de UWB com o fantôma do tumor, verde, colocado ao centro.

Os fantômas são posicionados no centro do tanque numa plataforma giratória (através de um motor). Num sistema clínico, seria a antena a rodar em torno da mama mas este protótipo não foi projectado nesse sentido, sendo o fantôma girado 144 vezes em rotações de  $2,5^\circ$  cada. Após cada rotação uma onda de frequência crescente de 1 a 6 GHz é emitida em direcção ao fantôma. Sendo o tanque de acrílico, irá reflectir o sinal sendo essas reflexões detectadas pela antena receptora. Para eliminar essas reflexões, é feita uma extracção de dados sem a presença do fantôma e o resultado é subtraído aquele obtido com a presença do fantôma. Assim, qualquer resposta que não seja do fantôma é ignorada. Todo o processo é descrito detalhadamente em [65].

Uma vez que os dados extraídos pelo ARV estão em função da frequência, é necessário aplicar a Transformada Inversa de Fourier para obter os dados no domínio do tempo. A assinatura do tumor (RTS) é então extraída a partir dos sinais no domínio do tempo. Tem-se então os dados prontos para a aplicação dos métodos para extracção de características relevantes e dos classificadores.

## 2.3 Pós-processamento do sinal

Para o presente trabalho foram utilizados ao todo três métodos para extracção de características relevantes – Análise de Componentes Principais (PCA), Análise de Componentes Independentes (ICA) e Transformada Discreta de Wavelet (DWT) – e três classificadores – Análise Discriminante Linear (LDA), Análise Discriminante Quadrática (QDA) e *Support Vector Machines* (SVM).

### 2.3.1 Métodos para Extracção de Características Relevantes

Muitas vezes, num processo de classificação, o excessivo número de variáveis e a dimensão de dados podem ser um problema. Para uma classificação mais fidedigna e também para diminuir o custo computacional é essencial aplicar um método para extracção de características relevantes. Estes métodos têm a capacidade de extrair as características mais significativas de cada sinal, reduzindo a sua dimensão. Os três métodos utilizados são o PCA, ICA e DWT, cujo funcionamento é explicado nos próximos sub-capítulos.

#### 2.3.1.1 Análise de Componentes Principais

O PCA, desenvolvido por Pearson, em 1901 e por Hotelling [76] em 1933, é provavelmente a técnica de redução da dimensão de um conjunto de dados mas utilizada globalmente. O objectivo do método é reduzir a dimensão de dados com variáveis multiplas, preservando o máximo de informação possível, eliminando informação desnecessária como ruído [77] e por vezes, até mesmo, revelando estruturas simplificadas escondidas [78]. O método é detalhadamente descrito por Jolliffe *et al.* [79] e uma breve explicação é feita de seguida.

O PCA utiliza álgebra linear para transformar os dados para um novo sistema de coordenadas. As novas variáveis, os componentes principais, são funções lineares das variáveis originais e não são correlacionadas [77]. As novas variáveis que são utilizadas são menos que as originais e assim reduz-se a dimensão dos dados, preservando a informação essencial [79]. O conjunto de dados tem uma variância total, que é a soma das variâncias individuais. Os componentes principais são ordenados por ordem decrescente de variância. Assim, a primeira variável será aquela que contém mais informação e as últimas poderão ser descartadas por não acrescentarem informação significativa ao nosso conjunto de dados [80]. Cabe ao utilizador depois definir o número de componentes a utilizar.

A selecção de bases apropriadas é um importante passo para a classificação porque define as características candidatas a componentes principais. Quanto menos componentes principais obtivermos, mais pequeno será o nosso conjunto de dados e por isso mais baixo será o custo computacional. Para obter os componentes principais de uma matriz  $X$  representada por  $(m \times n)$ , em que  $m$  é o número de medições e  $n$  o número de amostras, subtrai-se a média das amostras e os dados centrados são representados pela sua expansão de Karhunen-Loève:

$$X - E\{X\} = \sum_n^{N_M} \theta_m h_m \quad \text{Equação 2.5}$$

em que  $h_m$  é a base formada pelos vectores próprios da matrix de co-variância  $C$ ,  $\theta_m$  representa os coeficientes das bases de expansão e  $N_M$  é a dimensão do problema [42].

### 2.3.1.2 Análise de Componentes Independentes

O ICA pode ser visto como uma extensão do PCA na medida em que o PCA só descorrelaciona os sinais (estatística de 2ª ordem) enquanto o ICA reduz a sua dependência em ordens mais elevadas, com o objectivo de atingir independência, obtendo componentes linear e não linearmente descorrelacionados [81, 82]. Uma explicação mais detalhada sobre estes termos pode ser encontrada em [81]. Enquanto o PCA trabalha com dados gaussianos, o objetivo do ICA é encontrar uma representação de dados não-gaussianos para que os componentes sejam estatisticamente o mais independentes possível. Este método apresenta como principais limitações o facto de não ser possível determinar nem as variâncias dos componentes independentes, nem a ordem dos mesmos [81].

Para explicar o ICA tomemos então  $n$  dados observáveis  $x_1, \dots, x_N$  que resultam de uma combinação linear de  $N$  variáveis aleatórias  $s_1, \dots, s_N$  estatisticamente independentes entre si:

$$x_i = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{iN} s_N, \forall i = 1, \dots, N \quad \text{Equação 2.6}$$

Em notação matricial tem-se:

$$x = As \quad \text{Equação 2.7}$$

A matriz  $A$ , composta pelos coeficientes  $a_{ij}$ , é a matriz de mistura do problema, que relaciona  $x$  com  $s$ . Neste modelo, apenas as variáveis aleatórias  $x_i$  são conhecidas, e a partir delas serão estimados os coeficientes de mistura  $a_{ij}$  e as componentes independentes  $s_j$  [81, 83].



O ICA assume que os componentes  $s_i$  são estatisticamente independentes e garante que os componentes independentes têm que ter uma distribuição não-gaussiana. A matriz  $A$  é então estimada de modo a minimizar a informação em comum dos componentes independentes com distribuição não-gaussiana. Os componentes independentes são obtidos através da matriz  $W$ , inversa de  $A$ :

$$s = Wx \quad \text{Equação 2.8}$$

Existem vários algoritmos desenvolvidos para calcular eficientemente a matriz  $A$ . O *FastICA*, baseia-se num esquema de iteração de pontos fixos para que a distribuição seja o mais não-gaussiana possível [81] – este algoritmo é utilizado neste trabalho. Uma descrição detalhada do seu funcionamento pode ser encontrada em [81].

### 2.3.1.3 Transformada Discreta de Wavelet

A Transformada Discreta de Wavelet, baseada em wavelets adaptativas, é outro método de extracção de características cujo objectivo é a redução e optimização da informação discriminada presente. A transformada DWT é uma decomposição do sinal através da translação e dilatação da wavelet mãe. O método produz coeficientes de wavelets, que são usados para a classificação [84].

As wavelets têm sido utilizadas como métodos para extracção de características nas últimas décadas e têm provado ser bastante eficientes. Uma wavelet é uma pequena porção de uma onda. Mais especificamente, são funções base localizadas que são versões transladadas e dilatadas de uma wavelet mãe. A principal característica das wavelets é a de serem capazes de providenciar informação localizada de frequência de uma função ou sinal. É esta informação que dá um contributo importante para o processo de classificação [84].

Quando o DWT é aplicado ao conjunto das assinaturas dos tumores (RTS), os coeficientes das wavelets são obtidos pela decomposição de um filtro passa-baixo e a decomposição de um filtro passa-alto. Consequentemente, a banda passa-baixo pode ser dividida outra vez por um filtro passa-baixo e um filtro passa-alto. Este processo pode-se repetir o número de vezes desejado pelo utilizador, tendo em conta que por cada iteração o número de amostras é reduzido a metade. Os coeficientes finais, dados por uma sub-banda de frequência específica, são fornecidos aos classificadores.

Existe uma grande variedade de wavelets, o que torna complexa a escolha da wavelet a utilizar em cada problema a resolver. As wavelets estão organizadas por famílias, contendo cada família um número variável de wavelets. Conceição [27] estudou algumas dessas famílias e observou que, para os dados a analisar, a wavelet

com melhor performance foi a *Coiflet 5*. De realçar que este método só foi utilizado para os sinais referentes aos modelos de tumor numéricos, com a mesma base de dados que em [27], e não para os modelos reais. A *Coiflet 5* foi escolhida depois de serem testadas 126 wavelets com o classificador Análise de Discriminantes Linear, descrito em 2.3.2.1.

A frequência de banda utilizada corresponde à dos coeficientes da wavelet obtidos de uma banda passa-baixo obtida depois da descomposição de um filtro passa-baixo depois de dois níveis de decomposição. Esta banda permitiu melhores performances do que outras sub-bandas avaliadas até três e quatro níveis de decomposição [27].

### 2.3.2 Classificadores

Depois do processamento de redução de dimensão feito pelos métodos de extracção de características, os dados estão pronto para ser analisados pelo classificador. Foram utilizados três classificadores: LDA, QDA e SVM.

#### 2.3.2.1 Análise Discriminante Linear

A análise discriminante linear é uma técnica multivariada muito utilizada para classificação. Cada amostra do conjunto de treino ou do conjunto de teste pode ser chamada de objecto. Cada objecto do conjunto de treino está associado a um grupo e cada objecto do conjunto de teste irá ser classificado como pertencendo a um dos grupos predefinidos pelo conjunto de treino. Os grupos podem também ser denominados de classes. Para a aplicação da análise discriminante é necessário que os grupos para os quais cada objecto possa ser classificado sejam predefinidos, conhecidos *a priori*. É então elaborada uma função matemática, chamada de função discriminante, utilizada para classificar uma nova amostra como pertencendo a um dos grupos existentes [85]. O LDA tem a vantagem de ser facilmente visualizável em termos geométricos (até 3 dimensões) e de permitir discriminar mais do que dois grupos, ou classes, diferentes sem grande complexidade. Este método permite uma discriminação de grupos com uma distribuição normal multivariada e com matriz de covariâncias muito semelhantes [86-88]. Numa análise discriminante, tem-se variáveis dependentes, que representam a classe dos objectos, e variáveis independentes, que são as características dos objectos que definem cada classe [86].

A separação entre grupos é uma linha no caso dos objectos serem definidos por duas características, um plano se forem definidos por três características e um hiperplano se os objectos forem definidos por mais do que três características. A classificação de um objecto é feita consoante o valor resultante da função discriminante

aplicada ao objecto em estudo. A função discriminante é uma combinação linear de duas ou mais variáveis independentes [86]. O valor obtido para cada objecto,  $k$ , resulta da multiplicação de cada variável independente,  $X_{ik}$ , pelos seus pesos,  $W_i$ . Conforme apresentado por Hair *et al.* [86], a função discriminante apresenta a seguinte forma:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots + W_nX_{nk} \quad \text{Equação 2.9}$$

em que  $a$  é o intercepto (constante) e  $n$  o número total de variáveis independentes.

Caso hajam mais do que dois grupos a serem discriminados, a função de classificação é derivada uma ou mais vezes, consoante o número de grupos a serem discriminados. O critério para a definição dos pesos ou coeficientes da combinação linear é a maximização da variância entre os diferentes grupos em relação à variância dentro dos grupos [86].

Cada grupo tem um centróide associado. O centróide é a média dos valores  $Z$  de todos os objectos de um grupo. Os centróides são indispensáveis para a análise dos pesos e influenciam bastante o processo de classificação [86, 89].

### 2.3.2.2 Análise Discriminante Quadrática

A Análise Discriminante Quadrática requer as mesmas condições que as apresentadas no subcapítulo anterior para análise discriminante [85]. Também para o QDA, as variáveis dependentes representam as classes e as variáveis independentes representam as características dos objectos que definem cada classe. Esta técnica permite a discriminação de classes que tenham matrizes de co-variância significativamente diferentes, enquanto os objectos da classe apresentam uma distribuição normal multivariada com a mesma média [85, 88, 90]. Este classificador oferece maior flexibilidade do que o LDA, correndo o risco de se adaptar demasiado causando “*overfitting*” [90].

Para o QDA a regra de classificação consiste em atribuir ao objecto uma classe e através de uma condição, verificar se este pertence a essa classe. Caso não pertença irá ser feito o mesmo procedimento para as classes seguintes. Num cenário de duas classes, um objecto  $x$  é classificado como pertencendo ao Grupo 1 se se verificar:

$$\begin{aligned} & \mu'_1(\Sigma_2^{-1} - \Sigma_1^{-1})x - 2x'(\Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1) + (\mu'_2\Sigma_2^{-1}\mu_2 - \mu'_1\Sigma_1^{-1}\mu_1) \\ & \geq \ln \frac{|\Sigma_1|}{|\Sigma_2|} + 2 \ln \frac{|\pi_1|}{|\pi_2|} \end{aligned} \quad \text{Equação 2.10}$$

em que as médias das classes são dadas por  $\mu_1$  e  $\mu_2$ , e as matrizes de co-variância por  $\Sigma_1$  e  $\Sigma_2$ , para os Grupos 1 e 2, respectivamente.  $\pi_1$  e  $\pi_2$  são as probabilidades *a priori* do objecto pertencer ao Grupo 1 ou ao Grupo 2, respectivamente [86].

### 2.3.2.3 Support Vector Machines

O SVM é uma útil técnica de classificação que utiliza processos de *machine learning*, ou seja, aprende através dos dados introduzidos [91, 92].

Para um conjunto de dados linearmente inseparáveis, o SVM define essa separação através de um hiperplano com uma margem associada. Para obter um hiperplano ideal, a margem tem que ser maximizada. A margem é definida pelos vectores de cada classe mais próximos da margem, os chamados *support vectors* [92-94], como mostrado na Figura 2.5.

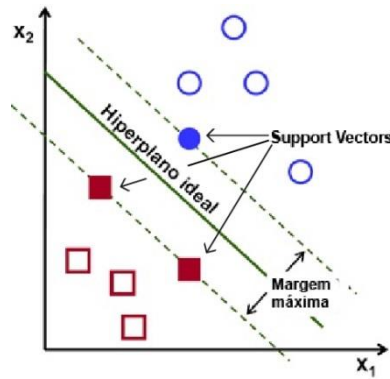


Figura 2.5 Demonstração geométrica a duas dimensões do conceito de hiperplano ideal, margem e *support vectors*.

A função genérica do hiperplano para um conjunto de dados representado por  $x$  é dada por:

$$wx + b = 0 \quad \text{Equação 2.11}$$

em que  $w$  é normal ao hiperplano e  $b$  é o termo independente [92].

No caso em que os dados são linearmente inseparáveis, para que possa ser calculado um hiperplano, os dados são transformados para um espaço de maiores dimensões (pode ser de dimensão infinita) para que os dados possam ser separáveis, tipicamente em dois grupos [91-95]. O SVM faz a transformação dimensional do espaço com o uso de uma função *Kernel* (K) [92, 95]. Um exemplo geométrico dessa transformação é apresentado na Figura 2.6.

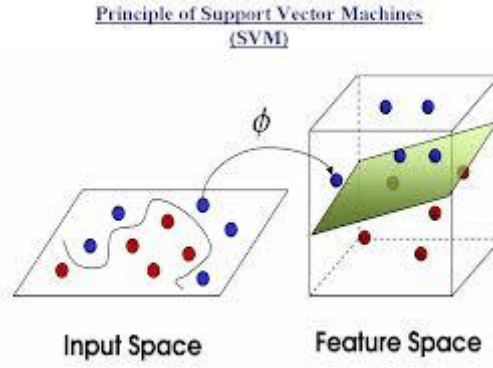


Figura 2.6 Transformação dimensional do espaço feita pela função *Kernel*.

Neste trabalho utilizou-se uma função de base radial (do inglês, *Radial Basis Function*, RBF) como *Kernel*. A equação da RBF é dada por:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \quad \text{Equação 2.12}$$

em que  $\gamma$  é o factor de escala da *Kernel* RBF [91]. Para o uso da *Kernel* é necessário que os dados possam ser representados pelo produto interno  $x_i \cdot x_j$  [92].

Com referido anteriormente, a margem é definida pelos vectores de cada grupo mais próximos do hiperplano. Porém, certos pontos, como *outliers* e ruído podem influenciar erradamente o cálculo da margem. *Outliers* são objectos com características bastante diferentes às características típicas dos outros objectos do seu grupo. Para reduzir o efeito dos *outliers* e do ruído é introduzido o conceito de variáveis de folga (em inglês, *soft margins*) [92]. Assim, o classificador será mais generalizado, sendo capaz de classificar conjuntos de teste independentes [94]. Considere-se um conjunto de objectos de treino, composto por pares  $(x_i, y_i)$ , em que  $i = 1, \dots, l$  representa cada amostra,  $x_i$  representa os vectores de entrada (as características) de cada amostra e  $y_i$  representa a respectiva classe. A optimização da margem é dada por:

$$\min_{w, b, \xi} \left[ \frac{1}{2}(w \cdot w) + C \sum_{i=1}^l \xi_i \right] \quad \text{Equação 2.13}$$

com as condições seguintes:  $y_i(w \cdot x_i + b) \geq 1 - \xi_i$  e  $\xi_i \geq 0$ , em que  $\xi_i$  é uma variável que representa o erro e que permite a flexibilização do hiperplano de separação [91, 94].

Para os casos sem variáveis de folga, a margem é sempre igual a 1, enquanto que com as variáveis de folga, a margem pode ser menor que 1, dada pela diferença  $1 - \xi_i$ . A constante  $C$  tem como objectivo equilibrar a minimização de  $w \cdot w$  (que maximiza a margem) com a garantia de que maior parte das amostras tenham uma margem de pelo menos 1 [94].

Os passos básicos da aplicação do SVM a um conjunto de dados são agora brevemente descritos. Primeiro é atribuído um número a cada classe. Por exemplo, se existem duas classes define-se uma delas com o valor -1 e a outra com o valor 1 [91]. No segundo passo, a função *Kernel* é aplicada, neste caso, a RBF. De seguida, testa-se o conjunto de treino, com o método de validação cruzada (explicado em 4.2.3), para as combinações de  $(C, \gamma)$  com o objectivo de encontrar a combinação que garanta os melhores resultados. Finalmente, o classificador é testado com um conjunto de teste independente com os valores de  $C$  e  $\gamma$  obtidos previamente [91].

## 3 Classificação de tumores da mama via simulação

Foram utilizados dois tipos de bases de dados no presente trabalho. Uma base de dados em que os tumores são simulações com modelos numéricos tais como as *Gaussian Random Spheres*, e uma outra construída a partir de modelos físicos (com material TX151). Neste capítulo 3 serão apresentados os resultados obtidos com os modelos numéricos.

### 3.1 Fantômas numéricos

Esta base de dados, construída a partir de tumores simulados por GRSs serviu de molde para a classificação feita com modelos físicos. Fez-se um tratamento dos dados e a sua classificação comparando com os resultados obtidos em [27] de modo a garantir que os classificadores posteriormente utilizados em bases de dados novas funcionariam correctamente.

Uma base de dados com mais tamanhos e formas de tumores é vantajosa no sentido em que se aproxima mais a um caso real. A principal vantagem das GRSs é a facilidade com que se podem alterar as características dos modelos, como por exemplo a nível do tamanho, da forma e da textura da sua superfície. Na maioria dos casos, os tumores benignos têm uma textura lisa e suave, com contornos ovais ou macrolobulados, enquanto os tumores malignos apresentam uma forma mais irregular, espiculada ou microlobulada e uma textura áspera. O tamanho também desempenha um papel importante sendo o tamanho médio dos tumores malignos cerca de três vezes maior que os benignos, contudo neste estudo, considera-se que são a forma e a textura que desempenham o principal papel na classificação de tumores [12, 13, 35, 37].

#### 3.1.1 Conceitos básicos das GRS

Muinonen *et al.* [96] desenvolveu um modelo matemático com o objectivo de simular asteróides e cometas. Em 2008, Davis *et al.* [42] adaptou esse modelo para simulação de modelos de tumores da mama

O tamanho e forma das GRS são especificados pela média do raio e a função de covariância do raio logarítmico.

A forma de uma GRS  $r = r(\vartheta, \varphi)$  é definida em coordenadas esféricas pelo raio logarítmico  $s = (s, \vartheta, \varphi)$ :

$$r(\vartheta, \varphi) = \alpha \exp \left[ s(\vartheta, \varphi) - \frac{1}{2} \beta^2 \right] \quad \text{Equação 3.1}$$

$$s(\vartheta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l s_{lm} Y_{lm}(\vartheta, \varphi) \quad \text{Equação 3.2}$$

Nas equações anteriores  $\alpha$  é o raio médio,  $\beta$  o desvio padrão do raio logarítmico,  $Y_{lm}$  as harmónicas esféricas ortonormais, e  $s_{lm}$  os coeficientes dos pesos das harmónicas esféricas em que  $l$  e  $m$  são o grau e a ordem de expansão, respectivamente.

A função de covariância do raio e do raio logarítmico  $\alpha^2 \Sigma_r$  e  $\Sigma_r$ , respectivamente, e as correspondentes variâncias  $\alpha^2 \sigma^2$  e  $\beta^2$  estão relacionadas através de:

$$\Sigma_r = \exp(\Sigma_s) - 1 \quad \text{Equação 3.3}$$

$$\sigma^2 = \exp(\beta^2) - 1 \quad \text{Equação 3.4}$$

A função de covariância do raio logarítmico é dada por:

$$\Sigma_s(\gamma) = \beta^2 C_s(\gamma), \quad \text{Equação 3.5}$$

onde  $\gamma$  é a distância angular entre duas direcções  $(\vartheta_1, \varphi_1)$  e  $(\vartheta_2, \varphi_2)$  e  $C_s$  a função de correlação do raio logarítmico. Na prática, as séries harmónicas têm que ser ajustadas a um certo ângulo suficientemente alto para manter uma boa precisão na geração das amostras esféricas.

A distância de correlação  $\ell$  e o ângulo de correlação  $\Gamma$  são definidas por:

$$\ell = \frac{1}{\sqrt{-C_s^{(2)}(0)}} \quad \text{Equação 3.6}$$

$$\Gamma = 2 \arcsin \left( \frac{1}{2} \ell \right) \quad \text{Equação 3.7}$$

em que  $C_s^{(2)}$  e  $\Sigma_s^{(2)}$  são a segunda derivada em ordem a  $\gamma$ .

Como foi referido anteriormente, a forma e tamanho dos modelos podem ser especificados com a alteração do raio médio  $\alpha$  e da função de covariância do raio logarítmico  $\Sigma_s$ . Manipulando estes parâmetros é possível simular tumores benignos e tumores malignos [97]. As bases do método para as GRS implementado neste trabalho



foram retiradas do “G-sphere” software desenvolvido por Muinonen *et al.* [98], traduzido em linguagem de programação Matlab.

Davis *et al.* [42] gerou modelos com três diferentes formas, lisos, microlobulados e espiculados. Conceição *et al.* [27] acrescentou modelos macrolobulados obtendo assim dois tipos de modelos benignos (lisos e macrolobulados), e dois tipos de malignos (microlobulados e espiculados). Neste trabalho a base de dados utilizada foi a mesma gerada em [27].

Os modelos lisos, macrolobulados e microlobulados são gerados variando o ângulo de correlação ( $\Gamma$ ) em intervalos de  $5^\circ$ . Quando o ângulo varia entre  $5^\circ$  e  $20^\circ$  são gerados modelos microlobulados, para os macrolobulados a variação é entre  $25^\circ$  e  $45^\circ$  e para os lisos entre  $50^\circ$  e  $90^\circ$ . Os modelos lisos têm uma superfície esférica ou elipsoidal com pequenas distorções. Os macrolobulados são como os lisos mas com largas protuberâncias ou têm formas elipsoidais com uma dimensão muito maior que a outra. Os modelos microlobulados têm pequenas protuberâncias lobulares de diferentes tamanhos na sua superfície. Os modelos espiculados são obtidos introduzindo o número desejado de espículos cónicos com 2 cm de altura a modelos lisos, como apresentado em [42]. O raio médio dos cones é o mesmo que o raio médio do corpo liso ao qual vão ser adicionados os espículos, logo, como o tamanho dos cones mantém-se constante, os espículos serão comparativamente mais prominentes em GRSs com um raio mais pequeno. O centro das GRSs é o mesmo que o centro da base dos espículos, mas os ângulos de projecção dos espículos são aleatórios.

### 3.1.2 Base de dados numérica

A base de dados estudada e trabalhada neste capítulo do trabalho é composta por 480 tumores homogéneos, de quatro tamanhos e quatro formas diferentes. Os tamanhos médios do raio de todos os tipos de tumores são 2.5, 5, 7.5 ou 10 mm. Para cada um dos tamanhos há tumores lisos, macrolobulados, microlobulados e espiculados. O número de espículos nos modelos espiculados pode ser 3, 5 ou 10.

São representados na Figura 3.1 exemplos de modelos de tumores benignos com raio médio de 5 mm.

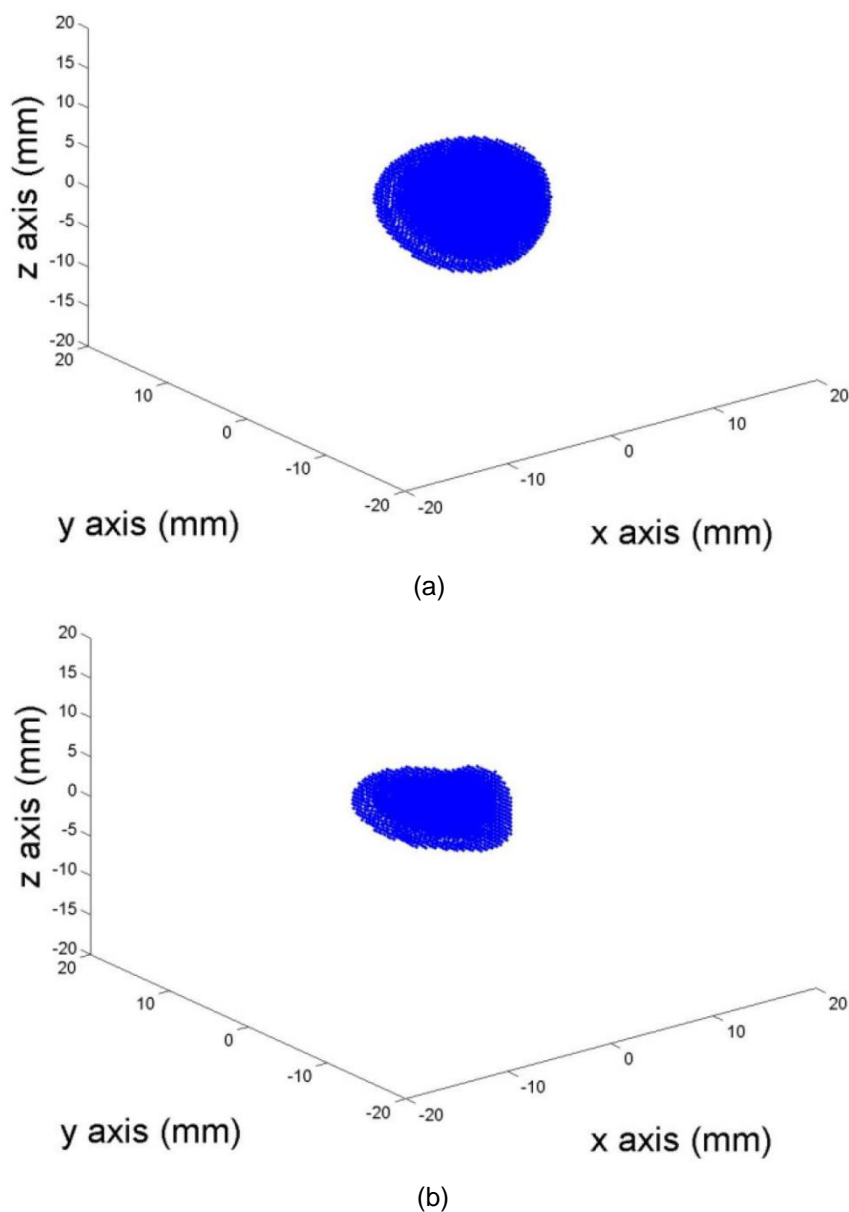
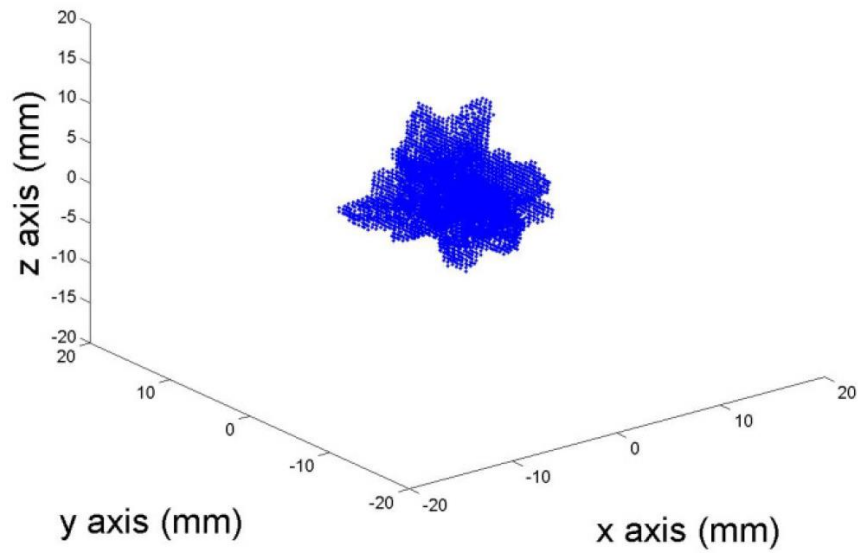
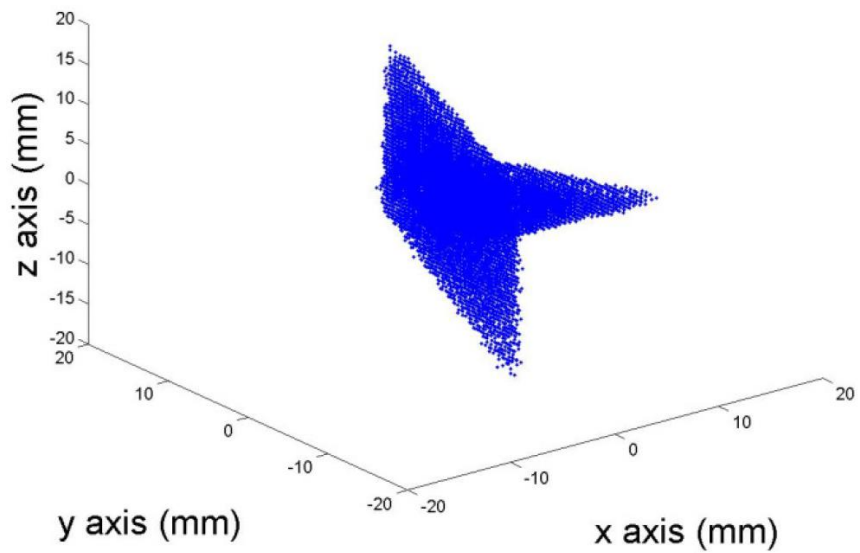


Figura 3.1 Amostras de GRS representando tumores benignos com raio médio de 5 mm. De cima para baixo: (a) liso e (b) macrobulado.

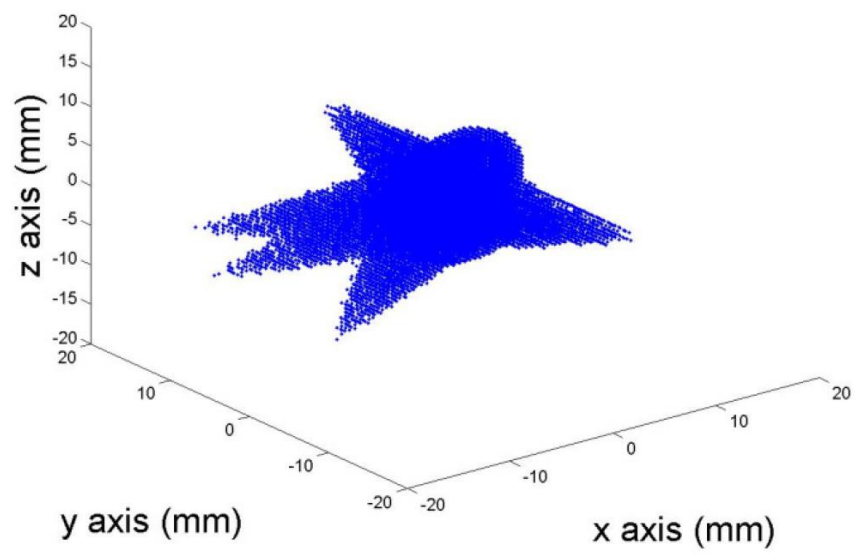
Na Figura 3.2 são mostrados alguns exemplos de modelos de tumores malignos com média de raio 5 mm. Mais especificamente um modelo macrolobulado, e três modelos espiculados, com 3, 5 e 10 espículos.



(a)



(b)



(c)

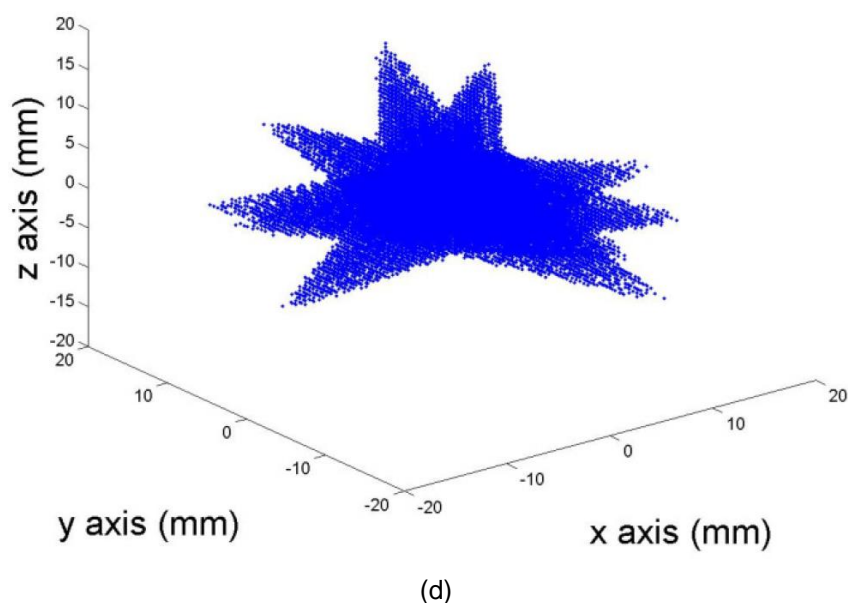


Figura 3.2 Amostras de GRS representando tumores malignos com raio médio de 5 mm. De cima para baixo: (a) microlobulado, (b) espiculado com 3 espículos, (c) espiculado com 5 espículos e (d) espiculado com 10 espículos

## 3.2 Arquitecturas

Os métodos de extração de características relevantes e os diferentes tipos de classificadores desempenham um papel fundamental na classificação mas não são os únicos factores a ter em conta. A arquitectura que o classificador irá seguir tem um papel importante na avaliação da classificação a testar, diferentes arquitecturas podem dar origem a resultados muito diferentes. Cada arquitectura diferente é definida pelas características que classifica, pela ordem pela qual são classificadas e em quantas classes divide um determinado grupo por etapa. Neste estudo, o SVM é tipicamente utilizado em classificadores binários enquanto o LDA e o QDA são utilizados tanto na divisão binária como numa divisão em quatro classes numa só separação [27].

Conceição *et al.* [27] provou que usando uma base de dados com tumores de quatro tamanhos e quatro formas diferentes, a classificação por tamanho demonstra melhores resultados que a classificação pela forma. Como o principal objectivo é classificar o tumor segundo a forma, por forma a saber se é benigno ou maligno, o ideal parece ser uma arquitectura que avalie primeiro tamanho e posteriormente a forma. Neste trabalho foram utilizadas seis arquitecturas, três delas só analisam uma característica (forma) e as restantes três classificam primeiro o tamanho e só depois a forma. Apesar de uma classificação entre benigno e maligno seja resultado espectável de um exame médico, uma classificação mais profunda que distinga os tumores lisos dos macrolobulados (no caso dos benignos) e os microlobulados dos espiculados (no

caso dos malignos) fornece uma informação clínica importante sobre o estado de desenvolvimento do tumor [43] e o tratamento escolhido será potencialmente o mais apropriado.

Cada arquitectura será agora descrita. A primeira arquitectura, Grosseira-Forma (GF), separa o conjunto das Radar Target Signatures (RTS) (assinatura no sinal de radar de um objecto) dos tumores numa só fase em dois grupos: benignos e malignos. A Detalhada-Forma (DF) faz a primeira divisão como a GF mas volta a separar cada um dos grupos em duas classes, os benignos em lisos e macrolobulados e os malignos em microlobulados e espiculados. Ambas as arquitecturas podem ser testadas com os classificadores LDA, QDA e SVM. Na Figura 3.3 são mostradas estas duas arquitecturas.

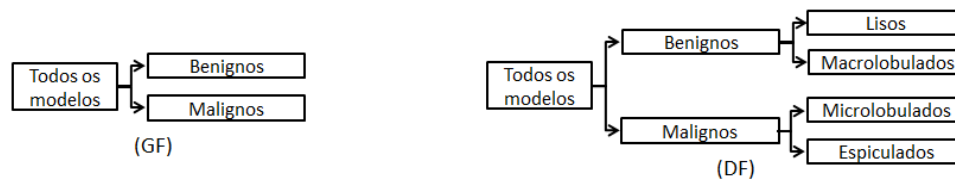


Figura 3.3 Arquitecturas baseadas só na forma dos modelos. À esquerda GF e à direita DF.

A separação Directa-Detalhada-Forma (DDF) separa assim como a DF o conjunto das RTS em lisos, macrolobulados, microlobulados e espiculados mas numa só passagem. A Directa-Detalhada-Tamanho-Directa-Detalhada-Forma (DDTDDF) faz uma primeira separação por tamanhos, dividindo os tumores pelo tamanho do seu raio em 2.5, 5, 7 e 10 mm directamente num só passo e em seguida uma separação detalhada da forma (como a DDF) a cada um destes grupos. Só o LDA e QDA foram utilizados nestas arquitecturas visto estas permitirem separações em quatro classes simultaneamente (e neste estudo o SVM não ser utilizado neste tipo de situação). As arquitecturas são mostradas na Figura 3.4.

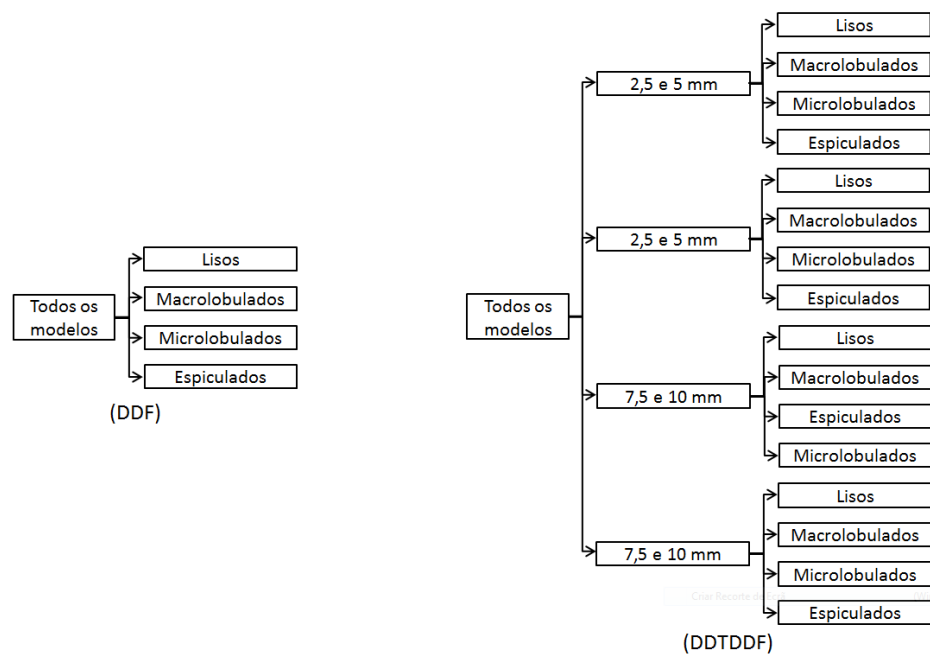


Figura 3.4 Arquitecturas que classificam detalhadamente a forma de cada tumor. À esquerda DDF e à direita DDTDDF.

Outra arquitectura utilizada foi a Grosseira-Tamanho-Grosseira-Forma (GTGF) que divide os tumores numa primeira separação em tamanhos mais pequenos (raio de 2.5 e 5 mm) num conjunto e tamanhos maiores (7.5 e 10mm) noutro. Em seguida separa cada um dos grupos resultantes em benignos e malignos. A Detalhada-Tamanho-Grosseira-Forma (DTGF) é uma arquitectura feita em três passos. O primeiro passo é idêntico ao da GTGF: no segundo passo o grupo com os tumores de 2.5 e 5 mm é dividido numa classe de tumores de raio 2.5 mm e noutra classe de tumores de raio 5 mm, o grupo com tumores de raio 7.5 e 10 mm é subdividido em cada um dos dois tamanhos: no terceiro passo os tumores são classificados quanto à sua forma grosseira: benigno ou malignos. Como todas as divisões são binárias, os três classificadores (LDA,QDA e SVM) foram testados nestas arquitecturas, que são mostradas na Figura 3.5.

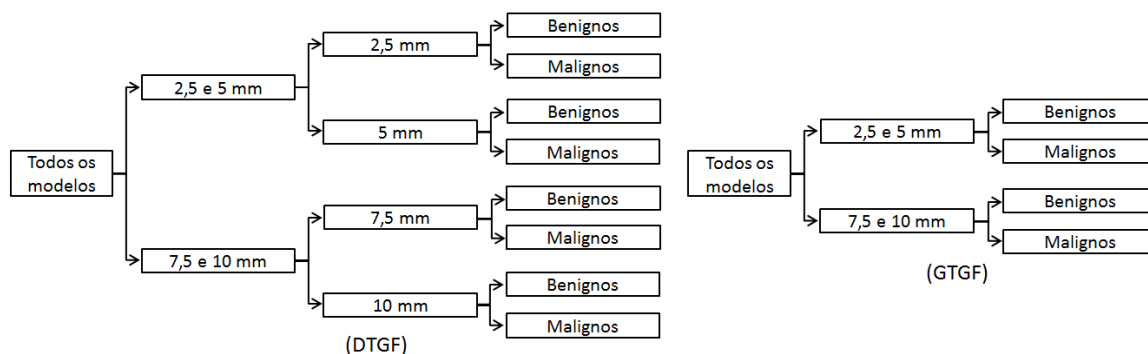


Figura 3.5 Arquitecturas que classificam os tumores em benignos e malignos com uma separação por tamanhos inicial. À Esquerda DTGF e à direita GTGF.

### 3.3 Resultados

#### 3.3.1 Características do PCA, ICA e DWT

Os métodos para extração de características relevantes servem para diminuir a dimensão dos dados a processar mas mantendo a informação mais importante e mais relevante. Assim, os processos algorítmicos de tratamento de dados são mais rápidos mas a performance do classificador não é significativamente afectada. Os três métodos utilizados nesta parte do trabalho foram o PCA, o ICA e o DWT. Neste sub-capítulo será mostrado como é feita a diminuição da dimensão dos sinais através de cada um deles. Começamos por analisar o caso do PCA para descobrir qual o número de componentes ideal a utilizar. Para isso foram utilizadas as arquitecturas Grosseira-Forma (GF), que separa os tumores em benignos ou malignos, e Grosseira-Tamanho, que separa os tumores mais pequenos dos maiores, (primeira divisão da DTGF e GTGF, respectivamente). Analisado o comportamento da curva que relaciona a performance do classificador e o número de componentes principais utilizados é possível perceber o número onde a partir do qual não compensa utilizar mais componentes (comparativamente ao tempo de processamento acrescido). Na Figura 3.6 está demonstrada essa relação para o LDA (linha azul) e para o QDA (linha preta). À esquerda é mostrado o gráfico com a Grosseira-Forma e à direita com a Grosseira-Tamanho.

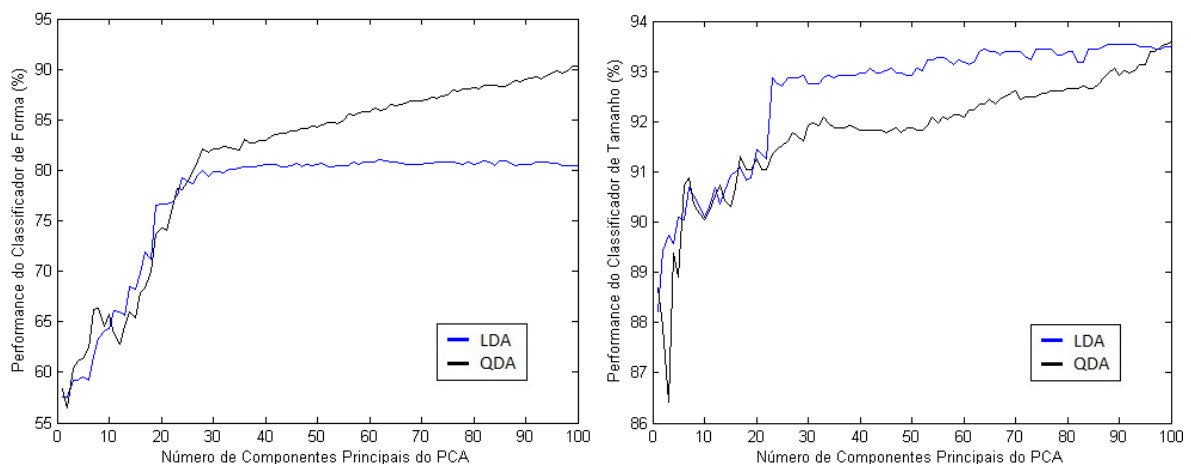


Figura 3.6 Relação entre número de componentes principais usados e a performance dos classificadores LDA (linha azul) e QDA (linha preta) para o GF (à esquerda) e o GT (à direita).

Com base nestes gráficos concluiu-se que o número ideal de componentes principais a utilizar seria um número à volta dos 30 componentes principais. Por isso, neste trabalho foram

utilizados os primeiros 30 componentes principais para a extracção de características relevantes com o PCA.

Quanto ao ICA, o algoritmo FastICA [99] produz 16 componentes independentes para representação de cada sinal.

Para o DWT, a *wavelet* utilizada foi a *Coiflet 5* e as melhores características são realçadas com uma decomposição feita em duas iterações. Ou seja, os coeficientes são obtidos depois dos sinais sofrerem uma decomposição por um filtro passa-baixo e, com o resultado dessa decomposição, repetir-se o mesmo processo. Como cada iteração reduz os sinais a metade, dos 100 iniciais ficamos com 25 para a classificação. Um esquema da decomposição feita pela DWT é apresentado na Figura 3.7.

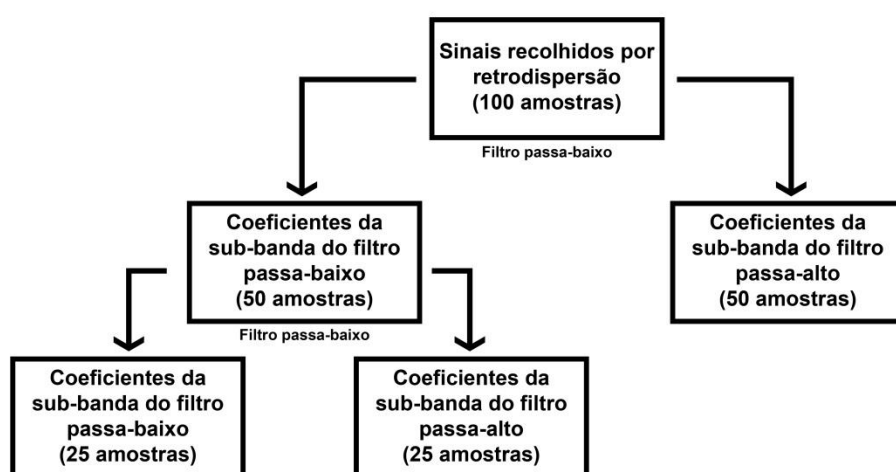


Figura 3.7 Esquema das decomposições do DWT com duas iterações adaptado de [27].

#### 3.3.2 Comparação entre os classificadores LDA e QDA

Neste sub-capítulo serão comparadas as classificações combinando os métodos de extracção de características relevantes PCA, ICA e DWT com os classificadores LDA e QDA para as seis diferentes arquitecturas descritas no sub-capítulo 3.2. Os resultados são apresentados na Tabela 3.1 em três colunas diferentes. A primeira coluna representa a performance parcial do classificador de tamanho, a segunda a performance parcial do classificador de forma e a terceira a performance global do classificador tamanho-forma, quando é o caso. As performances parciais indicam a capacidade que o classificador tem de dividir correctamente os tumores com base só numa característica enquanto que a performance global engloba a classificação feita por tamanho e posteriormente pela forma. O cálculo da performance global é feito multiplicando as performances parciais do classificador com base só no tamanho e só na forma. Para



cada arquitectura, o valor da combinação que obteve melhor resultado está realçado a azul.

Numa primeira análise à Tabela 3.1 verificamos que na maior parte dos casos os melhores valores resultam da combinação do DWT com o QDA. Observa-se, também, que os resultados de performance mais elevados obtêm-se quando o classificador classifica os tumores quanto ao tamanho e posteriormente quanto à forma.

Mais detalhadamente analisemos os métodos para extracção de características relevantes. Verifica-se que o DWT foi o método para extracção de características relevantes mais eficaz em todas as combinações destacadas a azul na Tabela 3.1. O DWT mostrou assim ser o método mais eficaz com uma média para o LDA e QDA de 75,57% de performance, ainda que sempre com valores muito próximos do PCA com uma média de 73,51%. O ICA não obteve mais que 62,79% de performance média para o LDA e QDA e provou ser o pior método para extracção de características relevantes.

No que diz respeito às classificações pelo tamanho ou pela forma individualmente várias observações podem ser feitas através da Tabela 3.1. Em concordância com a Figura 3.6, verifica-se que a performance grosseira de tamanho (analisada através da arquitectura Grosseira-Tamanho-Grosseira-Forma) é 9,01% superior à da grosseira de forma (92,92% contra 83,91%). Observa-se também que para o caso do tamanho o classificador mais eficiente é o LDA, e para a forma o classificador mais eficiente é o QDA, tanto nas arquitecturas grosseiras como nas detalhadas. Com as arquitecturas tamanho-forma, as performances globais foram muito semelhantes, com o QDA a obter média de performance 2,06% superior à do LDA (64,71% contra 62,65%).

Também sobre as arquitecturas dos classificadores se podem observar resultados interessantes. Uma separação por tamanhos antes da classificação de forma permite obter melhores resultados, o que se pode verificar quando comparados o valor da arquitectura grosseira-forma (83,91%) com a performance parcial de forma nas arquitecturas grosseira-tamanho-grosseira-forma (85,00%) e detalhada-tamanho-grosseira-forma (88,54%). O mesmo se comprova quando comparados o valor da arquitectura directa-detalhada-forma (62,76%) com o valor da mesma mas quando sujeita a uma divisão prévia directa-detalhada-tamanho (81,61%). A melhor separação por tamanhos é a grosseira-tamanho, obtendo sempre valores superiores a 90%, enquanto que as outras duas separações (directa-detalhada-tamanho e detalhada-tamanho) nunca apresentam valores superiores a 83,65%.

Em termos de performances globais, a arquitectura com uma melhor eficiência é a grosseira-forma, que só classifica baseada numa característica – a forma do tumor - e só numa etapa. Este resultado é esperado por ser um classificador com pouca especificidade (separação só entre tumores benignos e malignos) e por só ter uma

etapa. Nos classificadores com duas etapas quando classificam mal um tumor numa das características analisadas, mesmo que a classificação baseada na outra característica seja bem feita, a classificação final será afectada.

Tabela 3.1 Performance das classificações parciais de tamanho e forma e classificação geral de tamanho-forma combinando os métodos de extracção de características relevantes PCA, ICA e DWT com os classificadores LDA e QDA para as seis diferentes arquitecturas.

Arquitecturas dos Classificadores	Métodos para extracção para características relevantes	Classificador parcial de tamanho (%)		Classificador parcial de forma (%)		Classificador tamanho-forma (%)	
		LDA	QDA	LDA	QDA	LDA	QDA
Grosseira-Forma	PCA			79.84	82.14		
	ICA			69,80	67,86		
	DWT			80.26	<b>83.91</b>		
Detalhada-Forma	PCA			52.45	53.80		
	ICA			40,42	39,38		
	DWT			53.80	<b>59.38</b>		
Directa-Detalhada-Forma	PCA			54.06	56.51		
	ICA			42,24	38,59		
	DWT			54.95	<b>62.76</b>		
Directa-Detalhada-Tamanho-Directa-Detalhada-Forma	PCA	79.01	78.91	63.49	74.58	50.16	58.85
	ICA	73,28	73,28	48,65	61,04	35,65	44,73
	DWT	79.69	<b>81.41</b>	64.58	<b>81.61</b>	51.46	<b>66.44</b>
Grosseira-Tamanho-Grosseira-Forma	PCA	92.76	91.93	84.17	83.65	78.07	76.89
	ICA	90,99	90,68	72,19	74,01	65,68	67,11
	DWT	<b>92.92</b>	92.45	83.65	<b>85,00</b>	77.72	<b>78.58</b>
Detalhada-Tamanho-Grosseira-Forma	PCA	83.13	76.61	87.66	86.51	72.86	66.28
	ICA	75,83	72,55	77,14	73,80	58,49	53,54
	DWT	<b>83.65</b>	79.01	88.18	<b>88.54</b>	<b>73.76</b>	69.96

### 3.3.3 Avaliação global dos classificadores, dos métodos para extracção de características relevantes e das arquitecturas

Neste sub-capítulo irá ser feita uma avaliação global das performances dos classificadores, métodos de extracção de características e arquitecturas, analisando as médias para todas as combinações estudadas.

Primeiramente será analisado qual o melhor método para extracção de características para cada classificador. Para isso foi construída a Tabela 3.2 com as performances médias para todas as arquitecturas estudadas. Os valores mais elevados para cada classificador estão evidenciados a azul.

Tabela 3.2 Média dos resultados para cada combinação entre métodos de extracção de características relevantes e os classificadores.

Métodos para extracção de características relevantes	Performance com os classificadores (%)	
	LDA	QDA
PCA	73,14	73,89
ICA	62,53	63,05
DWT	73,72	77,42

Apesar de se obterem resultados semelhantes, o melhor método para extracção de características a utilizar com o LDA é o DWT, obtendo uma performance superior em 0,58% que o PCA. Também para o QDA, o DWT demonstra ser o melhor método para extracção de características com uma performance 3,53% superior que o PCA. PCA e DWT mostram ter resultados muito semelhantes enquanto o ICA parece ser a pior opção, com valores substancialmente inferiores.

Quanto às arquitecturas, são apresentados na Tabela 3.3 os valores médios das classificações feitas com cada arquitectura (média dos valores com LDA e QDA).

Tabela 3.3 Performances médias global resultantes de todas as combinações para cada arquitectura.

Arquitecturas	Performance Média (%)
Grosseira-Forma	77,30
Detalhada-Forma	49,87
Directa-Detalhada-Forma	51,52
Directa-Detalhada-Tamanho-Directa-Detalhada-Forma	64,82
Grosseira-Tamanho-Grosseira-Forma	<b>82,14</b>
Detalhada-Tamanho-Grosseira-Forma	75,97

Da Tabela 3.3 verifica-se que a arquitectura global com melhor performance é a Grosseira-Tamanho-Grosseira-Forma (82,14%), a mesma que tinha obtido melhores resultados absolutos quando combinando DWT e QDA na Tabela 3.1. De notar também que as arquitecturas faseadas em tamanho-forma têm notoriamente melhores resultados que as arquitecturas que avaliam só uma característica.

## **4 Classificação de tumores da mama com uso de um Protótipo de UWB**

Neste capítulo será feito um estudo parecido ao do capítulo anterior mas com uma base de dados composta por modelos físicos de tumores e a aquisição de dados feita através de um protótipo de UWB. Os testes com fantômas físicos são o passo seguinte para justificar o teste da técnica em mamas reais. Este teste é de grande importância porque é utilizado um protótipo da tecnologia (e não são testados apenas cenários de simulação).

Irá ser utilizada apenas uma porção das arquiteturas e métodos para extracção de características anteriormente utilizados com os modelos numéricos, devido ao facto de ser uma base de dados mais pequena e por não se poderem fabricar tamanhos e formas tão específicos de modelos de tumores manualmente.

Depois de apresentados os resultados e feita a discussão são feitas as conclusões. As limitações do trabalho e as perspectivas futuras são também apresentadas no final do capítulo.

### **4.1 Fantômas Físicos**

Foram fabricados, através do processo explicado no próximo sub-capítulo, 32 tumores físicos – 19 benignos e 13 malignos – com tamanhos entre 0,5 e 2 cm de raio. Para o número de modelos benignos e malignos ser o mesmo, foram escolhidos os 13 modelos benignos com tamanhos mais próximos aos dos malignos. Na parte anterior do trabalho foram só estudados modelos homogêneos da mama. Nesta parte, foi adicionada uma porção modelada de tecido fibroglandular junto do modelo do tumor, por forma a considerar alguma heterogeneidade no modelo da mama. A distância entre o tecido fibroglandular e o tumor é em média 3 cm, mas esta distância é variável durante as medições para cada modelo de tumor. Como os modelos são moldados à mão não é possível distinguir os modelões de tumores benignos entre lisos e macrolobulados ou, para os modelos malignos, entre microlobulados e espículados.

Nesta parte do trabalho a classificação só será feita entre malignos e benignos. O tamanho também não é considerado para fins de classificação porque a variação dos tamanhos dos modelos é pequena e não seria possível separá-los por grupos. Assim, a

única arquitectura utilizada é a Grosseira-Forma (apresentada na Figura 3.3 à esquerda), que separa directamente todos os tumores da base de dados em benignos ou malignos. Alguns dos modelos são mostrados na Figura 4.1.



Figura 4.1 Modelos físicos de tumores. Em cima, modelos de tumores malignos. Em baixo: modelos de tumores benignos.

#### 4.1.1 Fabrico dos fantômas

Os materiais utilizados para o fantôma imitam as verdadeiras propriedades dos tecidos da mama [100]. O fantôma da mama foi simulado usando um cilindro de acrilonitrilo de estireno com um diâmetro de 13 cm e de altura 35 cm cheio com glicerina, visto que a permitividade eléctrica da glicerina imita a média de valores encontrados nas regiões de baixa densidade da mama [72].

Os fantômas do tumor foram simulados com uma mistura de pó solidificador TX 151 numa proporção de volume de 6:1 de água para pó, Figura 4.1. Para simular o tecido fibroglandular é feito um modelo também com o pó TX 151 mas em proporções de volume de 4:1 de água para o pó, obtendo-se propriedades semelhantes às do tecido fibroglandular reportadas em [72].

## 4.2 Resultados

#### 4.2.1 Número de components principais do PCA

Neste parte do trabalho o único método para extracção de características relevantes utilizado foi o PCA. O ICA não foi utilizado por ter sido o pior método com os

modelos numéricos no capítulo anterior e o DWT porque o sinal relevante para classificar era demasiado curto (em termos de amostras de tempo) para que o DWT pudesse ser utilizado. Para o PCA foram utilizados 13 componentes principais, devido à impossibilidade do programa Matlab trabalhar com mais para a base de dados em estudo, como mostra a Figura 4.2., onde está representada a relação entre o número de componentes principais usados e a performance dos classificadores LDA (a azul) e QDA (a preto) para a classificação de tumores em os modelos homogéneos da mama (à esquerda) e em modelos heterogéneos da mama (à direita).

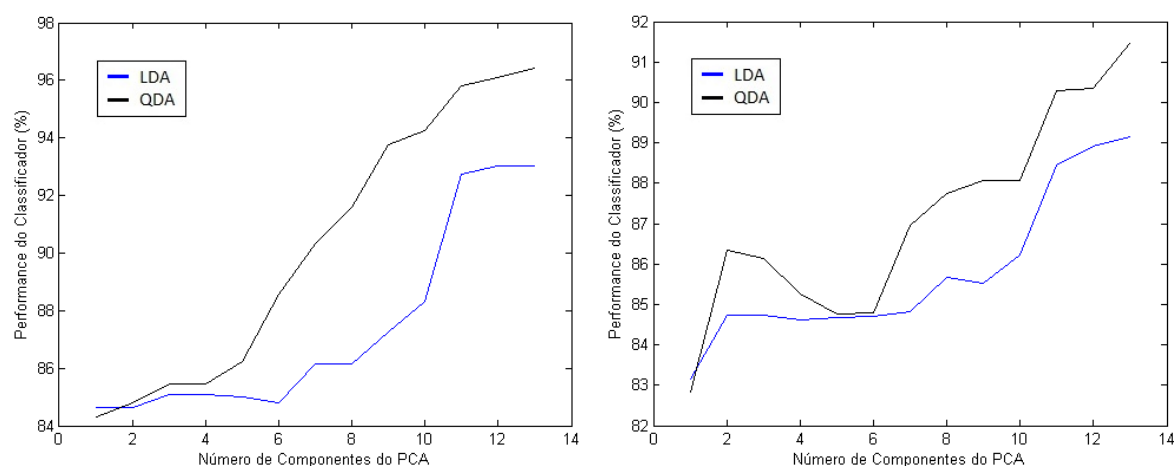


Figura 4.2 Relação entre número de componentes principais usados e a performance dos classificadores LDA (linha azul) e QDA (linha preta) para os modelos homogéneos (à esquerda) e heterogéneos (à direita).

#### 4.2.2 Número de antenas

O exame com o protótipo foi feito com um total de 144 antenas receptoras de sinal. Nem todas as antenas fornecem a mesma quantidade e qualidade de informação e por isso foi feito um estudo para tentar otimizar a informação e diminuir o número de antenas a estudar por forma a diminuir o tempo de processamento do sinal através do algoritmo de classificação. Na Figura 4.3 é mostrado o gráfico resultante desse estudo, representando a performance do classificador em relação ao número de antenas estudado para os modelos homogéneos (à esquerda) e para os modelos heterogéneos da mama (à direita). As antenas estão ordenadas da mais próxima ao tumor (antena 1) à mais longe do tumor (antena 144).

#### 4. Classificação de tumores da mama com uso de um Protótipo de UWB

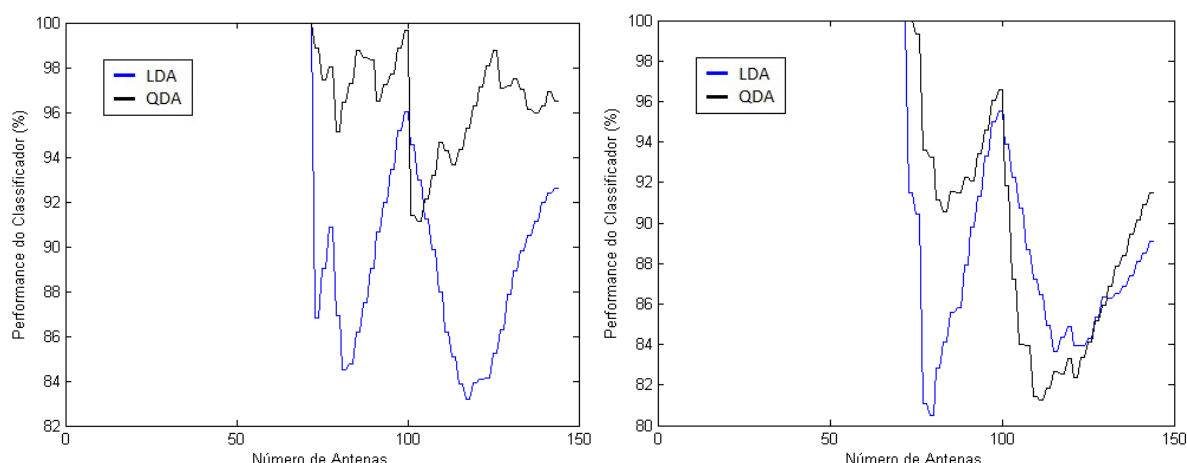


Figura 4.3 Relação entre número de antenas usadas para a classificação e a performance dos classificadores LDA (linha azul) e QDA (linha preta) para os modelos homogêneos (à esquerda) e heterogêneos da mama (à direita).

Foi também feita uma classificação só com a antena mais perto - antena 1 - e a antena mais distante do tumor - antena 73 - (a antena 2 e a antena 144 estavam à mesma distância do tumor). Na Tabela 4.1 estão mostrados os resultados dessa classificação para os modelos homogêneos e heterogêneos da mama.

Tabela 4.1 Classificação utilizando a antena mais próxima e a antena mais distante do tumor para os modelos homogêneos e heterogêneos da mama.

Antena a usar para a classificação	Classificação para os modelos homogêneos (%)		Classificação para os modelos heterogêneos (%)	
	LDA	QDA	LDA	QDA
<b>Antena 1</b>	100	100	100	100
<b>Antena 73</b>	86,83	98,89	91,52	99,95

Analisando estes dois cenários de classificação seria de esperar uma utilização inferior a 75 antenas (as mais próximas do modelo físico de tumor) para obter melhores resultados de classificação. No entanto, para o presente trabalho, foram utilizadas as 144 antenas devido à base de dados ser muito pequena e, por isso, o classificador se ajustar facilmente só com a classificação das antenas mais próximas ao modelo físico de tumor. Usando as 144 antenas simula-se com mais rigor uma situação real uma vez que nessa situação não se sabe onde o tumor está localizado e por isso quais as antenas mais próximas do mesmo. Utilizando as antenas mais distantes, as medições ficam também sujeitas a mais ruído, que estará presente num qualquer teste a um tumor real.



### 4.2.3 Comparação entre LDA, QDA e SVM

Neste sub-capítulo irão ser analisados os resultados das classificações dos modelos físicos de tumores combinando o PCA como método para extracção de características relevantes e os classificadores LDA, QDA e SVM. Só foi utilizada a arquitectura Grosseira-Forma (GF) como explicado em 4.1.

Como foi mencionado anteriormente, nesta parte do trabalho será utilizado um classificador de *machine learning*, *Support Vector Machines*. A performance da classificação por SVM depende da optimização de dois parâmetros,  $\gamma$  e  $C$ , como descrito em 2.3.2.3. Tanto para o caso dos fantômas da mama homogêneos como heterogêneos, os parâmetros óptimos são os seguintes:  $\gamma=2^2$  e  $C=2^3$ . Foram feitos testes com várias combinações de valores e foi analisado que a performance máxima se verificava com estes valores.

Todos os métodos de classificação (LDA, QDA e SVM) são aplicados aos dados previamente processados pelo método de validação cruzada (do inglês, *cross-validation*), mais precisamente *K-fold cross-validation*, para que a performance de cada classificador seja avaliada usando conjuntos de teste independentes dos conjuntos de treino [91]. Para o *K-fold cross-validation*, o conjunto dos tumores é dividido em K grupos (13 neste caso), cada um deles contendo um tumor benigno e um maligno. Cada grupo é testado tendo em conta os outros K-1 grupos que constituem o grupo de treino, obtendo-se K diferentes performances. No final, é feita a média das performances, obtendo-se então a performance final.

Para o cálculo da performance podem ser usadas duas abordagens diferentes, analisando os sinais individualmente ou agrupando-os. Para cada tumor foram recolhidos 144 sinais, cada um correspondente a uma antena, e cada sinal atribui uma classificação ao tumor. Se, para o cálculo da performance final, se considerar a classificação de cada um dos sinais individualmente, diz-se que os sinais não estão agrupados. Se, para cada tumor, se considerar a classificação obtida pela maioria dos sinais, diz-se que os sinais estão agrupados.

Na Tabela 4.2 são mostrados os resultados para os modelos homogêneos e heterogêneos da mama, para o caso em que os sinais estão agrupados e não agrupados.

Tabela 4.2 Resultados das classificações (com sinais agrupados e não agrupados) dos modelos físicos de tumores combinando o PCA com os classificadores LDA, QDA e SVM para os modelos homogêneos e heterogêneos da mama. A arquitetura utilizada foi a Grosseira-Forma.

Tipo de classificação	Classificação para os modelos homogêneos (%)			Classificação para os modelos heterogêneos (%)		
	LDA	QDA	SVM	LDA	QDA	SVM
<b>Sinais agrupados</b>	88,46	92,31	96,15	84,62	88,46	92,31
<b>Sinais não agrupados</b>	87,10	89,34	90,95	84,29	86,59	87,07

Numa análise global à Tabela 4.2 observa-se que se obtiveram melhores resultados para os modelos homogêneos da mama que para os heterogêneos, como era de esperar. Quanto aos classificadores também se verifica o previsto, o SVM é o classificador com os resultados de performance mais elevados, seguido do QDA e por fim o LDA, ainda que com diferenças pequenas nas performances entre todos.

Analisemos com mais detalhe os classificadores. O classificador de *machine learning* SVM foi o classificador com as performances mais elevadas. Para os modelos homogêneos da mama, o SVM, relativamente ao QDA e LDA, respectivamente, obteve uma performance superior em 3,84% e 7,69% quando os sinais foram agrupados e em 1,61% e 3,85% no caso dos sinais não agrupados.

No caso dos modelos heterogêneos da mama, o SVM também mostrou ser o melhor classificador, tanto para os sinais agrupados como para o sinais não agrupados. Para os sinais agrupados, o SVM obteve resultados 3,85% superiores aos do QDA e 7,69% aos do LDA. Numa avaliação com os sinais não agrupados, o SVM obteve resultados 0,48% e 2,78% superiores em relação ao QDA e LDA, respectivamente. Numa perspectiva geral, fazendo a média dos resultados de todas as classificações para cada classificador verifica-se mais uma vez que o SVM é o melhor classificador com uma média de 91,62%, sendo somente 2,45% superior à média de 89,18% do QDA, e 5,5% superior à média de 86,12% do LDA.

Analisando os resultados para os modelos homogêneos e heterogêneos da mama, como fazia antever pela Figura 4.2 e Figura 4.3, percebe-se que a classificação é mais eficaz para o caso dos modelos homogêneos da mama, obtendo sempre performances mais elevadas para todos os classificadores. Para uma classificação com os sinais não agrupados os resultados são superiores no cenário homogêneo, em 2,81%, 2,75% e 3,88% para o LDA, QDA e SVM, respectivamente. No caso da classificação com os sinais agrupados, a diferença para os três classificadores entre os

modelos homogêneos e heterogêneos da mama é sempre superior em 3,84% no cenário homogêneo, o que corresponde a mais um tumor ser bem classificado no conjunto de 26 tumores a serem testados. Em termos globais, calculando a média de todas as classificações feitas para os modelos homogêneos e heterogêneos da mama obtemos uma performance média de 90,72% para o caso homogêneo, 3,50% superior aos 87,22% obtidos no caso heterogêneo.

No que diz respeito ao tipo de classificação, entre sinais agrupados e sinais não agrupados, a maior diferença de resultados verifica-se para o SVM com a classificação com sinais agrupados a obter uma performance de 5% superior tanto para os modelos homogêneos da mama como para os heterogêneos. Para o LDA e QDA verifica-se o mesmo, mas com valores mais baixos, entre os 0,33% e os 2,97%. Na globalidade, a média de todas as classificações feitas com os sinais agrupados é 2,8% superior à média para os sinais não agrupados (90,39% *versus* 87,56%).

Observando a classificação para cada tumor foi possível verificar que os tumores benignos são mais vezes classificados como malignos do que o oposto, ou seja, tumores malignos serem classificados como benignos.

De notar que apesar das diferenças entre classificadores e tipos de tumores, todas as classificações obtiveram performances acima dos 84%, resultados bastante satisfatórios e promissores.



## 5 Considerações finais

### 5.1 Conclusões

O primeiro objectivo deste trabalho era a classificação em termos de tamanho e forma de 480 modelos numéricos de tumores de formas e tamanhos diferentes. Para tal, foram estudados três métodos para extracção de características relevantes (PCA, ICA e DWT), dois classificadores (LDA e QDA) e seis arquitecturas de classificação. Para cada arquitectura foram obtidos resultados de performance de classificação com as combinações classificador *versus* método para extracção de características possíveis. Após analisados os resultados, foram feitas algumas conclusões, apresentadas de seguida.

No caso do método PCA, os melhores resultados mostraram ser aqueles obtidos utilizando 30 componentes principais. Para o DWT, a melhor performance foi obtida quando as características foram extraídas da sub-banda de baixa frequência depois de aplicada duas vezes um filtro passa-baixo com a wavelet *Coiflet 5*.

No que diz respeito a arquitecturas, para uma classificação que distinga os tumores só em benignos e malignos, a arquitectura com melhor performance é a grosseira-forma. Porém, em termos globais, como mostra a Tabela 3.3, a arquitectura com melhor média de performances é a grosseira-tamanho-grosseira-forma. Em termos de performances parciais, em todas as arquitecturas estudadas a performance do classificador parcial da forma é melhor quando precedida de uma pré-classificação por tamanho.

O método para extracção de características relevantes com o qual a classificação é mais eficiente é o DWT, com resultados superiores mas muito próximos aos obtidos com PCA. O ICA mostrou ser o pior método para extracção de características.

Quanto aos classificadores (LDA e QDA) nenhum obteve resultados significativamente melhores que o outro. O QDA apresenta melhores resultados na globalidade mas por diferenças percentuais muito pequenas. É de notar que para a classificação por tamanho o LDA apresenta melhores resultados e, para a classificação por forma é o QDA o mais eficiente.

Estes resultados são promissores para diagnóstico de cancro da mama numa fase inicial (situação de rastreio). De notar que não foi utilizado nenhum classificador mais

eficaz nesta fase do trabalho, por exemplo de *machine learning*, que potencialmente iria aumentar as performances do classificador.

O objectivo principal da segunda parte desta dissertação concentrava-se na classificação com base na forma de uma base de dados de fantômas físicos de tumores para o cenário de um fantôma da mama homogéneo e heterogéneo (adicionando uma porção modelada de tecido fibroglandular). Esta base de dados é constituída por 26 fantômas físicos de tumores, 13 com a forma de tumores benignos e 13 com a forma de tumores malignos. Foi utilizado um protótipo pré-clínico de imagem de UWB para a obtenção dos dados descrito em 2.2.1. Como se pretende uma classificação com base só na forma, foi usada a arquitectura Grosseira-Forma (GF). O método utilizado para extracção de características relevantes foi o PCA e os classificadores testados: LDA, QDA e SVM. Com base nos resultados apresentados e discutidos em 4.2.3 foram feitas as conclusões apresentadas de seguida.

De um modo geral, todos os classificadores apresentaram resultados bastante satisfatórios com destaque para o SVM que apresentou uma média de performances de 91,62%, o que representa aproximadamente uma classificação correcta de 24 tumores dos 26 da base de dados. O SVM mostrou, assim, ser o classificador mais eficiente para o presente estudo, tanto para o caso dos fantômas homogéneos da mama como para os modelos heterogéneos. Este classificador de *machine learning* classificou em média mais um tumor correctamente em relação aos outros dois classificadores, o que, numa base de dados de 26 tumores, é estatisticamente relevante.

No que diz respeito ao diferente tipo de tumores era esperado que os melhores resultados fossem obtidos com os modelos homogéneos da mama. Quando comparadas as performances obtidas para os modelos homogéneos da mama e para os modelos heterogéneos, conclui-se que os melhores resultados são para o cenário menos complexo, os modelos homogéneos, com uma performance média de 90,72% *versus* 88,15%, relativa aos modelos heterogéneos. Este resultado era esperado tendo em conta que no caso dos modelos heterogéneos da mama estava presente uma porção modelada de tecido fibroglandular junto do modelo do tumor, o que inevitavelmente interferiu nos sinais analisados. Porém, quando analisados os resultados das performances observou-se que a diferença é somente de 2,57% em média, o que corresponde a menos do que um tumor mal classificado. Este resultado é bastante optimista porque o caso dos modelos heterogéneos da mama é o que se assemelha mais com a realidade (ainda que distante). Estes valores mostram que o tecido fibroglandular, nas condições estudadas, apesar de interferir nos resultados, não faz com que as performances dos classificadores baixem a um nível considerável. No

caso estudado o tecido fibroglandular estava numa pequena porção e afastado do tumor cerca de 3 cm. Num caso real a quantidade de tecido fibroglandular em redor do tumor pode ser maior assim como a distância entre os dois mais curta, podendo perturbar consideravelmente os resultados da performance do classificador (ou os sinais de retrodispersão dos tumores).

Quanto aos diferentes tipos de classificação, com sinais agrupados ou não agrupados, visto a diferença de performances ser pouca (menos de 2% em média), conclui-se que as classificações de cada posição da antena para cada tumor são iguais na maior parte dos casos. Algumas posições da antena podem fazer uma classificação sobre um tumor diferente do que as restantes, no entanto é raro que tal aconteça num número elevado de posições da antena.

Numa perspectiva geral, os objectivos propostos foram atingidos e com resultados bastante optimistas e motivadores para a continuação do estudo e desenvolvimento do protótipo de UWB. Assim como com os modelos numéricos de tumores, também o estudo com os modelos físicos de tumores é um bom indicador de que este método poderá evoluir ao ponto de ser utilizado em casos clínicos.

## **5.2 Limitações e perspectivas futuras**

Apesar dos bons resultados obtidos, este estudo teve algumas limitações. A maior limitação deste trabalho foi o tamanho da base de dados para os modelos físicos de tumores. Uma base de dados com 26 tumores é suficiente para construir um classificador e obter resultados através dela, como foi feito neste estudo, mas não é suficiente para conceder a credibilidade e validade estatística que um classificador precisa de ter. O método de validação cruzada foi utilizado para colmatar o baixo número de tumores, no entanto, uma base com mais tumores irá reforçar a credibilidade dos resultados.

Outra limitação enfrentada no decorrer da recolha de dados deve-se ao tamanho dos fantômas físicos dos tumores, não havendo uma grande variedade de tamanhos. Devido a este facto não foi possível fazer uma classificação também por tamanho.

Devido ao sinal no domínio do tempo estar muito discretizado, não foi possível implementar um método para extracção de características que inicialmente estava previsto, o DWT.

Para estudos futuros seria desejável uma base de dados consideravelmente maior e com uma variedade maior de tamanhos dos modelos de tumores por forma a consolidar a credibilidade do método de classificação e do próprio protótipo. Aumentar a

quantidade de tecido fibroglandular e a posição do mesmo poderá ser também importante para uma melhor modelação de um caso real. Seria também desejável uma amostragem em termos de tempo menos discretizada para cada sinal (uma maior resolução). A utilização de outros métodos para extracção de características relevantes e de outros classificadores, como as redes neuronais, poderiam também contribuir para uma classificação mais precisa.



# Referências Bibliográficas

1. Li, C., et al., *In vivo Breast Sound-Speed Imaging with Ultrasound Tomography*. Ultrasound in Medicine & Biology, 2009. **35**(10): p. 1615-1628.
2. Ferlay, J., et al. *GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10*. 2013 ; Disponível em: <http://globocan.iarc.fr>.
3. WHO | Cancer. 2013 ; Disponível em: <http://www.who.int/cancer/en/>.
4. WHO | The top 10 causes of death. 2013 ; Disponível em: <http://www.who.int/mediacentre/factsheets/fs310/en/index.html>.
5. Cancer), G.I.A.f.R.o. *Cancer Fact Sheets: Breast Cancer (Breast Cancer Incidence, Mortality and Prevalence Worldwide in 2008)*. 2008; Disponível em: <http://globocan.iarc.fr/factsheet.asp>.
6. Evron, E., et al., *Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR*. Lancet, 2001. **357**(9265): p. 1335-6.
7. Bird, R.E., T. Wallace, and B. Yankaskas, *Analysis of Cancers Missed at Screening Mammography*. Radiology, 1992. **184**: p. 613-617.
8. Huynh, P.T., A.M. Jarolimek, and S. Daye, *The False-Negative Mammogram*. RadioGraphics, 1998. **18**: p. 1137-1154.
9. Nass, S.L., I.C. Henderson, and J.C. Lashof, *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer*. 2001: National Academy Press.
10. Bindu, G., et al., *Active Microwave Imaging for Breast Cancer Detection*. Progress In Electromagnetics Research, 2006. **58**: p. 149-169.
11. Chen, S.C., et al., *Analysis of sonographic features for the differentiation of benign and malignant breast tumors of different sizes*. Ultrasound in Obstetrics and Gynecology, 2004. **23**(2): p. 188-193.
12. Nguyen, T.M. and R.M. Rangayyan. *Shape Analysis of Breast Masses in Mammograms Via the Fractal Dimension*. in *Engineering in Medicine and Biology 27th Annual Conference*. 2005. Shangai, China: IEEE.
13. Guliato, D., et al., *Polygonal Modeling of Contours of Breast Tumors with the Preservation of Spicules*. IEEE Transactions on Biomedical Engineering, 2008. **55**(1): p. 14-20.
14. Guliato, D., et al., *Feature Extraction from a Signature Based on the Turning Angle Function for the Classification of Breast Tumors*. Journal of Digital Imaging, 2008. **21**(2): p. 129-144.
15. Warner, E., et al., *Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer*. J Clin Oncol, 2001. **19**(15): p. 3524-31.
16. Lee, C.H., et al., *Breast Cancer Screening With Imaging: Recommendations From the Society of Breast Imaging and the ACR on the Use of Mammography, Breast MRI, Breast Ultrasound, and Other Technologies for the Detection of Clinically Occult Breast Cancer*. Journal of the American College of Radiology : JACR, 2010. **7**(1): p. 18-27.
17. Chang, R.-F., et al., *Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors*. Breast Cancer Research and Treatment, 2005. **89**(2): p. 179-185.
18. Chang, R.-F., et al., *Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis*. Ultrasound in Medicine & Biology, 2003. **29**(5): p. 679-686.
19. Edelman, R.R. and S. Warach, *Magnetic Resonance Imaging*. New England Journal of Medicine, 1993. **328**(10): p. 708-716.
20. Jacobs, M.A., et al., *Benign and malignant breast lesions: diagnosis with multiparametric MR imaging*. Radiology, 2003. **229**(1): p. 225-32.
21. Avril, N., et al., *Glucose metabolism of breast cancer assessed by 18F-FDG PET: histologic and immunohistochemical tissue analysis*. J Nucl Med, 2001. **42**(1): p. 9-16.
22. Avril, N., et al., *Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose*. J Clin Oncol, 1996. **14**(6): p. 1848-57.
23. Suárez, M., et al., *Early diagnosis of recurrent breast cancer with FDG-PET in patients with progressive elevation of serum tumor markers*. The quarterly journal of nuclear

- medicine : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), 2002. **46**(2): p. 113-121.
24. Muehllehner, G. and J.S. Karp, *Positron emission tomography*. 2006.
  25. Hagness, S.C., A. Taflove, and J.E. Bridges, *Two Dimensional FDTD Analysis of a Pulsed Microwave Confocal System for Breast Cancer Detection: Fixed-Focus and Antenna-Array Sensors*. IEEE Transactions on Biomedical Engineering, 1998. **45**: p. 1470-1479.
  26. Bridges, J.E., et al., *Microwave Discrimination between Malignant and Benign Breast Tumours*, 2002, L.L.C. Interstitial: USA.
  27. Conceicao, R.C., *The Development of Ultra Wideband Scanning Techniques for Detection and Classification of Breast Cancer*, 2010.
  28. Hagness, S.C., A. Taflove, and J.E. Bridges, *Three-Dimensional FDTD Analysis of a Pulsed Microwave Confocal System for Breast Cancer Detection: Design of an Antenna-Array Element*. IEEE Transactions on Antennas and Propagation, 1999. **47**(5): p. 783-791.
  29. Joines, W.T., *Frequency-Dependent Absorption of Electromagnetic Energy in Biological Tissue*. IEEE Transactions on Biomedical Engineering, 1984. **BME-31**(1): p. 17-20.
  30. Foster, K.R. and J.L. Schepps, *Dielectric properties of tumor and normal tissues at radio through microwave frequencies*. The Journal of microwave power, 1981. **16**(2): p. 107-119.
  31. Lazebnik, M., et al., *A Large-Scale Study of the Ultrawideband Microwave Dielectric Properties of Normal Breast Tissue Obtained from Reduction Surgeries*. Physics in Medicine and Biology, 2007. **52**: p. 2637–2656.
  32. Lazebnik, M., et al., *A Large-Scale Study of the Ultrawideband Microwave Dielectric Properties of Normal, Benign and Malignant Breast Tissues Obtained from Cancer Surgeries*. Physics in Medicine and Biology, 2007. **52**: p. 6093–6115.
  33. Conceição, R.C., et al., *Effects of Dielectric Heterogeneity in the Performance of Breast Tumour Classifiers*. Progress In Electromagnetics Research M, 2011. **17**: p. 73-86.
  34. Fear, E.C., et al., *Confocal Microwave Imaging for Breast Cancer Detection: Localization of Tumors in Three Dimensions*. IEEE Transactions on Biomedical Engineering, 2002. **49**(8): p. 812-822.
  35. Conceição, R.C., et al. *Classification of Suspicious Regions within Ultrawideband Radar Images of the Breast*. in *16th IET Irish Signals and Systems Conference, ISSC 2008*. 2008. Galway, Ireland: Instn Engg & Tech, UK.
  36. Chen, Y., et al. *Effect of Lesion Morphology on Microwave Signature in Ultra-Wideband Breast Imaging: A Preliminary Two-dimensional Investigation*. in *Antennas and Propagation Society International Symposium, 2007 IEEE*. 2007.
  37. Rangayyan, R.M., et al., *Measures of Acutance and Shape for Classification of Breast Tumors*. IEEE Transactions on Medical Imaging, 1997. **16**(6): p. 799-810.
  38. Chen, Y., et al., *Effect of Lesion Morphology on Microwave Signature in 2-D Ultra-Wideband Breast Imaging*. IEEE Transactions on Biomedical Engineering, 2008. **55**(8): p. 2011-2021.
  39. Chen, Y., et al. *Application of the MIMO Radar Technique for Lesion Classification in UWB Breast Cancer Detection*. in *17th European Signal Processing Conference (EUSIPCO)*. 2009. Glasgow, Scotland.
  40. Chen, Y., et al., *Multiple-Input Multiple-Output Radar for Lesion Classification in Ultrawideband Breast Imaging*. IEEE Journal of Selected Topics in Signal Processing, 2010. **4**(1): p. 187-201.
  41. Teo, J., et al., *Breast Lesion Classification Using Ultrawideband Early Time Breast Lesion Response*. IEEE Transactions on Antennas and Propagation, 2010. **58**(8): p. 2604-2613.
  42. Davis, S.K., et al., *Breast Tumor Characterization Based on Ultrawideband Microwave Backscatter*. IEEE Transactions on Biomedical Engineering, 2008. **55**(1): p. 237-246.
  43. Conceição, R.C., et al., *Investigation of Classifiers for Early-Stage Breast Cancer Based on Radar Target Signatures*. Progress In Electromagnetics Research, 2010. **105**: p. 295-311.
  44. Conceição, R.C., et al., *Support Vector Machines for the Classification of Early-Stage Breast Cancer Based on Radar Target Signatures*. Progress In Electromagnetics Research B, 2010. **23**: p. 311-327.

45. Conceição, R.C., et al., *Evaluation of Features and Classifiers for Classification of Early-Stage Breast Cancer*. Journal of Electromagnetic Waves and Applications, 2011. **25**(1): p. 1-14.
46. Ramsay, D.T., et al., *Anatomy of the lactating human breast redefined with ultrasound imaging*. J Anat, 2005. **206**(6): p. 525-34.
47. Bland, K.I., et al., *The Breast: Comprehensive Management of Benign and Malignant Disorders*, ed. Elsevier. Vol. 1., 2004.
48. Jossinet, J., *The Impedivity of Freshly Excised Human Breast Tissue*. Physiological Measurement, 1998. **19**(1): p. 61-75.
49. Prendergast, P., *Anatomy of the Breast*, in *Cosmetic Surgery*, M.A. Shiffman and A. Di Giuseppe, Editors. 2013, Springer Berlin Heidelberg. p. 47-55.
50. O'Neill, J.S., R.A. Elton, and W.R. Miller, *Aromatase activity in adipose tissue from breast quadrants: a link with tumour site*. Br Med J (Clin Res Ed), 1988. **296**(6624): p. 741-3.
51. Cancer, W. 2013; Disponível em: <http://www.who.int/cancer/en/>.
52. Institute, N.C. *Cancer Topics*. 2013; Available from: <http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer>.
53. UK, C.R. *About cancer*. 2013; Disponível em: <http://www.cancerresearchuk.org/cancer-help/about-cancer/what-is-cancer/cells/what-cancer-is>.
54. Chaudhary, S.S., et al., *Dielectric Properties of Normal and Malignant Human Breast Tissues at Radiowave and Microwave Frequencies*. Indian Journal of Biochemistry & Biophysics, 1984. **21**: p. 76-79.
55. Surowiec, A.J., et al., *Dielectric Properties of Breast Carcinoma and the Surrounding Tissues*. IEEE Transactions on Biomedical Engineering, 1988. **35**(4): p. 257-263.
56. Morimoto, T., et al., *Measurement of the electrical bio-impedance of breast tumors*. Eur Surg Res, 1990. **22**(2): p. 86-92.
57. Campbell, A.M. and D.V. Land, *Dielectric Properties of Female Human Breast Tissue Measured In Vitro at 3.2 GHz*. Physics in Medicine and Biology, 1992. **37**(1): p. 193-210.
58. Joines, W.T., et al., *The Measured Electrical Properties of Normal and Malignant Human Tissues from 50 to 900 MHz*. Medical Physics, 1994. **21**(4).
59. Choi, J.W., et al., *Microwave Detection of Metastasized Breast Cancer Cells in the Lymph Node; Potential Application for Sentinel Lymphadenectomy*. Breast Cancer Research and Treatment, 2004. **86**: p. 107-115.
60. Sha, L., E.R. Ward, and B. Stroy. *A review of dielectric properties of normal and malignant breast tissue*. in *SoutheastCon, 2002. Proceedings IEEE*. 2002.
61. Halter, R.J., et al., *The Correlation of In Vivo and Ex Vivo Tissue Dielectric Properties to Validate Electromagnetic Breast Imaging: Initial Clinical Experience*. Physiological Measurement, 2009. **30**(6): p. S121-S136.
62. Salvador, S.M. and G. Vecchi, *Experimental Tests of Microwave Breast Cancer Detection on Phantoms*. Antennas and Propagation, IEEE Transactions on, 2009. **57**(6): p. 1705-1712.
63. Lai, J.C.Y., et al., *UWB microwave imagin for breast cancer detection - experiments with heterogeneous breast phantoms*. Progress In Electromagnetics Research 2011. **16**: p. 19-29.
64. Klemm, M., et al. *Clinical Trials of a UWB Imaging Radar for Breast Cancer*. in *Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP)*. 2010. Barcelona, Spain.
65. Flores-Tapia, D. and S. Pistorius, *Real Time Breast Microwave Radar Image Reconstruction Using Circular Holography: A Study of Experimental Feasibility*. Medical Physics, 2011. **38**(10): p. 5420-5431.
66. Bourqui, J., J.M. Sill, and E.C. Fear, *A Prototype System for Measuring Microwave Frequency Reflections from the Breast*. International Journal of Biomedical Imaging, 2012. **2012**(Article ID 851234): p. 12 pages.
67. Chen, Y., I.J. Craddock, and P. Kosmas, *Feasibility Study of Lesion Classification via Contrast-Agent-Aided UWB Breast Imaging*. IEEE Transactions on Biomedical Engineering, 2010. **57**(5): p. 1003-1007.
68. O'Halloran, M., et al., *FDTD Modeling of the Breast: A Review*. Progress In Electromagnetics Research B, 2009. **18**: p. 1-24.
69. Yee, K.S., *Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media*. IEEE Transactions on Antennas and Propagation, 1966. **14**(3): p. 302-307.

70. Taflov, A. and S.C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Second Edition ed. 2000, Boston: Artech House.
71. Luebbers, R., et al., *A Frequency-Dependent Finite-Difference Time-Domain Formulation for Dispersive Materials*. IEEE Transactions on Electromagnetic Compability, 1990. **32**(3): p. 222-227.
72. Meaney, P.M., et al., *Initial Clinical Experience with Microwave Breast Imaging in Women with Normal Mammography*. Academic Radiology, 2007. **14**(2): p. 207-218.
73. R. C. Conceição, H.M., M. O'Halloran, D. Rodriguez-Herrera, D. Flores-Tapia, S. Pistorius, *Initial Classification of Breast Tumour Phantoms Using a UWB Radar Prototype*. 2013.
74. Sill, J.M. and E.C. Fear, *Tissue Sensing Adaptive Radar for Breast Cancer Detection — Experimental Investigation of Simple Tumor Models*. IEEE Transactions on Microwave Theory and Techniques, 2005. **53**(11): p. 3312-3319.
75. Byrne, D., et al., *Data Independent Radar Beamforming Algorithms for Breast Cancer Detection*. Progress In Electromagnetics Research, 2010. **107**: p. 331-348.
76. Hotelling, H., *Analysis of a complex of statistical variables into principal components*. Journal of Educational Psychology, 1933. **24**(6): p. 417-441.
77. Sewell, M., *Principal Component Analysis*. 2007.
78. Shlens, J. *A Tutorial on Principal Component Analysis*. 05/07/2010]; Disponível em: [http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition\\_jp.pdf](http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf).
79. Jolliffe, I., *Principal Component Analysis*, in *Encyclopedia os Statistics in Behavioral Science* 2005.
80. Daultrey, S., *Principal Component Analysis*. Concepts and techniques in modern geography, ed. U.o.E.A. Geo Abstracts. Vol. 8. 1976.
81. Hyvärinen, A. and E. Oja, *Independent Component Analysis: Algorithms and Applications*. Neural Networks, 2000. **13**(4-5): p. 411-430.
82. Lee, T.-W., *Independent Component Analysis*, in *Independent Component Analysis*. 1998, Springer US. p. 27-66.
83. Leite, L., *Análise de Componentes Independentes Aplicada à Identificação de Regiões Lesionadas em Mamografias*, 2005, Universidade Federal do Rio de Janeiro.
84. Mallet, Y., et al., *Classification Using Adaptive Wavelets for Feature Extraction*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997. **19**(10): p. 1058-1066.
85. Reis, E., *Estatística Multivariada Aplicada*. 2 ed, ed. E. Silabo. 2001.
86. Hair, J.F., et al., *Multivariate Data Analysis*. Sixth Edition ed. 2006, Upper Saddle River, New Jersey, USA: Prentice Hall.
87. Fukunaga, K. and R.R. Hayes, *Effects of sample size in classifier design*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1989. **11**(8): p. 873-885.
88. Friedman, J.H., *Regularized Discriminant Analysis*. Journal of the American Statistical Association, 1989. **84**(405): p. 165-175.
89. Júnior, I.J.D.N. and S.R. de Faria, *Proposição de Modelo de Análise Discriminante para Avaliação de Desempenho de Empresas de Sucesso e Fracasso do Setor de Indústrias Manufatureiras*.
90. Everitt, B.S. and G. Dunn, *Applied Multivariate Data Analysis*. Second Edition ed. 2001, New York: Arnold Publishers.
91. Hsu, C., C. Chang, and C. Lin. *A Practical Guide to Support Vector Classification*. 03/04/2010]; Disponível em: [www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf](http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf).
92. Campbell, C. *Introduction to Support Vector Machines*. 2008 05/02/2008]; Disponível em: [http://videolectures.net/epsrws08\\_campbell\\_isvm/](http://videolectures.net/epsrws08_campbell_isvm/).
93. Cortes, C. and V. Vapnik, *Support-Vector Networks*. Machine Learning, 1995. **20**(3): p. 273-297.
94. Ng, A. *Support Vector Machines (Part V of CS229 Machine Learning Course Materials)*. 13/05/2010]; Disponível em: <http://www.stanford.edu/class/cs229/notes/cs229-notes3.pdf>.
95. Bennett, K.P. and C. Campbell, *Support Vector Machines: Hype or Hallelujah?* ACM SIGKDD Explorations Newsletter, 2000. **2**(2): p. 1-13.
96. Muinonen, K., *Introducing the Gaussian Shape Hypothesis for Asteroids and Comets*. Astronomy and Astrophysics, 1998. **332**: p. 1087-1098.
97. Muinonen, K., *Chapter 11: Light Scattering by Stochastically Shaped Particles*, in *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*, M.I. Mishchenko, J.W. Hovenier, and L.D. Travis, Editors. 2000, Academic Press.

98. Muinonen, K. *Gaussian Random Sphere Program G-sphere*. 2002 23/05/2008]; 2002-12-16:[Disponível em: [www.astro.helsinki.fi/psr/](http://www.astro.helsinki.fi/psr/).
99. Gävert, H., et al. *The FastICA Package for MATLAB*. 20/07/2010]; Disponível em: [www.cis.hut.fi/projects/ica/fastica/](http://www.cis.hut.fi/projects/ica/fastica/).
100. Flores-Tapia, D., et al. *Microwave radar imaging of inhomogeneous breast phantoms using circular holography*. in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. 2012.